ICON/UXB Operating System Reference Manual

Volume 2

ICON INTERNATIONAL

P.O. Box 340
Orem, Utah 84059
(801) 225-6888

OPERATING SYSTEM REFERENCE MANUAL

ICON/UXB

Supplementary Documents

Volume 2

This manual reflects system enhancements made at Berkeley and sponsored in part by NSF Grants MCS-7807291, MCS-8005144, and MCS-74-07644-A04; DOE Contract DE-AT03-76SF00034 and Project Agreement DE-AS03-79ER10358; and by Defense Advanced Research Projects Agency (DoD) ARPA Order No. 4031, Monitored by Naval Electronics Systems Command under Contract No. N00039-80-K-0649.

This manual was prepared by the Documentation Group of Icon International, Inc., P.O. Box 340, Orem, UT 84057-0340. A form for reader's comments has been provided at the back of this publication. Comments are welcomed and may be sent to the above address. Users who respond will be entitled to free updates of this manual for one year.

Revision B

Order Number 172-022-003 (Manual Assembly)
Order Number 171-070-003 (Pages Only)

Printed in the U.S.A.

Change Record Page

Manual Part No. 172-022-003

Date	Revision	Description	Pages Affected
Jan. 1987	A	Initial production release	All
Nov. 1987	B	Incorporate additions of new supplementary documentation included in Releases 2.16, 3.0, and 3.1 of the ICON/UXB Operating System	Main cover, titlepage, Table of Contents

ICON/UXB Operating System Reference Manual

Volume 2 — Supplementary Documents

Icon International, Inc.

October, 1987

This volume contains documents which supplement the information in Volume 1 of the *ICON/UXB Operating System Reference Manual,* for the ICON version of the UNIX® operating system as distributed by U.C. Berkeley. The documents within this volume are grouped into the areas of general works and summaries, basic information on the UNIX® operating system, document preparation, programming, and miscellaneous information.

General Works

1. The UNIX Time-Sharing System. D.M. Ritchie and K. Thompson.
   The original UNIX® operating system paper, reprinted from CACM.
2. Bug Fixes and changes in 4.2BSD. S.J. Leffler.
   A brief discussion of the major user-visible changes made to the system since the last release.
3. UNIX/32V — Summary.
   A concise summary of the facilities available in the 32V Version of the UNIX® operating system.
4. 7th Edition UNIX — Summary.
   A concise summary of the facilities available in the 7th edition of the UNIX® operating system.

Getting Started

5. UNIX for Beginners — Second Edition. B.W. Kernighan.
   An introduction to the most basic use of the system.
6. Learn — Computer Aided Instruction on UNIX. M.E. Lesk and B.W. Kernighan.
   Describes a computer-aided instruction program that walks new users through the basics of files, the editor, and document preparation software.
7. An Introduction to the UNIX Shell. S.R. Bourne.
   An introduction to the capabilities of the command interpreter, the shell.
8. An Introduction to the C Shell. W. Joy.
   Introducing a popular command interpreter and many of the commonly used commands, assuming little prior knowledge of the UNIX® operating system.

9. An Introduction to Display Editing with Vi. W. Joy.
   The document to read to learn to use the *vi* screen-oriented display text editor.

10. Vi Command & Function Reference. A.P.W. Hewett.
    The reference manual for *vi.*

11. A Tutorial Introduction to the UNIX Text Editor. B.W. Kernighan.
    An easy way to get started with the editor.

12. Edit: A tutorial (Revised). R. Blau and J. Joyce.
    For those who prefer line oriented editing, an introduction assuming no previous knowledge of the UNIX operating system or of text editing.

13. Ex Reference Manual (Version 3.5 — Sept. 1980). W. Joy and M. Horton.
    The final reference for the *ex* editor, which underlies both *edit* and *vi.*

14. Ex Changes — Version 2.0 to 3.5
    A quick guide to what is new in version 3.5 of *ex* and *vi*, for those who have used version 2.0 through 3.1. Includes an update to the *vi* Tutorial and a command summary for *ex/edit*, version 2.0.

15. Advance Editing on UNIX. B.W. Kernighan.
    The next step.

16. Mail Reference Manual (Revised). K. Shoens and C. Leres.
    Complete details on the mail processing program.

Document preparation

17. Typing Documents on the UNIX System: Using the —ms Macros with Troff and Nroff. M.E. Lesk.
    Describes the basic use of the formatting tools and the formatting requests that can be used to lay out most documents, including thoses in this volume. Also includes *A Guide to Preparing Documents with —ms*, a quick summary to the *-ms* macro commands.

18. A Revised Version of —ms. B. Tuthill.
    A quick description of the revisions made to the —ms formatting macros for *nroff* and *troff.*

19. Writing Papers with *nroff* using —me. E.P. Allman.
    A popular macro package for *nroff.*

20. —me Reference Manual. E.P. Allman.
    The final word on —me.

21. Writing Tools — The Style and Diction Programs. L.L. Cherry and W. Vesterman
    Description of programs which help you understand and improve your writing style.

22. NROFF/TROFF User's Manual. J.F. Ossanna.
    The basic text formatting program.

23. A TROFF Tutorial. B.W. Kernighan.
    An introduction to TROFF for those who really want to know such things.

24. Refer — A Bibliography System. B. Tuthill.
    An introduction to the tools used to maintain bibliographic databases. The major program, *refer*, is used to automatically retrieve and format references based on document citations.

25. Some Applications of Inverted Indexes on the UNIX System. M.E. Lesk.
    Describes, among other things, the program *refer* which fills in bibliographic citations from a database automatically.

26. Updating Publication Lists. M.E. Lesk.
    Using *refer* to update a bibliographic database.

27. TBL — A Program to Format Tables. M.E. Lesk.
    A program to permit easy specification of tabular material for typesetting. Easy to learn and use.

28. A System for Typsetting Mathematics. B.W. Kernighan and L.L. Cherry.
    Describes EQN, an easy-to-learn language for doing high-quality mathematical typesetting.

29. Typesetting Mathematics — User's Guide (Second Edition). B.W. Kernighan and L.L. Cherry.
    The EQN User's Guide for typesetting mathematics.

Programming

30. UNIX Programming. B.W. Kernighan and D.M. Ritchie.
    Describes the programming interface to the operating system and the standard I/O library.

31. Make — A Program for Maintaining Computer Programs. S.I. Feldman.
    An indispensable tool for making sure that large programs are properly compiled with minimal effort.

32. System V/68 Assembler User's Guide.
    For compiler writers using the 68000 series microprocessors.

33. Screen Updating and Cursor Movement Optimization: A Library Package. K.C.R.C. Arnold.
    An aide for writing screen-oriented, terminal independent programs.

34. A Tutorial Introduction to ADB. J.F. Maranzano and S.R. Bourne.
    How to use the ADB debugger.

35. An Introduction to the Source Code Control System. E. Allman.
    A useful introductory article for those users who are licensed for SCCS.

Miscellaneous

36. A Guide to the Dungeons of Doom (Revised). M.C. Toy and K.C.R.C. Arnold.
    An introduction to the popular game of *rogue*.

37. STAR TREK. E. Allman.
    What's UNIX without a "trekkie" to accompany us?

38. A 4.2BSD Interprocess Communication Primer. S.J. Leffler, R.S. Fabry and W.N. Joy.
    An introduction to the interprocess communication facilities included in the 4.2BSD release of the s-1UNIX® operating system.

39. gprof: a Call Graph Execution Profiler. S.L. Graham, P.B. Kessler and M.K. McKusick.
    A description of the gprof profile used to account for the running time of called routines in the running time of the routines that called them.

The UNIX Time-Sharing System*

D. M. Ritchie and K. Thompson

ABSTRACT

UNIX† is a general-purpose, multi-user, interactive operating system for the larger Digital Equipment Corporation PDP-11 and the Interdata 8/32 computers. It offers a number of features seldom found even in larger operating systems, including

- i A hierarchical file system incorporating demountable volumes,
- ii Compatible file, device, and inter-process I/O,
- iii The ability to initiate asynchronous processes,
- iv System command language selectable on a per-user basis,
- v Over 100 subsystems including a dozen languages,
- vi High degree of portability.

This paper discusses the nature and implementation of the file system and of the user command interface.

1. INTRODUCTION

There have been four versions of the UNIX time-sharing system. The earliest (circa 1969-70) ran on the Digital Equipment Corporation PDP-7 and -9 computers. The second version ran on the unprotected PDP-11/20 computer. The third incorporated multiprogramming and ran on the PDP-11/34, /40, /45, /60, and /70 computers; it is the one described in the previously published version of this paper, and is also the most widely used today. This paper describes only the fourth, current system that runs on the PDP-11/70 and the Interdata 8/32 computers. In fact, the differences among the various systems is rather small; most of the revisions made to the originally published version of this paper, aside from those concerned with style, had to do with details of the implementation of the file system.

Since PDP-11 UNIX became operational in February, 1971, over 600 installations have been put into service. Most of them are engaged in applications such as computer science education, the preparation and formatting of documents and other textual material, the collection and processing of trouble data from various switching machines within the Bell System, and recording and checking telephone service orders. Our own installation is used mainly for research in operating systems, languages, computer networks, and other topics in computer science, and also for document preparation.

Perhaps the most important achievement of UNIX is to demonstrate that a powerful operating system for interactive use need not be expensive either in equipment or in human effort: it can run on hardware costing as little as $40,000, and less than two man-years were spent on the main system software. We hope, however, that users find that the most important characteristics of the

 This is a revised version of an article that appeared in Communications of the ACM, *17*, No. 7 (July 1974), pp. 365-375. That article was a revised version of a paper presented at the Fourth ACM Symposium on Operating Systems Principles, IBM Thomas J. Watson Research Center, Yorktown Heights, New York, October 15-17, 1973.

† UNIX is a trademark of Bell Laboratories.

system are its simplicity, elegance, and ease of use.

Besides the operating system proper, some major programs available under UNIX are

C compiler
Text editor based on QED[1]
Assembler, linking loader, symbolic debugger
Phototypesetting and equation setting programs[2, 3]
Dozens of languages including Fortran 77, Basic, Snobol, APL, Algol 68, M6, TMG, Pascal

There is a host of maintenance, utility, recreation and novelty programs, all written locally. The UNIX user community, which numbers in the thousands, has contributed many more programs and languages. It is worth noting that the system is totally self-supporting. All UNIX software is maintained on the system; likewise, this paper and all other documents in this issue were generated and formatted by the UNIX editor and text formatting programs.

II. HARDWARE AND SOFTWARE ENVIRONMENT

The PDP-11/70 on which the Research UNIX system is installed is a 16-bit word (8-bit byte) computer with 768K bytes of core memory; the system kernel occupies 90K bytes about equally divided between code and data tables. This system, however, includes a very large number of device drivers and enjoys a generous allotment of space for I/O buffers and system tables; a minimal system capable of running the software mentioned above can require as little as 96K bytes of core altogether. There are even larger installations; see the description of the PWB/UNIX systems,[4, 3] for example. There are also much smaller, though somewhat restricted, versions of the system.[3]

Our own PDP-11 has two 200-Mb moving-head disks for file system storage and swapping. There are 20 variable-speed communications interfaces attached to 300- and 1200-baud data sets, and an additional 12 communication lines hard-wired to 9600-baud terminals and satellite computers. There are also several 2400- and 4800-baud synchronous communication interfaces used for machine-to-machine file transfer. Finally, there is a variety of miscellaneous devices including nine-track magnetic tape, a line printer, a voice synthesizer, a phototypesetter, a digital switching network, and a chess machine.

The preponderance of UNIX software is written in the abovementioned C language.[5] Early versions of the operating system were written in assembly language, but during the summer of 1973, it was rewritten in C. The size of the new system was about one-third greater than that of the old. Since the new system not only became much easier to understand and to modify but also included many functional improvements, including multiprogramming and the ability to share reentrant code among several user programs, we consider this increase in size quite acceptable.

III. THE FILE SYSTEM

The most important role of the system is to provide a file system. From the point of view of the user, there are three kinds of files: ordinary disk files, directories, and special files.

3.1 Ordinary files

A file contains whatever information the user places on it, for example, symbolic or binary (object) programs. No particular structuring is expected by the system. A file of text consists simply of a string of characters, with lines demarcated by the newline character. Binary programs are sequences of words as they will appear in core memory when the program starts executing. A few user programs manipulate files with more structure; for example, the assembler generates, and the loader expects, an object file in a particular format. However, the structure of files is controlled by the programs that use them, not by the system.

3.2 Directories

Directories provide the mapping between the names of files and the files themselves, and thus induce a structure on the file system as a whole. Each user has a directory of his own files; he may also create subdirectories to contain groups of files conveniently treated together. A directory behaves exactly like an ordinary file except that it cannot be written on by unprivileged programs, so that the system controls the contents of directories. However, anyone with appropriate permission may read a directory just like any other file.

The system maintains several directories for its own use. One of these is the **root** directory. All files in the system can be found by tracing a path through a chain of directories until the desired file is reached. The starting point for such searches is often the **root**. Other system directories contain all the programs provided for general use; that is, all the *commands*. As will be seen, however, it is by no means necessary that a program reside in one of these directories for it to be executed.

Files are named by sequences of 14 or fewer characters. When the name of a file is specified to the system, it may be in the form of a *path name*, which is a sequence of directory names separated by slashes, "/", and ending in a file name. If the sequence begins with a slash, the search begins in the root directory. The name **/alpha/beta/gamma** causes the system to search the root for directory **alpha**, then to search **alpha** for **beta**, finally to find **gamma** in **beta**. **gamma** may be an ordinary file, a directory, or a special file. As a limiting case, the name "/" refers to the root itself.

A path name not starting with "/" causes the system to begin the search in the user's current directory. Thus, the name **alpha/beta** specifies the file named **beta** in subdirectory **alpha** of the current directory. The simplest kind of name, for example, **alpha**, refers to a file that itself is found in the current directory. As another limiting case, the null file name refers to the current directory.

The same non-directory file may appear in several directories under possibly different names. This feature is called *linking*; a directory entry for a file is sometimes called a link. The UNIX system differs from other systems in which linking is permitted in that all links to a file have equal status. That is, a file does not exist within a particular directory; the directory entry for a file consists merely of its name and a pointer to the information actually describing the file. Thus a file exists independently of any directory entry, although in practice a file is made to disappear along with the last link to it.

Each directory always has at least two entries. The name "**.**" in each directory refers to the directory itself. Thus a program may read the current directory under the name "**.**" without knowing its complete path name. The name "**..**" by convention refers to the parent of the directory in which it appears, that is, to the directory in which it was created.

The directory structure is constrained to have the form of a rooted tree. Except for the special entries "**.**" and "**..**", each directory must appear as an entry in exactly one other directory, which is its parent. The reason for this is to simplify the writing of programs that visit subtrees of the directory structure, and more important, to avoid the separation of portions of the hierarchy. If arbitrary links to directories were permitted, it would be quite difficult to detect when the last connection from the root to a directory was severed.

3.3 Special files

Special files constitute the most unusual feature of the UNIX file system. Each supported I/O device is associated with at least one such file. Special files are read and written just like ordinary disk files, but requests to read or write result in activation of the associated device. An entry for each special file resides in directory **/dev**, although a link may be made to one of these files just as it may to an ordinary file. Thus, for example, to write on a magnetic tape one may write on the file **/dev/mt**. Special files exist for each communication line, each disk, each tape drive, and for physical main memory. Of course, the active disks and the memory special file are protected from indiscriminate access.

There is a threefold advantage in treating I/O devices this way: file and device I/O are as similar as possible; file and device names have the same syntax and meaning, so that a program expecting a file name as a parameter can be passed a device name; finally, special files are subject to the same protection mechanism as regular files.

3.4 Removable file systems

Although the root of the file system is always stored on the same device, it is not necessary that the entire file system hierarchy reside on this device. There is a **mount** system request with two arguments: the name of an existing ordinary file, and the name of a special file whose associated storage volume (e.g., a disk pack) should have the structure of an independent file system containing its own directory hierarchy. The effect of **mount** is to cause references to the heretofore ordinary file to refer instead to the root directory of the file system on the removable volume. In effect, **mount** replaces a leaf of the hierarchy tree (the ordinary file) by a whole new subtree (the hierarchy stored on the removable volume). After the **mount**, there is virtually no distinction between files on the removable volume and those in the permanent file system. In our installation, for example, the root directory resides on a small partition of one of our disk drives, while the other drive, which contains the user's files, is mounted by the system initialization sequence. A mountable file system is generated by writing on its corresponding special file. A utility program is available to create an empty file system, or one may simply copy an existing file system.

There is only one exception to the rule of identical treatment of files on different devices: no link may exist between one file system hierarchy and another. This restriction is enforced so as to avoid the elaborate bookkeeping that would otherwise be required to assure removal of the links whenever the removable volume is dismounted.

3.5 Protection

Although the access control scheme is quite simple, it has some unusual features. Each user of the system is assigned a unique user identification number. When a file is created, it is marked with the user ID of its owner. Also given for new files is a set of ten protection bits. Nine of these specify independently read, write, and execute permission for the owner of the file, for other members of his group, and for all remaining users.

If the tenth bit is on, the system will temporarily change the user identification (hereafter, user ID) of the current user to that of the creator of the file whenever the file is executed as a program. This change in user ID is effective only during the execution of the program that calls for it. The set-user-ID feature provides for privileged programs that may use files inaccessible to other users. For example, a program may keep an accounting file that should neither be read nor changed except by the program itself. If the set-user-ID bit is on for the program, it may access the file although this access might be forbidden to other programs invoked by the given program's user. Since the actual user ID of the invoker of any program is always available, set-user-ID programs may take any measures desired to satisfy themselves as to their invoker's credentials. This mechanism is used to allow users to execute the carefully written commands that call privileged system entries. For example, there is a system entry invokable only by the "super-user" (below) that creates an empty directory. As indicated above, directories are expected to have entries for "**.**" and "**..**". The command which creates a directory is owned by the super-user and has the set-user-ID bit set. After it checks its invoker's authorization to create the specified directory, it creates it and makes the entries for "**.**" and "**..**".

Because anyone may set the set-user-ID bit on one of his own files, this mechanism is generally available without administrative intervention. For example, this protection scheme easily solves the MOO accounting problem posed by "Aleph-null."[6]

The system recognizes one particular user ID (that of the "super-user") as exempt from the usual constraints on file access; thus (for example), programs may be written to dump and reload the file system without unwanted interference from the protection system.

3.6 I/O calls

The system calls to do I/O are designed to eliminate the differences between the various devices and styles of access. There is no distinction between "random" and "sequential" I/O, nor is any logical record size imposed by the system. The size of an ordinary file is determined by the number of bytes written on it; no predetermination of the size of a file is necessary or possible.

To illustrate the essentials of I/O, some of the basic calls are summarized below in an anonymous language that will indicate the required parameters without getting into the underlying complexities. Each call to the system may potentially result in an error return, which for simplicity is not represented in the calling sequence.

To read or write a file assumed to exist already, it must be opened by the following call:

filep = open (name, flag)

where **name** indicates the name of the file. An arbitrary path name may be given. The **flag** argument indicates whether the file is to be read, written, or "updated," that is, read and written simultaneously.

The returned value **filep** is called a *file descriptor*. It is a small integer used to identify the file in subsequent calls to read, write, or otherwise manipulate the file.

To create a new file or completely rewrite an old one, there is a **create** system call that creates the given file if it does not exist, or truncates it to zero length if it does exist; **create** also opens the new file for writing and, like **open**, returns a file descriptor.

The file system maintains no locks visible to the user, nor is there any restriction on the number of users who may have a file open for reading or writing. Although it is possible for the contents of a file to become scrambled when two users write on it simultaneously, in practice difficulties do not arise. We take the view that locks are neither necessary nor sufficient, in our environment, to prevent interference between users of the same file. They are unnecessary because we are not faced with large, single-file data bases maintained by independent processes. They are insufficient because locks in the ordinary sense, whereby one user is prevented from writing on a file that another user is reading, cannot prevent confusion when, for example, both users are editing a file with an editor that makes a copy of the file being edited.

There are, however, sufficient internal interlocks to maintain the logical consistency of the file system when two users engage simultaneously in activities such as writing on the same file, creating files in the same directory, or deleting each other's open files.

Except as indicated below, reading and writing are sequential. This means that if a particular byte in the file was the last byte written (or read), the next I/O call implicitly refers to the immediately following byte. For each open file there is a pointer, maintained inside the system, that indicates the next byte to be read or written. If n bytes are read or written, the pointer advances by n bytes.

Once a file is open, the following calls may be used:

n = read (filep, buffer, count)
n = write (filep, buffer, count)

Up to **count** bytes are transmitted between the file specified by **filep** and the byte array specified by **buffer**. The returned value **n** is the number of bytes actually transmitted. In the **write** case, **n** is the same as **count** except under exceptional conditions, such as I/O errors or end of physical medium on special files; in a **read**, however, **n** may without error be less than **count**. If the read pointer is so near the end of the file that reading **count** characters would cause reading beyond the end, only sufficient bytes are transmitted to reach the end of the file; also, typewriter-like terminals never return more than one line of input. When a **read** call returns with **n** equal to zero, the end of the file has been reached. For disk files this occurs when the read pointer becomes equal to the current size of the file. It is possible to generate an end-of-file from a terminal by use of an escape sequence that depends on the device used.

Bytes written affect only those parts of a file implied by the position of the write pointer and the count; no other part of the file is changed. If the last byte lies beyond the end of the file, the file is made to grow as needed.

To do random (direct-access) I/O it is only necessary to move the read or write pointer to the appropriate location in the file.

location = lseek (filep, offset, base)

The pointer associated with **filep** is moved to a position **offset** bytes from the beginning of the file, from the current position of the pointer, or from the end of the file, depending on **base. offset** may be negative. For some devices (e.g., paper tape and terminals) seek calls are ignored. The actual offset from the beginning of the file to which the pointer was moved is returned in **location**.

There are several additional system entries having to do with I/O and with the file system that will not be discussed. For example: close a file, get the status of a file, change the protection mode or the owner of a file, create a directory, make a link to an existing file, delete a file.

IV. IMPLEMENTATION OF THE FILE SYSTEM

As mentioned in Section 3.2 above, a directory entry contains only a name for the associated file and a pointer to the file itself. This pointer is an integer called the *i-number* (for index number) of the file. When the file is accessed, its i-number is used as an index into a system table (the *i-list*) stored in a known part of the device on which the directory resides. The entry found thereby (the file's *i-node*) contains the description of the file:

i the user and group-ID of its owner

ii its protection bits

iii the physical disk or tape addresses for the file contents

iv its size

v time of creation, last use, and last modification

vi the number of links to the file, that is, the number of times it appears in a directory

vii a code indicating whether the file is a directory, an ordinary file, or a special file.

The purpose of an **open** or **create** system call is to turn the path name given by the user into an i-number by searching the explicitly or implicitly named directories. Once a file is open, its device, i-number, and read/write pointer are stored in a system table indexed by the file descriptor returned by the **open** or **create**. Thus, during a subsequent call to read or write the file, the descriptor may be easily related to the information necessary to access the file.

When a new file is created, an i-node is allocated for it and a directory entry is made that contains the name of the file and the i-node number. Making a link to an existing file involves creating a directory entry with the new name, copying the i-number from the original file entry, and incrementing the link-count field of the i-node. Removing (deleting) a file is done by decrementing the link-count of the i-node specified by its directory entry and erasing the directory entry. If the link-count drops to 0, any disk blocks in the file are freed and the i-node is deallocated.

The space on all disks that contain a file system is divided into a number of 512-byte blocks logically addressed from 0 up to a limit that depends on the device. There is space in the i-node of each file for 13 device addresses. For nonspecial files, the first 10 device addresses point at the first 10 blocks of the file. If the file is larger than 10 blocks, the 11 device address points to an indirect block containing up to 128 addresses of additional blocks in the file. Still larger files use the twelfth device address of the i-node to point to a double-indirect block naming 128 indirect blocks, each pointing to 128 blocks of the file. If required, the thirteenth device address is a triple-indirect block. Thus files may conceptually grow to $[(10+128+128^2+128^3)\cdot 512]$ bytes. Once opened, bytes numbered below 5120 can be read with a single disk access; bytes in the range 5120 to 70,656 require two accesses; bytes in the range 70,656 to 8,459,264 require three accesses; bytes from there to the largest file (1,082,201,088) require four accesses. In practice, a device cache mechanism (see below) proves effective in eliminating most of the indirect fetches.

The foregoing discussion applies to ordinary files. When an I/O request is made to a file whose i-node indicates that it is special, the last 12 device address words are immaterial, and the first specifies an internal *device name*, which is interpreted as a pair of numbers representing, respectively, a device type and subdevice number. The device type indicates which system routine will deal with I/O on that device; the subdevice number selects, for example, a disk drive attached to a particular controller or one of several similar terminal interfaces.

In this environment, the implementation of the **mount** system call (Section 3.4) is quite straightforward. **mount** maintains a system table whose argument is the i-number and device name of the ordinary file specified during the **mount**, and whose corresponding value is the device name of the indicated special file. This table is searched for each i-number/device pair that turns up while a path name is being scanned during an **open** or **create**; if a match is found, the i-number is replaced by the i-number of the root directory and the device name is replaced by the table value.

To the user, both reading and writing of files appear to be synchronous and unbuffered. That is, immediately after return from a **read** call the data are available; conversely, after a **write** the user's workspace may be reused. In fact, the system maintains a rather complicated buffering mechanism that reduces greatly the number of I/O operations required to access a file. Suppose a **write** call is made specifying transmission of a single byte. The system will search its buffers to see whether the affected disk block currently resides in main memory; if not, it will be read in from the device. Then the affected byte is replaced in the buffer and an entry is made in a list of blocks to be written. The return from the **write** call may then take place, although the actual I/O may not be completed until a later time. Conversely, if a single byte is read, the system determines whether the secondary storage block in which the byte is located is already in one of the system's buffers; if so, the byte can be returned immediately. If not, the block is read into a buffer and the byte picked out.

The system recognizes when a program has made accesses to sequential blocks of a file, and asynchronously pre-reads the next block. This significantly reduces the running time of most programs while adding little to system overhead.

A program that reads or writes files in units of 512 bytes has an advantage over a program that reads or writes a single byte at a time, but the gain is not immense; it comes mainly from the avoidance of system overhead. If a program is used rarely or does no great volume of I/O, it may quite reasonably read and write in units as small as it wishes.

The notion of the i-list is an unusual feature of UNIX. In practice, this method of organizing the file system has proved quite reliable and easy to deal with. To the system itself, one of its strengths is the fact that each file has a short, unambiguous name related in a simple way to the protection, addressing, and other information needed to access the file. It also permits a quite simple and rapid algorithm for checking the consistency of a file system, for example, verification that the portions of each device containing useful information and those free to be allocated are disjoint and together exhaust the space on the device. This algorithm is independent of the directory hierarchy, because it need only scan the linearly organized i-list. At the same time the notion of the i-list induces certain peculiarities not found in other file system organizations. For example, there is the question of who is to be charged for the space a file occupies, because all directory entries for a file have equal status. Charging the owner of a file is unfair in general, for one user may create a file, another may link to it, and the first user may delete the file. The first user is still the owner of the file, but it should be charged to the second user. The simplest reasonably fair algorithm seems to be to spread the charges equally among users who have links to a file. Many installations avoid the issue by not charging any fees at all.

V. PROCESSES AND IMAGES

An *image* is a computer execution environment. It includes a memory image, general register values, status of open files, current directory and the like. An image is the current state of a pseudo-computer.

A *process* is the execution of an image. While the processor is executing on behalf of a process, the image must reside in main memory; during the execution of other processes it remains in main memory unless the appearance of an active, higher-priority process forces it to be swapped out to the disk.

The user-memory part of an image is divided into three logical segments. The program text segment begins at location 0 in the virtual address space. During execution, this segment is write-protected and a single copy of it is shared among all processes executing the same program. At the first hardware protection byte boundary above the program text segment in the virtual address space begins a non-shared, writable data segment, the size of which may be extended by a system call. Starting at the highest address in the virtual address space is a stack segment, which automatically grows downward as the stack pointer fluctuates.

5.1 Processes

Except while the system is bootstrapping itself into operation, a new process can come into existence only by use of the **fork** system call:

processid = fork ( )

When **fork** is executed, the process splits into two independently executing processes. The two processes have independent copies of the original memory image, and share all open files. The new processes differ only in that one is considered the parent process: in the parent, the returned **processid** actually identifies the child process and is never 0, while in the child, the returned value is always 0.

Because the values returned by **fork** in the parent and child process are distinguishable, each process may determine whether it is the parent or child.

5.2 Pipes

Processes may communicate with related processes using the same system **read** and **write** calls that are used for file-system I/O. The call:

filep = pipe ( )

returns a file descriptor **filep** and creates an inter-process channel called a *pipe*. This channel, like other open files, is passed from parent to child process in the image by the **fork** call. A **read** using a pipe file descriptor waits until another process writes using the file descriptor for the same pipe. At this point, data are passed between the images of the two processes. Neither process need know that a pipe, rather than an ordinary file, is involved.

Although inter-process communication via pipes is a quite valuable tool (see Section 6.2), it is not a completely general mechanism, because the pipe must be set up by a common ancestor of the processes involved.

5.3 Execution of programs

Another major system primitive is invoked by

execute (file, arg_1, arg_2, . . . , arg_n)

which requests the system to read in and execute the program named by **file**, passing it string arguments $\mathbf{arg}_1$, $\mathbf{arg}_2$, . . . , $\mathbf{arg}_n$. All the code and data in the process invoking **execute** is replaced from the **file**, but open files, current directory, and inter-process relationships are unaltered. Only if the call fails, for example because **file** could not be found or because its execute-permission bit was not set, does a return take place from the **execute** primitive; it resembles a "jump" machine instruction rather than a subroutine call.

5.4 Process synchronization

Another process control system call:

processid = wait (status)

causes its caller to suspend execution until one of its children has completed execution. Then **wait** returns the **processid** of the terminated process. An error return is taken if the calling process has no descendants. Certain status from the child process is also available.

5.5 Termination

Lastly:

exit (status)

terminates a process, destroys its image, closes its open files, and generally obliterates it. The parent is notified through the **wait** primitive, and **status** is made available to it. Processes may also terminate as a result of various illegal actions or user-generated signals (Section VII below).

VI. THE SHELL

For most users, communication with the system is carried on with the aid of a program called the shell. The shell is a command-line interpreter: it reads lines typed by the user and interprets them as requests to execute other programs. (The shell is described fully elsewhere,[3] so this section will discuss only the theory of its operation.) In simplest form, a command line consists of the command name followed by arguments to the command, all separated by spaces:

command arg_1 arg_2 ... arg_n

The shell splits up the command name and the arguments into separate strings. Then a file with name **command** is sought; **command** may be a path name including the "/" character to specify any file in the system. If **command** is found, it is brought into memory and executed. The arguments collected by the shell are accessible to the command. When the command is finished, the shell resumes its own execution, and indicates its readiness to accept another command by typing a prompt character.

If file **command** cannot be found, the shell generally prefixes a string such as **/ bin /** to **command** and attempts again to find the file. Directory **/ bin** contains commands intended to be generally used. (The sequence of directories to be searched may be changed by user request.)

6.1 Standard I/O

The discussion of I/O in Section III above seems to imply that every file used by a program must be opened or created by the program in order to get a file descriptor for the file. Programs executed by the shell, however, start off with three open files with file descriptors 0, 1, and 2. As such a program begins execution, file 1 is open for writing, and is best understood as the standard output file. Except under circumstances indicated below, this file is the user's terminal. Thus programs that wish to write informative information ordinarily use file descriptor 1. Conversely, file 0 starts off open for reading, and programs that wish to read messages typed by the user read this file.

The shell is able to change the standard assignments of these file descriptors from the user's terminal printer and keyboard. If one of the arguments to a command is prefixed by ">", file descriptor 1 will, for the duration of the command, refer to the file named after the ">". For example:

ls

ordinarily lists, on the typewriter, the names of the files in the current directory. The command:

ls >there

creates a file called **there** and places the listing there. Thus the argument >**there** means "place output on **there**." On the other hand:

```
ed
```

ordinarily enters the editor, which takes requests from the user via his keyboard. The command

```
ed <script
```

interprets **script** as a file of editor commands; thus <**script** means "take input from **script**."

Although the file name following "<" or ">" appears to be an argument to the command, in fact it is interpreted completely by the shell and is not passed to the command at all. Thus no special coding to handle I/O redirection is needed within each command; the command need merely use the standard file descriptors 0 and 1 where appropriate.

File descriptor 2 is, like file 1, ordinarily associated with the terminal output stream. When an output-diversion request with ">" is specified, file 2 remains attached to the terminal, so that commands may produce diagnostic messages that do not silently end up in the output file.

6.2 Filters

An extension of the standard I/O notion is used to direct output from one command to the input of another. A sequence of commands separated by vertical bars causes the shell to execute all the commands simultaneously and to arrange that the standard output of each command be delivered to the standard input of the next command in the sequence. Thus in the command line:

```
ls | pr −2 | opr
```

ls lists the names of the files in the current directory; its output is passed to **pr**, which paginates its input with dated headings. (The argument "−2" requests double-column output.) Likewise, the output from **pr** is input to **opr**; this command spools its input onto a file for off-line printing.

This procedure could have been carried out more clumsily by:

```
ls >temp1
pr −2 <temp1 >temp2
opr <temp2
```

followed by removal of the temporary files. In the absence of the ability to redirect output and input, a still clumsier method would have been to require the **ls** command to accept user requests to paginate its output, to print in multi-column format, and to arrange that its output be delivered off-line. Actually it would be surprising, and in fact unwise for efficiency reasons, to expect authors of commands such as **ls** to provide such a wide variety of output options.

A program such as **pr** which copies its standard input to its standard output (with processing) is called a *filter*. Some filters that we have found useful perform character transliteration, selection of lines according to a pattern, sorting of the input, and encryption and decryption.

6.3 Command separators; multitasking

Another feature provided by the shell is relatively straightforward. Commands need not be on different lines; instead they may be separated by semicolons:

```
ls; ed
```

will first list the contents of the current directory, then enter the editor.

A related feature is more interesting. If a command is followed by "&," the shell will not wait for the command to finish before prompting again; instead, it is ready immediately to accept a new command. For example:

```
as source >output &
```

causes **source** to be assembled, with diagnostic output going to **output**; no matter how long the assembly takes, the shell returns immediately. When the shell does not wait for the completion of a command, the identification number of the process running that command is printed. This identification may be used to wait for the completion of the command or to terminate it. The "**&**" may be used several times in a line:

as source >output & ls >files &

does both the assembly and the listing in the background. In these examples, an output file other than the terminal was provided; if this had not been done, the outputs of the various commands would have been intermingled.

The shell also allows parentheses in the above operations. For example:

(date; ls) >x &

writes the current date and time followed by a list of the current directory onto the file **x**. The shell also returns immediately for another request.

6.4 The shell as a command; command files

The shell is itself a command, and may be called recursively. Suppose file **tryout** contains the lines:

```
as source
mv a.out testprog
testprog
```

The **mv** command causes the file **a.out** to be renamed **testprog. a.out** is the (binary) output of the assembler, ready to be executed. Thus if the three lines above were typed on the keyboard, **source** would be assembled, the resulting program renamed **testprog**, and **testprog** executed. When the lines are in **tryout**, the command:

sh <tryout

would cause the shell **sh** to execute the commands sequentially.

The shell has further capabilities, including the ability to substitute parameters and to construct argument lists from a specified subset of the file names in a directory. It also provides general conditional and looping constructions.

6.5 Implementation of the shell

The outline of the operation of the shell can now be understood. Most of the time, the shell is waiting for the user to type a command. When the newline character ending the line is typed, the shell's **read** call returns. The shell analyzes the command line, putting the arguments in a form appropriate for **execute**. Then **fork** is called. The child process, whose code of course is still that of the shell, attempts to perform an **execute** with the appropriate arguments. If successful, this will bring in and start execution of the program whose name was given. Meanwhile, the other process resulting from the **fork**, which is the parent process, **wait**s for the child process to die. When this happens, the shell knows the command is finished, so it types its prompt and reads the keyboard to obtain another command.

Given this framework, the implementation of background processes is trivial; whenever a command line contains "**&**," the shell merely refrains from waiting for the process that it created to execute the command.

Happily, all of this mechanism meshes very nicely with the notion of standard input and output files. When a process is created by the **fork** primitive, it inherits not only the memory image of its parent but also all the files currently open in its parent, including those with file descriptors 0, 1, and 2. The shell, of course, uses these files to read command lines and to write its prompts and diagnostics, and in the ordinary case its children—the command programs—inherit them automatically. When an argument with "<" or ">" is given, however, the offspring process, just before it performs **execute,** makes the standard I/O file descriptor (0 or 1, respectively) refer to the named file. This is easy because, by agreement, the smallest unused file descriptor is assigned when a new file is **opened** (or **created**); it is only necessary to close file 0 (or 1) and open the named file. Because the process in which the command program runs simply terminates when it is through, the association between a file specified after "<" or ">" and file descriptor 0 or 1 is ended automatically when the process dies. Therefore the shell need not know the actual names of the files that are its own standard input and output, because it need never reopen them.

Filters are straightforward extensions of standard I/O redirection with pipes used instead of files.

In ordinary circumstances, the main loop of the shell never terminates. (The main loop includes the branch of the return from **fork** belonging to the parent process; that is, the branch that does a **wait**, then reads another command line.) The one thing that causes the shell to terminate is discovering an end-of-file condition on its input file. Thus, when the shell is executed as a command with a given input file, as in:

sh <comfile

the commands in **comfile** will be executed until the end of **comfile** is reached; then the instance of the shell invoked by **sh** will terminate. Because this shell process is the child of another instance of the shell, the **wait** executed in the latter will return, and another command may then be processed.

6.6 Initialization

The instances of the shell to which users type commands are themselves children of another process. The last step in the initialization of the system is the creation of a single process and the invocation (via **execute**) of a program called **init**. The role of **init** is to create one process for each terminal channel. The various subinstances of **init** open the appropriate terminals for input and output on files 0, 1, and 2, waiting, if necessary, for carrier to be established on dial-up lines. Then a message is typed out requesting that the user log in. When the user types a name or other identification, the appropriate instance of **init** wakes up, receives the log-in line, and reads a password file. If the user's name is found, and if he is able to supply the correct password, **init** changes to the user's default current directory, sets the process's user ID to that of the person logging in, and performs an **execute** of the shell. At this point, the shell is ready to receive commands and the logging-in protocol is complete.

Meanwhile, the mainstream path of **init** (the parent of all the subinstances of itself that will later become shells) does a **wait**. If one of the child processes terminates, either because a shell found an end of file or because a user typed an incorrect name or password, this path of **init** simply recreates the defunct process, which in turn reopens the appropriate input and output files and types another log-in message. Thus a user may log out simply by typing the end-of-file sequence to the shell.

6.7 Other programs as shell

The shell as described above is designed to allow users full access to the facilities of the system, because it will invoke the execution of any program with appropriate protection mode. Sometimes, however, a different interface to the system is desirable, and this feature is easily arranged for.

Recall that after a user has successfully logged in by supplying a name and password, **init** ordinarily invokes the shell to interpret command lines. The user's entry in the password file may contain the name of a program to be invoked after log-in instead of the shell. This program is free to interpret the user's messages in any way it wishes.

For example, the password file entries for users of a secretarial editing system might specify that the editor **ed** is to be used instead of the shell. Thus when users of the editing system log in, they are inside the editor and can begin work immediately; also, they can be prevented from invoking programs not intended for their use. In practice, it has proved desirable to allow a temporary escape from the editor to execute the formatting program and other utilities.

Several of the games (e.g., chess, blackjack, 3D tic-tac-toe) available on the system illustrate a much more severely restricted environment. For each of these, an entry exists in the password file specifying that the appropriate game-playing program is to be invoked instead of the shell. People who log in as a player of one of these games find themselves limited to the game and unable to investigate the (presumably more interesting) offerings of the UNIX system as a whole.

VII. TRAPS

The PDP-11 hardware detects a number of program faults, such as references to non-existent memory, unimplemented instructions, and odd addresses used where an even address is required. Such faults cause the processor to trap to a system routine. Unless other arrangements have been made, an illegal action causes the system to terminate the process and to write its image on file **core** in the current directory. A debugger can be used to determine the state of the program at the time of the fault.

Programs that are looping, that produce unwanted output, or about which the user has second thoughts may be halted by the use of the **interrupt** signal, which is generated by typing the "delete" character. Unless special action has been taken, this signal simply causes the program to cease execution without producing a **core** file. There is also a **quit** signal used to force an image file to be produced. Thus programs that loop unexpectedly may be halted and the remains inspected without prearrangement.

The hardware-generated faults and the interrupt and quit signals can, by request, be either ignored or caught by a process. For example, the shell ignores quits to prevent a quit from logging the user out. The editor catches interrupts and returns to its command level. This is useful for stopping long printouts without losing work in progress (the editor manipulates a copy of the file it is editing). In systems without floating-point hardware, unimplemented instructions are caught and floating-point instructions are interpreted.

VIII. PERSPECTIVE

Perhaps paradoxically, the success of the UNIX system is largely due to the fact that it was not designed to meet any predefined objectives. The first version was written when one of us (Thompson), dissatisfied with the available computer facilities, discovered a little-used PDP-7 and set out to create a more hospitable environment. This (essentially personal) effort was sufficiently successful to gain the interest of the other author and several colleagues, and later to justify the acquisition of the PDP-11/20, specifically to support a text editing and formatting system. When in turn the 11/20 was outgrown, the system had proved useful enough to persuade management to invest in the PDP-11/45, and later in the PDP-11/70 and Interdata 8/32 machines, upon which it developed to its present form. Our goals throughout the effort, when articulated at all, have always been to build a comfortable relationship with the machine and to explore ideas and inventions in operating systems and other software. We have not been faced with the need to satisfy someone else's requirements, and for this freedom we are grateful.

Three considerations that influenced the design of UNIX are visible in retrospect.

First: because we are programmers, we naturally designed the system to make it easy to write, test, and run programs. The most important expression of our desire for programming convenience was that the system was arranged for interactive use, even though the original version only supported one user. We believe that a properly designed interactive system is much more productive and satisfying to use than a "batch" system. Moreover, such a system is rather easily adaptable to noninteractive use, while the converse is not true.

Second: there have always been fairly severe size constraints on the system and its software. Given the partially antagonistic desires for reasonable efficiency and expressive power, the size constraint has encouraged not only economy, but also a certain elegance of design. This may be a thinly disguised version of the "salvation through suffering" philosophy, but in our case it worked.

Third: nearly from the start, the system was able to, and did, maintain itself. This fact is more important than it might seem. If designers of a system are forced to use that system, they quickly become aware of its functional and superficial deficiencies and are strongly motivated to correct them before it is too late. Because all source programs were always available and easily modified on-line, we were willing to revise and rewrite the system and its software when new ideas were invented, discovered, or suggested by others.

The aspects of UNIX discussed in this paper exhibit clearly at least the first two of these design considerations. The interface to the file system, for example, is extremely convenient from a programming standpoint. The lowest possible interface level is designed to eliminate distinctions between the various devices and files and between direct and sequential access. No large "access method" routines are required to insulate the programmer from the system calls; in fact, all user programs either call the system directly or use a small library program, less than a page long, that buffers a number of characters and reads or writes them all at once.

Another important aspect of programming convenience is that there are no "control blocks" with a complicated structure partially maintained by and depended on by the file system or other system calls. Generally speaking, the contents of a program's address space are the property of the program, and we have tried to avoid placing restrictions on the data structures within that address space.

Given the requirement that all programs should be usable with any file or device as input or output, it is also desirable to push device-dependent considerations into the operating system itself. The only alternatives seem to be to load, with all programs, routines for dealing with each device, which is expensive in space, or to depend on some means of dynamically linking to the routine appropriate to each device when it is actually needed, which is expensive either in overhead or in hardware.

Likewise, the process-control scheme and the command interface have proved both convenient and efficient. Because the shell operates as an ordinary, swappable user program, it consumes no "wired-down" space in the system proper, and it may be made as powerful as desired at little cost. In particular, given the framework in which the shell executes as a process that spawns other processes to perform commands, the notions of I/O redirection, background processes, command files, and user-selectable system interfaces all become essentially trivial to implement.

Influences

The success of UNIX lies not so much in new inventions but rather in the full exploitation of a carefully selected set of fertile ideas, and especially in showing that they can be keys to the implementation of a small yet powerful operating system.

The **fork** operation, essentially as we implemented it, was present in the GENIE time-sharing system.[7] On a number of points we were influenced by Multics, which suggested the particular form of the I/O system calls[8] and both the name of the shell and its general functions. The notion that the shell should create a process for each command was also suggested to us by the early design of Multics, although in that system it was later dropped for efficiency reasons. A similar scheme is used by TENEX.[9]

IX. STATISTICS

The following numbers are presented to suggest the scale of the Research UNIX operation. Those of our users not involved in document preparation tend to use the system for program development, especially language work. There are few important "applications" programs.

Overall, we have today:

125	user population
33	maximum simultaneous users
1,630	directories
28,300	files
301,700	512-byte secondary storage blocks used

There is a "background" process that runs at the lowest possible priority; it is used to soak up any idle CPU time. It has been used to produce a million-digit approximation to the constant e, and other semi-infinite problems. Not counting this background work, we average daily:

13,500	commands
9.6	CPU hours
230	connect hours
62	different users
240	log-ins

X. ACKNOWLEDGMENTS

The contributors to UNIX are, in the traditional but here especially apposite phrase, too numerous to mention. Certainly, collective salutes are due to our colleagues in the Computing Science Research Center. R. H. Canaday contributed much to the basic design of the file system. We are particularly appreciative of the inventiveness, thoughtful criticism, and constant support of R. Morris, M. D. McIlroy, and J. F. Ossanna.

References

1. L. P. Deutsch and B. W. Lampson, "An online editor," *Comm. Assoc. Comp. Mach.*, vol. 10, no. 12, pp. 793-799, 803, December 1967.
2. B. W. Kernighan and L. L. Cherry, "A System for Typesetting Mathematics," *Comm. Assoc. Comp. Mach.*, vol. 18, pp. 151-157, Bell Laboratories, Murray Hill, New Jersey, March 1975.
3. This issue, B. W. Kernighan, M. E. Lesk, and J. F. Ossanna, "UNIX Time-Sharing System: Document Preparation," *Bell Sys. Tech. J.*, vol. 57, no. 6, pp. 2115-2135, 1978.
4. T. A. Dolotta and J. R. Mashey, "An Introduction to the Programmer's Workbench," *Proc. 2nd Int. Conf. on Software Engineering*, pp. 164-168, October 13-15, 1976.
5. B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.
6. Aleph-null, "Computer Recreations," *Software Practice and Experience*, vol. 1, no. 2, pp. 201-204, April-June 1971.
7. L. P. Deutsch and B. W. Lampson, "SDS 930 time-sharing system preliminary reference manual," Doc. 30.10.10, Project GENIE, Univ. Cal. at Berkeley, April 1965.
8. R. J. Feiertag and E. I. Organick, "The Multics input-output system," *Proc. Third Symposium on Operating Systems Principles*, pp. 35-41, October 18-20, 1971.
9. D. G. Bobrow, J. D. Burchfiel, D. L. Murphy, and R. S. Tomlinson, "TENEX, a Paged Time Sharing System for the PDP-10," *Comm. Assoc. Comp. Mach.*, vol. 15, no. 3, pp. 135-143, March 1972.

Bug fixes and changes in 4.2BSD

July 28, 1983

Samuel J. Leffler

Computer Systems Research Group
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, California 94720
(415) 642-7780

ABSTRACT

This document describes briefly the changes in the Berkeley system for the VAX between the 4.1BSD distribution of April 1981 and this, its revision of July 1983. It attempts to summarize, without going into great detail, the changes which have been made.

Notable improvements

- The file system organization has been redesigned to provide at least an order of magnitude improvement in disk bandwidth.
- The system now provides full support for the DOD Standard TCP/IP network communication protocols. This support has been integrated into the system in a manner which allows the development and concurrent use of other communication protocols. Hardware support and routing have been isolated from the protocols to allow sharing between varying network architectures. Software support is provided for 10 different hardware devices including 3 different 10 Mb/s Ethernet modules.
- A new set of interprocess communication facilities has replaced the old multiplexed file mechanism. These new facilities allow unrelated processes to exchange messages in either a connection-oriented or connection-less manner. The interprocess communication facilities have been integrated with the networking facilities (described above) to provide a single user interface which may be used in constructing applications which operate on one or more machines.
- A new signal package which closely models the hardware interrupt facilities found on the VAX replaces the old signals and jobs library of 4.1BSD. The new signal package provides for automatic masking of signals, sophisticated signal stack management, and reliable protection of critical regions.
- File names are now almost arbitrary length (up to 255 characters) and a new file type, symbolic link, has been added. Symbolic links provide a "symbolic referencing" mechanism similar to that found in Multics. They are interpolated during pathname expansion and allow users to create links to files and directories which span file systems.
- The system supports advisory locking on files. Files can have "shared" or "exclusive" locks applied by processes. Multiple processes may apply shared locks, but only one process at any time may have an exclusive lock on a file. Further, when an exclusive lock is present on a file, shared locks are disallowed. Locking requests normally block a process until they can be

completed, or they may be indicated as "non-blocking" in which case an error is returned if the lock can not be immediately obtained.

- The group identifier notion has been extended to a "group set". When users log in to the system they are placed in all their groups. Access control is now done based on the group set rather than just a single group id. This has obviated the need for the newgrp command.
- Per-user, per-filesystem disk quotas are now part of the system. Soft and hard limits may be specified on a per user and per filesystem basis to control the number of files and amount of disk space allocated to a user. Users who exceed a soft limit are warned and if, after three login sessions, their disk usage has not dropped below the soft limit, their soft limit is treated as a hard limit. Utilities exist for the creation, maintenance, and reporting of disk quotas.
- System time is now available in microsecond precision and millisecond accuracy. Users are provided with 3 high-resolution timers which may be set up to automatically reload on expiration. The timers operate in real time, user time, and process virtual time (for profiling). All statistics and times returned to users are now given in a standard format with seconds and microseconds separated. This eliminates program dependence on the line clock frequency.
- A new system call to rename files in the same file system has been added. This call eliminates many of the anomalies which could occur in older versions of the system due to lack of atomicity in removing and renaming files.
- A new system call to truncate files to a specific length has been added. This call improves the performance of the Fortran I/O library.
- Swap space configuration has been improved by allowing multiple swap partition of varying sizes to be interleaved. These partitions are sized at boot time to minimize configuration dependencies.
- The Fortran 77 compiler and associated I/O library have undergone extensive changes to improve reliability and performance. Compilation may, optionally, include optimization phases to improve code density and decrease execution time.
- A new symbolic debugger, dbx, replaces the old symbolic debugger sdb. Dbx works on both C and Fortran 77 programs and allows users to set break points and trace execution by source code line numbers, references to memory locations, procedure entry, etc. Dbx allows users to reference structured and local variables using the program's programming language syntax.
- The delivermail program has been replaced by sendmail. Sendmail provides full internetwork routing, domain style naming as defined in the DARPA Request For Comments document #833, and eliminates the compiled in configuration database previously used by delivermail. Further, sendmail uses the DARPA standard Simple Mail Transfer Protocol (SMTP) for mail delivery.
- The system contains a new line printer system. Multiple line printers and spooling queues are supported through a printer database file. Printers on serial lines, raster printing devices, and laser printers are supported through a series of filter programs which interface to the standard line printer "core programs". A line printer control program, lpc, allows printers and printer queues to be manipulated. Spooling to remote printers is supported in a transparent fashion.
- Cu has been replaced by a new program tip. Tip supports a number of auto-call units and allows destination sites to be specified by name rather than phone number. Tip also supports file transfer to non-UNIX machines and can be used with sites which require half-duplex and/or odd-even parity.
- Uucp now supports many auto-call units other than the DN11. Spooling has been reorganized into multiple directories to cut down on system overhead. Several new utilities and shell scripts exist for use in adminstrating uucp traffic. Operation has been greatly improved by numerous bug fixes.

Bug fixes and changes

Section 1

adb Support has been added for interpreting kernel data structures on a running system and in post mortem crash dumps created by savecore. A **-k** option causes adb to map addresses according to the system and current process page tables. A new command, $p, can be used to switch between process contexts. Many scripts are available for symbolically displaying kernel data structures, searching for a process' context by process ID, etc. A new document, "Using ADB to Debug the UNIX Kernel", supplies hints in the use of adb with system crash dumps.

addbib Is a new utility for creating and extending bibliographic data bases for use with refer.

apply Is a new program which may be used to apply a command to a set of arguments.

ar Has a new key, 'o', for preserving a file's modification time when it is extracted from an archive.

cc Supports the additional symbol information used by dbx. The old symbol information, used by the defunct sdb debugger, is available by specifying the **-go** flag. A new flag, **-pg**, creates executable programs which collect profiling information to be interpreted by the new gprof program. A bug in the C preprocessor, which caused line numbers to be incorrect for macros with formal parameters with embedded newlines has been fixed. The C preprocessor now properly handles hexadecimal constants in "#if" constructs and checks for missing "#endif" statements.

chfn Now works interactively in changing a user's information field in the password file.

chgrp Is now in section 1 and may be executed by anyone. Users other than the super-user may change group ownership of a file they own to any group in their group access list.

cp Now has a **-r** flag to copy recursively down a file system tree.

csh A bug which caused backquoted commands to wedge the terminal when interrupted has been fixed. Job identifiers are now globbed. A bug which caused the "wait" command to uninteruptible in certain cases has been fixed. History may now be saved and restored across terminal sessions with the *savehist* variable. The newgrp command has been deleted due to the new group facilities.

ctags Now handles C **typedefs**.

cu Exists only in the form of a "compatible front-end" to the new tip program.

dbx Is a new symbolic debugger replacing sdb. Dbx handles C and Fortran programs.

delivermail
Has been replaced by the new sendmail program.

df Understands the new file system organization and reports all disk space totals in kilobytes.

du Now reports disk usage in kilobytes and uses the new field in the inode structure which contains the actual number of blocks allocated to a file to increase accuracy of calculations.

dump Has been moved to section 8.

error Has been taught about the error message formats of troff.

eyacc A bug which caused the generated parser to not recognize valid null statements has been fixed.

f77 Has undergone major changes.

The i/o library has been extensively tested and debugged. Sequential files are now opened at the BEGINNING by default; previously they were opened at the end.

Compilation of data statements has been substantially sped up. Significant new optimization is optionally available (this is still a bit buggy and should be used with caution). Even without optimization, however, single precision computations execute much faster.

The new debugger, dbx, has replaced sdb for debugging Fortran programs; sdb is no longer supported.

Files with ".F" suffixes are preprocessed by the C preprocessor. This allows C-style "#include" and "#define" constructs to be used. The compiler has been modified to print error messages with sensible line numbers. Make also understands the ".F" suffix. Note that when using the C preprocessor, the 72 column convention is not followed.

The **-I** option for specifying short integers has been changed to **-i**. The **-I** option is now used to specify directory search paths for "#include" statements. A **-pg** option for creating executable images which collect profiling information for gprof has been added.

fed Is a font editor of dubious value.

file Now understands symbolic links.

find Has a new **-type** value, 'l', for finding symbolic links.

fp Is a new compiler/interpreter for the Functional Programming language. A supporting document is present in Volume 2C of the UNIX Programmer's Manual.

fpr Is a new program for printing Fortran files with embedded Fortran carriage controls.

fsplit Is a new program for splitting a multi-function Fortran file into individual files.

ftp Is a new program which supports the ARPA standard File Transfer Protocol.

gcore Is a new program which creates a core dump of a running process.

gprof Is a new profiling tool which displays execution time for the dynamic call graph of a program. Gprof works on C, Fortran, and Pascal programs compiled with the **-pg** option. Gprof may also be used in creating a call graph profile for the operating system. A supporting document, "gprof: A Call Graph Execution Profiler" is included in Volume 2C of the UNIX Programmer's Manual.

groups Is a new program which displays a user's group access list.

hostid Is a new program which displays the system's unique identifier as returned by the new gethostid system call. The super-user uses this program to set the host identifier at boot time.

hostname Is a new program which displays the system's name as returned by the new gethostname system call. The super-user uses this program to set the host name at boot time.

indent Is a new program for formatting C program source.

install Is a shell script used in installing software.

iostat Now reports kilobytes per second transferred for each disk. This is useful as the unit of information transferred is no longer a constant one kilobytes.

last Now displays the remote host from which a user logged in (when accessing a machine across a network). The pseudo user "ftp" may be specified to find out information about FTP file transfer sessions.

lastcomm Now displays flags for each command indicating if the program dumped core, used PDP-11 mode, executed with a set-user-ID, or was created as the result of a fork (with no following exec).

learn Now has lessons for vi (this is user contributed software which is not part of the standard system).

lint Has a new −C flag for creating lint libraries from C source code. Has improved type checking on static variables.

lisp Has been ported to several 68000 UNIX systems, the relevant code is included in the distribution. A new vector data type and a form of "closure" have been added.

ln Has a new flag, **−s**, for creating symbolic links.

login Has been extensively modified for use with the rlogind and telnetd network servers.

lpq Is totally new, see lpr.

lpr And its related programs are totally new. The line printer system supports multiple printers of many different characteristics. A master data base, /etc/printcap, describes both local printers and printers accessable across a network. A document describing the line printer system is now part of Volume 2C of the UNIX Programmer's Manual.

lprm Is totally new, see lpr.

ls Has been rewritten for the new directory format. It understands symbolic links and uses the new inode field which contains the actual number of blocks allocated to a file when the **−s** flag is supplied. Many rarely used options have been deleted.

m4 A bug which caused m4 to dump core when keywords were undefined then redefined has been fixed.

Mail Now supports mail folders in the style of the Rand MH system. Has been reworked to cooperate with sendmail in understanding the new mail address formats. Allows users to defined message header fields which are not be displayed when a messages is viewed. Many other changes are described in a revised version of the user manual.

make Understands not to unlink directories when interrupted. Understands the new ".F" suffix for Fortran source files which are to be run through the C preprocessor. Has a new predefined macro MFLAGS which contains the flags supplied to make on the command line (useful in creating hierarchies of makefiles).

mkdir Now uses the mkdir system call to run faster.

mt Has a new command, status, which shows the current state of a tape drive.

mv Has been rewritten to use the new rename system call. As a result, multiple directories may now be moved in a single command, the restrictions on having ".." in a pathname are no longer present, and everything runs faster.

net And all related Berknet programs are no longer part of the standard distribution. These programs live on in /usr/src/old for those who can not do without them.

netstat Is a new program which displays network statistics and active connections.

oldcsh No longer exists.

od Has gobs of new formats options.

pagesize Is a new program which prints the system page size for use in constructing portable shell scripts.

passwd Now reliably interlocks with chsh, chfn, and vipw, in guarding against concurrent updates to the password file.

pc/pi **For** loops are now done according to the standard. Files may now be dynamically allocated and disposed. Records and variant records are now aligned to correspond to C structures and unions (this was falsely claimed before). Several obscure bugs involving formal routines have been fixed. Three new library routines support random access file i/o, see /usr/include/pascal for details.

pc **For** loop variables and **with** pointers are now allocated to registers. Separate compilation type checking can now be done without reference to the source file; this permits movement (including distribution) of .o files and creation of libraries. Display entries are saved only when needed (a speed optimization).

pdx Is a new debugger for use with pi. Pdx is invoked automatically by the interpreter if a run-time error is encountered. Future work is planned to extend the new dbx debugger to understand code generated by the Pascal compiler pc.

ps Has been changed to work with the new kernel and is no longer dependent on system page size. All process segment sizes are now shown in kilobytes. Understands that the old "using new signal facilities" bit in the process structure now means "using old 4.1BSD signal facilities".

pwd Now simply calls the *getwd*(3) routine.

rcp Is a new program for copying files across a network. The complete syntax of cp is supported, including recursive directory copying.

refer Has had many bugs fixed in it and the associated -ms macro package support made to work.

reset Now resets all the special characters to the system defaults specified in the include file <sys/ttychars.h>.

rlogin Is a new program for logging in to a machine across a network. Rlogin uses the files /etc/hosts.equiv and .rhosts in the users login directory to allow logins to be performed without a password. Rlogin supports proper handling of ^S/^Q and flushing of output when an interrupt is typed at the terminal. Its '~' escape sequences are reminiscent of the old cu program (as it is based on the same source code).

rmdir Now uses the rmdir system call to run more efficiently and not require root privileges. Unfortunately, this means arguments which end in one or more "/" characters are no longer legal.

roffbib Is a new program for running off bibiliographic databases.

rsh Is a new program which supports remote command execution across a network.

ruptime Is a new program which displays system status information for clusters of machines attached to a local area network.

rwho Is a new program which displays users logged in on clusters of machines attached to a local area network.

script Has been rewritten to use pseudo-terminals. This allows the C shell job control facilities (among other things) to be used while scripting. A side effect of this change is that scripts now contain everything typed at a terminal.

sdb Has been replaced by dbx; it still lives on in /usr/src/old for those with a personal attachment.

sendbug Is a new command for submitting bug reports on 4.2BSD in a standard format suitable for automatic filing by the bugfiler program.

sh No longer has a newgrp command due to the new groups facilities.

sortbib Is a new command for sorting bibliographic databases.

strip Has been made blindingly fast by using the new truncate system call (thereby eliminating the old method of copying the file).

stty The default system erase, kill, and interrupt characters have been made the DEC standard values of DEL ('^?'), '^U', and '^C'. This is not expected to gain much popularity, but was done in the interest of compatibility with many other standard operating systems.

su Has been changed to do a "full login" when starting up the subshell. A new flag, **-f**, does a "fast" su for when a system is heavily loaded. Extra arguments supplied to su are now treated as a command line and executed directly instead of creating an interactive shell.

sysline Is a new program for maintaining system status information on terminals which support a "status line"; a poor man's alternative to a window manager (or emacs).

tail Has a larger buffer so that "tail -r" and similar show more.

talk Is a new program which provides a screen-oriented write facility. Users may be "talked to"across a network, though satellite response times have indicated overseas conversations are still best done by phone. Can be very obnoxious when engaged in important work.

tar Now allocates its internal buffers dynamically so that the block size can be specified to be very large for streaming tape drives. Also, now avoids many core-core copy operations. Has a new **–C** option for forcing chdir operations in the middle of operation (thereby allowing multiple disjoint subtrees to be easily placed in a single file, each with short relative pathnames). Has a new flag, 'B', for forcing 20 block records to be read and written; useful in joining two tar commands with a remote shell to transfer large amounts of data across a network.

telnet Is a new program which supports the ARPA standard Telnet protocol.

tip Replaces cu as the standard mechanism for connecting to machines across a phone line or through a hardwired connection. Tip uses a database of system descriptions, supports many different auto-call units, and understands many nuances required to talk to non-UNIX systems. Files may be transferred to and from non-UNIX systems in a simple fashion.

ul A bug which sometimes caused an extra blank line to be printed after reaching end of file has been fixed.

uucp And related programs have been extensively enhanced to support many different auto-call units and multiple spooling directories (among other things). A large number of bugs and performance enhancements have been made.

uusnap Is a new program which gives a snap-shot of the uucp spooling area.

vfontinfo Is a program used to inspect and print information about fonts.

vgrind Now uses a regular expression language to describe formatting. A **–f** flag forces vgrind to act as a filter, generating output suitable for inclusion in troff and/or nroff documents. Language descriptions exist for C, Pascal, Model, C shell, Bourne shell, Ratfor, and Icon programs.

vi A bug which caused the ^B command to place the cursor on the wrong line has been fixed. A bug which caused vi to believe a file had been modified when an i/o error occurred has been fixed. A bug which allowed "hardtabs" to be set to 0 causing a divide by zero fault has been fixed.

vlp Is a new program for pretty printing Lisp programs.

vmstat Has had one new piece of information added to **–s** summary, the number of fast page reclaims performed. The fields related to paging activity are now all given in kilobytes.

vpr And associated programs for spooling and printing files on Varian and Versatec printers are now shell scripts which use the new line printer support.

vwidth Is a new program for making troff width tables for a font.

wc Is once again identical to the version 7 program. That is, the **–v**, **–t**, **–b**, **–s**, and **–u** flags have been deleted.

whereis Understands the new directory organization for the source code.

which Now understands how to handle aliases.

who Now displays the remote machine from which a user is logged in.

Section 2.

The most important change in section 2 is that the documentation has been significantly improved. Manual page entries now indicate the possible error codes which may be returned and how to interpret them. The introduction to section 2 now includes a glossary of terms used throughout the section. The terminology and formatting have been made consistent. Many manual pages now have "NOTES" or "CAVEATS" providing useful information heretofore left out for the sake of brevity. As always the manual pages are still for the programmer; they are terse and extremely concise. The "4.2BSD System Manual" is likewise concise, but a bit more verbose in providing an overall picture of the system facilities.

With regard to changes in the facilities, these fall into three major categories: interprocess communication, signals, and file system related calls. The interprocess communication facilities center around the *socket* mechanism described in the "A 4.2BSD Interprocess Communication Primer". The new signals do not have an accompanying document, so the manual pages should be studied carefully. The new file system calls pretty much stand on their own, with a late section of the document "A Fast File System for UNIX" supplying a quick overview of the most important new file system facilities. Finally, it should be noted that the job control facilities introduced in 4.1BSD have been adopted as a standard part of 4.2BSD. No special distinction is given to these calls (in 4.1BSD they were earmarked "2J").

Many of the new system calls have both a "set" and a "get" form. Only the "get" forms are indicated below. Consult the manual for details on the "set" form.

intro Has been updated to reflect the new list of possible error codes. Now includes a glossary of terminology used in section 2.

access Now has symbolic definitions for the *mode* parameter defined in *<sys/file.h>*.

bind Is a new interprocess communication system call for binding names to sockets.

connect Is a new interprocess communication system call for establishing a connection between two sockets.

creat Has been obsoleted by the new *open* interface.

fchmod Is a new system call which does a *chmod* operation given a file descriptor; useful in conjunction with the new advisory locking facilities.

fchown Is a new system call which does a *chown* operation given a file descriptor; useful in conjunction with the new advisory locking facilities.

fcntl Is a new system call which is useful in controlling how i/o is performed on a file descriptor (non-blocking i/o, signal drive i/o). This interface is compatible with the System III fcntl interface.

flock Is a new system call for manipulating advisory locks on files. Locks may be shared or exclusive and locking operations may be indicated as being non-blocking, in which case a process is not blocked if the requested lock is currently in use.

fstat Now returns a larger stat buffer; see below under stat.

fsync Is a new system call for synchronizing a file's in-core state with that on disk. Its intended use is in building transaction oriented facilities.

ftruncate Is a new system call which does a *truncate* operation given a file descriptor; useful in conjunction with the new advisory locking facilities.

getdtablesize Is a new system call which returns the size of the descriptor table.

getgroups Is a new system call which returns the group access list for the caller.

gethostid Is a new system call which returns the unique (hopefully) identifier for the current host.

gethostname
Is a new system call which returns the name of the current host.

getitimer Is a new system call which gets the current value for an interval timer.

getpagesize Is a new system call which returns the system page size.

getpriority Is a new system call which returns the current scheduling priority for a specific process, a group of processes, or all processes owned by a user. In the latter two cases, the priority returned is the highest (lowest numerical value) enjoyed by any of the specified processes.

getrlimit Is a new system call which returns information about a resource limit. The getrlimit and setrlimit calls replace the old vlimit call from 4.1BSD.

getrusage Is a new system call which returns information about resource utilitization of a child process or the caller. This call replaces the vtimes call of 4.1BSD.

getsockopt Is a new interprocess communication system call which returns the current options present on a socket.

gettimeofday
Is a new system call which returns the current Greenwich date and time, and the current timezone in which the machine is operating. Time is returned in seconds and microseconds since January 1, 1970.

ioctl Has been changed to encode the size of parameters and whether they are to be copied in, out, or in and out of the user address space in the *request*. The symbolic names for the various ioctl requests remain the same, only the numeric values have changed. A number of new ioctls exist for use with sockets and the network facilities. The old LINTRUP request has been replaced by a call to fcntl and the SIGIO signal.

killpg Has now been made a system call; in 4.1BSD it was a library routine.

listen Is a new interprocess communication system call used to indicate a socket will be used to listen for incoming connection requests.

lseek Now has symbolic definitions for its *whence* parameter defined in *<sys/file.h>*.

mkdir Is a new system call which creates a directory.

mpx The multiplexed file facilities are no longer part of the system. They have been replaced by the socket, and related, system calls.

open Is different, now taking an optional third parameter and supporting file creation, automatic truncation, automatic append on write, and "exclusive" opens. The open interface has been made compatible with System III with the exception that non-blocking opens on terminal lines requiring carrier are not supported.

profil Now returns statistical information based on a 100 hz clock rate.

quota Is a new system call which is part of the disk quota facilities. Quota is used to manipulate disk quotas for a specific user, as well as perform certain random chores such as syncing quotas to disk.

read Now automatically restarts when a read on a terminal is interrupted by a signal before any data is read.

readv Is a new system call which supports scattering of read data into (possibly) disjoint areas of memory.

readlink Is a new system call for reading the value of a symbolic link.

recv Is a new interprocess communication system call used to receive a message on a connected socket.

recvfrom Is a new interprocess communication system call used to receive a message on a (possibly) unconnected socket.

recvmsg Is a new interprocess communication system call used to receive a message on a (possibly) unconnected socket which may have access rights included. When using on-machine communication, recvmsg and sendmsg may be used to pass file descriptors between processes.

rename Is a new system call which changes the name of an entry in the file system (plain file, directory, character special file, etc.). Rename has an important property in that it guarantees the target will always exist, even if the system crashes in the middle of the operation. Rename only works with source and destination in the same file system.

rmdir Is a new system call for removing a directory.

select Is a new system call (mainly for interprocess communication) which provides facility for synchronous i/o multiplexing. Sets of file descriptors may be queried for readability, writability, and if any exceptional conditions are present (such as out of band data on a socket). An optional timeout may also be supplied in which case the select operation will return after a specified period of time should no descriptor satisfy the requests.

send Is a new interprocess communication system call for sending a message on a connected socket.

sendto Is a new interprocess communication system call for sending a message on a (possibly) unconnected socket.

sendmsg Is a new interprocess communication system call for sending a message on a (possibly) unconnected socket which may included access rights.

setquota Is a new system call for enabling or disabling disk quotas on a file system.

setregid Is a new system call which replaces the 4.1BSD setgid system call. Setregid allows the real and effective group ID's of a process to be set separately.

setreuid Is a new system call which replaces the 4.1BSD setuid system call. Setreuid allows the real and effective user ID's of a process to be set separately.

shutdown Is a new interprocess communication system call for shutting down part or all of full-duplex connection.

sigblock Is a new system call for blocking signals during a critical section of code.

sigpause Is a new system call for blocking a set of signals and then pausing indefinitely for a signal to arrive.

sigsetmask Is a new system call for setting the set of signals which are currently blocked from delivery to a process.

sigstack Is a new system call for defining an alternate stack on which signals are to be processed.

sigsys Is no longer supported. The new signal facilities are a superset of those which sigsys provided.

sigvec Is the new system call interface for defining signal actions. For each signal (except SIGSTOP and SIGKILL), sigvec allows a "signal vector" to be defined. The signal vector is comprised of a handler, a mask of signals to be blocked while the handler executes, and an indication of whether or not the handler should execute on a signal stack defined by a sigstack call. The old signal interface is provided as a library routine with several important caveats. First, signal actions are no longer reset to their default value after a signal is delivered to a process. Second, while a signal handler is executing the signal which is being processed is blocked until the handler returns. To simulate the old signal interface, the user must explicitly reset the signal action to be the default value and unblock the signal being processed.

Four new signals have been added for the interprocess communication and interval timer facilities. SIGIO is delivered to a process when an fcntl call enables signal driven i/o and input is present on a terminal (and a signal handler is defined).

SIGURG is delivered when an urgent condition arises on a socket (and a handler is defined). SIGPROF and SIGVTALRM are associated with the ITIMER_PROF and ITIMER_VIRTUAL interval timers, and delivered to a process when such a timer expires (the SIGALRM signal is used for the ITIMER_REAL interval timer). The old SIGTINT signal is replaced by SIGIO.

socket Is a new interprocess communication system call for creating a socket.

socketpair Is a new interprocess communication system call for creating a pair of connected and unnamed sockets.

stat Now returns a larger structure. New fields are present indicating the optimal blocking factor in which i/o should be performed (for disk files the block size of the underlying file system) and the actual number of disk blocks allocated to the file. Inode numbers are now 32-bit quantities. Several spare fields have been allocated for future expansion. These include space for 64-bit file sizes and 64-bit time stamps. Two new file types may be returned, S_IFLNK for symbolic links, and S_IFSOCK for sockets residing in the file system.

swapon Has been renamed from the vswapon call of 4.1BSD.

symlink Is a new system call for creating a symbolic link.

truncate Is a new system call for truncating a file to a specific size.

unlink Should no longer be used for removing directories. Directories should only be created with mkdir and removed with rmdir. Creating hard links to directories can cause disastrous results.

utime Is defunct, replaced by utimes.

utimes Is a new system call which uses the new time format in setting the accessed and updated times on a file.

vfork Is still present, but definitely on its way out. Future plans include implementing fork with a scheme in which pages are initially shared read-only. On the first attempt by a process to write on a page the parent and child would receive separate writable copies of the page.

vlimit Is no longer supported. Vlimit is replaced by the getrlimit and setrlimit calls.

vread Is no longer supported in the system.

vswapon Has been renamed swapon.

vtimes Is no longer supported. Vtimes is replaced by the getrusage call.

vwrite Is no longer supported in the system.

wait Now is automatically restarted when interrupted by a signal before status could be returned.

wait3 Returns resource usage in a different format than that which was returned in 4.1BSD. This structure is compatible with the new getrusage system call. Wait3 is now automatically restarted when interrupted by a signal before status could be returned.

write Now is automatically restarted when writing on a terminal and interrupted by a signal before any i/o was completed.

writev Is a new version of the write system call which supports gathering of data in (possibly) discontiguous areas of memory

Section 3

The section 3 documentation has been reorganized to group manual entries by library. Introductory sections for each logical and physical library contain lists of the entry points in the library.

A number of routines which had been system calls under 4.1BSD are now user-level library routines in 4.2BSD. These routines have been grouped under section "3C" headings, "C" for compatibility. Further, certain routines present in the standard C run-time library which do not easily categorize as part of one of the standard libraries, have been group under "3X" headings.

curses A number of bug fixes have been incorporated, and the documentation has been revised.

stdio The standard i/o library has been modified to block i/o operations to disk files according to the block size of the underlying file system. This is accomplished using the new *st_blksize* value returned by *fstat*. The resultant performance improvement is significant as the old 1 kilobyte buffer size often resulted in 7 memory-to-memory copy operations by the system on 8 kilobyte block file systems.

End-of-file marks now "stick". That is, all input requests on a stdio channel after encountering end-of-file will return end-of-file until a *clearerr* call is made. This has implications for programs which use stdio to read from a terminal and do not process end-of-file as a terminating keystroke.

Two new functions may be used to control i/o buffering. The *setlinebuf* routine is used to change **stdout** or **stderr** from block buffered or unbuffered to line buffered. The *setbuffer* routine is an alternate form of *setbuf* which can be used after a stream has been opened, but before it is read or written.

bstring Three new routines, *bcmp*, *bcopy*, and *bzero* have been added to the library. These routines use the VAX string instructions to manipulate binary byte strings of a known size.

ctime Now uses the *gettimeofday* system call and supports time conversion in six different time zones. Daylight savings calculations are also performed in each time zone when appropriate.

isprint Now considers space a printing character; as the manual page has always indicated.

directory Is a new directory interface package which provides a portable interface to reading directories. A version of this library which operates under 4.1BSD is also available.

getpass Now properly handles being unable to open /dev/tty.

getwd Has been moved from the old jobs library to the standard C run-time library. It now returns an error string rather than printing on the standard error when unable to decipher the current working directory.

perror Now uses the writev system call to pass all its arguments to the system in a single system call. This has profound effects on programs which transmit error messages across a network.

psignal And sys_siglist are routines for printing signal names in an equivalent manner to perror.

qsort Has been greatly sped up by choosing a random element with which to apply its divide and conquer algorithm.

random Is a successor to rand which generates much better random numbers. The old rand routine is still available and most programs have not been switched over to random as doing so would make certain facilities such encrypted mail unable to operate on existing data files.

setjmp And longjmp now save and restore the signal mask so that non-local exit from a signal handler is transparent. The old semantics are available with _setjmp and _longjmp.

net Is a new set of routines for accessing database files for the DARPA Internet. Four databases exist: one for host names, one for network names, one for protocol numbers, and one for network services. The latter returns an Internet port and protocol to be used in accessing a given network service.

An additional collection of routines, all prefaced with "inet_" may be used to manipulate Internet addresses, and interpret and convert between Internet addresses and ASCII representations in the Internet standard "dot" notation.

Finally, routines are available for converting 16 and 32 bit quantities between host and network order (on high-ender machines these routines are defined to be noops).

fstab The routines for manipulating /etc/fstab have been rewritten to return arbitrary length null-terminated strings.

Section 4

The system now supports the 11/730, the new 64Kbit RAM memory controllers for the 11/750 and 11/780, and the second UNIBUS adapter for the 11/750. Several new character and/or block device drivers have been added, as well as support for many hardware devices which are accessible only through the network facilities. Each new piece of hardware supported is listed below.

New manual entries in section 4 have been created to describe communications protocols, and network architectures supported. At present the only network architecture fully supported is the DARPA Internet with the TCP, IP, UDP, and ICMP protocols.

acc A network driver for the ACC LH/DH IMP interface.

ad A driver for the Data Translation A/D converter.

arp The Address Resolution Protocol for dynamically mapping betwee 32-bit DARPA Internet addresses and 48-bit Xerox 10Mb/s Ethernet addresses.

css A network driver for the DEC IMP-11A LH/DH IMP interface.

dmc A network interface driver for the DEC DMC-11/DMR-11 point-to-point communications device.

ec A network interface driver for the 3Com 10Mb/s Ethernet controller.

en A network interface driver for the Xerox 3Mb/s experimental Ethernet controller.

hy A network interface driver for the Network Systems Hyperchannel Adapter.

ik A driver for an Ikonas frame buffer graphics device interface.

il A network interface driver for the Interlan 10Mb/s Ethernet interface.

imp A network interface driver for the standard 1822 interface to an IMP; used in conjunction with either acc or css hardware.

kg A driver for a KL-11/DL-11W used as an alternate real time clock source for gathering kernel statistics and profiling information.

lo A software loopback network interface for protocol testing and performance analysis.

pcl A network interface driver for the DEC PCL-11B communications controller.

ps A driver for an Evans and Sutherland Picture System 2 graphics device connected with a DMA interface.

pty Now includes a simple packet protocol to support flow controlled operation with mechanisms for flushing data to be read and/or written.

rx A driver for the DEC dual RX02 floppy disk unit.

ts Now supports TU80 tape drives.

tu The VAX-11/750 console cassette interface has been made somewhat usable when operating in single-user mode. The device driver employs assembly language pseudo-dma code for the reception of incoming packets from the cassette.

uda Now supports RA81, RA80, and RA60 disk drives.

un A network interface driver for an Ungermann-Bass network interface unit connected to the host via a DR-11W.

up Now supports ECC correction and bad sector handling. Also has improved logic for recognizing many different kinds of disk drives automatically at boot time.

uu A driver for DEC dual TU58 tape cartridges connected via a DL-11W interface.

va The Varian driver has been rewritten so that it may coexist on the same UNIBUS with devices which require exclusive use of the bus; i.e. RK07's.

vv A network interface driver for the Proteon proNET 10Mb/s ring network controller.

Section 5

dir Reflects the new directory format.

disktab Is a new file for maintaining disk geometry information. This is a temporary scheme until the information stored in this file for each disk is recorded on the disk pack itself.

dump Is a superset of that used in 4.1BSD.

fs Reflects the new file system organization.

gettytab Is a new file which idescribes terminal characteristics. Each entry in the file describes one of the possible arguments to the getty program.

hosts Is a database for mapping between host names and DARPA Internet host addresses.

mtab Has been modified to include a "type" field indicating whether the file system is mounted read-only, read-write, or read-write with disk quotas enabled.

networks Is a database for mapping between network names and DARPA standard network numbers.

phones Is a phone number data base for tip.

printcap Is a termcap clone for configuring printers.

protocols Is a database for mapping between protocol names and DARPA Internetwork standard protocol numbers.

remote Is a database of remote hosts for use with tip.

services Is a database in which DARPA Internet services are recorded. The information contained in this file indicates the name of the service, the protocol which is required to access it, and the port number at which a client should connect to utilize the service.

tar Is a new entry describing the format of a tar tape.

utmp Has been augmented to include a remote host from which a login session originates. The wtmp file is also used to record FTP sessions.

vgrindefs Is a file describing how to vgrind programs written in many languages.

Section 6

aardvark Does not work because it requires the "Dungeon Definition Language" processor which is a binary image requiring 4.1BSD compatibility mode; the DDL source is still present.

aliens The aliens have returned home, the game is no longer included in the distribution.

backgammon
Is now screen oriented. A new program, teachgammon, instructs the new backgammon player. The old version is now called btlgammon.

canfield Is a new game which plays a brand of the popular game of solitaire. Betting is included, the program cfscores may be used to find out your current debt.

ching Now pipes its output through more. Thus the hacker placates the seekers.

chase No longer exists because the binary does not work under 4.2BSD.

factor Is a rewrite in C of the old version 7 assembly language program which finds the prime factors of a number.

fortune Has yet more adages.

hangman Is now screen oriented.

mille Now plays more intelligently.

primes Is a rewrite in C of the old version 7 assembly language program which finds prime numbers within a specified range.

rogue Has been made more of a scoundrel. The supplementary document "A Guide to the Dungeons of Doom", has been updated as well, and is now part of Volume 2C of the programmer's manual.

sail Is a new game which simulates sea battles of yore. The manual page is large enough to be a separate document and so has been left in its source directory.

trek The original trek has returned; trekies rejoice.

Section 7

hier Has been updated to reflect the reorganization to the user and system source.

mailaddr Is a new entry describing mail addressing syntax under sendmail (possibly too Berkeley specific).

ms The –ms macros have been extended to allow automatic creation of a table of contents. Support for the refer preprocessor is improved. Several bugs related to multicolumn output and floating keeps have been fixed. Extensions to the accent mark string set are available by including the .AM macro. Footnotes can now be automatically numbered (in superscript) by –ms and referenced in the text with a ** string register. The manual page includes a summary of important number and string registers. A new document "Changes to –ms" is included in Volume 2C of the programmer's manual.

Section 8

Major changes affecting system operations include:

- The system now supports disk quotas. These allow system administrators to control users' disk space and file allocation on a per-file system basis. Utilities in this section exist for fixing, summarizing, and editing disk quota summary files.
- File systems are now made with a new program, newfs, which acts as front end to the old mkfs program. There no longer is a need to remember disk partition sizes, as newfs gets this information automatically from the /etc/disktab file. In addition, newfs attempts to lay out file systems according to the characteristics of the underlying disk drive (taking into account disk geometry information).
- DEC standard bad block forwarding is now supported on the RP06 and second vendor UNIBUS storage module disks. The bad144 program can now be used to mark sectors bad on many disks, though inclusion in the bad sector table is still somewhat risky due to requirements in the ordering of entries in the table.
- A new program, format, should be used to initialize all non-DEC storage modules before creating file systems. Format formats the sector headers and creates a bad sector table which is used in normal system operation. Format runs in a standalone mode.

- Getty has been rewritten to use a description file, /etc/gettytab. This allows sites to tailor terminal operation and configuration without making modifications to getty.
- The line printer system is totally new. A program to administer the operation of printers, lpc, is supplied, and printer accounting has been consolidated into a single program, pac.
- The program used to restore files from dump tapes is now called *restore*. This name change was done to reinforce the fact that it is completely rewritten and operates in a very different way than the old restor program. Restore operates on mounted file systems and uses only normal file system operations to restore files. Versions of both dump and restore which operate across a network are included as rdump and rrestore. Dump and restore (and their network oriented counterparts) now perform so efficiently (mostly because of the new file system), that disk to disk backups should no longer be an attractive alternative.

arff No longer asks if you want to clobber the floppy when manipulating archives which are not on the floppy.

bad144 Has been modified to use the /etc/disktab file. Can be used to create bad sector tables for the DEC RP06 and several new Winchester disk drives. Consult the source code for details and use with extreme care.

badsect Has been modified to work with the new file system and now must interact with fsck to perform its duties. Consult the manual page for more information.

bugfiler Is a new program for automatic filing and acknowledgement of bug reports submitted by the sendbug program. Intended to operate with the Rand MH software which is part of the user contributed software. Used at Berkeley to process bug reports on 4.2BSD.

chgrp Has been moved to section 1.

comsat Has been changed to filter the noise lines in message headers when displaying incoming mail. No longer uses a second process watchdog as it uses the more reliable socket facilities instead of the old mpx facilities.

config Has been extensively modified to handle the new root and swap device specification syntax. A new document, "Configuring 4.2BSD UNIX Systems with Config", describes its use, as well as other important information needed in configuring system images; this is part of Volume 2C of the programmer's manual.

diskpart Is a new program which may be used to generate disk partition tables according to the rules used at Berkeley. Can automatically generate entries required for device drivers and for the /etc/diskpart file. (Does not handle the new DEC DSA style drives properly because it tries to reserve space for the bad sector table.)

drtest Is a new standalone program which is useful in testing standalone disk device drivers and for pinpointing bad sectors on a disk.

dump Has been modified for the new file system organization. Mainly due to the new file system, it runs virtually at tape speed. Properly handles locking on the dumpdates file when multiple dumps are performed concurrently on the same machine.

dumpfs Is a new program for dumping out information about a file system such as the block size and disk layout information.

edquota Is a new program for editing user quotas. Operates by invoking your favorite editor on an ASCII representation of the information stored in the binary quota files. Edquota also has a "replication" mode whereby a quota template may be used to create quotas for a group of users.

fastboot Is a new shell script which reboots the system without checking the file systems; should be used with extreme care.

fasthalt Is a new script which is similar to fastboot.

format Is a new standalone program for formatting non-DEC storage modules and creating the appropriate bad sector table on the disk.

fsck Has been changed for the new file system. Fsck is more paranoid then ever in checking the disks, and has been sped up significantly. The accompanying Volume 2C document has been updated to reflect the new file system organization.

ftpd Is the DARPA File Transfer Protocol server program. It supports C shell style globbing of arguments and a large set of the commands in the specification (except the ABORT command!).

gettable Is a new program which can be used in aquiring up to date DARPA Internet host database files.

getty Has been rewritten to use a terminal description database, /etc/gettytab. Consult the manual entries for *getty*(8) and *gettytab*(5) for more information.

icheck Has been modified for the new file system.

init Has been significantly modified to use the new signal facilities. In doing so, several race conditions related to signal delivery have been fixed.

kgmon Is a new program for controlling running systems which have been created with kernel profiling. Using kgmon, profiling can be turned on or off and internal profiling buffers can dumped into a gmon.out file suiitable for interpretation by gprof.

lpc Is a new program controlling line printers and their associated spooling queues. Lpc can be used to enable and disable printers and/or their spooling queues. Lpc can also be used to rearrange existing jobs in a queue.

lpd Has been rewritten and now runs as a "server", using the interprocess communication facilities to service print requests. A supplementary document describing the line printer system is now part of Volume 2C of the programmer's manual.

MAKEDEV
Is a new shell script which resides in /dev and is used to create special files there. MAKEDEV keeps commands for creating and manipulating local devices in a separate file MAKEDEV.local.

mkfs Has been virtually rewritten for the new file system. The arguments supplied are very different. For the most part, users now use the newfs program when creating file systems. Mkfs now automatically creates the lost+found directory.

mount Now indicates file systems which are mounted read-only or have disk quotas enabled.

newfs Is a new front-end to the mkfs program. Newfs figures out the appropriate parameters to supply to mkfs, invokes it, and then, if necessary, installs the boot blocks necessary to bootstrap UNIX on 11/750's.

pac Is a new program which can be used to do printer accounting on any printer. It subsumes the vpac program.

quot Now uses the information in the inode of each file to find out how many blocks are allocated to it.

quotacheck
Is a new program which performs consistency checks on disk quota files. Quotacheck is normally run from the /etc/rc.local file after a system is rebooted, though it can also be run on mounted on file systems which are not in use.

quotaon Is a new program which enables disk quotas on file systems. A link to quotaon, named quotaoff, is used to disable disk quotas on file systems.

pstat Has been modified to understand new kernel data structures.

rc Has had system dependent startup commands moved to /etc/rc.local.

rdump Is a new program to dump file systems across a network.

renice Has been rewritten to use the new setpriority system call. As a result, you can now renice users and process groups.

repquota Is a new program which summarizes disk quotas on one or more file systems.

restor No longer exists. A new program, restore, is its successor.

restore Replaces restor. Restore operates on mounted file systems; it contains an interactive mode and can be used to restore files by name. Restore has become almost as flexible to use as tar in retrieving files from tape.

rexecd Is a network server for the *rexec*(3X) library routine. Supports remote command execution where authentication is performed using user accounts and passwords.

rlogind Is a network server for the *rlogin*(1C) command. Supports remote login sessions where authentication is performed using privileged port numbers and two files, /etc/hosts.equiv and .rhosts (in each users home directory).

rmt Is a program used by rrestore and rdump for doing remote tape operations.

route Is a program for manually manipulating network routing tables.

routed Is a routing daemon which uses a variant of the Xerox Routing Information Protocol to automatically maintain up to date routing tables.

rrestore Is a version of restore which works across a network.

rshd Is a server for the *rsh*(1C) command. It supports remote command execution using privileged port numbers and the /etc/hosts.equiv and .rhosts files in users' home directories.

rwhod Is a server which generates and listens for host status information on local networks. The information stored by rwhod is used by the *rwho*(1C) and*ruptime*(1C) programs.

rxformat Is a program for formatting floppy disks (this uses the *rx* device driver, not the console floppy interface).

savecore Has been modified to get many pieces of information from the running system and crash dump to avoid compiled in constants.

sendmail Is a new program replacing delivermail; it provides fully internetwork mail forwarding capabilities. Sendmail uses the DARPA standard SMTP protocol to send and receive mail. Sendmail uses a configuration file to control its operation, eliminating the compiled in description used in delivermail.

setifaddr Is a new program used to set a network interface's address. Calls to this program are normally placed in the /etc/rc.local file to configure the network hardware present on a machine.

syslog Is a server which receives system logging messages. Currently, only the sendmail program uses this server.

telnetd Is a server for the DARPA standard TELNET protocol.

tftpd Is a server for the DARPA Trivial File Transfer Protocol.

trpt Is a program used in debugging TCP. Trpt transliterates protocol trace information recorded by TCP in a circular buffer in kernel memory.

tunefs Is a program for modifying certain parameters in the super block of file systems.

vipw Is no longer a shell script and properly interacts with passwd, chsh, and chfn in locking the password file.

UNIX/32V — Summary

March 9, 1979

A. What's new: highlights of the UNIX†/32V System

32-bit world. UNIX/32V handles 32-bit addresses and 32-bit data. Devices are addressable to 2^{31} bytes, files to 2^{30} bytes.

Portability. Code of the operating system and most utilities has been extensively revised to minimize its dependence on particular hardware. UNIX/32V is highly compatible with UNIX version 7.

Fortran 77. F77 compiler for the new standard language is compatible with C at the object level. A Fortran structurer, STRUCT, converts old, ugly Fortran into RATFOR, a structured dialect usable with F77.

Shell. Completely new SH program supports string variables, trap handling, structured programming, user profiles, settable search path, multilevel file name generation, etc.

Document preparation. TROFF phototypesetter utility is standard. NROFF (for terminals) is now highly compatible with TROFF. MS macro package provides canned commands for many common formatting and layout situations. TBL provides an easy to learn language for preparing complicated tabular material. REFER fills in bibliographic citations from a data base.

UNIX-to-UNIX file copy. UUCP performs spooled file transfers between any two machines.

Data processing. SED stream editor does multiple editing functions in parallel on a data stream of indefinite length. AWK report generator does free-field pattern selection and arithmetic operations.

Program development. MAKE controls re-creation of complicated software, arranging for minimal recompilation.

Debugging. ADB does postmortem and breakpoint debugging.

C language. The language now supports definable data types, generalized initialization, block structure, long integers, unions, explicit type conversions. The LINT verifier does strong type checking and detection of probable errors and portability problems even across separately compiled functions.

Lexical analyzer generator. LEX converts specification of regular expressions and semantic actions into a recognizing subroutine. Analogous to YACC.

Graphics. Simple graph-drawing utility, graphic subroutines, and generalized plotting filters adapted to various devices are now standard.

Standard input-output package. Highly efficient buffered stream I/O is integrated with formatted input and output.

Other. The operating system and utilities have been enhanced and freed of restrictions in many other ways too numerous to relate.

† UNIX is a Trademark of Bell Laboratories.

B. Hardware

The UNIX/32V operating system runs on a DEC VAX-11/780* with at least the following equipment:

memory: 256K bytes or more.

disk: RP06, RM03, or equivalent.

tape: any 9-track MASSBUS-compatible tape drive.

The following equipment is strongly recommended:

communications controller such as DZ11 or DL11.

full duplex 96-character ASCII terminals.

extra disk for system backup.

The system is normally distributed on 9-track tape. The minimum memory and disk space specified is enough to run and maintain UNIX/32V, and to keep all source on line. More memory will be needed to handle a large number of users, big data bases, diversified complements of devices, or large programs. The resident code occupies 40-55K bytes depending on configuration; system data also occupies 30-55K bytes.

C. Software

Most of the programs available as UNIX/32V commands are listed. Source code and printed manuals are distributed for all of the listed software except games. Almost all of the code is written in C. Commands are self-contained and do not require extra setup information, unless specifically noted as "interactive." Interactive programs can be made to run from a prepared script simply by redirecting input. Most programs intended for interactive use (e.g., the editor) allow for an escape to command level (the Shell). Most file processing commands can also go from standard input to standard output ("filters"). The piping facility of the Shell may be used to connect such filters directly to the input or output of other programs.

1. Basic Software

This includes the time-sharing operating system with utilities, and a compiler for the programming language C—enough software to write and run new applications and to maintain or modify UNIX/32V itself.

1.1. Operating System

□ UNIX — The basic resident code on which everything else depends. Supports the system calls, and maintains the file system. A general description of UNIX design philosophy and system facilities appeared in the Communications of the ACM, July, 1974. A more extensive survey is in the Bell System Technical Journal for July-August 1978. Capabilities include:

- ○Reentrant code for user processes.
- ○"Group" access permissions for cooperative projects, with overlapping memberships.
- ○Alarm-clock timeouts.
- ○Timer-interrupt sampling and interprocess monitoring for debugging and measurement.
- ○Multiplexed I/O for machine-to-machine communication.

□ DEVICES — All I/O is logically synchronous. I/O devices are simply files in the file system. Normally, invisible buffering makes all physical record structure and device characteristics transparent and exploits the hardware's ability to do overlapped I/O. Unbuffered physical record I/O is available for unusual applications. Drivers

*VAX is a Trademark of Digital Equipment Corporation.

for these devices are available:
○Asynchronous interfaces: DZ11, DL11. Support for most common ASCII terminals.
○Automatic calling unit interface: DN11.
○Printer/plotter: Versatek.
○Magnetic tape: TE16.
○Pack type disk: RP06, RM03; minimum-latency seek scheduling.
○Physical memory of VAX-11, or mapped memory in resident system.
○Null device.
○Recipies are supplied to aid the construction of drivers for:
Asynchronous interface: DH11.
Synchronous interface: DU11.
DECtape: TC11.
Fixed head disk: RS11, RS03 and RS04.
Cartridge-type disk: RK05.
Phototypesetter: Graphic Systems System/1 through DR11C.

□ BOOT Procedures to get UNIX/32V started.

1.2. User Access Control

□ LOGIN Sign on as a new user.
○Verify password and establish user's individual and group (project) identity.
○Adapt to characteristics of terminal.
○Establish working directory.
○Announce presence of mail (from MAIL).
○Publish message of the day.
○Execute user-specified profile.
○Start command interpreter or other initial program.

□ PASSWD Change a password.
○User can change his own password.
○Passwords are kept encrypted for security.

□ NEWGRP Change working group (project). Protects against unauthorized changes to projects.

1.3. Terminal Handling

□ TABS Set tab stops appropriately for specified terminal type.

□ STTY Set up options for optimal control of a terminal. In so far as they are deducible from the input, these options are set automatically by LOGIN.
○Half vs. full duplex.
○Carriage return+line feed vs. newline.
○Interpretation of tabs.
○Parity.
○Mapping of upper case to lower.
○Raw vs. edited input.
○Delays for tabs, newlines and carriage returns.

1.4. File Manipulation

□ CAT Concatenate one or more files onto standard output. Particularly used for unadorned printing, for inserting data into a pipeline, and for buffering output that comes in dribs and drabs. Works on any file regardless of contents.

□ CP	Copy one file to another, or a set of files to a directory. Works on any file regardless of contents.
□ PR	Print files with title, date, and page number on every page. ○Multicolumn output. ○Parallel column merge of several files.
□ LPR	Off-line print. Spools arbitrary files to the line printer.
□ CMP	Compare two files and report if different.
□ TAIL	Print last *n* lines of input ○May print last *n* characters, or from *n* lines or characters to end.
□ SPLIT	Split a large file into more manageable pieces. Occasionally necessary for editing (ED).
□ DD	Physical file format translator, for exchanging data with foreign systems, especially IBM 370's.
□ SUM	Sum the words of a file.

1.5. Manipulation of Directories and File Names

□ RM	Remove a file. Only the name goes away if any other names are linked to the file. ○Step through a directory deleting files interactively. ○Delete entire directory hierarchies.
□ LN	"Link" another name (alias) to an existing file.
□ MV	Move a file or files. Used for renaming files.
□ CHMOD	Change permissions on one or more files. Executable by files' owner.
□ CHOWN	Change owner of one or more files.
□ CHGRP	Change group (project) to which a file belongs.
□ MKDIR	Make a new directory.
□ RMDIR	Remove a directory.
□ CD	Change working directory.
□ FIND	Prowl the directory hierarchy finding every file that meets specified criteria. ○Criteria include: name matches a given pattern, creation date in given range, date of last use in given range, given permissions, given owner, given special file characteristics, boolean combinations of above. ○Any directory may be considered to be the root. ○Perform specified command on each file found.

1.6. Running of Programs

□ SH	The Shell, or command language interpreter. ○Supply arguments to and run any executable program. ○Redirect standard input, standard output, and standard error files.

○Pipes: simultaneous execution with output of one process connected to the input of another.
○Compose compound commands using:
    if ... then ... else,
    case switches,
    while loops,
    for loops over lists,
    break, continue and exit,
    parentheses for grouping.
○Initiate background processes.
○Perform Shell programs, i.e., command scripts with substitutable arguments.
○Construct argument lists from all file names satisfying specified patterns.
○Take special action on traps and interrupts.
○User-settable search path for finding commands.
○Executes user-settable profile upon login.
○Optionally announces presence of mail as it arrives.
○Provides variables and parameters with default setting.

□ TEST — Tests for use in Shell conditionals.
○String comparison.
○File nature and accessibility.
○Boolean combinations of the above.

□ EXPR — String computations for calculating command arguments.
○Integer arithmetic
○Pattern matching

□ WAIT — Wait for termination of asynchronously running processes.

□ READ — Read a line from terminal, for interactive Shell procedure.

□ ECHO — Print remainder of command line. Useful for diagnostics or prompts in Shell programs, or for inserting data into a pipeline.

□ SLEEP — Suspend execution for a specified time.

□ NOHUP — Run a command immune to hanging up the terminal.

□ NICE — Run a command in low (or high) priority.

□ KILL — Terminate named processes.

□ CRON — Schedule regular actions at specified times.
○Actions are arbitrary programs.
○Times are conjunctions of month, day of month, day of week, hour and minute. Ranges are specifiable for each.

□ AT — Schedule a one-shot action for an arbitrary time.

□ TEE — Pass data between processes and divert a copy into one or more files.

1.7. Status Inquiries

□ LS — List the names of one, several, or all files in one or more directories.
○Alphabetic or temporal sorting, up or down.
○Optional information: size, owner, group, date last modified, date last accessed, permissions, i-node number.

□ FILE — Try to determine what kind of information is in a file by consulting the file system index and by reading the file itself.

- □ DATE — Print today's date and time. Has considerable knowledge of calendric and horological peculiarities.
  ○May set UNIX/32V's idea of date and time.
- □ DF — Report amount of free space on file system devices.
- □ DU — Print a summary of total space occupied by all files in a hierarchy.
- □ QUOT — Print summary of file space usage by user id.
- □ WHO — Tell who's on the system.
  ○List of presently logged in users, ports and times on.
  ○Optional history of all logins and logouts.
- □ PS — Report on active processes.
  ○List your own or everybody's processes.
  ○Tell what commands are being executed.
  ○Optional status information: state and scheduling info, priority, attached terminal, what it's waiting for, size.
- □ IOSTAT — Print statistics about system I/O activity.
- □ TTY — Print name of your terminal.
- □ PWD — Print name of your working directory.

1.8. Backup and Maintenance

- □ MOUNT — Attach a device containing a file system to the tree of directories. Protects against nonsense arrangements.
- □ UMOUNT — Remove the file system contained on a device from the tree of directories. Protects against removing a busy device.
- □ MKFS — Make a new file system on a device.
- □ MKNOD — Make an i-node (file system entry) for a special file. Special files are physical devices, virtual devices, physical memory, etc.
- □ TP
- □ TAR — Manage file archives on magnetic tape or DECtape. TAR is newer.
  ○Collect files into an archive.
  ○Update DECtape archive by date.
  ○Replace or delete DECtape files.
  ○Print table of contents.
  ○Retrieve from archive.
- □ DUMP — Dump the file system stored on a specified device, selectively by date, or indiscriminately.
- □ RESTOR — Restore a dumped file system, or selectively retrieve parts thereof.
- □ SU — Temporarily become the super user with all the rights and privileges thereof. Requires a password.
- □ DCHECK
- □ ICHECK
- □ NCHECK — Check consistency of file system.
  ○Print gross statistics: number of files, number of directories, number of special files, space used, space free.

○Report duplicate use of space.
○Retrieve lost space.
○Report inaccessible files.
○Check consistency of directories.
○List names of all files.

□ CLRI — Peremptorily expunge a file and its space from a file system. Used to repair damaged file systems.

□ SYNC — Force all outstanding I/O on the system to completion. Used to shut down gracefully.

1.9. Accounting

The timing information on which the reports are based can be manually cleared or shut off completely.

□ AC — Publish cumulative connect time report.
○Connect time by user or by day.
○For all users or for selected users.

□ SA — Publish Shell accounting report. Gives usage information on each command executed.
○Number of times used.
○Total system time, user time and elapsed time.
○Optional averages and percentages.
○Sorting on various fields.

1.10. Communication

□ MAIL — Mail a message to one or more users. Also used to read and dispose of incoming mail. The presence of mail is announced by LOGIN and optionally by SH.
○Each message can be disposed of individually.
○Messages can be saved in files or forwarded.

□ CALENDAR — Automatic reminder service for events of today and tomorrow.

□ WRITE — Establish direct terminal communication with another user.

□ WALL — Write to all users.

□ MESG — Inhibit receipt of messages from WRITE and WALL.

□ CU — Call up another time-sharing system.
○Transparent interface to remote machine.
○File transmission.
○Take remote input from local file or put remote output into local file.
○Remote system need not be UNIX/32V.

□ UUCP — UNIX to UNIX copy.
○Automatic queuing until line becomes available and remote machine is up.
○Copy between two remote machines.
○Differences, mail, etc., between two machines.

1.11. Basic Program Development Tools

Some of these utilities are used as integral parts of the higher level languages described in section 2.

□ AR — Maintain archives and libraries. Combines several files into one for housekeeping efficiency.

○Create new archive.
○Update archive by date.
○Replace or delete files.
○Print table of contents.
○Retrieve from archive.

□ AS — Assembler.
○Creates object program consisting of
  code, normally read-only and sharable,
  initialized data or read-write code,
  uninitialized data.
○Relocatable object code is directly executable without further transformation.
○Object code normally includes a symbol table.
○"Conditional jump" instructions become branches or branches plus jumps depending on distance.

□ Library — The basic run-time library. These routines are used freely by all software.
○Buffered character-by-character I/O.
○Formatted input and output conversion (SCANF and PRINTF) for standard input and output, files, in-memory conversion.
○Storage allocator.
○Time conversions.
○Number conversions.
○Password encryption.
○Quicksort.
○Random number generator.
○Mathematical function library, including trigonometric functions and inverses, exponential, logarithm, square root, bessel functions.

□ ADB — Interactive debugger.
○Postmortem dumping.
○Examination of arbitrary files, with no limit on size.
○Interactive breakpoint debugging with the debugger as a separate process.
○Symbolic reference to local and global variables.
○Stack trace for C programs.
○Output formats:
  1-, 2-, or 4-byte integers in octal, decimal, or hex
  single and double floating point
  character and string
  disassembled machine instructions
○Patching.
○Searching for integer, character, or floating patterns.

□ OD — Dump any file. Output options include any combination of octal or decimal or hex by words, octal by bytes, ASCII, opcodes, hexadecimal.
○Range of dumping is controllable.

□ LD — Link edit. Combine relocatable object files. Insert required routines from specified libraries.
○Resulting code is sharable by default.

□ LORDER — Places object file names in proper order for loading, so that files depending on others come after them.

□ NM — Print the namelist (symbol table) of an object program. Provides control over the style and order of names that are printed.

□ SIZE Report the memory requirements of one or more object files.

□ STRIP Remove the relocation and symbol table information from an object file to save space.

□ TIME Run a command and report timing information on it.

□ PROF Construct a profile of time spent per routine from statistics gathered by time-sampling the execution of a program.
○Subroutine call frequency and average times for C programs.

□ MAKE Controls creation of large programs. Uses a control file specifying source file dependencies to make new version; uses time last changed to deduce minimum amount of work necessary.
○Knows about CC, YACC, LEX, etc.

1.12. UNIX/32V Programmer's Manual

□ Manual Machine-readable version of the UNIX/32V Programmer's Manual.
○System overview.
○All commands.
○All system calls.
○All subroutines in C and assembler libraries.
○All devices and other special files.
○Formats of file system and kinds of files known to system software.
○Boot and maintenance procedures.

□ MAN Print specified manual section on your terminal.

1.13. Computer-Aided Instruction

□ LEARN A program for interpreting CAI scripts, plus scripts for learning about UNIX/32V by using it.
○Scripts for basic files and commands, editor, advanced files and commands, EQN, MS macros, C programming language.

2. Languages

2.1. The C Language

□ CC Compile and/or link edit programs in the C language. The UNIX/32V operating system, most of the subsystems and C itself are written in C. For a full description of C, read *The C Programming Language,* Brian W. Kernighan and Dennis M. Ritchie, Prentice-Hall, 1978.
○General purpose language designed for structured programming.
○Data types include character, integer, float, double, pointers to all types, functions returning above types, arrays of all types, structures and unions of all types.
○Operations intended to give machine-independent control of full machine facility, including to-memory operations and pointer arithmetic.
○Macro preprocessor for parameterized code and inclusion of standard files.
○All procedures recursive, with parameters by value.
○Machine-independent pointer manipulation.
○Object code uses full addressing capability of the VAX-11.
○Runtime library gives access to all system facilities.
○Definable data types.

○Block structure

□ LINT — Verifier for C programs. Reports questionable or nonportable usage such as:
  Mismatched data declarations and procedure interfaces.
  Nonportable type conversions.
  Unused variables, unreachable code, no-effect operations.
  Mistyped pointers.
  Obsolete syntax.
○Full cross-module checking of separately compiled programs.

□ CB — A beautifier for C programs. Does proper indentation and placement of braces.

2.2. Fortran

□ F77 — A full compiler for ANSI Standard Fortran 77.
○Compatible with C and supporting tools at object level.
○Optional source compatibility with Fortran 66.
○Free format source.
○Optional subscript-range checking, detection of uninitialized variables.
○All widths of arithmetic: 2- and 4-byte integer; 4- and 8-byte real; 8- and 16-byte complex.

□ RATFOR — Ratfor adds rational control structure à la C to Fortran.
○Compound statements.
○If-else, do, for, while, repeat-until, break, next statements.
○Symbolic constants.
○File insertion.
○Free format source
○Translation of relationals like >, >=.
○Produces genuine Fortran to carry away.
○May be used with F77.

□ STRUCT — Converts ordinary ugly Fortran into structured Fortran (i.e., Ratfor), using statement grouping, if-else, while, for, repeat-until.

2.3. Other Algorithmic Languages

□ DC — Interactive programmable desk calculator. Has named storage locations as well as conventional stack for holding integers or programs.
○Unlimited precision decimal arithmetic.
○Appropriate treatment of decimal fractions.
○Arbitrary input and output radices, in particular binary, octal, decimal and hexadecimal.
○Reverse Polish operators:
    + − * /
    remainder, power, square root,
    load, store, duplicate, clear,
    print, enter program text, execute.

□ BC — A C-like interactive interface to the desk calculator DC.
○All the capabilities of DC with a high-level syntax.
○Arrays and recursive functions.
○Immediate evaluation of expressions and evaluation of functions upon call.
○Arbitrary precision elementary functions: exp, sin, cos, atan.
○Go-to-less programming.

2.4. Macroprocessing

□ M4 — A general purpose macroprocessor.
- ○Stream-oriented, recognizes macros anywhere in text.
- ○Syntax fits with functional syntax of most higher-level languages.
- ○Can evaluate integer arithmetic expressions.

2.5. Compiler-compilers

□ YACC — An LR(1)-based compiler writing system. During execution of resulting parsers, arbitrary C functions may be called to do code generation or semantic actions.
- ○BNF syntax specifications.
- ○Precedence relations.
- ○Accepts formally ambiguous grammars with non-BNF resolution rules.

□ LEX — Generator of lexical analyzers. Arbitrary C functions may be called upon isolation of each lexical token.
- ○Full regular expression, plus left and right context dependence.
- ○Resulting lexical analysers interface cleanly with YACC parsers.

3. Text Processing

3.1. Document Preparation

□ ED — Interactive context editor. Random access to all lines of a file.
- ○Find lines by number or pattern. Patterns may include: specified characters, don't care characters, choices among characters, repetitions of these constructs, beginning of line, end of line.
- ○Add, delete, change, copy, move or join lines.
- ○Permute or split contents of a line.
- ○Replace one or all instances of a pattern within a line.
- ○Combine or split files.
- ○Escape to Shell (command language) during editing.
- ○Do any of above operations on every pattern-selected line in a given range.
- ○Optional encryption for extra security.

□ PTX — Make a permuted (key word in context) index.

□ SPELL — Look for spelling errors by comparing each word in a document against a word list.
- ○25,000-word list includes proper names.
- ○Handles common prefixes and suffixes.
- ○Collects words to help tailor local spelling lists.

□ LOOK — Search for words in dictionary that begin with specified prefix.

□ CRYPT — Encrypt and decrypt files for security.

3.2. Document Formatting

□ TROFF

□ NROFF — Advanced typesetting. TROFF drives a Graphic Systems phototypesetter; NROFF drives ascii terminals of all types. This summary was typeset using TROFF. TROFF and NROFF are capable of elaborate feats of formatting, when appropriately programmed. TROFF and NROFF accept the same input language.

- Completely definable page format keyed to dynamically planted "interrupts" at specified lines.
- Maintains several separately definable typesetting environments (e.g., one for body text, one for footnotes, and one for unusually elaborate headings).
- Arbitrary number of output pools can be combined at will.
- Macros with substitutable arguments, and macros invocable in mid-line.
- Computation and printing of numerical quantities.
- Conditional execution of macros.
- Tabular layout facility.
- Positions expressible in inches, centimeters, ems, points, machine units or arithmetic combinations thereof.
- Access to character-width computation for unusually difficult layout problems.
- Overstrikes, built-up brackets, horizontal and vertical line drawing.
- Dynamic relative or absolute positioning and size selection, globally or at the character level.
- Can exploit the characteristics of the terminal being used, for approximating special characters, reverse motions, proportional spacing, etc.

The Graphic Systems typesetter has a vocabulary of several 102-character fonts (4 simultaneously) in 15 sizes. TROFF provides terminal output for rough sampling of the product.

NROFF will produce multicolumn output on terminals capable of reverse line feed, or through the postprocessor COL.

High programming skill is required to exploit the formatting capabilities of TROFF and NROFF, although unskilled personnel can easily be trained to enter documents according to canned formats such as those provided by MS, below. TROFF and EQN are essentially identical to NROFF and NEQN so it is usually possible to define interchangeable formats to produce approximate proof copy on terminals before actual typesetting. The preprocessors MS, TBL, and REFER are fully compatible with TROFF and NROFF.

□ MS — A standardized manuscript layout package for use with NROFF/TROFF. This document was formatted with MS.

- Page numbers and draft dates.
- Automatically numbered subheads.
- Footnotes.
- Single or double column.
- Paragraphing, display and indentation.
- Numbered equations.

□ EQN — A mathematical typesetting preprocessor for TROFF. Translates easily readable formulas, either in-line or displayed, into detailed typesetting instructions. Formulas are written in a style like this:

sigma sup 2 ~=~ 1 over N sum from i=1 to N (x sub i - x bar) sup 2

which produces:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

- Automatic calculation of size changes for subscripts, sub-subscripts, etc.
- Full vocabulary of Greek letters and special symbols, such as 'gamma', 'GAMMA', 'integral'.
- Automatic calculation of large bracket sizes.
- Vertical "piling" of formulae for matrices, conditional alternatives, etc.
- Integrals, sums, etc., with arbitrarily complex limits.
- Diacriticals: dots, double dots, hats, bars, etc.

○Easily learned by nonprogrammers and mathematical typists.

□ NEQN — A version of EQN for NROFF; accepts the same input language. Prepares formulas for display on any terminal that NROFF knows about, for example, those based on Diablo printing mechanism.
○Same facilities as EQN within graphical capability of terminal.

□ TBL — A preprocessor for NROFF/TROFF that translates simple descriptions of table layouts and contents into detailed typesetting instructions.
○Computes column widths.
○Handles left- and right-justified columns, centered columns and decimal-point alignment.
○Places column titles.
○Table entries can be text, which is adjusted to fit.
○Can box all or parts of table.

□ REFER — Fills in bibliographic citations in a document from a data base (not supplied).
○References may be printed in any style, as they occur or collected at the end.
○May be numbered sequentially, by name of author, etc.

□ TC — Simulate Graphic Systems typesetter on Tektronix 4014 scope. Useful for checking TROFF page layout before typesetting.

□ COL — Canonicalize files with reverse line feeds for one-pass printing.

□ DEROFF — Remove all TROFF commands from input.

□ CHECKEQ — Check document for possible errors in EQN usage.

4. Information Handling

□ SORT — Sort or merge ASCII files line-by-line. No limit on input size.
○Sort up or down.
○Sort lexicographically or on numeric key.
○Multiple keys located by delimiters or by character position.
○May sort upper case together with lower into dictionary order.
○Optionally suppress duplicate data.

□ TSORT — Topological sort — converts a partial order into a total order.

□ UNIQ — Collapse successive duplicate lines in a file into one line.
○Publish lines that were originally unique, duplicated, or both.
○May give redundancy count for each line.

□ TR — Do one-to-one character translation according to an arbitrary code.
○May coalesce selected repeated characters.
○May delete selected characters.

□ DIFF — Report line changes, additions and deletions necessary to bring two files into agreement.
○May produce an editor script to convert one file into another.
○A variant compares two new versions against one old one.

□ COMM — Identify common lines in two sorted files. Output in up to 3 columns shows lines present in first file only, present in both, and/or present in second only.

□ JOIN — Combine two files by joining records that have identical keys.

□ GREP — Print all lines in a file that satisfy a pattern as used in the editor ED.
○May print all lines that fail to match.

○May print count of hits.
○May print first hit in each file.

□ LOOK — Binary search in sorted file for lines with specified prefix.

□ WC — Count the lines, "words" (blank-separated strings) and characters in a file.

□ SED — Stream-oriented version of ED. Can perform a sequence of editing operations on each line of an input stream of unbounded length.
○Lines may be selected by address or range of addresses.
○Control flow and conditional testing.
○Multiple output streams.
○Multi-line capability.

□ AWK — Pattern scanning and processing language. Searches input for patterns, and performs actions on each line of input that satisfies the pattern.
○Patterns include regular expressions, arithmetic and lexicographic conditions, boolean combinations and ranges of these.
○Data treated as string or numeric as appropriate.
○Can break input into fields; fields are variables.
○Variables and arrays (with non-numeric subscripts).
○Full set of arithmetic operators and control flow.
○Multiple output streams to files and pipes.
○Output can be formatted as desired.
○Multi-line capabilities.

5. Graphics

The programs in this section are predominantly intended for use with Tektronix 4014 storage scopes.

□ GRAPH — Prepares a graph of a set of input numbers.
○Input scaled to fit standard plotting area.
○Abscissae may be supplied automatically.
○Graph may be labeled.
○Control over grid style, line style, graph orientation, etc.

□ SPLINE — Provides a smooth curve through a set of points intended for GRAPH.

□ PLOT — A set of filters for printing graphs produced by GRAPH and other programs on various terminals. Filters provided for 4014, DASI terminals, Versatec printer/plotter.

6. Novelties, Games, and Things That Didn't Fit Anywhere Else

□ BACKGAMMON — A player of modest accomplishment.

□ BCD — Converts ascii to card-image form.

□ CAL — Print a calendar of specified month and year.

□ CHING — The *I Ching*. Place your own interpretation on the output.

□ FORTUNE — Presents a random fortune cookie on each invocation. Limited jar of cookies included.

□ UNITS — Convert amounts between different scales of measurement. Knows hundreds of units. For example, how many km/sec is a parsec/megayear?

- ARITHMETIC
  Speed and accuracy test for number facts.
- QUIZ Test your knowledge of Shakespeare, Presidents, capitals, etc.
- WUMP Hunt the wumpus, thrilling search in a dangerous cave.
- HANGMAN Word-guessing game. Uses a dictionary supplied with SPELL.
- FISH Children's card-guessing game.

7th Edition UNIX—Summary

September 6, 1978

Bell Laboratories
Murray Hill, New Jersey 07974

A. What's new: highlights of the 7th Edition UNIX† System

Aimed at larger systems. Devices are addressable to 2^{31} bytes, files to 2^{30} bytes. 128K memory (separate instruction and data space) is needed for some utilities.

Portability. Code of the operating system and most utilities has been extensively revised to minimize its dependence on particular hardware.

Fortran 77. F77 compiler for the new standard language is compatible with C at the object level. A Fortran structurer, STRUCT, converts old, ugly Fortran into RATFOR, a structured dialect usable with F77.

Shell. Completely new SH program supports string variables, trap handling, structured programming, user profiles, settable search path, multilevel file name generation, etc.

Document preparation. TROFF phototypesetter utility is standard. NROFF (for terminals) is now highly compatible with TROFF. MS macro package provides canned commands for many common formatting and layout situations. TBL provides an easy to learn language for preparing complicated tabular material. REFER fills in bibliographic citations from a data base.

UNIX-to-UNIX file copy. UUCP performs spooled file transfers between any two machines.

Data processing. SED stream editor does multiple editing functions in parallel on a data stream of indefinite length. AWK report generator does free-field pattern selection and arithmetic operations.

Program development. MAKE controls re-creation of complicated software, arranging for minimal recompilation.

Debugging. ADB does postmortem and breakpoint debugging, handles separate instruction and data spaces, floating point, etc.

C language. The language now supports definable data types, generalizes initialization, block structure, long integers, unions, explicit type conversions. The LINT verifier does strong type checking and detection of probable errors and portability problems even across separately compiled functions.

Lexical analyzer generator. LEX converts specification of regular expressions and semantic actions into a recognizing subroutine. Analogous to YACC.

Graphics. Simple graph-drawing utility, graphic subroutines, and generalized plotting filters adapted to various devices are now standard.

Standard input-output package. Highly efficient buffered stream I/O is integrated with formated input and output.

Other. The operating system and utilities have been enhanced and freed of restrictions in many other ways too numerous to relate.

† UNIX is a trademark of Bell Laboratories.

B. Hardware

The 7th edition UNIX operating system runs on DEC PDP-11/45 or 11/70* with at least the following equipment:

123K to 2M words of managed memory; parity not used.

disk: RP03, RP04, RP06, RK05 (more than 1 RK05) or equivalent.

console typewriter.

clock: KW11-L or KW11-P.

The following equipment is strongly recommended:

communications controller such as DL11 or DH11.

full duplex 96-character ASCII terminals.

9-track tape or extra disk for system backup.

The system is normally distributed on 9-track tape. The minimum memory and disk space specified is enough to run and maintain UNIX. More will be needed to keep all source on line or to handle a large number of users, big data bases, diversified complements of devices, or large programs. The resident code occupies 12-20K words depending on configuration; system data occupies 10-28K words.

There is no commitment to provide 7th Edition UNIX on PDP-11/34, 11/40 and 11/60 hardware.

C. Software

Most of the programs available as UNIX commands are listed. Source code and printed manuals are distributed for all of the listed software except games. Almost all of the code is written in C. Commands are self-contained and do not require extra setup information, unless specifically noted as "interactive." Interactive programs can be made to run from a prepared script simply by redirecting input. Most programs intended for interactive use (e.g., the editor) allow for an escape to command level (the Shell). Most file processing commands can also go from standard input to standard output ("filters"). The piping facility of the Shell may be used to connect such filters directly to the input or output of other programs.

1. Basic Software

This includes the time-sharing operating system with utilities, a machine language assembler and a compiler for the programming language C—enough software to write and run new applications and to maintain or modify UNIX itself.

1.1. Operating System

UNIX — The basic resident code on which everything else depends. Supports the system calls, and maintains the file system. A general description of UNIX design philosophy and system facilities appeared in the Communications of the ACM, July, 1984. A more extensive survey is in the Bell System Technical Journal for July-August 1978. Capabilities include:

- Reentrant code for user processes.
- Separate instruction and data spaces.
- "Group" access permissions for cooperative projects, with overlapping memberships.
- Alarm-clock timeouts.

* PDP is a Trademark of Digital Equipment Corporation.

- Timer-interrupt sampling and interprocess monitoring for debugging and measurement.
- Multiplexed I/O for machine-to-machine communication.

DEVICES All I/O is logically synchronous. I/O devices are simply files in the file system. Normally, invisible buffering makes all physical record structure and device characteristics transparent and exploits the hardware's ability to do overlapped I/O. Unbuffered physical record I/O is available for unusual applications. Drivers for these devices are available; others can be easily written:

- Asynchronous interfaces: DH11, DL11. Support for most common ASCII terminals.
- Synchronous interface: DP11.
- Automatic calling unit interface: DN11.
- Line printer: LP11.
- Magnetic tape: TU10 and TU16.
- DECtape: TC11.
- Fixed head disk: RS11, RS03 and RS04.
- Pack type disk: RP03, RP04, RP06; minimum-latency seek scheduling.
- Cartridge-type disk: RK05, one or more physical devices per logical device.
- Null device.
- Physical memory of PDP-11, or mapped memory in resident system.
- Phototypesetter: Graphic Systems System/I through DR11C.

BOOT Procedures to get UNIX started.

MKCONF Tailor device-dependent system code to hardware configuration. As distributed, UNIX can be brought up directly on any acceptable CPU with any acceptable disk, any sufficient amount of core, and either clock. Other changes, such as optimal assignment of directories to devices, inclusion of floating point simulator, or installation of device names in file system, can then be made at leisure.

1.2. User Access Control

LOGIN Sign on as a new user.

- Verify password and establish user's individual and group (project) identity.
- Adapt to characteristics of terminal.
- Establish working directory.
- Announce presence of mail (from MAIL).
- Publish message of the day.
- Execute user-specified profile.
- Start command interpreter or other initial program.

PASSWD Change a password.

- User can change his own password.
- Passwords are kept encrypted for security.

NEWGRP Change working group (project). Protects against unauthorized changes to projects.

1.3. Terminal Handling

TABS — Set tab stops appropriately for specified terminal type.

STTY — Set up options for optimal control of a terminal. In so far as they are deducible from the input, these options are set automatically by LOGIN.

- Half vs. full duplex.
- Carriage return + line feed vs. newline.
- Interpretation of tabs.
- Parity.
- Mapping of upper case to lower.
- Raw vs. edited input.
- Delays for tabs, newlines and carriage returns.

1.4. File Manipulation

CAT — Concatenate one or more files onto standard output. Particularly used for unadorned printing, for inserting data into a pipeline, and for buffering output that comes in dribs and drabs. Works on any file regardless of contents.

CP — Copy one file to another, or a set of files to a directory. Works on any file regardless of contents.

PR — Print files with title, date, and page number on every page.

- Multicolumn output.
- Parallel column merge of several files.

LPR — Off-line print. Spools arbitrary files to the line printer.

CMP — Compare two files and report if different.

TAIL — Print last *n* lines of input

- May print last *n* characters, or from *n* lines or characters to end.

SPLIT — Split a large file into more manageable pieces. Occasionally necessary for editing (ED).

DD — Physical file format translator, for exchanging data with foreign systems, especially IBM 370s.

SUM — Sum the words of a file.

1.5. Manipulation of Directories and File Names

RM — Remove a file. Only the name goes away if any other names are linked to the file.

- Step through a directory deleting files interactively.
- Delete entire directory hierarchies.

LN — "Link" another name (alias) to an existing file.

MV — Move a file or files. Used for renaming files.

CHMOD — Change permissions on one or more files. Executable by files' owner.

CHOWN — Change owner of one or more files.

CHGRP Change group (project) to which a file belongs.

MKDIR Make a new directory.

RMDIR Remove a directory.

CD Change working directory.

FIND Prowl the directory hierarchy finding every title that meets specified criteria.

- Criteria include:

  name matches a given pattern.
  creation date in given range.
  date of last use in give range.
  given permissions.
  given owner.
  given special file characteristics.
  boolean combinations of above.
- Any directory may be considered to be the root.
- Perform specified command on each file found.

1.6. Running of Programs

SH The Shell, or command language interpreter.

- Supply arguments to and run any executable program.
- Redirect standard input, standard output, and standard error files.
- Pipes: simultaneous execution of one process connected to the input of another.
- Compose compound commands using:

  if ... then ... else.
  case switches.
  while loops.
  for loops over lists.
  break, continue and exit.
  parentheses for grouping.
- Initiate background processes.
- Perform Shell programs, i.e., command scripts with substitutable arguments.
- Construct argument lists from all file names satisfying specified patterns.
- Take special action on traps and interrupts.
- User-settable search path for finding commands.
- Executes user-settable profile upon login.
- Optionally announces presence of mail as it arrives.
- Provides variables and parameters with default setting.

TEST	Tests for use in Shell conditionals. • String comparison. • File nature and accessibility. • Boolean combinations of the above.
EXPR	String computations for calculating command arguments. • Integer arithmetic • Pattern matching
WAIT	Wait for termination of asynchronously running processes.
READ	Read from a terminal, for interactive Shell procedure.
ECHO	Print remainder of command line. Useful for diagnostics or prompts in Shell programs, or for inserting data into a pipeline.
SLEEP	Suspend execution for a specified time.
NOHUP	Run a command immune to hanging up the terminal.
NICE	Run a command in low (or high) priority.
KILL	Terminate named process.
CRON	Schedule regular actions at specified times. • Actions are arbitrary programs. • Times are conjunctions of month, day of month, day of week, hour and minute. Ranges are specifiable for each.
AT	Schedule a one-shot action for an arbitrary time.
TEE	Pass data between processes and divert a copy into one or more files.

1.7. Status Inquiries

LS	List the names of one, several or all files in one or more directories. • Alphabetic or temporal sorting, up or down. • Optional information: size, owner, group, date last modified, date last accessed, permissions, i-node number.
FILE	Try to determine what kind of information is in a file by consuling the file system index and by reading the file itself.
DATE	Print today's date and time. Has considerable knowledge of calendric and horological peculiarities. • May set UNIX's idea of date and time.
DF	Report amount of free space on file system devices.
DU	Print a summary of total space occupied by all files in a hierarchy.
QUOT	Print summary of file space usage by user id.
WHP	Tell who's on the system. • List of presently logged in users, ports and times on. • Optional history of all logins and logouts.

PS — Report on active processes.

- List your own or everybody's processes.
- Tell what commands are being executed.
- Optional status information: state and scheduling info, priority, attached terminal, what it's waiting for, size.

IOSTAT — Print statistics about system I/O activity.

TTY — Print name of your terminal.

PWD — Print name of your working directory.

1.8. Backup and Maintenance

MOUNT — Attach a device containing a file system to the tree of directories. Protects against nonsense arrangements.

UMOUNT — Remove the file system contained on a device from the tree of directories. Protects against removing a busy device.

MKFS — Make a new file system on a device.

MKNOD — Make an i-node (file system entry) for a special file. Special files are physical devices, virtual devices, physical memory, etc.

TP

TAR — Manage file archives on magnetic tape or DECtape. TAR is newer.

- Collect files into an archive.
- Update DECtape archive by date.
- Replace or delete DECtape files.
- Print table of contents.
- Retrieve from archive.

DUMP — Dump the file system stored on a specified device, selectively by date, or indiscriminately.

RESTOR — Restore a dumped file system, or selectively retrieve parts thereof.

SU — Temporarily become the super user with all the rights and privileges thereof. Requires a password.

DCHECK

ICHECK

NCHECK — Check consistency of file system.

- Print gross statistics: number of files, number of directories, number of special files, space used, space free.
- Report duplicate use of space.
- Retrieve lost space.
- Report inaccessible files.
- Check consistency of directories.
- List names of all files.

CLRI — Peremptorily expunge a file and its space from a file system. Used to repair damaged file systems.

SYNC — Force all outstanding I/O on the system to completion. Used to shut down gracefully.

1.9. Accounting

The timing information on which the reports are based can be manually cleared or shut off completely.

AC — Publish cumulative connect time report.

- Connect time by user or by day.
- For all users or for selected users.

SA — Publish Shell accounting report. Gives usage information on each command executed.

- Number of times used.
- Total system time, user time and elapsed time.
- Optional averages and percentages.
- Sorting on various fields.

1.10. Communication

MAIL — Mail a message to one or more users. Also used to read and dispose of incoming mail. The presence of mail is announced by LOGIN and optionally by SH.

- Each message can be disposed of individually.
- Messages can be saved in files or forwarded.

CALENDAR — Automatic reminder service for events of today and tomorrow.

WRITE — Establish direct terminal communication with another user.

WALL — Write to all users.

MESG — Inhibit receipt of messages from WRITE and WALL.

CU — Call up another time-sharing system.

- Transparent interface to remote machine.
- File transmission.
- Take remote input from local file or put remote output into local file.
- Remote system need not be UNIX.

UUCP — UNIX to UNIX copy.

- Automatic queuing until line becomes available and remote machine is up.
- Copy between two remote machines.
- Differences, mail, etc., between two machines.

1.11. Basic Program Development Tools

Some of these utilities are used as integral parts of the higher level languages described in section 2.

AR — Maintain archives and libraries. Combines several files into one for housekeeping efficiency.

- Create new archive.
- Update archive by date.
- Replace or delete files.
- Print table of contents.
- Retrieve from archive.

AS — Assembler. Similar to PAL-11, but different in detail.

- Creates object program consisting of
  - code, possibly read-only,
  - initialized data or read-write code,
  - uninitialized data.
- Relocatable object code is directly executable without further transformation.
- Object code normally includes a symbol table.
- Multiple source files.
- Local labels.
- Conditional assembly.
- "Conditional jump" instructions become branches or branches plus jumps depending on distance.

Library — The basic run-time library. These routines are used freely by all software.

- Buffered character-by-character I/O.
- Formatted input and output conversion (SCANF and PRINTF) for standard input and output, files, in-memory conversion.
- Storage allocator.
- Time conversions.
- Number conversions.
- Password encryption.
- Quicksort.
- Random number generator.
- Mathematical function library, including trigonometric functions and inverses, exponential, logarithm, square root, bassel functions.

ADB — Interactive debugger.

- Postmortem dumping.
- Examination of arbitrary files, with no limit on size.
- Interactive breakpoint debugging with the debugger as a separate process.
- Symbolic reference to local and global variables.
- Stack trace for C programs.

- Output formats:

  1-, 2-, or 4-byte integers in octal, decimal, or hex
  single and double floating point
  character and string
  disassembled machine instructions
- Patching.
- Searching for integer, character, or floating patterns.
- Handles separated instruction and data space.

OD — Dump any file. Output options include any combination of octal or decimal by words, octal by bytes, ASCII, opcodes, hexadecimal.

- Range of dumping in controllable.

LD — Link edit. Combine relocatable object files. Insert required routines from specified libraries.

- Resulting code may be sharable.
- Resulting code may have separate instruction and data spaces.

LORDER — Places object file names in proper order for loading, so that files depending on others come after them.

NM — Print the namelist (symbol table) of an object program. Provides control over the style and order of names that are printed.

SIZE — Report the core requirements of one or more object files.

STRIP — Remove the relocation and symbol table information from an object file to save space.

TIME — Run a command and report timing information on it.

PROF — Construct a profile of time spent per routine from statistics gathered by time sampling the execution of a program. Uses floating point.

- Subroutine call frequency and average times for C programs.

MAKE — Controls creation of large programs. Uses a control file specifying source file dependencies to make new version; uses time last changed to deduce minimum amount of work necessary.

- Knows about CC, YAC, LEX, etc.

1.12. UNIX Programmer's Manual

Manual — Machine-readable version of the UNIX Programmer's Manual.

- System overview.
- All commands.
- All system calls.
- All subroutines in C and assembler libraries.
- All devices and other special files.
- Formats of file system and kinds of files known to system software.
- Boot and maintenance procedures.

1.13. Computer-Aided Instruction

LEARN — A program for interpreting CAI scripts, plus scripts for learning about UNIX by using it.

- Scripts for basic files and commands, editor, advanced files and commands, EQN, MS macros, C programming language.

2. Languages

2.1. The C Language

CC — Compile and/or link edit programs in the C language. The UNIX operating system, most of the subsystems and C itself are written in C. For a full description of C, read *The C Programming Language,* Brian W Kernighan and Dennis M. Ritchie, Prentice-Hall, 1978.

- General purpose language designed for structured programming.
- Data types include characters, integer, float, double, pointers to all types, functions returning above types, arrays of all types, structures and unions of all types.
- Operations intended to give machine-independent control of full machine facility, including to memory operations and pointer arithmetic.
- Macro preprocessor for parameterized code and inclusion of standard files.
- All procedures recursive, with parameters by value.
- Machine-independent pointer manipulation.
- Object code uses full addressing capability of the PDP-11.
- Runtime library gives access to all system facilities.
- Definable data types.
- Block structure.

LINT — Verifier for C programs. Reports questionable or nonportable usage such as:

Mismatched data declarations and procedure interfaces.

Nonportable type conversions.

Unused variables, unreachable code, no-effect operations.

Mistyped pointers.

Obsolete syntax.

- Full cross-module checking of separately compiled programs.

CB — A beautifier for C programs. Does proper indentation and placement of braces.

2.2. Fortran

F77 — A full compiler for ANSI Standard Fortran 77.

- Compatible with C and supporting tools at object level.
- Optional source compatibility with Fortran 66.
- Free format source.
- Optional subscript-range checking, detection of uninitialized variables.
- All widths of arithmetic: 2- and 4-byte integer; 4- and 8-byte real; 8- and 16-byte complex.

RATFOR — Ratfor adds rational control structure a la C to Fortran.

- Compound statements.
- If-else, do, for, while, repeat-until, break, next statements.
- Symbolic constants.
- File insertion.
- Free format source.
- Translation of relationals like >, >=.
- Produces genuine Fortran to carry away.
- May be used with F77.

STRUCT — Converts ordinary ugly Fortran into structured Fortran (i.e., Ratfor), using statement grouping, if-else, while, for, repeat-until.

2.3. Other Algorithmic Languages

BAS — An interactive interpreter, similar in style to BASIC. Interpret unnumbered statements immediately, numbered statements upon "run".

- Statements include:

  comment,
  dump,
  for...next,
  goto,
  if...else...fi,
  list,
  print,
  prompt,
  return,
  run,
  save.
- All calculations double precision.
- Recursive function defining and calling.
- Built-in functions include log, exp, sin, cos, atn, int, sqr, abs, rnd.
- Escape to ED for complex program editing.

DC — Interactive programmable desk calculator. Has named storage locations as well as conventional stack for holding integers or programs.

- Unlimited precision decimal arithmetic.
- Appropriate treatment of decimal fractions.
- Arbitrary input and output radices, in particular binary, octal, decimal and hexadecimal.
- Reverse Polish operators:

  + - * /
  remainder, power, square root,
  load, store, duplicate, clear,
  print, enter program text, execute.

BC — A C-like interactive interface to the desk calculator DC.

- All the capabilities of DC with a high-level syntax.
- Arrays and recursive functions.
- Immediate evaluation of expressions and evaluation of functions upon call.
- Arbitrary precision elementary functions: exp, sin, cos, atn.
- Go-to-less programming.

M4 — A general purpose macroprocessor.

- Stream-oriented, recognizes macros anywhere in text.
- Syntax fits with functional syntax of most higher-level languages.
- Can evaluate integer arithmetic expressions.

2.4. Compiler-compilers

YACC — An LR-based compiler writing system. During execution of resulting parsers, arbitrary C functions may be called to do code generation or semantic actions.

- BNF syntax specifications.
- Precedence relations.
- Accepts formally ambiguous grammars with non-BNF resolution rules.

LEX — Generator of lexical analyzers. Arbitrary C functions may be called upon isolation of each lexical token.

- Full regular expression, plus left and right context dependence.
- Resulting lexical analyzers interface cleanly with YACC parsers.

3. Text Processing

3.1 Document Preparation

ED — Interactive context editor. Random access to all lines of a file.

- Find lines by number or pattern. Patterns may include: specified characters, don't care characters, choices among characters, repetitions of these constructs, beginning of line, end of line.
- Add, delete, change, copy, move or join lines.
- Permute or split contents of a line.
- Replace one or all instances of a pattern within a line.
- Combine or split files.
- Escape to Shell (command language) during editing.
- Do any of above operations on every pattern-selected line in a given range.
- Optional encryption for extra security.

PTX — Make a permuted (key word in context) index.

SPELL — Look for spelling errors by comparing each word in a document against a word list.

- 25,000-word list includes proper names.
- Handles common prefixes and suffixes.
- Collects words to help tailor local spelling lists.

LOOK — Search for words in dictionary that begin with specified prefix.

TYPO — Look for spelling errors by a statistical technique; not limited to English.

CRYPT — Encrypt and decrypt files for security.

3.2. Document Formatting

ROFF — A typesetting program for terminals. Easy for nontechnical people to learn, and good for simple documents. Input consists of data lines intermixed with control lines, such as

.sp 2 insert two lines of space

.ce center the next line

ROFF is deemed to be obsolete; it is intended only for casual use.

- Justification of either or both margins.
- Automatic hyphenation.
- Generalized running heads and feet, with even-odd page capability, numbering, etc.
- Definable macros for frequently used control sequences (no substitutable arguments).
- All 4 margins and page size dynamically adjustable.
- Hanging indents and one-line indents.
- Absolute and relative parameter settings.
- Optional legal-style numbering of output lines.
- Multiple file capability.
- Not usable as a filter.

TROFF

NROFF — Advanced typesetting, TROFF drives a Graphic Systems phototypesetter; NROFF drives ASCII terminals of all types. This summary was typeset using TROFF. TROFF and NROFF style is similar to ROFF, but they are capable of much more elaborate feats of formatting, when appropriately programmed. TROFF and NROFF accept the same input language.

- All ROFF capabilities available or definable.
- Completely definable page format keyed to dynamically planted "interrupts" at specified lines.
- Maintains several separately definable typesetting environments (e.g., one for body text, one for footnotes, and one for unusually elaborate headings).
- Arbitrary number of output pools can be combined at will.
- Macros with substitutable arguments, and macros invocable in mid-line.
- Computation and printing of numerical quantities.
- Conditional execution of macros.
- Tabular layout facility.
- Positions expressible in inches, centimeters, ems, points, machine units or arithmetic combinations thereof.
- Access to character-width computation for unusually difficult layout problems.
- Overstrikes, built-up rackets, horizontal and vertical line drawing.

- Dynamic relative or absolute positioning and size selection, globally or at the character level.
- Can exploit the characteristics of the terminal being used, for approximating special characters, reverse motions, proportional spacing, etc.

The Graphic Systems typesetter has a vocabulary of several 102-character fonts (4 simultaneously) in 15 sizes. TROFF provides terminal output for rough sampling of the product.

NROFF will produce multicolumn output on terminals capable of reverse line feed, or through the postprocessor COL.

High programming skill is required to exploit the formatting capabilities of TROFF and NROFF, although unskilled personnel can easily be trained to enter documents according to canned formats such as those provided by MS, below. TROFF and EQN are essentially identical to NROFF and NEQN so it is usually possible to define interchangeable formats to product approximate proof copy on terminals before actual typesetting. The preprocessors MS, TBL, and REFER are fully compatible with TROFF and NROFF.

MS — A standardized manuscript layout package for use with NROFF/TROFF. This document was formated with MS.

- Page numbers and draft dates.
- Automatically numbered subheads.
- Footnotes.
- Single or double column.
- Paragraphing, display and indentation.
- Numbered equations.

EQN — A mathematical typesetting preprocessor for TROFF. Translates easily readable formulas, either in-line or displayed, into detailed typesetting instructions. Formulas are written in a style like this:

sigma sup 2 "=" 1 over N sum from i=1 to N (x sub i - x bar) sup 2

which produces:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - xbar)^2$$

- Automatic calculation of size changes for subscripts, sub-subscripts, etc.
- Full vocabulary of Greek letters and special symbols, such as "gamma", "GAMMA", and "integral".
- Automatic calculation of large bracket sizes.
- Vertical "piling" of formulae for matrices, conditional alternatives, etc.
- Integrals, sums, etc., with arbitrarily complex limits.
- Diacriticals: dots, double dots, hats, bars, etc.
- Easily learned by nonprogrammers and mathematical typists.

NEQN — A version of EQN for NROFF; accepts the same input language. Prepares formulas for display on any terminal that NROFF knows about, for example, those based on Diablo printing mechanism.

- Same facilities as EQN within graphical capability of terminal.

TBL — A preprocessor for NROFF/TROFF that translates simple descriptions of table layouts and contents into detailed typesetting instructions.

- Computes column widths.
- Handles left- and right-justified columns, centered columns and decimal-point alignment.
- Places column titles.
- Table entries can be text, which is adjusted to fit.
- Can box all or parts of table.

REFER — Fills in bibliographic citations in a document from a data base (not supplied).

- References may be printed in any style, as they occur or collected at the end.
- May be numbered sequentially, by name of author, etc.

TC — Simulate Graphic Systems typesetter on Tektronix 4014 scope. Useful for checking TROFF page layout before typesetting.

GREEK — Fancy printing on Diablo-mechanism terminals like DASI-300 and DASI-450, and on Tektronix 4014.

- Gives half-line forward and reverse motions.
- Approximates Greek letters and other special characters by overstriking.

COL — Canonicalize files with reverse line feeds for one-pass printing.

DEROFF — Remove all TROFF commands from input.

CHECKEQ — Check document for possible errors in EQN usage.

4. Information Handling

SORT — Sort or merge ASCII files line-by-line. No limit on input size.

- Sort up or down.
- Sort lexicographically or on numeric key.
- Multiple keys located by delimiters or by character position.
- May sort upper case together with lower into dictionary order.
- Optionally suppress duplicate data.

TSORT — Topological sort—converts a partial order into a total order.

UNIQ — Collapse successive duplicate lines in a file into one line.

- Publish lines that were originally unique, duplicated, or both.
- May give redundancy count for each line.

TR — Do one-to-one character translation according to an arbitrary code.

- May coalesce selected repeated characters.
- May delete selected characters.

DIFF — Report line changes, additions and deletions necessary to bring two files into agreement.

- May produce an editor script to convert one file into another.
- A variant compares two new versions against one old one.

COMM — Identify common lines in two sorted files. Output in up to 3 columns shows lines present in first file only, present in both, and/or present in second only.

JOIN — Combine two files by joining records that have identical keys.

GREP — Print all lines in a file that satisfy a pattern as used in the editor ED.

- May print all lines that fail to match.
- May print count of hits.
- May print first hit in each file.

LOOK — Binary search in sorted file for lines with specified prefix.

WC — Count the lines, "words" (blank-separated strings) and characters in a file.

SED — Stream-oriented version of ED. Can perform a sequence of editing operations on each line of an input stream of unbounded length.

- Lines may be selected by address or range of addresses.
- Control flow and conditional testing.
- Multiple output streams.
- Multi-line capability.

AWK — Pattern scanning and processing language. Searches input for patterns, and performs actions on each line of input that satisfies the pattern.

- Patterns include regular expressions, arithmetic and lexicographic conditions, boolean combinations and ranges of these.
- Data treated as string or numeric as appropriate.
- Can break input into fields; fields are variables.
- Variables and arrays (with non-numeric subscripts).
- Full set of arithmetic operators and control flow.
- Multiple output streams to files and pipes.
- Output can be formatted as desired.
- Multi-line capabilities.

5. Graphics

The programs in this section are predominantly for use with Tektronix 4014 storage scopes.

GRAPH — Prepares a graph of a set of input numbers.

- Input scaled to fit standard plotting area.
- Abscissae may be supplied automatically.
- Graph may be labeled.
- Control over grid style, line style, graph orientation, etc.

SPLINE — Provides a smooth curve through a set of points intended for GRAPH.

PLOT — A set of filters for printing graphs produced by GRAPH and other programs on various terminals. Filters provided for 4014, DASI terminals, Versatec printer/plotter.

6. Novelties, Games, and Things That Didn't Fit Anywhere Else

BACKGAMMON	A player of modest accomplishment.
CHESS	Plays good class D chess.
CHECKERS	Ditto for checkers.
BCD	Converts ASCII to card-image form.
PPT	Converts ASCII to paper tape form.
BJ	A blackjack dealer.
CUBIC	An accomplished player of 4x4x4 tic-tac-toe.
MAZE	Constructs random mazes for you to solve.
MOO	A fascinating number-guessing game.
CAL	Print a calendar of specified month and year.
BANNER	Print output in huge letters.
CHING	The *I ching*. Place your own interpretation on the output.
FORTUNE	Presents a random fortune cookie on each invocation. Limited jar of cookies included.
UNITS	Convert amounts between different scales of measurement. Knows hundreds of units. For example, how many km/sec is a parsec/megayear?
TT	A tic-tac-toe program that learns. It never makes the same mistake twice.
ARITHMETIC	Speed and accuracy test for number facts.
FACTOR	Factor large integers.
QUIZ	Test your knowledge of Shakespeare, Presidents, capitals, etc.
WUMP	Hunt the wumpus, thrilling search in a dangerous cave.
REVERSI	A two person board game, isomorphic to Othello®.
HANGMAN	Word guessing game. Uses the dictionary supplied with SPELL.
FISH	Children's card-guessing game.

UNIX For Beginners — Second Edition

Brian W. Kernighan

ABSTRACT

This paper is meant to help new users get started on the UNIX† operating system. It includes:

- basics needed for day-to-day use of the system — typing commands, correcting typing mistakes, logging in and out, mail, inter-terminal communication, the file system, printing files, redirecting I/O, pipes, and the shell.
- document preparation — a brief discussion of the major formatting programs and macro packages, hints on preparing documents, and capsule descriptions of some supporting software.
- UNIX programming — using the editor, programming the shell, programming in C, other languages and tools.
- An annotated UNIX bibliography.

September 12, 1986

† UNIX is a trademark of Bell Laboratories.

UNIX For Beginners — Second Edition

Brian W. Kernighan

INTRODUCTION

From the user's point of view, the UNIX operating system is easy to learn and use, and presents few of the usual impediments to getting the job done. It is hard, however, for the beginner to know where to start, and how to make the best use of the facilities available. The purpose of this introduction is to help new users get used to the main ideas of the UNIX system and start making effective use of it quickly.

You should have a couple of other documents with you for easy reference as you read this one. The most important is *The UNIX Programmer's Manual*; it's often easier to tell you to read about something in the manual than to repeat its contents here. The other useful document is *A Tutorial Introduction to the UNIX Text Editor,* which will tell you how to use the editor to get text — programs, data, documents — into the computer.

A word of warning: the UNIX system has become quite popular, and there are several major variants in widespread use. Of course details also change with time. So although the basic structure of UNIX and how to use it is common to all versions, there will certainly be a few things which are different on your system from what is described here. We have tried to minimize the problem, but be aware of it. In cases of doubt, this paper describes Version 7 UNIX.

This paper has five sections:

1. Getting Started: How to log in, how to type, what to do about mistakes in typing, how to log out. Some of this is dependent on which system you log into (phone numbers, for example) and what terminal you use, so this section must necessarily be supplemented by local information.
2. Day-to-day Use: Things you need every day to use the system effectively: generally useful commands; the file system.
3. Document Preparation: Preparing manuscripts is one of the most common uses for UNIX systems. This section contains advice, but not extensive instructions on any of the formatting tools.
4. Writing Programs: UNIX is an excellent system for developing programs. This section talks about some of the tools, but again is not a tutorial in any of the programming languages provided by the system.
5. A UNIX Reading List. An annotated bibliography of documents that new users should be aware of.

I. GETTING STARTED

Logging In

You must have a UNIX login name, which you can get from whoever administers your system. You also need to know the phone number, unless your system uses permanently connected terminals. The UNIX system is capable of dealing with a wide variety of terminals: Terminet 300's; Execuport, TI and similar portables; video (CRT) terminals like the HP2640, etc.; high-priced graphics terminals like the Tektronix 4014; plotting terminals like those from GSI and DASI; and even the venerable Teletype in its various forms. But note: UNIX is strongly oriented towards devices with *lower case.* If your terminal produces only upper case (e.g., model 33 Teletype, some video and portable terminals), life will be so difficult that you should look for another terminal.

Be sure to set the switches appropriately on your device. Switches that might need to be adjusted include the speed, upper/lower case mode, full duplex, even parity, and any others that local wisdom advises. Establish a connection using whatever magic is needed for your terminal; this may involve dialing a telephone call or merely flipping a switch. In either case, UNIX should type "**login:**" at you. If it types garbage, you may be at the wrong speed; check the switches. If that fails, push the "break" or "interrupt" key a few times, slowly. If that fails to produce a login message, consult a guru.

When you get a **login:** message, type your login name *in lower case.* Follow it by a RETURN; the system will not do anything until you type a RETURN. If a password is required, you will be asked for it, and (if possible) printing will be turned off while you type it. Don't forget

RETURN.

The culmination of your login efforts is a "prompt character," a single character that indicates that the system is ready to accept commands from you. The prompt character is usually a dollar sign **$** or a percent sign %. (You may also get a message of the day just before the prompt character, or a notification that you have mail.)

Typing Commands

Once you've seen the prompt character, you can type commands, which are requests that the system do something. Try typing

date

followed by RETURN. You should get back something like

Mon Jan 16 14:17:10 EST 1978

Don't forget the RETURN after the command, or nothing will happen. If you think you're being ignored, type a RETURN; something should happen. RETURN won't be mentioned again, but don't forget it — it has to be there at the end of each line.

Another command you might try is **who**, which tells you everyone who is currently logged in:

who

gives something like

mb	**tty01**	**Jan 16**	**09:11**
ski	**tty05**	**Jan 16**	**09:33**
gam	**tty11**	**Jan 16**	**13:07**

The time is when the user logged in; "ttyxx" is the system's idea of what terminal the user is on.

If you make a mistake typing the command name, and refer to a non-existent command, you will be told. For example, if you type

whom

you will be told

whom: not found

Of course, if you inadvertently type the name of some other command, it will run, with more or less mysterious results.

Strange Terminal Behavior

Sometimes you can get into a state where your terminal acts strangely. For example, each letter may be typed twice, or the RETURN may not cause a line feed or a return to the left margin. You can often fix this by logging out and logging back in. Or you can read the description of the command **stty** in section I of the manual. To get intelligent treatment of tab characters (which are much used in UNIX) if your terminal doesn't have tabs, type the command

stty −tabs

and the system will convert each tab into the right number of blanks for you. If your terminal does have computer-settable tabs, the command **tabs** will set the stops correctly for you.

Mistakes in Typing

If you make a typing mistake, and see it before RETURN has been typed, there are two ways to recover. The sharp-character # erases the last character typed; in fact successive uses of # erase characters back to the beginning of the line (but not beyond). So if you type badly, you can correct as you go:

dd#atte##e

is the same as **date**.

The at-sign @ erases all of the characters typed so far on the current input line, so if the line is irretrievably fouled up, type an @ and start the line over.

What if you must enter a sharp or at-sign as part of the text? If you precede either # or @ by a backslash \, it loses its erase meaning. So to enter a sharp or at-sign in something, type \# or \@. The system will always echo a newline at you after your at-sign, even if preceded by a backslash. Don't worry — the at-sign has been recorded.

To erase a backslash, you have to type two sharps or two at-signs, as in \##. The backslash is used extensively in UNIX to indicate that the following character is in some way special.

Read-ahead

UNIX has full read-ahead, which means that you can type as fast as you want, whenever you want, even when some command is typing at you. If you type during output, your input characters will appear intermixed with the output characters, but they will be stored away and interpreted in the correct order. So you can type several commands one after another without waiting for the first to finish or even begin.

Stopping a Program

You can stop most programs by typing the character "DEL" (perhaps called "delete" or "rubout" on your terminal). The "interrupt" or "break" key found on most terminals can also be used. In a few programs, like the text editor, DEL stops whatever the program is doing but leaves you in that program. Hanging up the phone will stop most programs.

Logging Out

The easiest way to log out is to hang up the phone. You can also type

login

and let someone else use the terminal you were on. It is usually not sufficient just to turn off the terminal. Most UNIX systems do not use a time-out mechanism, so you'll be there forever unless you hang up.

Mail

When you log in, you may sometimes get the message

You have mail.

UNIX provides a postal system so you can communicate with other users of the system. To read your mail, type the command

mail

Your mail will be printed, one message at a time, most recent message first. After each message, **mail** waits for you to say what to do with it. The two basic responses are **d**, which deletes the message, and RETURN, which does not (so it will still be there the next time you read your mailbox). Other responses are described in the manual. (Earlier versions of **mail** do not process one message at a time, but are otherwise similar.)

How do you send mail to someone else? Suppose it is to go to "joe" (assuming "joe" is someone's login name). The easiest way is this:

mail joe
now type in the text of the letter
on as many lines as you like ...
After the last line of the letter
type the character "control-d",
that is, hold down "control" and type
a letter "d".

And that's it. The "control-d" sequence, often called "EOF" for end-of-file, is used throughout the system to mark the end of input from a terminal, so you might as well get used to it.

For practice, send mail to yourself. (This isn't as strange as it might sound — mail to oneself is a handy reminder mechanism.)

There are other ways to send mail — you can send a previously prepared letter, and you can mail to a number of people all at once. For more details see **mail**(1). (The notation **mail**(1) means the command **mail** in section 1 of the *UNIX Programmer's Manual.*)

Writing to other users

At some point, out of the blue will come a message like

Message from joe tty07...

accompanied by a startling beep. It means that Joe wants to talk to you, but unless you take explicit action you won't be able to talk back. To respond, type the command

write joe

This establishes a two-way communication path. Now whatever Joe types on his terminal will appear on yours and vice versa. The path is slow, rather like talking to the moon. (If you are in the middle of something, you have to get to a state where you can type a command. Normally, whatever program you are running has to terminate or be terminated. If you're editing, you can escape temporarily from the editor — read the editor tutorial.)

A protocol is needed to keep what you type from getting garbled up with what Joe types. Typically it's like this:

Joe types **write smith** and waits.
Smith types **write joe** and waits.
Joe now types his message (as many lines as he likes). When he's ready for a reply, he signals it by typing **(o)**, which stands for "over".
Now Smith types a reply, also terminated by **(o)**.
This cycle repeats until someone gets tired; he then signals his intent to quit with **(oo)**, for "over and out".
To terminate the conversation, each side must type a "control-d" character alone on a line. ("Delete" also works.) When the other person types his "control-d", you will get the message **EOF** on your terminal.

If you write to someone who isn't logged in, or who doesn't want to be disturbed, you'll be told. If the target is logged in but doesn't answer after a decent interval, simply type "control-d".

On-line Manual

The *UNIX Programmer's Manual* is typically kept on-line. If you get stuck on something, and can't find an expert to assist you, you can print on your terminal some manual section that might help. This is also useful for getting the most up-to-date information on a command. To print a manual section, type "man command-name". Thus to read up on the **who** command, type

man who

and, of course,

man man

tells all about the **man** command.

Computer Aided Instruction

Your UNIX system may have available a program called **learn**, which provides computer aided instruction on the file system and basic commands, the editor, document preparation, and even C programming. Try typing the command

learn

If **learn** exists on your system, it will tell you what to do from there.

II. DAY-TO-DAY USE

Creating Files — The Editor

If you have to type a paper or a letter or a program, how do you get the information stored in the machine? Most of these tasks are done with the UNIX "text editor" **ed**. Since **ed** is thoroughly documented in **ed**(1) and explained in *A Tutorial Introduction to the UNIX Text Editor,* we won't spend any time here describing how to use it. All we want it for right now is to make some *files*. (A file is just a collection of information stored in the machine, a simplistic but adequate definition.)

To create a file called **junk** with some text in it, do the following:

ed junk	(invokes the text editor)
a	(command to "ed", to add text)
now type in	
whatever text you want ...	
.	(signals the end of adding text)

The "**.**" that signals the end of adding text must be at the beginning of a line by itself. Don't forget it, for until it is typed, no other **ed** commands will be recognized — everything you type will be treated as text to be added.

At this point you can do various editing operations on the text you typed in, such as correcting spelling mistakes, rearranging paragraphs and the like. Finally, you must write the information you have typed into a file with the editor command **w**:

w

ed will respond with the number of characters it wrote into the file **junk**.

Until the **w** command, nothing is stored permanently, so if you hang up and go home the information is lost.† But after **w** the information is there permanently; you can re-access it any time by typing

ed junk

Type a **q** command to quit the editor. (If you try to quit without writing, **ed** will print a **?** to remind you. A second **q** gets you out regardless.)

Now create a second file called **temp** in the same manner. You should now have two files, **junk** and **temp**.

What files are out there?

The **ls** (for "list") command lists the names (not contents) of any of the files that UNIX knows about. If you type

ls

the response will be

junk
temp

which are indeed the two files just created. The names are sorted into alphabetical order automatically, but other variations are possible. For example, the command

ls −t

causes the files to be listed in the order in which they were last changed, most recent first. The **−l** option gives a "long" listing:

ls −l

will produce something like

```
-rw-rw-rw- 1 bwk  41 Jul 22 2:56 junk
-rw-rw-rw- 1 bwk  78 Jul 22 2:57 temp
```

The date and time are of the last change to the file. The 41 and 78 are the number of characters (which should agree with the numbers you got from **ed**). **bwk** is the owner of the file, that is, the person who created it. The **−rw−rw−rw−** tells who has permission to read and write the file, in this case everyone.

Options can be combined: **ls −lt** gives the same thing as **ls −l**, but sorted into time order. You can also name the files you're interested in, and **ls** will list the information about them only. More details can be found in **ls**(1).

The use of optional arguments that begin with a minus sign, like **−t** and **−lt**, is a common convention for UNIX programs. In general, if a program accepts such optional arguments, they precede any filename arguments. It is also vital that you separate the various arguments with spaces: **ls−l** is not the same as **ls −l**.

† This is not strictly true — if you hang up while editing, the data you were working on is saved in a file called **ed.hup**, which you can continue with at your next session.

Printing Files

Now that you've got a file of text, how do you print it so people can look at it? There are a host of programs that do that, probably more than are needed.

One simple thing is to use the editor, since printing is often done just before making changes anyway. You can say

> **ed junk**
> **1,$p**

ed will reply with the count of the characters in **junk** and then print all the lines in the file. After you learn how to use the editor, you can be selective about the parts you print.

There are times when it's not feasible to use the editor for printing. For example, there is a limit on how big a file **ed** can handle (several thousand lines). Secondly, it will only print one file at a time, and sometimes you want to print several, one after another. So here are a couple of alternatives.

First is **cat**, the simplest of all the printing programs. **cat** simply prints on the terminal the contents of all the files named in a list. Thus

> **cat junk**

prints one file, and

> **cat junk temp**

prints two. The files are simply concatenated (hence the name "**cat**") onto the terminal.

pr produces formatted printouts of files. As with **cat**, **pr** prints all the files named in a list. The difference is that it produces headings with date, time, page number and file name at the top of each page, and extra lines to skip over the fold in the paper. Thus,

> **pr junk temp**

will print **junk** neatly, then skip to the top of a new page and print **temp** neatly.

pr can also produce multi-column output:

> **pr –3 junk**

prints **junk** in 3-column format. You can use any reasonable number in place of "3" and **pr** will do its best. **pr** has other capabilities as well; see **pr**(1).

It should be noted that **pr** is *not* a formatting program in the sense of shuffling lines around and justifying margins. The true formatters are **nroff** and **troff**, which we will get to in the section on document preparation.

There are also programs that print files on a high-speed printer. Look in your manual under **opr** and **lpr**. Which to use depends on what equipment is attached to your machine.

Shuffling Files About

Now that you have some files in the file system and some experience in printing them, you can try bigger things. For example, you can move a file from one place to another (which amounts to giving it a new name), like this:

> **mv junk precious**

This means that what used to be "junk" is now "precious". If you do an **ls** command now, you will get

> **precious**
> **temp**

Beware that if you move a file to another one that already exists, the already existing contents are lost forever.

If you want to make a *copy* of a file (that is, to have two versions of something), you can use the **cp** command:

> **cp precious temp1**

makes a duplicate copy of **precious** in **temp1**.

Finally, when you get tired of creating and moving files, there is a command to remove files from the file system, called **rm**.

> **rm temp temp1**

will remove both of the files named.

You will get a warning message if one of the named files wasn't there, but otherwise **rm**, like most UNIX commands, does its work silently. There is no prompting or chatter, and error messages are occasionally curt. This terseness is sometimes disconcerting to newcomers, but experienced users find it desirable.

What's in a Filename

So far we have used filenames without ever saying what's a legal name, so it's time for a couple of rules. First, filenames are limited to 14 characters, which is enough to be descriptive. Second, although you can use almost any character in a filename, common sense says you should stick to ones that are visible, and that you should probably avoid characters that might be used with other meanings. We have already seen, for example, that in the **ls** command, **ls –t** means to list in time order. So if you had a file whose name was **–t**, you would have a tough time listing it by name. Besides the minus sign, there are other characters which have special meaning. To avoid pitfalls, you would do well to use only letters, numbers and the period until you're familiar with the situation.

On to some more positive suggestions. Suppose you're typing a large document like a book. Logically this divides into many small pieces, like chapters and perhaps sections. Physically it must be divided too, for **ed** will not handle really big files. Thus you should type the document as a number of files. You might have a separate file for each chapter, called

```
chap1
chap2
etc...
```

Or, if each chapter were broken into several files, you might have

```
chap1.1
chap1.2
chap1.3
...
chap2.1
chap2.2
...
```

You can now tell at a glance where a particular file fits into the whole.

There are advantages to a systematic naming convention which are not obvious to the novice UNIX user. What if you wanted to print the whole book? You could say

pr chap1.1 chap1.2 chap1.3 ......

but you would get tired pretty fast, and would probably even make mistakes. Fortunately, there is a shortcut. You can say

pr chap*

The * means "anything at all," so this translates into "print all files whose names begin with **chap**", listed in alphabetical order.

This shorthand notation is not a property of the **pr** command, by the way. It is system-wide, a service of the program that interprets commands (the "shell," **sh**(1)). Using that fact, you can see how to list the names of the files in the book:

ls chap*

produces

```
chap1.1
chap1.2
chap1.3
...
```

The * is not limited to the last position in a filename — it can be anywhere and can occur several times. Thus

rm *junk* *temp*

removes all files that contain **junk** or **temp** as any part of their name. As a special case, * by itself matches every filename, so

pr *

prints all your files (alphabetical order), and

rm *

removes *all files*. (You had better be *very* sure that's what you wanted to say!)

The * is not the only pattern-matching feature available. Suppose you want to print only chapters 1 through 4 and 9. Then you can say

pr chap[12349]*

The [...] means to match any of the characters inside the brackets. A range of consecutive letters or digits can be abbreviated, so you can also do this with

pr chap[1–49]*

Letters can also be used within brackets: **[a–z]** matches any character in the range **a** through **z**.

The **?** pattern matches any single character, so

ls ?

lists all files which have single-character names, and

ls –l chap?.1

lists information about the first file of each chapter (**chap1.1**, **chap2.1**, etc.).

Of these niceties, * is certainly the most useful, and you should get used to it. The others are frills, but worth knowing.

If you should ever have to turn off the special meaning of *, ?, etc., enclose the entire argument in single quotes, as in

ls '?'

We'll see some more examples of this shortly.

What's in a Filename, Continued

When you first made that file called **junk**, how did the system know that there wasn't another **junk** somewhere else, especially since the person in the next office is also reading this tutorial? The answer is that generally each user has a private *directory*, which contains only the files that belong to him. When you log in, you are "in" your directory. Unless you take special action, when you create a new file, it is made in the directory that you are currently in; this is most often your own directory, and thus the file is unrelated to any other file of the same name that might exist in someone else's directory.

The set of all files is organized into a (usually big) tree, with your files located several branches into the tree. It is possible for you to "walk" around this tree, and to find any file in the system, by starting at the root of the tree and walking along the proper set of branches. Conversely, you can start where you are and walk toward the root.

Let's try the latter first. The basic tools is the command **pwd** ("print working directory"), which prints the name of the directory you are currently in.

Although the details will vary according to the system you are on, if you give the command **pwd**, it will print something like

/usr/your-name

This says that you are currently in the directory **your-name**, which is in turn in the directory **/usr**, which is in turn in the root directory called by convention just /. (Even if it's not called **/usr** on your system, you will get something analogous. Make the corresponding changes and read on.)

If you now type

ls /usr/your-name

you should get exactly the same list of file names as you get from a plain **ls**: with no arguments, **ls** lists the contents of the current directory; given the name of a directory, it lists the contents of that directory.

Next, try

ls /usr

This should print a long series of names, among which is your own login name **your-name**. On many systems, **usr** is a directory that contains the directories of all the normal users of the system, like you.

The next step is to try

ls /

You should get a response something like this (although again the details may be different):

bin
dev
etc
lib
tmp
usr

This is a collection of the basic directories of files that the system knows about; we are at the root of the tree.

Now try

cat /usr/your-name/junk

(if **junk** is still around in your directory). The name

/usr/your-name/junk

is called the **pathname** of the file that you normally think of as "junk". "Pathname" has an obvious meaning: it represents the full name of the path you have to follow from the root through the tree of directories to get to a particular file. It is a universal rule in the UNIX system that anywhere you can use an ordinary filename, you can use a pathname.

Here is a picture which may make this clearer:

```
                (root)
                / | \
               /  |  \
      bin    etc  usr   dev   tmp
     / | \  / | \ / | \ / | \ / | \
                 /  |  \
             adam  eve  mary
             /    / \      \
                 /   \     junk
              junk  temp
```

Notice that Mary's **junk** is unrelated to Eve's.

This isn't too exciting if all the files of interest are in your own directory, but if you work with someone else or on several projects concurrently, it becomes handy indeed. For example, your friends can print your book by saying

pr /usr/your-name/chap*

Similarly, you can find out what files your neighbor has by saying

ls /usr/neighbor-name

or make your own copy of one of his files by

cp /usr/your-neighbor/his-file yourfile

If your neighbor doesn't want you poking around in his files, or vice versa, privacy can be arranged. Each file and directory has read-write-execute permissions for the owner, a group, and everyone else, which can be set to control access. See **ls**(1) and **chmod**(1) for details. As a matter of observed fact, most users most of the time find openness of more benefit than privacy.

As a final experiment with pathnames, try

ls /bin /usr/bin

Do some of the names look familiar? When you run a program, by typing its name after the prompt character, the system simply looks for a file of that name. It normally looks first in your directory (where it typically doesn't find it), then in **/bin** and finally in **/usr/bin**. There is nothing magic about commands like **cat** or **ls**, except that they have been collected into a couple of places to be easy to find and administer.

What if you work regularly with someone else on common information in his directory? You could just log in as your friend each time you want to, but you can also say "I want to work on his files instead of my own". This is done by changing the directory that you are currently in:

cd /usr/your-friend

(On some systems, **cd** is spelled **chdir**.) Now when you use a filename in something like **cat** or **pr**, it refers to the file in your friend's directory. Changing directories doesn't affect any permissions associated with a file — if you couldn't access a file from your own directory, changing to another directory won't alter that fact. Of course, if you forget what directory you're in, type

pwd

to find out.

It is usually convenient to arrange your own files so that all the files related to one thing are in a directory separate from other projects. For example, when you write your book, you might want to keep all the text in a directory called **book**. So make one with

mkdir book

then go to it with

cd book

then start typing chapters. The book is now found in (presumably)

/usr/your-name/book

To remove the directory **book**, type

rm book/*
rmdir book

The first command removes all files from the directory; the second removes the empty directory.

You can go up one level in the tree of files by saying

cd ..

"**..**" is the name of the parent of whatever directory you are currently in. For completeness, "**.**" is an alternate name for the directory you are in.

Using Files instead of the Terminal

Most of the commands we have seen so far produce output on the terminal; some, like the editor, also take their input from the terminal. It is universal in UNIX systems that the terminal can be replaced by a file for either or both of input and output. As one example,

ls

makes a list of files on your terminal. But if you say

ls >filelist

a list of your files will be placed in the file **filelist** (which will be created if it doesn't already exist, or overwritten if it does). The symbol > means "put the output on the following file, rather than on the terminal." Nothing is produced on the terminal. As another example, you could combine several files into one by capturing the output of **cat** in a file:

cat f1 f2 f3 >temp

The symbol >> operates very much like > does, except that it means "add to the end of." That is,

cat f1 f2 f3 >>temp

means to concatenate **f1**, **f2** and **f3** to the end of whatever is already in **temp**, instead of overwriting the existing contents. As with >, if **temp** doesn't exist, it will be created for you.

In a similar way, the symbol < means to take the input for a program from the following file, instead of from the terminal. Thus, you could make up a script of commonly used editing commands and put them into a file called **script**. Then you can run the script on a file by saying

ed file <script

As another example, you can use **ed** to prepare a letter in file **let**, then send it to several people with

mail adam eve mary joe <let

Pipes

One of the novel contributions of the UNIX system is the idea of a *pipe*. A pipe is simply a way to connect the output of one program to the input of another program, so the two run as a sequence of processes — a pipeline.

For example,

pr f g h

will print the files **f**, **g**, and **h**, beginning each on a new page. Suppose you want them run together instead. You could say

cat f g h >temp
pr <temp
rm temp

but this is more work than necessary. Clearly what we want is to take the output of **cat** and connect it to the input of **pr**. So let us use a pipe:

cat f g h | pr

The vertical bar | means to take the output from

cat, which would normally have gone to the terminal, and put it into **pr** to be neatly formatted.

There are many other examples of pipes. For example,

ls | pr −3

prints a list of your files in three columns. The program **wc** counts the number of lines, words and characters in its input, and as we saw earlier, **who** prints a list of currently-logged on people, one per line. Thus

who | wc

tells how many people are logged on. And of course

ls | wc

counts your files.

Any program that reads from the terminal can read from a pipe instead; any program that writes on the terminal can drive a pipe. You can have as many elements in a pipeline as you wish.

Many UNIX programs are written so that they will take their input from one or more files if file arguments are given; if no arguments are given they will read from the terminal, and thus can be used in pipelines. **pr** is one example:

pr −3 a b c

prints files **a**, **b** and **c** in order in three columns. But in

cat a b c | pr −3

pr prints the information coming down the pipeline, still in three columns.

The Shell

We have already mentioned once or twice the mysterious "shell," which is in fact **sh**(1). The shell is the program that interprets what you type as commands and arguments. It also looks after translating *, etc., into lists of filenames, and <, >, and | into changes of input and output streams.

The shell has other capabilities too. For example, you can run two programs with one command line by separating the commands with a semicolon; the shell recognizes the semicolon and breaks the line into two commands. Thus

date; who

does both commands before returning with a prompt character.

You can also have more than one program running *simultaneously* if you wish. For example, if you are doing something time-consuming, like the editor script of an earlier section, and you don't want to wait around for the results before starting something else, you can say

ed file <script &

The ampersand at the end of a command line says "start this command running, then take further commands from the terminal immediately," that is, don't wait for it to complete. Thus the script will begin, but you can do something else at the same time. Of course, to keep the output from interfering with what you're doing on the terminal, it would be better to say

ed file <script >script.out &

which saves the output lines in a file called **script.out**.

When you initiate a command with **&**, the system replies with a number called the process number, which identifies the command in case you later want to stop it. If you do, you can say

kill process-number

If you forget the process number, the command **ps** will tell you about everything you have running. (If you are desperate, **kill 0** will kill all your processes.) And if you're curious about other people, **ps a** will tell you about *all* programs that are currently running.

You can say

(command-1; command-2; command-3) &

to start three commands in the background, or you can start a background pipeline with

command-1 | command-2 &

Just as you can tell the editor or some similar program to take its input from a file instead of from the terminal, you can tell the shell to read a file to get commands. (Why not? The shell, after all, is just a program, albeit a clever one.) For instance, suppose you want to set tabs on your terminal, and find out the date and who's on the system every time you log in. Then you can put the three necessary commands (**tabs**, **date**, **who**) into a file, let's call it **startup**, and then run it with

sh startup

This says to run the shell with the file **startup** as input. The effect is as if you had typed the contents of **startup** on the terminal.

If this is to be a regular thing, you can eliminate the need to type **sh**: simply type, once only, the command

chmod +x startup

and thereafter you need only say

startup

to run the sequence of commands. The **chmod**(1) command marks the file executable; the shell recognizes this and runs it as a sequence of commands.

If you want **startup** to run automatically every time you log in, create a file in your login directory called **.profile**, and place in it the line **startup**. When the shell first gains control when you log in, it looks for the **.profile** file and does whatever commands it finds in it. We'll get back to the shell in the section on programming.

III. DOCUMENT PREPARATION

UNIX systems are used extensively for document preparation. There are two major formatting programs, that is, programs that produce a text with justified right margins, automatic page numbering and titling, automatic hyphenation, and the like. **nroff** is designed to produce output on terminals and line-printers. **troff** (pronounced "tee-roff") instead drives a phototypesetter, which produces very high quality output on photographic paper. This paper was formatted with **troff**.

Formatting Packages

The basic idea of **nroff** and **troff** is that the text to be formatted contains within it "formatting commands" that indicate in detail how the formatted text is to look. For example, there might be commands that specify how long lines are, whether to use single or double spacing, and what running titles to use on each page.

Because **nroff** and **troff** are relatively hard to learn to use effectively, several "packages" of canned formatting requests are available to let you specify paragraphs, running titles, footnotes, multi-column output, and so on, with little effort and without having to learn **nroff** and **troff**. These packages take a modest effort to learn, but the rewards for using them are so great that it is time well spent.

In this section, we will provide a hasty look at the "manuscript" package known as **−ms**. Formatting requests typically consist of a period and two upper-case letters, such as **.TL**, which is used to introduce a title, or **.PP** to begin a new paragraph.

A document is typed so it looks something like this:

```
.TL
title of document
.AU
author name
.SH
section heading
.PP
paragraph ...
.PP
another paragraph ...
.SH
another section heading
.PP
etc.
```

The lines that begin with a period are the formatting requests. For example, **.PP** calls for starting a new paragraph. The precise meaning of **.PP** depends on what output device is being used (typesetter or terminal, for instance), and on what publication the document will appear in. For example, **−ms** normally assumes that a paragraph is preceded by a space (one line in **nroff**, ½ line in **troff**), and the first word is indented. These rules can be changed if you like, but they are changed by changing the interpretation of **.PP**, not by re-typing the document.

To actually produce a document in standard format using **−ms**, use the command

troff −ms files ...

for the typesetter, and

nroff −ms files ...

for a terminal. The **−ms** argument tells **troff** and **nroff** to use the manuscript package of formatting requests.

There are several similar packages; check with a local expert to determine which ones are in common use on your machine.

Supporting Tools

In addition to the basic formatters, there is a host of supporting programs that help with document preparation. The list in the next few paragraphs is far from complete, so browse through the manual and check with people around you for other possibilities.

eqn and **neqn** let you integrate mathematics into the text of a document, in an easy-to-learn language that closely resembles the way you would speak it aloud. For example, the **eqn** input

sum from i=0 to n x sub i ~=~ pi over 2

produces the output

$$\sum_{i=0}^{n} x_i = \frac{\pi}{2}$$

The program **tbl** provides an analogous service for preparing tabular material; it does all the computations necessary to align complicated columns with elements of varying widths.

refer prepares bibliographic citations from a data base, in whatever style is defined by the formatting package. It looks after all the details of numbering references in sequence, filling in page and volume numbers, getting the author's initials and the journal name right, and so on.

spell and **typo** detect possible spelling mistakes in a document. **spell** works by comparing the words in your document to a dictionary, printing those that are not in the dictionary. It knows enough about English spelling to detect plurals and the like, so it does a very good job. **typo** looks for words which are "unusual", and prints those. Spelling mistakes tend to be more unusual, and thus show up early when the most unusual words are printed first.

grep looks through a set of files for lines that contain a particular text pattern (rather like the editor's context search does, but on a bunch of files). For example,

grep 'ing$' chap*

will find all lines that end with the letters **ing** in the files **chap***. (It is almost always a good practice to put single quotes around the pattern you're searching for, in case it contains characters like * or $ that have a special meaning to the shell.) **grep** is often useful for finding out in which of a set of files the misspelled words detected by **spell** are actually located.

diff prints a list of the differences between two files, so you can compare two versions of something automatically (which certainly beats proofreading by hand).

wc counts the words, lines and characters in a set of files. **tr** translates characters into other characters; for example it will convert upper to lower case and vice versa. This translates upper into lower:

tr A–Z a–z <input >output

sort sorts files in a variety of ways; **cref** makes cross-references; **ptx** makes a permuted index (keyword-in-context listing). **sed** provides many of the editing facilities of **ed**, but can apply them to arbitrarily long inputs. **awk** provides the ability to do both pattern matching and numeric computations, and to conveniently process fields within lines. These programs are for more advanced users, and they are not limited to document preparation. Put them on your list of things to learn about.

Most of these programs are either independently documented (like **eqn** and **tbl**), or are sufficiently simple that the description in the *UNIX Programmer's Manual* is adequate explanation.

Hints for Preparing Documents

Most documents go through several versions (always more than you expected) before they are finally finished. Accordingly, you should do whatever possible to make the job of changing them easy.

First, when you do the purely mechanical operations of typing, type so that subsequent editing will be easy. Start each sentence on a new line. Make lines short, and break lines at natural places, such as after commas and semicolons, rather than randomly. Since most people change documents by rewriting phrases and adding, deleting and rearranging sentences, these precautions simplify any editing you have to do later.

Keep the individual files of a document down to modest size, perhaps ten to fifteen thousand characters. Larger files edit more slowly, and of course if you make a dumb mistake it's better to have clobbered a small file than a big one. Split into files at natural boundaries in the document, for the same reasons that you start each sentence on a new line.

The second aspect of making change easy is to not commit yourself to formatting details too early. One of the advantages of formatting packages like **–ms** is that they permit you to delay decisions to the last possible moment. Indeed, until a document is printed, it is not even decided whether it will be typeset or put on a line printer.

As a rule of thumb, for all but the most trivial jobs, you should type a document in terms of a set of requests like **.PP**, and then define them appropriately, either by using one of the canned packages (the better way) or by defining your own **nroff** and **troff** commands. As long as you have entered the text in some systematic way, it can always be cleaned up and re-formatted by a judicious combination of editing commands and request definitions.

IV. PROGRAMMING

There will be no attempt made to teach any of the programming languages available but a few words of advice are in order. One of the reasons why the UNIX system is a productive programming environment is that there is already a rich set of tools available, and facilities like pipes, I/O redirection, and the capabilities of the shell often make it possible to do a job by pasting together programs that already exist instead of writing from scratch.

The Shell

The pipe mechanism lets you fabricate quite complicated operations out of spare parts that already exist. For example, the first draft of the **spell** program was (roughly)

cat ...	*collect the files*
|tr ...	*put each word on a new line*
|tr ...	*delete punctuation, etc.*
|sort	*into dictionary order*
|uniq	*discard duplicates*
|comm	*print words in text but not in dictionary*

More pieces have been added subsequently, but this goes a long way for such a small effort.

The editor can be made to do things that would normally require special programs on other systems. For example, to list the first and last lines of each of a set of files, such as a book, you could laboriously type

```
ed
e chap1.1
1p
$p
e chap1.2
1p
$p
etc.
```

But you can do the job much more easily. One way is to type

```
ls chap* >temp
```

to get the list of filenames into a file. Then edit this file to make the necessary series of editing commands (using the global commands of **ed**), and write it into **script**. Now the command

```
ed <script
```

will produce the same output as the laborious hand typing. Alternately (and more easily), you can use the fact that the shell will perform loops, repeating a set of commands over and over again for a set of arguments:

```
for i in chap*
do
        ed $i <script
done
```

This sets the shell variable **i** to each file name in turn, then does the command. You can type this command at the terminal, or put it in a file for later execution.

Programming the Shell

An option often overlooked by newcomers is that the shell is itself a programming language, with variables, control flow (**if-else**, **while**, **for**, **case**), subroutines, and interrupt handling. Since there are many building-block programs, you can sometimes avoid writing a new program merely by piecing together some of the building blocks with shell command files.

We will not go into any details here; examples and rules can be found in *An Introduction to the UNIX Shell*, by S. R. Bourne.

Programming in C

If you are undertaking anything substantial, C is the only reasonable choice of programming language: everything in the UNIX system is tuned to it. The system itself is written in C, as are most of the programs that run on it. It is also a easy language to use once you get started. C is introduced and fully described in *The C Programming Language* by B. W. Kernighan and D. M. Ritchie (Prentice-Hall, 1978). Several sections of the manual describe the system interfaces, that is, how you do I/O and similar functions. Read *UNIX Programming* for more complicated things.

Most input and output in C is best handled with the standard I/O library, which provides a set of I/O functions that exist in compatible form on most machines that have C compilers. In general, it's wisest to confine the system interactions in a program to the facilities provided by this library.

C programs that don't depend too much on special features of UNIX (such as pipes) can be moved to other computers that have C compilers. The list of such machines grows daily; in addition to the original PDP-11, it currently includes at least Honeywell 6000, IBM 370, Interdata 8/32, Data General Nova and Eclipse, HP 2100, Harris /7, VAX 11/780, SEL 86, and Zilog Z80. Calls to the standard I/O library will work on all of these machines.

There are a number of supporting programs that go with C. **lint** checks C programs for potential portability problems, and detects errors such as mismatched argument types and uninitialized variables.

For larger programs (anything whose source is on more than one file) **make** allows you to specify the dependencies among the source files and the processing steps needed to make a new version; it then checks the times that the pieces were last changed and does the minimal amount of recompiling to create a consistent updated version.

The debugger **adb** is useful for digging through the dead bodies of C programs, but is rather hard to learn to use effectively. The most effective debugging tool is still careful thought, coupled with judiciously placed print statements.

The C compiler provides a limited instrumentation service, so you can find out where programs spend their time and what parts are worth optimizing. Compile the routines with the **−p** option; after the test run, use **prof** to print an execution profile. The command **time** will give you the gross run-time statistics of a program, but they are not super accurate or reproducible.

Other Languages

If you *have* to use Fortran, there are two possibilities. You might consider Ratfor, which gives you the decent control structures and free-form input that characterize C, yet lets you write code that is still portable to other environments. Bear in mind that UNIX Fortran tends to produce large and relatively slow-running programs. Furthermore, supporting software like **adb**, **prof**, etc., are all virtually useless with Fortran programs. There may also be a Fortran 77 compiler on your system. If so, this is a viable alternative to Ratfor, and has the non-trivial advantage that it is compatible with C and related programs. (The Ratfor processor and C tools can be used with Fortran 77 too.)

If your application requires you to translate a language into a set of actions or another language, you are in effect building a compiler, though probably a small one. In that case, you should be using the **yacc** compiler-compiler, which helps you develop a compiler quickly. The **lex** lexical analyzer generator does the same job for the simpler languages that can be expressed as regular expressions. It can be used by itself, or as a front end to recognize inputs for a **yacc**-based program. Both **yacc** and **lex** require some sophistication to use, but the initial effort of learning them can be repaid many times over in programs that are easy to change later on.

Most UNIX systems also make available other languages, such as Algol 68, APL, Basic, Lisp, Pascal, and Snobol. Whether these are useful depends largely on the local environment: if someone cares about the language and has worked on it, it may be in good shape. If not, the odds are strong that it will be more trouble than it's worth.

V. UNIX READING LIST

General:

K. L. Thompson and D. M. Ritchie, *The UNIX Programmer's Manual,* Bell Laboratories, 1978. Lists commands, system routines and interfaces, file formats, and some of the maintenance procedures. You can't live without this, although you will probably only need to read section 1.

Documents for Use with the UNIX Time-sharing System. Volume 2 of the Programmer's Manual. This contains more extensive descriptions of major commands, and tutorials and reference manuals. All of the papers listed below are in it, as are descriptions of most of the programs mentioned above.

D. M. Ritchie and K. L. Thompson, "The UNIX Time-sharing System," CACM, July 1974. An overview of the system, for people interested in operating systems. Worth reading by anyone who programs. Contains a remarkable number of one-sentence observations on how to do things right.

The Bell System Technical Journal (BSTJ) Special Issue on UNIX, July/August, 1978, contains many papers describing recent developments, and some retrospective material.

The 2nd International Conference on Software Engineering (October, 1976) contains several papers describing the use of the Programmer's Workbench (PWB) version of UNIX.

Document Preparation:

B. W. Kernighan, "A Tutorial Introduction to the UNIX Text Editor" and "Advanced Editing on UNIX," Bell Laboratories, 1978. Beginners need the introduction; the advanced material will help you get the most out of the editor.

M. E. Lesk, "Typing Documents on UNIX," Bell Laboratories, 1978. Describes the **−ms** macro package, which isolates the novice from the vagaries of **nroff** and **troff**, and takes care of most formatting situations. If this specific package isn't available on your system, something similar probably is. The most likely alternative is the PWB/UNIX macro package **−mm**; see your local guru if you use PWB/UNIX.

B. W. Kernighan and L. L. Cherry, "A System for Typesetting Mathematics," Bell Laboratories Computing Science Tech. Rep. 17.

M. E. Lesk, "Tbl — A Program to Format Tables," Bell Laboratories CSTR 49, 1976.

J. F. Ossanna, Jr., "NROFF/TROFF User's Manual," Bell Laboratories CSTR 54, 1976. **troff** is the basic formatter used by **−ms**, **eqn** and **tbl**. The reference manual is indispensable if you are going to write or maintain these or similar programs. But start with:

B. W. Kernighan, "A TROFF Tutorial," Bell Laboratories, 1976. An attempt to unravel the intricacies of **troff**.

Programming:

B. W. Kernighan and D. M. Ritchie, *The C Programming Language,* Prentice-Hall, 1978. Contains a tutorial introduction, complete discussions of all language features, and the reference manual.

B. W. Kernighan and D. M. Ritchie, "UNIX Programming," Bell Laboratories, 1978. Describes

how to interface with the system from C programs: I/O calls, signals, processes.

S. R. Bourne, "An Introduction to the UNIX Shell," Bell Laboratories, 1978. An introduction and reference manual for the Version 7 shell. Mandatory reading if you intend to make effective use of the programming power of this shell.

S. C. Johnson, "Yacc — Yet Another Compiler-Compiler," Bell Laboratories CSTR 32, 1978.

M. E. Lesk, "Lex — A Lexical Analyzer Generator," Bell Laboratories CSTR 39, 1975.

S. C. Johnson, "Lint, a C Program Checker," Bell Laboratories CSTR 65, 1977.

S. I. Feldman, "MAKE — A Program for Maintaining Computer Programs," Bell Laboratories CSTR 57, 1977.

J. F. Maranzano and S. R. Bourne, "A Tutorial Introduction to ADB," Bell Laboratories CSTR 62, 1977. An introduction to a powerful but complex debugging tool.

S. I. Feldman and P. J. Weinberger, "A Portable Fortran 77 Compiler," Bell Laboratories, 1978. A full Fortran 77 for UNIX systems.

LEARN — Computer-Aided Instruction on UNIX (Second Edition)

Brian W. Kernighan

Michael E. Lesk

ABSTRACT

This paper describes the second version of the *learn* program for interpreting CAI scripts on the UNIX† operating system, and a set of scripts that provide a computerized introduction to the system.

Six current scripts cover basic commands and file handling, the editor, additional file handling commands, the *eqn* program for mathematical typing, the "–ms" package of formatting macros, and an introduction to the C programming language. These scripts now include a total of about 530 lessons.

Many users from a wide variety of backgrounds have used *learn* to acquire basic UNIX skills. Most usage involves the first two scripts, an introduction to UNIX files and commands, and the UNIX editor.

The second version of *learn* is about four times faster than the previous one in CPU utilization, and much faster in perceived time because of better overlap of computing and printing. It also requires less file space than the first version. Many of the lessons have been revised; new material has been added to reflect changes and enhancements in UNIX itself. Script-writing is also easier because of revisions to the script language.

1. Introduction.

Learn is a driver for CAI scripts. It is intended to permit the easy composition of lessons and lesson fragments to teach people computer skills. Since it is teaching the same system on which it is implemented, it makes direct use of UNIX facilities to create a controlled UNIX environment. The system includes two main parts: (1) a driver that interprets the lesson scripts; and (2) the lesson scripts themselves. At present there are seven scripts:

— basic file handling commands

— the UNIX text editors *ed* and *vi*

— advanced file handling

— the *eqn* language for typing mathematics

— the "ms" macro package for document formatting

— the C programming language

The purported advantages of CAI scripts for training in computer skills include the following:

(a) students are forced to perform the exercises that are in fact the basis of training in any case;

† UNIX is a trademark of Bell Laboratories.

(b) students receive immediate feedback and confirmation of progress;

(c) students may progress at their own rate;

(d) no schedule requirements are imposed; students may study at any time convenient for them;

(e) the lessons may be improved individually and the improvements are immediately available to new users;

(f) since the student has access to a computer for the CAI script there is a place to do exercises;

(g) the use of high technology will improve student motivation and the interest of their management.

Opposed to this, of course, is the absence of anyone to whom the student may direct questions. If CAI is used without a "counselor" or other assistance, it should properly be compared to a textbook, lecture series, or taped course, rather than to a seminar. CAI has been used for many years in a variety of educational areas.[1,2,3] The use of a computer to teach computer use itself, however, offers unique advantages. The skills developed to get through the script are exactly those needed to use the computer; there is no waste effort.

The scripts written so far are based on some familiar assumptions about education; these assumptions are outlined in the next section. The remaining sections describe the operation of the script driver and the particular scripts now available. The driver puts few restrictions on the script writer, but the current scripts are of a rather rigid and stereotyped form in accordance with the theory in the next section and practical limitations.

2. Educational Assumptions and Design.

First, the way to teach people how to do something is to have them do it. Scripts should not contain long pieces of explanation; they should instead frequently ask the student to do some task. So teaching is always by example: the typical script fragment shows a small example of some technique and then asks the user to either repeat that example or produce a variation on it. All are intended to be easy enough that most students will get most questions right, reinforcing the desired behavior.

Most lessons fall into one of three types. The simplest presents a lesson and asks for a yes or no answer to a question. The student is given a chance to experiment before replying. The script checks for the correct reply. Problems of this form are sparingly used.

The second type asks for a word or number as an answer. For example a lesson on files might say

> *How many files are there in the current directory? Type "answer N", where N is the number of files.*

The student is expected to respond (perhaps after experimenting) with

> *answer 17*

or whatever. Surprisingly often, however, the idea of a substitutable argument (i.e., replacing *N* by 17) is difficult for non-programmer students, so the first few such lessons need real care.

The third type of lesson is open-ended — a task is set for the student, appropriate parts of the input or output are monitored, and the student types *ready* when the task is done. Figure 1 shows a sample dialog that illustrates the last of these, using two lessons about the *cat* (concatenate, i.e., print) command taken from early in the script that teaches file handling. Most *learn* lessons are of this form.

After each correct response the computer congratulates the student and indicates the lesson number that has just been completed, permitting the student to restart the script after that lesson. If the answer is wrong, the student is offered a chance to repeat the lesson. The "speed" rating of the student (explained in section 5) is given after the lesson number when the lesson is completed successfully; it is printed only for the aid of script authors checking out possible errors in the lessons.

Figure 1: Sample dialog from basic files script

(Student responses in italics; '$' is the prompt)

```
A file can be printed on your terminal
by using the "cat" command.  Just say
"cat file" where "file" is the file name.
For example, there is a file named
"food" in this directory.  List it
by saying "cat food"; then type "ready".
$ cat food
  this is the file
  named food.
$ ready

Good.  Lesson 3.3a (1)

Of course, you can print any file with "cat".
In particular, it is common to first use
"ls" to find the name of a file and then "cat"
to print it.  Note the difference between
"ls", which tells you the name of the file,
and "cat", which tells you the contents.
One file in the current directory is named for
a President.  Print the file, then type "ready".
$ cat President
cat: can't open President
$ ready

Sorry, that's not right.  Do you want to try again? yes
Try the problem again.
$ ls
.ocopy
X1
roosevelt
$ cat roosevelt
  this file is named roosevelt
  and contains three lines of
  text.
$ ready

Good.  Lesson 3.3b (0)

The "cat" command can also print several files
at once.  In fact, it is named "cat" as an abbreviation
for "concatenate"....
```

It is assumed that there is no foolproof way to determine if the student truly "understands" what he or she is doing; accordingly, the current *learn* scripts only measure performance, not comprehension. If the student can perform a given task, that is deemed to be "learning."[4]

The main point of using the computer is that what the student does is checked for correctness immediately. Unlike many CAI scripts, however, these scripts provide few facilities for dealing with wrong answers. In practice, if most of the answers are not right the script is a failure; the universal solution to student error is to provide a new, easier script. Anticipating possible wrong answers is an endless job, and it is really easier as well as better to provide a simpler script.

Along with this goes the assumption that anything can be taught to anybody if it can be broken into sufficiently small pieces. Anything not absorbed in a single chunk is just subdivided.

To avoid boring the faster students, however, an effort is made in the files and editor scripts to provide three tracks of different difficulty. The fastest sequence of lessons is aimed at roughly the bulk and speed of a typical tutorial manual and should be adequate for review and for well-prepared students. The next track is intended for most users and is roughly twice as long. Typically, for example, the fast track might present an idea and ask for a variation on the example shown; the normal track will first ask the student to repeat the example that was shown before attempting a variation. The third and slowest track, which is often three or four times the length of the fast track, is intended to be adequate for anyone. (The lessons of Figure 1 are from the third track.) The multiple tracks also mean that a student repeating a course is unlikely to hit the same series of lessons; this makes it profitable for a shaky user to back up and try again, and many students have done so.

The tracks are not completely distinct, however. Depending on the number of correct answers the student has given for the last few lessons, the program may switch tracks. The driver is actually capable of following an arbitrary directed graph of lesson sequences, as discussed in section 5. Some more structured arrangement, however, is used in all current scripts to aid the script writer in organizing the material into lessons. It is sufficiently difficult to write lessons that the three-track theory is not followed very closely except in the files and editor scripts. Accordingly, in some cases, the fast track is produced merely by skipping lessons from the slower track. In others, there is essentially only one track.

The main reason for using the *learn* program rather than simply writing the same material as a workbook is not the selection of tracks, but actual hands-on experience. Learning by doing is much more effective than pencil and paper exercises.

Learn also provides a mechanical check on performance. The first version in fact would not let the student proceed unless it received correct answers to the questions it set and it would not tell a student the right answer. This somewhat Draconian approach has been moderated in version 2. Lessons are sometimes badly worded or even just plain wrong; in such cases, the student has no recourse. But if a student is simply unable to complete one lesson, that should not prevent access to the rest. Accordingly, the current version of *learn* allows the student to skip a lesson that he cannot pass; a "no" answer to the "Do you want to try again?" question in Figure 1 will pass to the next lesson. It is still true that *learn* will not tell the student the right answer.

Of course, there are valid objections to the assumptions above. In particular, some students may object to not understanding what they are doing; and the procedure of smashing everything into small pieces may provoke the retort "you can't cross a ditch in two jumps." Since writing CAI scripts is considerably more tedious than ordinary manuals, however, it is safe to assume that there will always be alternatives to the scripts as a way of learning. In fact, for a reference manual of 3 or 4 pages it would not be surprising to have a tutorial manual of 20 pages and a (multi-track) script of 100 pages. Thus the reference manual will exist long before the scripts.

3. Scripts.

As mentioned above, the present scripts try at most to follow a three-track theory. Thus little of the potential complexity of the possible directed graph is employed, since care must be taken in lesson construction to see that every necessary fact is presented in every possible path through the units. In addition, it is desirable that every unit have alternate successors to deal with student errors.

In most existing courses, the first few lessons are devoted to checking prerequisites. For example, before the student is allowed to proceed through the editor script the script verifies that the student understands files and is able to type. It is felt that the sooner lack of student preparation is detected, the easier it will be on the student. Anyone proceeding through the scripts should be getting mostly correct answers; otherwise, the system will be unsatisfactory both because the wrong habits are being learned and because the scripts make little effort to deal with wrong answers. Unprepared students should not be encouraged to continue with scripts.

There are some preliminary items which the student must know before any scripts can be tried. In particular, the student must know how to connect to a UNIX system, set the terminal properly, log in, and execute simple commands (e.g., *learn* itself). In addition, the character erase and line kill conventions (# and @) should be known. It is hard to see how this much could be taught by computer-aided instruction, since a student who does not know these basic skills will not be able to run the learning program. A brief description on paper is provided (see Appendix A), although assistance will be needed for the first few minutes. This assistance, however, need not be highly skilled.

The first script in the current set deals with files. It assumes the basic knowledge above and teaches the student about the *ls*, *cat*, *mv*, *rm*, *cp* and *diff* commands. It also deals with the abbreviation characters *, ?, and [ ] in file names. It does not cover pipes or I/O redirection, nor does it present the many options on the *ls* command.

This script contains 31 lessons in the fast track; two are intended as prerequisite checks, seven are review exercises. There are a total of 75 lessons in all three tracks, and the instructional passages typed at the student to begin each lesson total 4,476 words. The average lesson thus begins with a 60-word message. In general, the fast track lessons have somewhat longer introductions, and the slow tracks somewhat shorter ones. The longest message is 144 words and the shortest 14.

The second script trains students in the use of the UNIX context editor *ed*, a sophisticated editor using regular expressions for searching.[5] All editor features except encryption, mark names and ';' in addressing are covered. The fast track contains 2 prerequisite checks, 93 lessons, and a review lesson. It is supplemented by 146 additional lessons in other tracks.

A comparison of sizes may be of interest. The *ed* description in the reference manual is 2,572 words long. The *ed* tutorial[6] is 6,138 words long. The fast track through the *ed* script is 7,407 words of explanatory messages, and the total *ed* script, 242 lessons, has 15,615 words. The average *ed* lesson is thus also about 60 words; the largest is 171 words and the smallest 10. The original *ed* script represents about three man-weeks of effort.

The advanced file handling script deals with *ls* options, I/O diversion, pipes, and supporting programs like *pr*, *wc*, *tail*, *spell* and *grep*. (The basic file handling script is a prerequisite.) It is not as refined as the first two scripts; this is reflected at least partly in the fact that it provides much less of a full three-track sequence than they do. On the other hand, since it is perceived as "advanced," it is hoped that the student will have somewhat more sophistication and be better able to cope with it at a reasonably high level of performance.

A fourth script covers the *eqn* language for typing mathematics. This script must be run on a terminal capable of printing mathematics, for instance the DASI 300 and similar Diablo-based terminals, or the nearly extinct Model 37 teletype. Again, this script is relatively short of tracks: of 76 lessons, only 17 are in the second track and 2 in the third track. Most of these provide additional practice for students who are having trouble in the first track.

The *-ms* script for formatting macros is a short one-track only script. The macro package it describes is no longer the standard, so this script will undoubtedly be superseded in the future. Furthermore, the linear style of a single learn script is somewhat inappropriate for the macros, since the macro package is composed of many independent features, and few users need all of them. It would be better to have a selection of short lesson sequences dealing with the features independently.

The script on C is in a state of transition. It was originally designed to follow a tutorial on C, but that document has since become obsolete. The current script has been partially converted to follow the order of presentation in *The C Programming Language*,[7] but this job is not complete. The C script was never intended to teach C; rather it is supposed to be a series of exercises for which the computer provides checking and (upon success) a suggested solution.

This combination of scripts covers much of the material which any UNIX user will need to know to make effective use of the system. With enlargement of the advanced files course to include more on the command interpreter, there will be a relatively complete introduction to UNIX available via *learn*. Although we make no pretense that *learn* will replace other instructional materials, it should provide a useful supplement to existing tutorials and reference manuals.

4. Experience with Students.

Learn has been installed on many different UNIX systems. Most of the usage is on the first two scripts, so these are more thoroughly debugged and polished. As a (random) sample of user experience, the *learn* program has been used at Bell Labs at Indian Hill for 10,500 lessons in a four month period. About 3600 of these are in the files script, 4100 in the editor, and 1400 in advanced files. The passing rate is about 80%, that is, about 4 lessons are passed for every one failed. There have been 86 distinct users of the files script, and 58 of the editor. On our system at Murray Hill, there have been nearly 2000 lessons over two weeks that include Christmas and New Year. Users have ranged in age from six up.

It is difficult to characterize typical sessions with the scripts; many instances exist of someone doing one or two lessons and then logging out, as do instances of someone pausing in a script for twenty minutes or more. In the earlier version of *learn*, the average session in the files course took 32 minutes and covered 23 lessons. The distribution is quite broad and skewed, however; the longest session was 130 minutes and there were five sessions shorter than five minutes. The average lesson took about 80 seconds. These numbers are roughly typical for non-programmers; a UNIX expert can do the scripts at approximately 30 seconds per lesson, most of which is the system printing.

At present working through a section of the middle of the files script took about 1.4 seconds of processor time per lesson, and a system expert typing quickly took 15 seconds of real time per lesson. A novice would probably take at least a minute. Thus a UNIX system could support ten students working simultaneously with some spare capacity.

5. The Script Interpreter.

The *learn* program itself merely interprets scripts. It provides facilities for the script writer to capture student responses and their effects, and simplifies the job of passing control to and recovering control from the student. This section describes the operation and usage of the driver program, and indicates what is required to produce a new script. Readers only interested in the existing scripts may skip this section.

The file structure used by *learn* is shown in Figure 2. There is one parent directory (named *lib*) containing the script data. Within this directory are subdirectories, one for each subject in which a course is available, one for logging (named *log*), and one in which user sub-directories are created (named *play*). The subject directory contains master copies of all lessons, plus any supporting material for that subject. In a given subdirectory, each lesson is a single text file. Lessons are usually named systematically; the file that contains lesson n is called Ln.

When *learn* is executed, it makes a private directory for the user to work in, within the *learn* portion of the file system. A fresh copy of all the files used in each lesson (mostly data for the student to operate upon) is made each time a student starts a lesson, so the script writer may assume that everything is reinitialized each time a lesson is entered. The student directory is deleted after each session; any permanent records must be kept elsewhere.

The script writer must provide certain basic items in each lesson:

(1) the text of the lesson;

(2) the set-up commands to be executed before the user gets control;

(3) the data, if any, which the user is supposed to edit, transform, or otherwise process;

Figure 2: Directory structure for *learn*

```
lib
        play
                        student1
                                        files for student1...
                        student2
                                        files for student2...
        files
                        L0.1a           lessons for files course
                        L0.1b
                        ...
        editor
                        ...
        (other courses)
        log
```

(4) the evaluating commands to be executed after the user has finished the lesson, to decide whether the answer is right; and

(5) a list of possible successor lessons.

Learn tries to minimize the work of bookkeeping and installation, so that most of the effort involved in script production is in planning lessons, writing tutorial paragraphs, and coding tests of student performance.

The basic sequence of events is as follows. First, *learn* creates the working directory. Then, for each lesson, *learn* reads the script for the lesson and processes it a line at a time. The lines in the script are: (1) commands to the script interpreter to print something, to create a files, to test something, etc.; (2) text to be printed or put in a file; (3) other lines, which are sent to the shell to be executed. One line in each lesson turns control over to the user; the user can run any UNIX commands. The user mode terminates when the user types *yes*, *no*, *ready*, or *answer*. At this point, the user's work is tested; if the lesson is passed, a new lesson is selected, and if not the old one is repeated.

Let us illustrate this with the script for the second lesson of Figure 1; this is shown in Figure 3.

Lines which begin with # are commands to the *learn* script interpreter. For example,

#print

causes printing of any text that follows, up to the next line that begins with a sharp.

#print file

prints the contents of *file*; it is the same as *cat file* but has less overhead. Both forms of *#print* have the added property that if a lesson is failed, the *#print* will not be executed the second time through; this avoids annoying the student by repeating the preamble to a lesson.

#create filename

creates a file of the specified name, and copies any subsequent text up to a # to the file. This is used for creating and initializing working files and reference data for the lessons.

#user

gives control to the student; each line he or she types is passed to the shell for execution. The *#user* mode is terminated when the student types one of *yes*, *no*, *ready* or *answer*. At that time,

```
Figure 3:  Sample Lesson

#print
Of course, you can print any file with "cat".
In particular, it is common to first use
"ls" to find the name of a file and then "cat"
to print it. Note the difference between
"ls", which tells you the name of the files,
and "cat", which tells you the contents.
One file in the current directory is named for
a President. Print the file, then type "ready".
#create roosevelt
  this file is named roosevelt
  and contains three lines of
  text.
#copyout
#user
#uncopyout
tail -3 .ocopy >X1
#cmp X1 roosevelt
#log
#next
3.2b 2
```

the driver resumes interpretation of the script.

#copyin
#uncopyin

Anything the student types between these commands is copied onto a file called *.copy*. This lets the script writer interrogate the student's responses upon regaining control.

#copyout
#uncopyout

Between these commands, any material typed at the student by any program is copied to the file *.ocopy*. This lets the script writer interrogate the effect of what the student typed, which true believers in the performance theory of learning usually prefer to the student's actual input.

#pipe
#unpipe

Normally the student input and the script commands are fed to the UNIX command interpreter (the "shell") one line at a time. This won't do if, for example, a sequence of editor commands is provided, since the input to the editor must be handed to the editor, not to the shell. Accordingly, the material between *#pipe* and *#unpipe* commands is fed continuously through a pipe so that such sequences work. If *copyout* is also desired the *copyout* brackets must include the *pipe* brackets.

There are several commands for setting status after the student has attempted the lesson.

#cmp file1 file2

is an in-line implementation of *cmp*, which compares two files for identity.

#match stuff

The last line of the student's input is compared to *stuff*, and the success or fail status is set according to it. Extraneous things like the word *answer* are stripped before the comparison is made. There may be several *#match* lines; this provides a convenient mechanism for handling multiple "right" answers. Any text up to a # on subsequent lines after a successful *#match* is printed; this is illustrated in Figure 4, another sample lesson.

Figure 4: Another Sample Lesson

```
#print
What command will move the current line
to the end of the file? Type
"answer COMMAND", where COMMAND is the command.
#copyin
#user
#uncopyin
#match m$
#match .m$
"m$" is easier.
#log
#next
63.1d 10
```

#bad stuff

This is similar to *#match*, except that it corresponds to specific failure answers; this can be used to produce hints for particular wrong answers that have been anticipated by the script writer.

#succeed
#fail

print a message upon success or failure (as determined by some previous mechanism).

When the student types one of the "commands" *yes*, *no*, *ready*, or *answer*, the driver terminates the *#user* command, and evaluation of the student's work can begin. This can be done either by the built-in commands above, such as *#match* and *#cmp*, or by status returned by normal UNIX commands, typically *grep* and *test*. The last command should return status true (0) if the task was done successfully and false (non-zero) otherwise; this status return tells the driver whether or not the student has successfully passed the lesson.

Performance can be logged:

#log file

writes the date, lesson, user name and speed rating, and a success/failure indication on *file*. The command

#log

by itself writes the logging information in the logging directory within the *learn* hierarchy, and is the normal form.

#next

is followed by a few lines, each with a successor lesson name and an optional speed rating on it. A typical set might read

```
25.1a  10
25.2a  5
25.3a  2
```

indicating that unit 25.1a is a suitable follow-on lesson for students with a speed rating of 10 units, 25.2a for student with speed near 5, and 25.3a for speed near 2. Speed ratings are maintained for each session with a student; the rating is increased by one each time the student gets a lesson right and decreased by four each time the student gets a lesson wrong. Thus the driver tries to maintain a level such that the users get 80% right answers. The maximum rating is limited to

10 and the minimum to 0. The initial rating is zero unless the student specifies a different rating when starting a session.

If the student passes a lesson, a new lesson is selected and the process repeats. If the student fails, a false status is returned and the program reverts to the previous lesson and tries another alternative. If it can not find another alternative, it skips forward a lesson. *bye*, bye, which causes a graceful exit from the *learn* system. Hanging up is the usual novice's way out.

The lessons may form an arbitrary directed graph, although the present program imposes a limitation on cycles in that it will not present a lesson twice in the same session. If the student is unable to answer one of the exercises correctly, the driver searches for a previous lesson with a set of alternatives as successors (following the *#next* line). From the previous lesson with alternatives one route was taken earlier; the program simply tries a different one.

It is perfectly possible to write sophisticated scripts that evaluate the student's speed of response, or try to estimate the elegance of the answer, or provide detailed analysis of wrong answers. Lesson writing is so tedious already, however, that most of these abilities are likely to go unused.

The driver program depends heavily on features of UNIX that are not available on many other operating systems. These include the ease of manipulating files and directories, file redirection, the ability to use the command interpreter as just another program (even in a pipeline), command status testing and branching, the ability to catch signals like interrupts, and of course the pipeline mechanism itself. Although some parts of *learn* might be transferable to other systems, some generality will probably be lost.

A bit of history: The first version of *learn* had fewer built-in words in the driver program, and made more use of the facilities of UNIX. For example, file comparison was done by creating a *cmp* process, rather than comparing the two files within *learn*. Lessons were not stored as text files, but as archives. There was no concept of the in-line document; even *#print* had to be followed by a file name. Thus the initialization for each lesson was to extract the archive into the working directory (typically 4-8 files), then *#print* the lesson text.

The combination of such things made *learn* slower. The new version is about 4 or 5 times faster. Furthermore, it appears even faster to the user because in a typical lesson, the printing of the message comes first, and file setup with *#create* can be overlapped with the printng, so that when the program finishes printing, it is really ready for the user to type at it.

It is also a great advantage to the script maintainer that lessons are now just ordinary text files. They can be edited without any difficulty, and UNIX text manipulation tools can be applied to them. The result has been that there is much less resistance to going in and fixing substandard lessons.

6. Conclusions

The following observations can be made about secretaries, typists, and other non-programmers who have used *learn*:

(a) A novice must have assistance with the mechanics of communicating with the computer to get through to the first lesson or two; once the first few lessons are passed people can proceed on their own.

(b) The terminology used in the first few lessons is obscure to those inexperienced with computers. It would help if there were a low level reference card for UNIX to supplement the existing programmer oriented bulky manual and bulky reference card.

(c) The concept of "substitutable argument" is hard to grasp, and requires help.

(d) They enjoy the system for the most part. Motivation matters a great deal, however.

It takes an hour or two for a novice to get through the script on file handling. The total time for a reasonably intelligent and motivated novice to proceed from ignorance to a reasonable ability to create new files and manipulate old ones seems to be a few days, with perhaps half of each day spent on the machine.

The normal way of proceeding has been to have students in the same room with someone who knows UNIX and the scripts. Thus the student is not brought to a halt by difficult questions. The burden on the counselor, however, is much lower than that on a teacher of a course. Ideally, the students should be encouraged to proceed with instruction immediately prior to their actual use of the computer. They should exercise the scripts on the same computer and the same kind of terminal that they will later use for their real work, and their first few jobs for the computer should be relatively easy ones. Also, both training and initial work should take place on days when the UNIX hardware and software are working reliably. Rarely is all of this possible, but the closer one comes the better the result. For example, if it is known that the hardware is shaky one day, it is better to attempt to reschedule training for another one. Students are very frustrated by machine downtime; when nothing is happening, it takes some sophistication and experience to distinguish an infinite loop, a slow but functioning program, a program waiting for the user, and a broken machine.*

One disadvantage of training with *learn* is that students come to depend completely on the CAI system, and do not try to read manuals or use other learning aids. This is unfortunate, not only because of the increased demands for completeness and accuracy of the scripts, but because the scripts do not cover all of the UNIX system. New users should have manuals (appropriate for their level) and read them; the scripts ought to be altered to recommend suitable documents and urge students to read them.

There are several other difficulties which are clearly evident. From the student's viewpoint, the most serious is that lessons still crop up which simply can't be passed. Sometimes this is due to poor explanations, but just as often it is some error in the lesson itself — a botched setup, a missing file, an invalid test for correctness, or some system facility that doesn't work on the local system in the same way it did on the development system. It takes knowledge and a certain healthy arrogance on the part of the user to recognize that the fault is not his or hers, but the script writer's. Permitting the student to get on with the next lesson regardless does alleviate this somewhat, and the logging facilities make it easy to watch for lessons that no one can pass, but it is still a problem.

The biggest problem with the previous *learn* was speed (or lack thereof) — it was often excruciatingly slow and made a significant drain on the system. The current version so far does not seem to have that difficulty, although some scripts, notably *eqn* , are intrinsically slow. *eqn* , for example, must do a lot of work even to print its introductions, let alone check the student responses, but delay is perceptible in all scripts from time to time.

Another potential problem is that it is possible to break *learn* inadvertently, by pushing interrupt at the wrong time, or by removing critical files, or any number of similar slips. The defenses against such problems have steadily been improved, to the point where most students should not notice difficulties. Of course, it will always be possible to break *learn* maliciously, but this is not likely to be a problem.

One area is more fundamental — some UNIX commands are sufficiently global in their effect that *learn* currently does not allow them to be executed at all. The most obvious is *cd* , which changes to another directory. The prospect of a student who is learning about directories inadvertently moving to some random directory and removing files has deterred us from even writing lessons on *cd* , but ultimately lessons on such topics probably should be added.

7. Acknowledgments

We are grateful to all those who have tried *learn,* for we have benefited greatly from their suggestions and criticisms. In particular, M. E. Bittrich, J. L. Blue, S. I. Feldman, P. A. Fox, and M. J. McAlpin have provided substantial feedback. Conversations with E. Z. Rothkopf also provided many of the ideas in the system. We are also indebted to Don Jackowski for serving as a

* We have even known an expert programmer to decide the computer was broken when he had simply left his terminal in local mode. Novices have great difficulties with such problems.

guinea pig for the second version, and to Tom Plum for his efforts to improve the C script.

References

1. D.L. Bitzer and D. Skaperdas, "The Economics of a Large Scale Computer Based Educational System: Plato IV," in *Computer Assisted Instruction, Testing and Guidance*, ed. Wayne Holtzman, pp. 17-29, Harper and Row, New York, 1970.
2. D.C. Gray, J.P. Hulskamp, J.H. Kumm, S. Lichtenstein, and N.E. Nimmervoll, "COALA - A Minicomputer CAI System," *IEEE Trans. Education*, vol. E-20(1), pp. 73-77, Feb. 1977.
3. P. Suppes, "On Using Computers to Individualize Instruction," in *The Computer in American Education*, ed. D.D. Bushnell and D.W. Allen, pp. 11-24, John Wiley, New York, 1967.
4. B.F. Skinner, "Why We Need Teaching Machines," *Harv. Educ. Review*, vol. 31, pp. 377-398, 1961. Reprinted in *Educational Technology*, ed. J.P. DeCecco, Holt Rinehart & Winston (New York, 1964)
5. K. Thompson and D. M. Ritchie, *UNIX Programmer's Manual*, Bell Laboratories, 1978. See section *ed* (1).
6. B.W. Kernighan, *A Tutorial Introduction to the UNIX text editor*, 1974. Bell Laboratories internal memorandum
7. B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

An Introduction to the UNIX Shell

S. R. Bourne

ABSTRACT

The ***shell*** is a command programming language that provides an interface to the UNIX† operating system. Its features include control-flow primitives, parameter passing, variables and string substitution. Constructs such as *while, if then else, case* and *for* are available. Two-way communication is possible between the ***shell*** and commands. String-valued parameters, typically file names or flags, may be passed to a command. A return code is set by commands that may be used to determine control-flow, and the standard output from a command may be used as shell input.

The ***shell*** can modify the environment in which commands run. Input and output can be redirected to files, and processes that communicate through 'pipes' can be invoked. Commands are found by searching directories in the file system in a sequence that can be defined by the user. Commands can be read either from the terminal or from a file, which allows command procedures to be stored for later use.

September 16, 1986

† UNIX is a trademark of Bell Laboratories.

An Introduction to the UNIX Shell

S. R. Bourne

1.0 Introduction

The shell is both a command language and a programming language that provides an interface to the UNIX operating system. This memorandum describes, with examples, the UNIX shell. The first section covers most of the everyday requirements of terminal users. Some familiarity with UNIX is an advantage when reading this section; see, for example, "UNIX for beginners".[1] Section 2 describes those features of the shell primarily intended for use within shell procedures. These include the control-flow primitives and string-valued variables provided by the shell. A knowledge of a programming language would be a help when reading this section. The last section describes the more advanced features of the shell. References of the form "see *pipe* (2)" are to a section of the UNIX manual.[2]

1.1 Simple commands

Simple commands consist of one or more words separated by blanks. The first word is the name of the command to be executed; any remaining words are passed as arguments to the command. For example,

```
who
```

is a command that prints the names of users logged in. The command

```
ls −l
```

prints a list of files in the current directory. The argument *−l* tells *ls* to print status information, size and the creation date for each file.

1.2 Background commands

To execute a command the shell normally creates a new *process* and waits for it to finish. A command may be run without waiting for it to finish. For example,

```
cc pgm.c &
```

calls the C compiler to compile the file *pgm.c*. The trailing & is an operator that instructs the shell not to wait for the command to finish. To help keep track of such a process the shell reports its process number following its creation. A list of currently active processes may be obtained using the *ps* command.

1.3 Input output redirection

Most commands produce output on the standard output that is initially connected to the terminal. This output may be sent to a file by writing, for example,

```
ls −l >file
```

The notation >*file* is interpreted by the shell and is not passed as an argument to *ls*. If *file* does not exist then the shell creates it; otherwise the original contents of *file* are replaced with the output from *ls*. Output may be appended to a file using the notation

```
ls −l >>file
```

In this case *file* is also created if it does not already exist.

The standard input of a command may be taken from a file instead of the terminal by writing, for example,

```
wc <file
```

The command *wc* reads its standard input (in this case redirected from *file*) and prints the number of characters, words and lines found. If only the number of lines is required then

```
wc -l <file
```

could be used.

1.4 Pipelines and filters

The standard output of one command may be connected to the standard input of another by writing the 'pipe' operator, indicated by |, as in,

```
ls -l | wc
```

Two commands connected in this way constitute a *pipeline* and the overall effect is the same as

```
ls -l >file; wc <file
```

except that no *file* is used. Instead the two processes are connected by a pipe (see *pipe* (2)) and are run in parallel. Pipes are unidirectional and synchronization is achieved by halting *wc* when there is nothing to read and halting *ls* when the pipe is full.

A *filter* is a command that reads its standard input, transforms it in some way, and prints the result as output. One such filter, *grep,* selects from its input those lines that contain some specified string. For example,

```
ls | grep old
```

prints those lines, if any, of the output from *ls* that contain the string *old.* Another useful filter is *sort.* For example,

```
who | sort
```

will print an alphabetically sorted list of logged in users.

A pipeline may consist of more than two commands, for example,

```
ls | grep old | wc -l
```

prints the number of file names in the current directory containing the string *old.*

1.5 File name generation

Many commands accept arguments which are file names. For example,

```
ls -l main.c
```

prints information relating to the file *main.c*.

The shell provides a mechanism for generating a list of file names that match a pattern. For example,

```
ls -l *.c
```

generates, as arguments to *ls,* all file names in the current directory that end in *.c*. The character * is a pattern that will match any string including the null string. In general *patterns* are specified as follows.

*	Matches any string of characters including the null string.
?	Matches any single character.
[...]	Matches any one of the characters enclosed. A pair of characters separated by a minus will match any character lexically between the pair.

For example,

    [a–z]*

matches all names in the current directory beginning with one of the letters *a* through *z*.

    /usr/fred/test/?

matches all names in the directory **/usr/fred/test** that consist of a single character. If no file name is found that matches the pattern then the pattern is passed, unchanged, as an argument.

This mechanism is useful both to save typing and to select names according to some pattern. It may also be used to find files. For example,

    echo /usr/fred/*/core

finds and prints the names of all *core* files in sub-directories of **/usr/fred**. (*echo* is a standard UNIX command that prints its arguments, separated by blanks.) This last feature can be expensive, requiring a scan of all sub-directories of **/usr/fred**.

There is one exception to the general rules given for patterns. The character '.' at the start of a file name must be explicitly matched.

    echo *

will therefore echo all file names in the current directory not beginning with '.'.

    echo .*

will echo all those file names that begin with '.'. This avoids inadvertent matching of the names '.' and '..' which mean 'the current directory' and 'the parent directory' respectively. (Notice that *ls* suppresses information for the files '.' and '..'.)

1.6 Quoting

Characters that have a special meaning to the shell, such as < > * ? | &, are called metacharacters. A complete list of metacharacters is given in appendix B. Any character preceded by a \ is *quoted* and loses its special meaning, if any. The \ is elided so that

    echo \?

will echo a single ?, and

    echo \\

will echo a single \. To allow long strings to be continued over more than one line the sequence **\newline** is ignored.

\ is convenient for quoting single characters. When more than one character needs quoting the above mechanism is clumsy and error prone. A string of characters may be quoted by enclosing the string between single quotes. For example,

    echo xx´****´xx

will echo

    xx****xx

The quoted string may not contain a single quote but may contain newlines, which are preserved. This quoting mechanism is the most simple and is recommended for casual use.

A third quoting mechanism using double quotes is also available that prevents interpretation of some but not all metacharacters. Discussion of the details is deferred to section 3.4.

1.7 Prompting

When the shell is used from a terminal it will issue a prompt before reading a command. By default this prompt is '**$** '. It may be changed by saying, for example,

PS1=yesdear

that sets the prompt to be the string *yesdear*. If a newline is typed and further input is needed then the shell will issue the prompt '> '. Sometimes this can be caused by mistyping a quote mark. If it is unexpected then an interrupt (DEL) will return the shell to read another command. This prompt may be changed by saying, for example,

PS2=more

1.8 The shell and login

Following *login* (1) the shell is called to read and execute commands typed at the terminal. If the user's login directory contains the file **.profile** then it is assumed to contain commands and is read by the shell before reading any commands from the terminal.

1.9 Summary

- **ls**
  Print the names of files in the current directory.
- **ls >file**
  Put the output from *ls* into *file*.
- **ls | wc -l**
  Print the number of files in the current directory.
- **ls | grep old**
  Print those file names containing the string *old*.
- **ls | grep old | wc -l**
  Print the number of files whose name contains the string *old*.
- **cc pgm.c &**
  Run *cc* in the background.

2.0 Shell procedures

The shell may be used to read and execute commands contained in a file. For example,

```
sh file [ args ... ]
```

calls the shell to read commands from *file*. Such a file is called a *command procedure* or *shell procedure*. Arguments may be supplied with the call and are referred to in *file* using the positional parameters **$1, $2, ...** For example, if the file *wg* contains

```
who | grep $1
```

then

```
sh wg fred
```

is equivalent to

```
who | grep fred
```

UNIX files have three independent attributes, *read, write* and *execute*. The UNIX command *chmod* (1) may be used to make a file executable. For example,

```
chmod +x wg
```

will ensure that the file *wg* has execute status. Following this, the command

```
wg fred
```

is equivalent to

```
sh wg fred
```

This allows shell procedures and programs to be used interchangeably. In either case a new process is created to run the command.

As well as providing names for the positional parameters, the number of positional parameters in the call is available as **$#**. The name of the file being executed is available as **$0**.

A special shell parameter **$*** is used to substitute for all positional parameters except **$0**. A typical use of this is to provide some default arguments, as in,

```
nroff -T450 -ms $*
```

which simply prepends some arguments to those already given.

2.1 Control flow - for

A frequent use of shell procedures is to loop through the arguments (**$1, $2, ...**) executing commands once for each argument. An example of such a procedure is *tel* that searches the file **/usr/lib/telnos** that contains lines of the form

```
...
fred mh0123
bert mh0789
...
```

The text of *tel* is

```
for i
do grep $i /usr/lib/telnos; done
```

The command

```
tel fred
```

prints those lines in **/usr/lib/telnos** that contain the string *fred*.

```
tel fred bert
```

prints those lines containing *fred* followed by those for *bert.*

The **for** loop notation is recognized by the shell and has the general form

> **for** *name* **in** *w1 w2* ...
> **do** *command-list*
> **done**

A *command-list* is a sequence of one or more simple commands separated or terminated by a newline or semicolon. Furthermore, reserved words like **do** and **done** are only recognized following a newline or semicolon. *name* is a shell variable that is set to the words *w1 w2* ... in turn each time the *command-list* following **do** is executed. If **in** *w1 w2* ... is omitted then the loop is executed once for each positional parameter; that is, **in $*** is assumed.

Another example of the use of the **for** loop is the *create* command whose text is

```
for i do >$i; done
```

The command

```
create alpha beta
```

ensures that two empty files *alpha* and *beta* exist and are empty. The notation >*file* may be used on its own to create or clear the contents of a file. Notice also that a semicolon (or newline) is required before **done.**

2.2 Control flow - case

A multiple way branch is provided for by the **case** notation. For example,

```
case $# in
    1)    cat >>$1 ;;
    2)    cat >>$2 <$1 ;;
    *)    echo ´usage: append [ from ] to´ ;;
esac
```

is an *append* command. When called with one argument as

```
append file
```

$# is the string *1* and the standard input is copied onto the end of *file* using the *cat* command.

```
append file1 file2
```

appends the contents of *file1* onto *file2.* If the number of arguments supplied to *append* is other than 1 or 2 then a message is printed indicating proper usage.

The general form of the **case** command is

> **case** *word* **in**
> *pattern*) *command-list* ;;
> ...
> **esac**

The shell attempts to match *word* with each *pattern,* in the order in which the patterns appear. If a match is found the associated *command-list* is executed and execution of the **case** is complete. Since * is the pattern that matches any string it can be used for the default case.

A word of caution: no check is made to ensure that only one pattern matches the case argument. The first match found defines the set of commands to be executed. In the example below the commands following the second * will never be executed.

```
case $# in
     *) . . . ;;
     *) . . . ;;
esac
```

Another example of the use of the **case** construction is to distinguish between different forms of an argument. The following example is a fragment of a *cc* command.

```
for i
do case $i in
   -[ocs]) . . . ;;
   -*)     echo 'unknown flag $i' ;;
   *.c)    /lib/c0 $i . . . ;;
   *) echo 'unexpected argument $i' ;;
   esac
done
```

To allow the same commands to be associated with more than one pattern the **case** command provides for alternative patterns separated by a | . For example,

```
case $i in
     -x |-y)     . . .
esac
```

is equivalent to

```
case $i in
     -[xy]) . . .
esac
```

The usual quoting conventions apply so that

```
case $i in
     \?)  . . .
```

will match the character **?**.

2.3 Here documents

The shell procedure *tel* in section 2.1 uses the file **/usr/lib/telnos** to supply the data for *grep*. An alternative is to include this data within the shell procedure as a *here* document, as in,

```
for i
do grep $i <<!
   . . .
   fred mh0123
   bert mh0789
   . . .
!
done
```

In this example the shell takes the lines between <<! and ! as the standard input for *grep*. The string ! is arbitrary, the document being terminated by a line that consists of the string following << .

Parameters are substituted in the document before it is made available to *grep* as illustrated by the following procedure called *edg* .

```
ed $3 <<%
g/$1/s//$2/g
w
%
```

The call

```
edg string1 string2 file
```

is then equivalent to the command

```
ed file <<%
g/string1/s//string2/g
w
%
```

and changes all occurrences of *string1* in *file* to *string2*. Substitution can be prevented using \ to quote the special character **$** as in

```
ed $3 <<+
1,\$s/$1/$2/g
w
+
```

(This version of *edg* is equivalent to the first except that *ed* will print a **?** if there are no occurrences of the string **$1**.) Substitution within a *here* document may be prevented entirely by quoting the terminating string, for example,

```
grep $i <<\#
...
#
```

The document is presented without modification to *grep*. If parameter substitution is not required in a *here* document this latter form is more efficient.

2.4 Shell variables

The shell provides string-valued variables. Variable names begin with a letter and consist of letters, digits and underscores. Variables may be given values by writing, for example,

```
user=fred box=m000 acct=mh0000
```

which assigns values to the variables **user, box** and **acct.** A variable may be set to the null string by saying, for example,

```
null=
```

The value of a variable is substituted by preceding its name with **$**; for example,

```
echo $user
```

will echo *fred*.

Variables may be used interactively to provide abbreviations for frequently used strings. For example,

```
b=/usr/fred/bin
mv pgm $b
```

will move the file *pgm* from the current directory to the directory **/usr/fred/bin**. A more general notation is available for parameter (or variable) substitution, as in,

```
echo ${user}
```

which is equivalent to

echo $user

and is used when the parameter name is followed by a letter or digit. For example,

tmp=/tmp/ps
ps a >${tmp}a

will direct the output of *ps* to the file **/tmp/psa,** whereas,

ps a >$tmpa

would cause the value of the variable **tmpa** to be substituted.

Except for **$?** the following are set initially by the shell. **$?** is set after executing each command.

$? The exit status (return code) of the last command executed as a decimal string. Most commands return a zero exit status if they complete successfully, otherwise a non-zero exit status is returned. Testing the value of return codes is dealt with later under **if** and **while** commands.

$# The number of positional parameters (in decimal). Used, for example, in the *append* command to check the number of parameters.

$$ The process number of this shell (in decimal). Since process numbers are unique among all existing processes, this string is frequently used to generate unique temporary file names. For example,

ps a >/tmp/ps$$
...
rm /tmp/ps$$

$! The process number of the last process run in the background (in decimal).

$– The current shell flags, such as **–x** and **–v**.

Some variables have a special meaning to the shell and should be avoided for general use.

$MAIL When used interactively the shell looks at the file specified by this variable before it issues a prompt. If the specified file has been modified since it was last looked at the shell prints the message *you have mail* before prompting for the next command. This variable is typically set in the file **.profile,** in the user's login directory. For example,

MAIL=/usr/mail/fred

$HOME The default argument for the *cd* command. The current directory is used to resolve file name references that do not begin with a **/**, and is changed using the *cd* command. For example,

cd /usr/fred/bin

makes the current directory **/usr/fred/bin**.

cat wn

will print on the terminal the file *wn* in this directory. The command *cd* with no argument is equivalent to

cd $HOME

This variable is also typically set in the the user's login profile.

$PATH A list of directories that contain commands (the *search path*). Each time a command is executed by the shell a list of directories is searched for an executable file. If **$PATH** is not set then the current directory, **/bin**, and **/usr/bin** are searched by default. Otherwise **$PATH** consists of directory names separated by **:** . For example,

PATH=:/usr/fred/bin:/bin:/usr/bin

specifies that the current directory (the null string before the first **:**), **/usr/fred/bin, /bin** and **/usr/bin** are to be searched in that order. In this way individual users can have their own 'private' commands that are accessible independently of the current directory. If the command name contains a / then this directory search is not used; a single attempt is made to execute the command.

$PS1 The primary shell prompt string, by default, '**$** '.

$PS2 The shell prompt when further input is needed, by default, '> '.

$IFS The set of characters used by ***blank interpretation*** (see section 3.4).

2.5 The test command

The *test* command, although not part of the shell, is intended for use by shell programs. For example,

test –f file

returns zero exit status if *file* exists and non-zero exit status otherwise. In general *test* evaluates a predicate and returns the result as its exit status. Some of the more frequently used *test* arguments are given here, see *test* (1) for a complete specification.

test s	true if the argument *s* is not the null string
test –f file	true if *file* exists
test –r file	true if *file* is readable
test –w file	true if *file* is writable
test –d file	true if *file* is a directory

2.6 Control flow - while

The actions of the **for** loop and the **case** branch are determined by data available to the shell. A **while** or **until** loop and an **if then else** branch are also provided whose actions are determined by the exit status returned by commands. A **while** loop has the general form

while *command-list$_1$*
do *command-list$_2$*
done

The value tested by the **while** command is the exit status of the last simple command following **while.** Each time round the loop *command-list$_1$* is executed; if a zero exit status is returned then *command-list$_2$* is executed; otherwise, the loop terminates. For example,

```
while test $1
do ...
   shift
done
```

is equivalent to

```
for i
do ...
done
```

shift is a shell command that renames the positional parameters **$2, $3, ...** as **$1, $2, ...** and loses **$1** .

Another kind of use for the **while/until** loop is to wait until some external event occurs and then run some commands. In an **until** loop the termination condition is reversed. For example,

```
until test −f file
do sleep 300; done
commands
```

will loop until *file* exists. Each time round the loop it waits for 5 minutes before trying again. (Presumably another process will eventually create the file.)

2.7 Control flow - if

Also available is a general conditional branch of the form,

> **if** *command-list*
> **then** *command-list*
> **else** *command-list*
> **fi**

that tests the value returned by the last simple command following **if**.

The **if** command may be used in conjunction with the *test* command to test for the existence of a file as in

```
if test −f file
then  process file
else  do something else
fi
```

An example of the use of **if**, **case** and **for** constructions is given in section 2.10.

A multiple test **if** command of the form

```
if ...
then ...
else  if ...
      then ...
      else  if ...
            ...
            fi
      fi
fi
```

may be written using an extension of the **if** notation as,

```
if ...
then ...
elif  ...
then ...
elif  ...
...
fi
```

The following example is the *touch* command which changes the 'last modified' time for a list of files. The command may be used in conjunction with *make* (1) to force recompilation of a list of files.

```
flag=
for i
do case $i in
   -c)      flag=N ;;
   *) if test -f $i
      then  ln $i junk$$; rm junk$$
      elif test $flag
      then  echo file \'$i\' does not exist
      else  >$i
      fi
   esac
done
```

The **-c** flag is used in this command to force subsequent files to be created if they do not already exist. Otherwise, if the file does not exist, an error message is printed. The shell variable *flag* is set to some non-null string if the **-c** argument is encountered. The commands

```
ln ...; rm ...
```

make a link to the file and then remove it thus causing the last modified date to be updated.

The sequence

```
if command1
then  command2
fi
```

may be written

```
command1 && command2
```

Conversely,

```
command1 || command2
```

executes *command2* only if *command1* fails. In each case the value returned is that of the last simple command executed.

2.8 Command grouping

Commands may be grouped in two ways,

{ *command-list* ; }

and

(*command-list*)

In the first *command-list* is simply executed. The second form executes *command-list* as a separate process. For example,

```
(cd x; rm junk )
```

executes *rm junk* in the directory **x** without changing the current directory of the invoking shell.

The commands

```
cd x; rm junk
```

have the same effect but leave the invoking shell in the directory **x**.

2.9 Debugging shell procedures

The shell provides two tracing mechanisms to help when debugging shell procedures. The first is invoked within the procedure as

set −v

(**v** for verbose) and causes lines of the procedure to be printed as they are read. It is useful to help isolate syntax errors. It may be invoked without modifying the procedure by saying

sh −v proc ...

where *proc* is the name of the shell procedure. This flag may be used in conjunction with the **−n** flag which prevents execution of subsequent commands. (Note that saying *set −n* at a terminal will render the terminal useless until an end-of-file is typed.)

The command

set −x

will produce an execution trace. Following parameter substitution each command is printed as it is executed. (Try these at the terminal to see what effect they have.) Both flags may be turned off by saying

set −

and the current setting of the shell flags is available as **$−** .

2.10 The man command

The following is the *man* command which is used to print sections of the UNIX manual. It is called, for example, as

man sh
man −t ed
man 2 fork

In the first the manual section for *sh* is printed. Since no section is specified, section 1 is used. The second example will typeset (**−t** option) the manual section for *ed.* The last prints the *fork* manual page from section 2.

```
cd /usr/man

: 'colon is the comment command'
: 'default is nroff ($N), section 1 ($s)'
N=n s=1

for i
do case $i in

   [1-9]*) s=$i ;;

   -t)     N=t ;;

   -n)     N=n ;;

   -*)     echo unknown flag \'$i\' ;;

   *) if test -f man$s/$i.$s
      then  ${N}roff man0/${N}aa man$s/$i.$s
      else  : 'look through all manual sections'
            found=no
            for j in 1 2 3 4 5 6 7 8 9
            do if test -f man$j/$i.$j
               then man $j $i
                    found=yes
               fi
            done
            case $found in
                 no) echo '$i: manual page not found'
            esac
      fi
   esac
done
```

Figure 1. A version of the man command

3.0 Keyword parameters

Shell variables may be given values by assignment or when a shell procedure is invoked. An argument to a shell procedure of the form *name=value* that precedes the command name causes *value* to be assigned to *name* before execution of the procedure begins. The value of *name* in the invoking shell is not affected. For example,

    user=fred command

will execute *command* with **user** set to *fred*. The **−k** flag causes arguments of the form *name=value* to be interpreted in this way anywhere in the argument list. Such *names* are sometimes called keyword parameters. If any arguments remain they are available as positional parameters **$1, $2, ...**

The *set* command may also be used to set positional parameters from within a procedure. For example,

    set – *

will set **$1** to the first file name in the current directory, **$2** to the next, and so on. Note that the first argument, –, ensures correct treatment when the first file name begins with a – .

3.1 Parameter transmission

When a shell procedure is invoked both positional and keyword parameters may be supplied with the call. Keyword parameters are also made available implicitly to a shell procedure by specifying in advance that such parameters are to be exported. For example,

    export user box

marks the variables **user** and **box** for export. When a shell procedure is invoked copies are made of all exportable variables for use within the invoked procedure. Modification of such variables within the procedure does not affect the values in the invoking shell. It is generally true of a shell procedure that it may not modify the state of its caller without explicit request on the part of the caller. (Shared file descriptors are an exception to this rule.)

Names whose value is intended to remain constant may be declared *readonly*. The form of this command is the same as that of the *export* command,

    readonly name ...

Subsequent attempts to set readonly variables are illegal.

3.2 Parameter substitution

If a shell parameter is not set then the null string is substituted for it. For example, if the variable **d** is not set

    echo $d

or

    echo ${d}

will echo nothing. A default string may be given as in

    echo ${d–.}

which will echo the value of the variable **d** if it is set and '.' otherwise. The default string is evaluated using the usual quoting conventions so that

    echo ${d–'*'}

will echo * if the variable **d** is not set. Similarly echo **${d−$1}**

will echo the value of **d** if it is set and the value (if any) of **$1** otherwise. A variable may be assigned a default value using the notation

echo **${d=.}**

which substitutes the same string as

echo **${d−.}**

and if **d** were not previously set then it will be set to the string '**.**'. (The notation **${...=...}** is not available for positional parameters.)

If there is no sensible default then the notation

echo **${d?message}**

will echo the value of the variable **d** if it has one, otherwise *message* is printed by the shell and execution of the shell procedure is abandoned. If *message* is absent then a standard message is printed. A shell procedure that requires some parameters to be set might start as follows.

: ${user?} ${acct?} ${bin?}

...

Colon (**:**) is a command that is built in to the shell and does nothing once its arguments have been evaluated. If any of the variables **user, acct** or **bin** are not set then the shell will abandon execution of the procedure.

3.3 Command substitution

The standard output from a command can be substituted in a similar way to parameters. The command *pwd* prints on its standard output the name of the current directory. For example, if the current directory is **/usr/fred/bin** then the command

d=\`pwd\`

is equivalent to

d=/usr/fred/bin

The entire string between grave accents (\`...\`) is taken as the command to be executed and is replaced with the output from the command. The command is written using the usual quoting conventions except that a \` must be escaped using a ****. For example,

ls \`echo "$1"\`

is equivalent to

ls **$1**

Command substitution occurs in all contexts where parameter substitution occurs (including *here* documents) and the treatment of the resulting text is the same in both cases. This mechanism allows string processing commands to be used within shell procedures. An example of such a command is *basename* which removes a specified suffix from a string. For example,

basename main.c .c

will print the string *main*. Its use is illustrated by the following fragment from a *cc* command.

```
case $A in
     ...
     *.c)  B=`basename $A .c`
     ...
esac
```

that sets **B** to the part of **$A** with the suffix **.c** stripped.

Here are some composite examples.

- **for i in `ls –t`; do ...**
  The variable **i** is set to the names of files in time order, most recent first.
- **set `date`; echo $6 $2 $3, $4**
  will print, e.g., *1977 Nov 1, 23:59:59*

3.4 Evaluation and quoting

The shell is a macro processor that provides parameter substitution, command substitution and file name generation for the arguments to commands. This section discusses the order in which these evaluations occur and the effects of the various quoting mechanisms.

Commands are parsed initially according to the grammar given in appendix A. Before a command is executed the following substitutions occur.

- parameter substitution, e.g. **$user**
- command substitution, e.g. **`pwd`**

  Only one evaluation occurs so that if, for example, the value of the variable **X** is the string **$*y*** then

  ```
  echo $X
  ```

  will echo **$*y***.
- blank interpretation

  Following the above substitutions the resulting characters are broken into non-blank words (*blank interpretation*). For this purpose 'blanks' are the characters of the string $IFS. By default, this string consists of blank, tab and newline. The null string is not regarded as a word unless it is quoted. For example,

  ```
  echo ''
  ```

  will pass on the null string as the first argument to *echo*, whereas

  ```
  echo $null
  ```

  will call *echo* with no arguments if the variable **null** is not set or set to the null string.
- file name generation

  Each word is then scanned for the file pattern characters *****, **?** and [...] and an alphabetical list of file names is generated to replace the word. Each such file name is a separate argument.

The evaluations just described also occur in the list of words associated with a **for** loop. Only substitution occurs in the *word* used for a **case** branch.

As well as the quoting mechanisms described earlier using \ and '...' a third quoting mechanism is provided using double quotes. Within double quotes parameter and command substitution occurs but file name generation and the interpretation of blanks does not. The following characters have a special meaning within double quotes and may be quoted using \.

$ parameter substitution
` command substitution
" ends the quoted string
\ quotes the special characters $ ` " \

For example,

echo "$x"

will pass the value of the variable **x** as a single argument to *echo*. Similarly,

echo "$*"

will pass the positional parameters as a single argument and is equivalent to

echo "$1 $2 ..."

The notation $@ is the same as $* except when it is quoted.

echo "$@"

will pass the positional parameters, unevaluated, to *echo* and is equivalent to

echo "$1" "$2" ...

The following table gives, for each quoting mechanism, the shell metacharacters that are evaluated.

	metacharacter					
	\	$	*	`	"	'
'	n	n	n	n	n	t
`	y	n	n	t	n	n
"	y	y	n	y	t	n

t terminator
y interpreted
n not interpreted

Figure 2. Quoting mechanisms

In cases where more than one evaluation of a string is required the built-in command *eval* may be used. For example, if the variable **X** has the value *$y*, and if **y** has the value *pqr* then

eval echo $X

will echo the string *pqr*.

In general the *eval* command evaluates its arguments (as do all commands) and treats the result as input to the shell. The input is read and the resulting command(s) executed. For example,

wg='eval who |grep'
$wg fred

is equivalent to

who |grep fred

In this example, *eval* is required since there is no interpretation of metacharacters, such as **|**, **following substitution.**

3.5 Error handling

The treatment of errors detected by the shell depends on the type of error and on whether the shell is being used interactively. An interactive shell is one whose input and output are connected to a terminal (as determined by *gtty* (2)). A shell invoked with the **−i** flag is also interactive.

Execution of a command (see also 3.7) may fail for any of the following reasons.

- Input output redirection may fail. For example, if a file does not exist or cannot be created.
- The command itself does not exist or cannot be executed.
- The command terminates abnormally, for example, with a "bus error" or "memory fault". See Figure 2 below for a complete list of UNIX signals.
- The command terminates normally but returns a non-zero exit status.

In all of these cases the shell will go on to execute the next command. Except for the last case an error message will be printed by the shell. All remaining errors cause the shell to exit from a command procedure. An interactive shell will return to read another command from the terminal. Such errors include the following.

- Syntax errors. e.g., **if ... then ... done**
- A signal such as interrupt. The shell waits for the current command, if any, to finish execution and then either exits or returns to the terminal.
- Failure of any of the built-in commands such as *cd.*

The shell flag **−e** causes the shell to terminate if any error is detected.

```
1    hangup
2    interrupt
3*   quit
4*   illegal instruction
5*   trace trap
6*   IOT instruction
7*   EMT instruction
8*   floating point exception
9    kill (cannot be caught or ignored)
10*  bus error
11*  segmentation violation
12*  bad argument to system call
13   write on a pipe with no one to read it
14   alarm clock
15   software termination (from kill (1))
```

Figure 3. UNIX signals

Those signals marked with an asterisk produce a core dump if not caught. However, the shell itself ignores quit which is the only external signal that can cause a dump. The signals in this list of potential interest to shell programs are 1, 2, 3, 14 and 15.

3.6 Fault handling

Shell procedures normally terminate when an interrupt is received from the terminal. The *trap* command is used if some cleaning up is required, such as removing temporary files. For example,

```
trap 'rm /tmp/ps$$; exit' 2
```

sets a trap for signal 2 (terminal interrupt), and if this signal is received will execute the commands

```
rm /tmp/ps$$; exit
```

exit is another built-in command that terminates execution of a shell procedure. The *exit* is required; otherwise, after the trap has been taken, the shell will resume executing the procedure at the place where it was interrupted.

UNIX signals can be handled in one of three ways. They can be ignored, in which case the signal is never sent to the process. They can be caught, in which case the process must decide what action to take when the signal is received. Lastly, they can be left to cause termination of the process without it having to take any further action. If a signal is being ignored on entry to the shell procedure, for example, by invoking it in the background (see 3.7) then *trap* commands (and the signal) are ignored.

The use of *trap* is illustrated by this modified version of the *touch* command (Figure 4). The cleanup action is to remove the file **junk$$**.

```
flag=
trap 'rm -f junk$$; exit' 1 2 3 15
for i
do case $i in
   -c)      flag=N ;;
   *) if test -f $i
      then  ln $i junk$$; rm junk$$
      elif test $flag
      then  echo file \'$i\' does not exist
      else  >$i
      fi
   esac
done
```

Figure 4. The touch command

The *trap* command appears before the creation of the temporary file; otherwise it would be possible for the process to die without removing the file.

Since there is no signal 0 in UNIX it is used by the shell to indicate the commands to be executed on exit from the shell procedure.

A procedure may, itself, elect to ignore signals by specifying the null string as the argument to trap. The following fragment is taken from the *nohup* command.

```
trap '' 1 2 3 15
```

which causes *hangup, interrupt, quit* and *kill* to be ignored both by the procedure and by invoked commands.

Traps may be reset by saying

```
trap 2 3
```

which resets the traps for signals 2 and 3 to their default values. A list of the current values of traps may be obtained by writing

```
trap
```

The procedure *scan* (Figure 5) is an example of the use of *trap* where there is no exit in the trap command. *scan* takes each directory in the current directory, prompts with its name, and then executes commands typed at the terminal until an end of file or an interrupt is received. Interrupts are ignored while executing the requested commands but cause termination when *scan* is waiting for input.

```
d=`pwd`
for i in *
do if test -d $d/$i
   then cd $d/$i
        while echo "$i:"
              trap exit 2
              read x
        do trap : 2; eval $x; done
   fi
done
```

Figure 5. The scan command

read x is a built-in command that reads one line from the standard input and places the result in the variable **x**. It returns a non-zero exit status if either an end-of-file is read or an interrupt is received.

3.7 Command execution

To run a command (other than a built-in) the shell first creates a new process using the system call *fork.* The execution environment for the command includes input, output and the states of signals, and is established in the child process before the command is executed. The built-in command *exec* is used in the rare cases when no fork is required and simply replaces the shell with a new command. For example, a simple version of the *nohup* command looks like

```
trap '' 1 2 3 15
exec $*
```

The *trap* turns off the signals specified so that they are ignored by subsequently created commands and *exec* replaces the shell by the command specified.

Most forms of input output redirection have already been described. In the following *word* is only subject to parameter and command substitution. No file name generation or blank interpretation takes place so that, for example,

```
echo ... >*.c
```

will write its output into a file whose name is ***.c**. Input output specifications are evaluated left to right as they appear in the command.

> *word* — The standard output (file descriptor 1) is sent to the file *word* which is created if it does not already exist.

>> *word* — The standard output is sent to file *word.* If the file exists then output is appended (by seeking to the end); otherwise the file is created.

< *word* — The standard input (file descriptor 0) is taken from the file *word.*

<< *word* — The standard input is taken from the lines of shell input that follow up to but not including a line consisting only of *word.* If *word* is quoted then no interpretation of the document occurs. If *word* is not quoted then parameter and command substitution occur and \ is used to quote the characters \ $ ` and the first character of *word.* In the latter case **\newline** is ignored (c.f. quoted strings).

>& *digit* — The file descriptor *digit* is duplicated using the system call *dup* (2) and the result is used as the standard output.

<& *digit* — The standard input is duplicated from file descriptor *digit.*

<&- — The standard input is closed.

>&- — The standard output is closed.

Any of the above may be preceded by a digit in which case the file descriptor created is that specified by the digit instead of the default 0 or 1. For example,

... 2>file

runs a command with message output (file descriptor 2) directed to *file*.

... 2>&1

runs a command with its standard output and message output merged. (Strictly speaking file descriptor 2 is created by duplicating file descriptor 1 but the effect is usually to merge the two streams.)

The environment for a command run in the background such as

list *.c | lpr &

is modified in two ways. Firstly, the default standard input for such a command is the empty file **/dev/null**. This prevents two processes (the shell and the command), which are running in parallel, from trying to read the same input. Chaos would ensue if this were not the case. For example,

ed file &

would allow both the editor and the shell to read from the same input at the same time.

The other modification to the environment of a background command is to turn off the QUIT and INTERRUPT signals so that they are ignored by the command. This allows these signals to be used at the terminal without causing background commands to terminate. For this reason the UNIX convention for a signal is that if it is set to 1 (ignored) then it is never changed even for a short time. Note that the shell command *trap* has no effect for an ignored signal.

3.8 Invoking the shell

The following flags are interpreted by the shell when it is invoked. If the first character of argument zero is a minus, then commands are read from the file **.profile**.

−c *string*
: If the **−c** flag is present then commands are read from *string*.

−s If the **−s** flag is present or if no arguments remain then commands are read from the standard input. Shell output is written to file descriptor 2.

−i If the **−i** flag is present or if the shell input and output are attached to a terminal (as told by *gtty*) then this shell is *interactive*. In this case TERMINATE is ignored (so that **kill 0** does not kill an interactive shell) and INTERRUPT is caught and ignored (so that **wait** is interruptable). In all cases QUIT is ignored by the shell.

Acknowledgements

The design of the shell is based in part on the original UNIX shell[3] and the PWB/UNIX shell,[4] some features having been taken from both. Similarities also exist with the command interpreters of the Cambridge Multiple Access System[5] and of CTSS.[6]

I would like to thank Dennis Ritchie and John Mashey for many discussions during the design of the shell. I am also grateful to the members of the Computing Science Research Center and to Joe Maranzano for their comments on drafts of this document.

References

1. B. W. Kernighan, *UNIX for Beginners*, 1978.
2. K. Thompson and D. M. Ritchie, *UNIX Programmer's Manual*, Bell Laboratories, 1978. Seventh Edition.

3. K. Thompson, "The UNIX Command Language," in *Structured Programming—Infotech State of the Art Report*, pp. 375-384, Infotech International Ltd., Nicholson House, Maidenhead, Berkshire, England, March 1975.

4. J. R. Mashey, *PWB/UNIX Shell Tutorial*, September 30, 1977.

5. D. F. Hartley (Ed.), *The Cambridge Multiple Access System - Users Reference Manual*, University Mathematical Laboratory, Cambridge, England, 1968.

6. P. A. Crisman (Ed.), *The Compatible Time-Sharing System*, M.I.T. Press, Cambridge, Mass., 1965.

Appendix A - Grammar

item: *word*
input-output
name = value

simple-command: *item*
simple-command item

command: *simple-command*
(*command-list*)
{ *command-list* }
for *name* **do** *command-list* **done**
for *name* **in** *word* ... **do** *command-list* **done**
while *command-list* **do** *command-list* **done**
until *command-list* **do** *command-list* **done**
case *word* **in** *case-part* ... **esac**
if *command-list* **then** *command-list* *else-part* **fi**

pipeline: *command*
pipeline | *command*

andor: *pipeline*
andor && *pipeline*
andor || *pipeline*

command-list: *andor*
command-list ;
command-list &
command-list ; *andor*
command-list & *andor*

input-output: > *file*
< *file*
>> *word*
<< *word*

file: *word*
& *digit*
& –

case-part: *pattern*) *command-list* ;;

pattern: *word*
pattern | *word*

else-part: **elif** *command-list* **then** *command-list* *else-part*
else *command-list*
empty

empty:

word: a sequence of non-blank characters

name: a sequence of letters, digits or underscores starting with a letter

digit: **0 1 2 3 4 5 6 7 8 9**

Appendix B - Meta-characters and Reserved Words

a) syntactic

- `|` pipe symbol
- `&&` 'andf' symbol
- `||` 'orf' symbol
- `;` command separator
- `;;` case delimiter
- `&` background commands
- `( )` command grouping
- `<` input redirection
- `<<` input from a here document
- `>` output creation
- `>>` output append

b) patterns

- `*` match any character(s) including none
- `?` match any single character
- `[...]` match any of the enclosed characters

c) substitution

- `${...}` substitute shell variable
- `` `...` `` substitute command output

d) quoting

- `\` quote the next character
- `'...'` quote the enclosed characters except for '
- `"..."` quote the enclosed characters except for $ ` \ "

e) reserved words

if then else elif fi
case in esac
for while until do done
{ }

An Introduction to the C Shell

William Joy

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, California 94720

ABSTRACT

Csh is a new command language interpreter for UNIX† systems. It incorporates good features of other shells and a *history* mechanism similar to the *redo* of INTERLISP. While incorporating many features of other shells which make writing shell programs (shell scripts) easier, most of the features unique to *csh* are designed more for the interactive UNIX user.

UNIX users who have read a general introduction to the system will find a valuable basic explanation of the shell here. Simple terminal interaction with *csh* is possible after reading just the first section of this document. The second section describes the shells capabilities which you can explore after you have begun to become acquainted with the shell. Later sections introduce features which are useful, but not necessary for all users of the shell.

Back matter includes an appendix listing special characters of the shell and a glossary of terms and commands introduced in this manual.

September 16, 1986

† UNIX is a trademark of Bell Laboratories.

An introduction to the C shell

William Joy

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, California 94720

Introduction

A *shell* is a command language interpreter. *Csh* is the name of one particular command interpreter on UNIX. The primary purpose of *csh* is to translate command lines typed at a terminal into system actions, such as invocation of other programs. *Csh* is a user program just like any you might write. Hopefully, *csh* will be a very useful program for you in interacting with the UNIX system.

In addition to this document, you will want to refer to a copy of the UNIX programmer's manual. The *csh* documentation in the manual provides a full description of all features of the shell and is a final reference for questions about the shell.

Many words in this document are shown in *italics*. These are important words; names of commands, and words which have special meaning in discussing the shell and UNIX. Many of the words are defined in a glossary at the end of this document. If you don't know what is meant by a word, you should look for it in the glossary.

Acknowledgements

Numerous people have provided good input about previous versions of *csh* and aided in its debugging and in the debugging of its documentation. I would especially like to thank Michael Ubell who made the crucial observation that history commands could be done well over the word structure of input text, and implemented a prototype history mechanism in an older version of the shell. Eric Allman has also provided a large number of useful comments on the shell, helping to unify those concepts which are present and to identify and eliminate useless and marginally useful features. Mike O'Brien suggested the pathname hashing mechanism which speeds command execution. Jim Kulp added the job control and directory stack primitives and added their documentation to this introduction.

1. Terminal usage of the shell

1.1. The basic notion of commands

A *shell* in UNIX acts mostly as a medium through which other *programs* are invoked. While it has a set of *builtin* functions which it performs directly, most commands cause execution of programs that are, in fact, external to the shell. The shell is thus distinguished from the command interpreters of other systems both by the fact that it is just a user program, and by the fact that it is used almost exclusively as a mechanism for invoking other programs.

Commands in the UNIX system consist of a list of strings or *words* interpreted as a *command name* followed by *arguments*. Thus the command

```
mail bill
```

consists of two words. The first word *mail* names the command to be executed, in this case the mail program which sends messages to other users. The shell uses the name of the command in attempting to execute it for you. It will look in a number of *directories* for a file with the name *mail* which is expected to contain the mail program.

The rest of the words of the command are given as *arguments* to the command itself when it is executed. In this case we specified also the argument *bill* which is interpreted by the *mail* program to be the name of a user to whom mail is to be sent. In normal terminal usage we might use the *mail* command as follows.

```
% mail bill
I have a question about the csh documentation.
My document seems to be missing page 5.
Does a page five exist?
        Bill
EOT
%
```

Here we typed a message to send to *bill* and ended this message with a ↑D which sent an end-of-file to the mail program. (Here and throughout this document, the notation "↑*x*" is to be read "control-*x*" and represents the striking of the *x* key while the control key is held down.) The mail program then echoed the characters 'EOT' and transmitted our message. The characters '% ' were printed before and after the mail command by the shell to indicate that input was needed.

After typing the '% ' prompt the shell was reading command input from our terminal. We typed a complete command 'mail bill'. The shell then executed the *mail* program with argument *bill* and went dormant waiting for it to complete. The mail program then read input from our terminal until we signalled an end-of-file via typing a ↑D after which the shell noticed that mail had completed and signaled us that it was ready to read from the terminal again by printing another '% ' prompt.

This is the essential pattern of all interaction with UNIX through the shell. A complete command is typed at the terminal, the shell executes the command and when this execution completes, it prompts for a new command. If you run the editor for an hour, the shell will patiently wait for you to finish editing and obediently prompt you again whenever you finish editing.

An example of a useful command you can execute now is the *tset* command, which sets the default *erase* and *kill* characters on your terminal – the erase character erases the last character you typed and the kill character erases the entire line you have entered so far. By default, the erase character is '#' and the kill character is '@'. Most people who use CRT displays prefer to use the backspace (↑H) character as their erase character since it is then easier to see what you have typed so far. You can make this be true by typing

```
tset –e
```

which tells the program *tset* to set the erase character, and its default setting for this character is a backspace.

1.2. Flag arguments

A useful notion in UNIX is that of a *flag* argument. While many arguments to commands specify file names or user names some arguments rather specify an optional capability of the command which you wish to invoke. By convention, such arguments begin with the character '–' (hyphen). Thus the command

ls

will produce a list of the files in the current *working directory*. The option *–s* is the size option, and

ls –s

causes *ls* to also give, for each file the size of the file in blocks of 512 characters. The manual section for each command in the UNIX reference manual gives the available options for each command. The *ls* command has a large number of useful and interesting options. Most other commands have either no options or only one or two options. It is hard to remember options of commands which are not used very frequently, so most UNIX utilities perform only one or two functions rather than having a large number of hard to remember options.

1.3. Output to files

Commands that normally read input or write output on the terminal can also be executed with this input and/or output done to a file.

Thus suppose we wish to save the current date in a file called 'now'. The command

date

will print the current date on our terminal. This is because our terminal is the default *standard output* for the date command and the date command prints the date on its standard output. The shell lets us *redirect* the *standard output* of a command through a notation using the *metacharacter* '>' and the name of the file where output is to be placed. Thus the command

date > now

runs the *date* command such that its standard output is the file 'now' rather than the terminal. Thus this command places the current date and time into the file 'now'. It is important to know that the *date* command was unaware that its output was going to a file rather than to the terminal. The shell performed this *redirection* before the command began executing.

One other thing to note here is that the file 'now' need not have existed before the *date* command was executed; the shell would have created the file if it did not exist. And if the file did exist? If it had existed previously these previous contents would have been discarded! A shell option *noclobber* exists to prevent this from happening accidentally; it is discussed in section 2.2.

The system normally keeps files which you create with '>' and all other files. Thus the default is for files to be permanent. If you wish to create a file which will be removed automatically, you can begin its name with a '#' character, this 'scratch' character denotes the fact that the file will be a scratch file.* The system will remove such files after a couple of days, or sooner if file space becomes very tight. Thus, in running the *date* command above, we don't really want to save the output forever, so we would more likely do

---

*Note that if your erase character is a '#', you will have to precede the '#' with a '\'. The fact that the '#' character is the old (pre-CRT) standard erase character means that it seldom appears in a file name, and allows this convention to be used for scratch files. If you are using a CRT, your erase character should be a ↑H, as we demonstrated in section 1.1 how this could be set up.

date > #now

1.4. Metacharacters in the shell

The shell has a large number of special characters (like '>') which indicate special functions. We say that these notations have *syntactic* and *semantic* meaning to the shell. In general, most characters which are neither letters nor digits have special meaning to the shell. We shall shortly learn a means of *quotation* which allows us to use *metacharacters* without the shell treating them in any special way.

Metacharacters normally have effect only when the shell is reading our input. We need not worry about placing shell metacharacters in a letter we are sending via *mail,* or when we are typing in text or data to some other program. Note that the shell is only reading input when it has prompted with '% '.

1.5. Input from files; pipelines

We learned above how to *redirect* the *standard output* of a command to a file. It is also possible to redirect the *standard input* of a command from a file. This is not often necessary since most commands will read from a file whose name is given as an argument. We can give the command

sort < data

to run the *sort* command with standard input, where the command normally reads its input, from the file 'data'. We would more likely say

sort data

letting the *sort* command open the file 'data' for input itself since this is less to type.

We should note that if we just typed

sort

then the sort program would sort lines from its *standard input.* Since we did not *redirect* the standard input, it would sort lines as we typed them on the terminal until we typed a ↑D to indicate an end-of-file.

A most useful capability is the ability to combine the standard output of one command with the standard input of another, i.e. to run the commands in a sequence known as a *pipeline.* For instance the command

ls –s

normally produces a list of the files in our directory with the size of each in blocks of 512 characters. If we are interested in learning which of our files is largest we may wish to have this sorted by size rather than by name, which is the default way in which *ls* sorts. We could look at the many options of *ls* to see if there was an option to do this but would eventually discover that there is not. Instead we can use a couple of simple options of the *sort* command, combining it with *ls* to get what we want.

The *–n* option of sort specifies a numeric sort rather than an alphabetic sort. Thus

ls –s | sort –n

specifies that the output of the *ls* command run with the option *–s* is to be *piped* to the command *sort* run with the numeric sort option. This would give us a sorted list of our files by size, but with the smallest first. We could then use the *–r* reverse sort option and the *head* command in combination with the previous command doing

ls –s | sort –n –r | head –5

Here we have taken a list of our files sorted alphabetically, each with the size in blocks. We have run this to the standard input of the *sort* command asking it to sort numerically in reverse order (largest first). This output has then been run into the command *head* which gives us the first few lines. In this case we have asked *head* for the first 5 lines. Thus this command gives us the names and sizes of our 5 largest files.

The notation introduced above is called the *pipe* mechanism. Commands separated by '|' characters are connected together by the shell and the standard output of each is run into the standard input of the next. The leftmost command in a pipeline will normally take its standard input from the terminal and the rightmost will place its standard output on the terminal. Other examples of pipelines will be given later when we discuss the history mechanism; one important use of pipes which is illustrated there is in the routing of information to the line printer.

1.6. Filenames

Many commands to be executed will need the names of files as arguments. UNIX *pathnames* consist of a number of *components* separated by '/'. Each component except the last names a directory in which the next component resides, in effect specifying the *path* of directories to follow to reach the file. Thus the pathname

/etc/motd

specifies a file in the directory 'etc' which is a subdirectory of the *root* directory '/'. Within this directory the file named is 'motd' which stands for 'message of the day'. A *pathname* that begins with a slash is said to be an *absolute* pathname since it is specified from the absolute top of the entire directory hierarchy of the system (the *root*). *Pathnames* which do not begin with '/' are interpreted as starting in the current *working directory*, which is, by default, your *home* directory and can be changed dynamically by the *cd* change directory command. Such pathnames are said to be *relative* to the working directory since they are found by starting in the working directory and descending to lower levels of directories for each *component* of the pathname. If the pathname contains no slashes at all then the file is contained in the working directory itself and the pathname is merely the name of the file in this directory. Absolute pathnames have no relation to the working directory.

Most filenames consist of a number of alphanumeric characters and '.'s (periods). In fact, all printing characters except '/' (slash) may appear in filenames. It is inconvenient to have most non-alphabetic characters in filenames because many of these have special meaning to the shell. The character '.' (period) is not a shell-metacharacter and is often used to separate the *extension* of a file name from the base of the name. Thus

prog.c prog.o prog.errs prog.output

are four related files. They share a *base* portion of a name (a base portion being that part of the name that is left when a trailing '.' and following characters which are not '.' are stripped off). The file 'prog.c' might be the source for a C program, the file 'prog.o' the corresponding object file, the file 'prog.errs' the errors resulting from a compilation of the program and the file 'prog.output' the output of a run of the program.

If we wished to refer to all four of these files in a command, we could use the notation

prog.*

This word is expanded by the shell, before the command to which it is an argument is executed, into a list of names which begin with 'prog.'. The character '*' here matches any sequence (including the empty sequence) of characters in a file name. The names which match are alphabetically sorted and placed in the *argument list* of the command. Thus the command

echo prog.*

will echo the names

prog.c prog.errs prog.o prog.output

Note that the names are in sorted order here, and a different order than we listed them above. The *echo* command receives four words as arguments, even though we only typed one word as as argument directly. The four words were generated by ***filename expansion*** of the one input word.

Other notations for ***filename expansion*** are also available. The character '?' matches any single character in a filename. Thus

echo ? ?? ???

will echo a line of filenames; first those with one character names, then those with two character names, and finally those with three character names. The names of each length will be independently sorted.

Another mechanism consists of a sequence of characters between '[' and ']'. This metasequence matches any single character from the enclosed set. Thus

prog.[co]

will match

prog.c prog.o

in the example above. We can also place two characters around a '–' in this notation to denote a range. Thus

chap.[1–5]

might match files

chap.1 chap.2 chap.3 chap.4 chap.5

if they existed. This is shorthand for

chap.[12345]

and otherwise equivalent.

An important point to note is that if a list of argument words to a command (an *argument list)* contains filename expansion syntax, and if this filename expansion syntax fails to match any existing file names, then the shell considers this to be an error and prints a diagnostic

No match.

and does not execute the command.

Another very important point is that files with the character '.' at the beginning are treated specially. Neither '*' or '?' or the '[' ']' mechanism will match it. This prevents accidental matching of the filenames '.' and '..' in the working directory which have special meaning to the system, as well as other files such as *.cshrc* which are not normally visible. We will discuss the special role of the file *.cshrc* later.

Another filename expansion mechanism gives access to the pathname of the ***home*** directory of other users. This notation consists of the character '~' (tilde) followed by another users' login name. For instance the word '~bill' would map to the pathname '/usr/bill' if the home directory for 'bill' was '/usr/bill'. Since, on large systems, users may have login directories scattered over many different disk volumes with different prefix directory names, this notation provides a reliable way of accessing the files of other users.

A special case of this notation consists of a '~' alone, e.g. '~/mbox'. This notation is expanded by the shell into the file 'mbox' in your ***home*** directory, i.e. into '/usr/bill/mbox' for me on Ernie Co-vax, the UCB Computer Science Department VAX machine, where this document was prepared. This can be very useful if you have used ***cd*** to change to another directory and have found a file you wish to copy using ***cp***. If I give the command

cp thatfile ~

the shell will expand this command to

cp thatfile /usr/bill

since my home directory is /usr/bill.

There also exists a mechanism using the characters '{' and '}' for abbreviating a set of words which have common parts but cannot be abbreviated by the above mechanisms because they are not files, are the names of files which do not yet exist, are not thus conveniently described. This mechanism will be described much later, in section 4.2, as it is used less frequently.

1.7. Quotation

We have already seen a number of metacharacters used by the shell. These metacharacters pose a problem in that we cannot use them directly as parts of words. Thus the command

echo *

will not echo the character '*'. It will either echo an sorted list of filenames in the current *working directory,* or print the message 'No match' if there are no files in the working directory.

The recommended mechanism for placing characters which are neither numbers, digits, '/', '.' or '–' in an argument word to a command is to enclose it with single quotation characters ''', i.e.

echo '*'

There is one special character '!' which is used by the *history* mechanism of the shell and which cannot be *escaped* by placing it within ''' characters. It and the character ''' itself can be preceded by a single '\' to prevent their special meaning. Thus

echo \'\!

prints

'!

These two mechanisms suffice to place any printing character into a word which is an argument to a shell command. They can be combined, as in

echo \''*'

which prints

'*

since the first '\' escaped the first ''' and the '*' was enclosed between ''' characters.

1.8. Terminating commands

When you are executing a command and the shell is waiting for it to complete there are several ways to force it to stop. For instance if you type the command

cat /etc/passwd

the system will print a copy of a list of all users of the system on your terminal. This is likely to continue for several minutes unless you stop it. You can send an INTERRUPT *signal* to the *cat* command by typing the DEL or RUBOUT key on your terminal.* Since *cat* does not take any precautions to avoid or otherwise handle this signal the INTERRUPT will cause it to terminate. The shell notices that *cat* has terminated and prompts you again with '% '. If you hit INTERRUPT again, the shell will just repeat its prompt since it handles INTERRUPT signals and chooses to continue to execute commands rather than terminating like *cat* did, which would have the effect of logging you out.

*Many users use *stty*(1) to change the interrupt character to ↑C.

Another way in which many programs terminate is when they get an end-of-file from their standard input. Thus the *mail* program in the first example above was terminated when we typed a ↑D which generates an end-of-file from the standard input. The shell also terminates when it gets an end-of-file printing 'logout'; UNIX then logs you off the system. Since this means that typing too many ↑D's can accidentally log us off, the shell has a mechanism for preventing this. This *ignoreeof* option will be discussed in section 2.2.

If a command has its standard input redirected from a file, then it will normally terminate when it reaches the end of this file. Thus if we execute

    mail bill < prepared.text

the mail command will terminate without our typing a ↑D. This is because it read to the end-of-file of our file 'prepared.text' in which we placed a message for 'bill' with an editor program. We could also have done

    cat prepared.text | mail bill

since the *cat* command would then have written the text through the pipe to the standard input of the mail command. When the *cat* command completed it would have terminated, closing down the pipeline and the *mail* command would have received an end-of-file from it and terminated. Using a pipe here is more complicated than redirecting input so we would more likely use the first form. These commands could also have been stopped by sending an INTERRUPT.

Another possibility for stopping a command is to suspend its execution temporarily, with the possibility of continuing execution later. This is done by sending a STOP signal via typing a ↑Z. This signal causes all commands running on the terminal (usually one but more if a pipeline is executing) to become suspended. The shell notices that the command(s) have been suspended, types 'Stopped' and then prompts for a new command. The previously executing command has been suspended, but otherwise unaffected by the STOP signal. Any other commands can be executed while the original command remains suspended. The suspended command can be continued using the *fg* command with no arguments. The shell will then retype the command to remind you which command is being continued, and cause the command to resume execution. Unless any input files in use by the suspended command have been changed in the meantime, the suspension has no effect whatsoever on the execution of the command. This feature can be very useful during editing, when you need to look at another file before continuing. An example of command suspension follows.

```
% mail harold
Someone just copied a big file into my directory and its name is
↑Z
Stopped
% ls
funnyfile
prog.c
prog.o
% jobs
[1]  + Stopped            mail harold
% fg
mail harold
funnyfile. Do you know who did it?
EOT
%
```

In this example someone was sending a message to Harold and forgot the name of the file he wanted to mention. The mail command was suspended by typing ↑Z. When the shell noticed that the mail program was suspended, it typed 'Stopped' and prompted for a new command. Then the *ls* command was typed to find out the name of the file. The *jobs* command was run to find out which command was suspended. At this time the *fg* command was typed to continue execution of the mail program. Input to the mail program was then continued and ended with a ↑D which indicated the end of the message at which time the mail program typed EOT. The *jobs* command will show which commands are suspended. The ↑Z should only be typed at the beginning of a line since everything typed on the current line is discarded when a signal is sent from the keyboard. This also happens on INTERRUPT, and QUIT signals. More information on suspending jobs and controlling them is given in section 2.6.

If you write or run programs which are not fully debugged then it may be necessary to stop them somewhat ungracefully. This can be done by sending them a QUIT signal, sent by typing a ↑\. This will usually provoke the shell to produce a message like:

    Quit (Core dumped)

indicating that a file 'core' has been created containing information about the program 'a.out's state when it terminated due to the QUIT signal. You can examine this file yourself, or forward information to the maintainer of the program telling him/her where the *core file* is.

If you run background commands (as explained in section 2.6) then these commands will ignore INTERRUPT and QUIT signals at the terminal. To stop them you must use the *kill* command. See section 2.6 for an example.

If you want to examine the output of a command without having it move off the screen as the output of the

    cat /etc/passwd

command will, you can use the command

    more /etc/passwd

The *more* program pauses after each complete screenful and types '—More—' at which point you can hit a space to get another screenful, a return to get another line, or a 'q' to end the *more* program. You can also use more as a filter, i.e.

    cat /etc/passwd | more

works just like the more simple more command above.

For stopping output of commands not involving *more* you can use the ↑S key to stop the typeout. The typeout will resume when you hit ↑Q or any other key, but ↑Q is normally used because it only restarts the output and does not become input to the program which is running. This works well on low-speed terminals, but at 9600 baud it is hard to type ↑S and ↑Q fast enough to paginate the output nicely, and a program like *more* is usually used.

An additional possibility is to use the ↑O flush output character; when this character is typed, all output from the current command is thrown away (quickly) until the next input read occurs or until the next shell prompt. This can be used to allow a command to complete without having to suffer through the output on a slow terminal; ↑O is a toggle, so flushing can be turned off by typing ↑O again while output is being flushed.

1.9. What now?

We have so far seen a number of mechanisms of the shell and learned a lot about the way in which it operates. The remaining sections will go yet further into the internals of the shell, but you will surely want to try using the shell before you go any further. To try it you can log in to UNIX and type the following command to the system:

    chsh myname /bin/csh

Here 'myname' should be replaced by the name you typed to the system prompt of 'login:' to get onto the system. Thus I would use 'chsh bill /bin/csh'. **You only have to do this once; it takes effect at next login.** You are now ready to try using *csh*.

Before you do the 'chsh' command, the shell you are using when you log into the system is '/bin/sh'. In fact, much of the above discussion is applicable to '/bin/sh'. The next section will introduce many features particular to *csh* so you should change your shell to *csh* before you begin reading it.

2. Details on the shell for terminal users

2.1. Shell startup and termination

When you login, the shell is started by the system in your *home* directory and begins by reading commands from a file *.cshrc* in this directory. All shells which you may start during your terminal session will read from this file. We will later see what kinds of commands are usefully placed there. For now we need not have this file and the shell does not complain about its absence.

A *login shell*, executed after you login to the system, will, after it reads commands from *.cshrc,* read commands from a file *.login* also in your home directory. This file contains commands which you wish to do each time you login to the UNIX system. My *.login* file looks something like:

```
set ignoreeof
set mail=(/usr/spool/mail/bill)
echo "${prompt}users" ; users
alias ts \
      'set noglob ; eval `tset -s -m dialup:c100rv4pna -m plugboard:?hp2621nl *`';
ts; stty intr ↑C kill ↑U crt
set time=15 history=10
msgs -f
if (-e $mail) then
      echo "${prompt}mail"
      mail
endif
```

This file contains several commands to be executed by UNIX each time I login. The first is a *set* command which is interpreted directly by the shell. It sets the shell variable *ignoreeof* which causes the shell to not log me off if I hit ↑D. Rather, I use the *logout* command to log off of the system. By setting the *mail* variable, I ask the shell to watch for incoming mail to me. Every 5 minutes the shell looks for this file and tells me if more mail has arrived there. An alternative to this is to put the command

```
biff y
```

in place of this *set;* this will cause me to be notified immediately when mail arrives, and to be shown the first few lines of the new message.

Next I set the shell variable 'time' to '15' causing the shell to automatically print out statistics lines for commands which execute for at least 15 seconds of CPU time. The variable 'history' is set to 10 indicating that I want the shell to remember the last 10 commands I type in its *history list*, (described later).

I create an *alias* "ts" which executes a *tset*(1) command setting up the modes of the terminal. The parameters to *tset* indicate the kinds of terminal which I usually use when not on a hardwired port. I then execute "ts" and also use the *stty* command to change the interrupt character to ↑C and the line kill character to ↑U.

I then run the 'msgs' program, which provides me with any system messages which I have not seen before; the '-f' option here prevents it from telling me anything if there are no new messages. Finally, if my mailbox file exists, then I run the 'mail' program to process my mail.

When the 'mail' and 'msgs' programs finish, the shell will finish processing my *.login* file and begin reading commands from the terminal, prompting for each with '% '. When I log off (by giving the *logout* command) the shell will print 'logout' and execute commands from the file '.logout' if it exists in my home directory. After that the shell will terminate and UNIX will log me off the system. If the system is not going down, I will receive a new login message. In any case, after the 'logout' message the shell is committed to terminating and will take no further input from my terminal.

2.2. Shell variables

The shell maintains a set of *variables*. We saw above the variables *history* and *time* which had values '10' and '15'. In fact, each shell variable has as value an array of zero or more *strings*. Shell variables may be assigned values by the set command. It has several forms, the most useful of which was given above and is

set name=value

Shell variables may be used to store values which are to be used in commands later through a substitution mechanism. The shell variables most commonly referenced are, however, those which the shell itself refers to. By changing the values of these variables one can directly affect the behavior of the shell.

One of the most important variables is the variable *path*. This variable contains a sequence of directory names where the shell searches for commands. The *set* command with no arguments shows the value of all variables currently defined (we usually say *set)* in the shell. The default value for path will be shown by *set* to be

```
% set
argv     ()
cwd      /usr/bill
home     /usr/bill
path     (. /usr/ucb /bin /usr/bin)
prompt   %
shell    /bin/csh
status   0
term     c100rv4pna
user     bill
%
```

This output indicates that the variable path points to the current directory '.' and then '/usr/ucb', '/bin' and '/usr/bin'. Commands which you may write might be in '.' (usually one of your directories). Commands developed at Berkeley, live in '/usr/ucb' while commands developed at Bell Laboratories live in '/bin' and '/usr/bin'.

A number of locally developed programs on the system live in the directory '/usr/local'. If we wish that all shells which we invoke to have access to these new programs we can place the command

set path=(. /usr/ucb /bin /usr/bin /usr/local)

in our file *.cshrc* in our home directory. Try doing this and then logging out and back in and do

set

again to see that the value assigned to *path* has changed.

One thing you should be aware of is that the shell examines each directory which you insert into your path and determines which commands are contained there. Except for the current directory '.', which the shell treats specially, this means that if commands are added to a directory in your search path after you have started the shell, they will not necessarily be found by the shell. If you wish to use a command which has been added in this way, you should give the command

rehash

to the shell, which will cause it to recompute its internal table of command locations, so that it will find the newly added command. Since the shell has to look in the current directory '.' on each command, placing it at the end of the path specification usually works equivalently and reduces overhead.

Other useful built in variables are the variable *home* which shows your home directory, *cwd* which contains your current working directory, the variable *ignoreeof* which can be set in your *.login* file to tell the shell not to exit when it receives an end-of-file from a terminal (as described above). The variable 'ignoreeof' is one of several variables which the shell does not care about the value of, only whether they are *set* or *unset*. Thus to set this variable you simply do

    set ignoreeof

and to unset it do

    unset ignoreeof

These give the variable 'ignoreeof' no value, but none is desired or required.

Finally, some other built-in shell variables of use are the variables *noclobber* and *mail*. The metasyntax

    > filename

which redirects the standard output of a command will overwrite and destroy the previous contents of the named file. In this way you may accidentally overwrite a file which is valuable. If you would prefer that the shell not overwrite files in this way you can

    set noclobber

in your *.login* file. Then trying to do

    date > now

would cause a diagnostic if 'now' existed already. You could type

    date >! now

if you really wanted to overwrite the contents of 'now'. The '>!' is a special metasyntax indicating that clobbering the file is ok.†

2.3. The shell's history list

The shell can maintain a *history list* into which it places the words of previous commands. It is possible to use a notation to reuse commands or words from commands in forming new commands. This mechanism can be used to repeat previous commands or to correct minor typing mistakes in commands.

The following figure gives a sample session involving typical usage of the history mechanism of the shell. In this example we have a very simple C program which has a bug (or two) in it in the file 'bug.c', which we 'cat' out on our terminal. We then try to run the C compiler on it, referring to the file again as '!$', meaning the last argument to the previous command. Here the '!' is the history mechanism invocation metacharacter, and the '$' stands for the last argument, by analogy to '$' in the editor which stands for the end of the line. The shell echoed the command, as it would have been typed without use of the history mechanism, and then executed it. The compilation yielded error diagnostics so we now run the editor on the file we were trying to compile, fix the bug, and run the C compiler again, this time referring to this command simply as '!c', which repeats the last command which started with the letter 'c'. If there were other commands starting with 'c' done recently we could have said '!cc' or even '!cc:p' which would have printed the last command starting with 'cc' without executing it.

After this recompilation, we ran the resulting 'a.out' file, and then noting that there still was a bug, ran the editor again. After fixing the program we ran the C compiler again, but tacked onto the command an extra '–o bug' telling the compiler to place the resultant binary in the file 'bug' rather than 'a.out'. In general, the history mechanisms may be used anywhere in the

†The space between the '!' and the word 'now' is critical here, as '!now' would be an invocation of the *history* mechanism, and have a totally different effect.

```
% cat bug.c
main()

{
	printf("hello);
}
% cc !$
cc bug.c
"bug.c", line 4: newline in string or char constant
"bug.c", line 5: syntax error
% ed !$
ed bug.c
29
4s/);/"&/p
	printf("hello");
w
30
q
% !c
cc bug.c
% a.out
hello% !e
ed bug.c
30
4s/lo/lo\\n/p
	printf("hello\n");
w
32
q
% !c -o bug
cc bug.c -o bug
% size a.out bug
a.out: 2784+364+1028 = 4176b = 0x1050b
bug: 2784+364+1028 = 4176b = 0x1050b
% ls -l !*
ls -l a.out bug
-rwxr-xr-x 1 bill      3932 Dec 19 09:41 a.out
-rwxr-xr-x 1 bill      3932 Dec 19 09:42 bug
% bug
hello
% num bug.c | spp
spp: Command not found.
% ↑spp↑ssp
num bug.c | ssp
   1  main()
   3  {
   4          printf("hello\n");
   5  }
% !! | lpr
num bug.c | ssp | lpr
%
```

formation of new commands and other characters may be placed before and after the substituted commands.

We then ran the 'size' command to see how large the binary program images we have created were, and then an 'ls –l' command with the same argument list, denoting the argument list '*'. Finally we ran the program 'bug' to see that its output is indeed correct.

To make a numbered listing of the program we ran the 'num' command on the file 'bug.c'. In order to compress out blank lines in the output of 'num' we ran the output through the filter 'ssp', but misspelled it as spp. To correct this we used a shell substitute, placing the old text and new text between '↑' characters. This is similar to the substitute command in the editor. Finally, we repeated the same command with '!!', but sent its output to the line printer.

There are other mechanisms available for repeating commands. The *history* command prints out a number of previous commands with numbers by which they can be referenced. There is a way to refer to a previous command by searching for a string which appeared in it, and there are other, less useful, ways to select arguments to include in a new command. A complete description of all these mechanisms is given in the C shell manual pages in the UNIX Programmers Manual.

2.4. Aliases

The shell has an *alias* mechanism which can be used to make transformations on input commands. This mechanism can be used to simplify the commands you type, to supply default arguments to commands, or to perform transformations on commands and their arguments. The alias facility is similar to a macro facility. Some of the features obtained by aliasing can be obtained also using shell command files, but these take place in another instance of the shell and cannot directly affect the current shells environment or involve commands such as *cd* which must be done in the current shell.

As an example, suppose that there is a new version of the mail program on the system called 'newmail' you wish to use, rather than the standard mail program which is called 'mail'. If you place the shell command

```
alias mail newmail
```

in your *.cshrc* file, the shell will transform an input line of the form

```
mail bill
```

into a call on 'newmail'. More generally, suppose we wish the command 'ls' to always show sizes of files, that is to always do '–s'. We can do

```
alias ls ls –s
```

or even

```
alias dir ls –s
```

creating a new command syntax 'dir' which does an 'ls –s'. If we say

```
dir ~bill
```

then the shell will translate this to

```
ls –s /mnt/bill
```

Thus the *alias* mechanism can be used to provide short names for commands, to provide default arguments, and to define new short commands in terms of other commands. It is also possible to define aliases which contain multiple commands or pipelines, showing where the arguments to the original command are to be substituted using the facilities of the history mechanism. Thus the definition

```
alias cd ′cd \!* ; ls ′
```

would do an *ls* command after each change directory *cd* command. We enclosed the entire alias definition in '' characters to prevent most substitutions from occurring and the character ';' from being recognized as a metacharacter. The '!' here is escaped with a '\' to prevent it from being interpreted when the alias command is typed in. The '\!*' here substitutes the entire argument list to the pre-aliasing *cd* command, without giving an error if there were no arguments. The ';' separating commands is used here to indicate that one command is to be done and then the next. Similarly the definition

alias whois 'grep \!↑ /etc/passwd'

defines a command which looks up its first argument in the password file.

Warning: The shell currently reads the *.cshrc* file each time it starts up. If you place a large number of commands there, shells will tend to start slowly. A mechanism for saving the shell environment after reading the *.cshrc* file and quickly restoring it is under development, but for now you should try to limit the number of aliases you have to a reasonable number... 10 or 15 is reasonable, 50 or 60 will cause a noticeable delay in starting up shells, and make the system seem sluggish when you execute commands from within the editor and other programs.

2.5. More redirection; >> and >&

There are a few more notations useful to the terminal user which have not been introduced yet.

In addition to the standard output, commands also have a *diagnostic output* which is normally directed to the terminal even when the standard output is redirected to a file or a pipe. It is occasionally desirable to direct the diagnostic output along with the standard output. For instance if you want to redirect the output of a long running command into a file and wish to have a record of any error diagnostic it produces you can do

command >& file

The '>&' here tells the shell to route both the diagnostic output and the standard output into 'file'. Similarly you can give the command

command |& lpr

to route both standard and diagnostic output through the pipe to the line printer daemon *lpr.*#

Finally, it is possible to use the form

command >> file

to place output at the end of an existing file.†

2.6. Jobs; Background, Foreground, or Suspended

When one or more commands are typed together as a pipeline or as a sequence of commands separated by semicolons, a single *job* is created by the shell consisting of these commands together as a unit. Single commands without pipes or semicolons create the simplest jobs. Usually, every line typed to the shell creates a job. Some lines that create jobs (one per line) are

---

#A command form

command >&! file

exists, and is used when *noclobber* is set and *file* already exists.

†If *noclobber* is set, then an error will result if *file* does not exist, otherwise the shell will create *file* if it doesn't exist. A form

command >>! file

makes it not be an error for file to not exist when *noclobber* is set.

sort < data
ls –s | sort –n | head –5
mail harold

If the metacharacter '&' is typed at the end of the commands, then the job is started as a *background* job. This means that the shell does not wait for it to complete but immediately prompts and is ready for another command. The job runs *in the background* at the same time that normal jobs, called *foreground* jobs, continue to be read and executed by the shell one at a time. Thus

du > usage &

would run the *du* program, which reports on the disk usage of your working directory (as well as any directories below it), put the output into the file 'usage' and return immediately with a prompt for the next command without out waiting for *du* to finish. The *du* program would continue executing in the background until it finished, even though you can type and execute more commands in the mean time. When a background job terminates, a message is typed by the shell just before the next prompt telling you that the job has completed. In the following example the *du* job finishes sometime during the execution of the *mail* command and its completion is reported just before the prompt after the *mail* job is finished.

```
% du > usage &
[1] 503
% mail bill
How do you know when a background job is finished?
EOT
[1] – Done                du > usage
%
```

If the job did not terminate normally the 'Done' message might say something else like 'Killed'. If you want the terminations of background jobs to be reported at the time they occur (possibly interrupting the output of other foreground jobs), you can set the *notify* variable. In the previous example this would mean that the 'Done' message might have come right in the middle of the message to Bill. Background jobs are unaffected by any signals from the keyboard like the STOP, INTERRUPT, or QUIT signals mentioned earlier.

Jobs are recorded in a table inside the shell until they terminate. In this table, the shell remembers the command names, arguments and the *process numbers* of all commands in the job as well as the working directory where the job was started. Each job in the table is either running *in the foreground* with the shell waiting for it to terminate, running *in the background,* or *suspended.* Only one job can be running in the foreground at one time, but several jobs can be suspended or running in the background at once. As each job is started, it is assigned a small identifying number called the *job number* which can be used later to refer to the job in the commands described below. Job numbers remain the same until the job terminates and then are re-used.

When a job is started in the backgound using '&', its number, as well as the process numbers of all its (top level) commands, is typed by the shell before prompting you for another command. For example,

```
% ls –s | sort –n > usage &
[2] 2034 2035
%
```

runs the 'ls' program with the '–s' options, pipes this output into the 'sort' program with the '–n' option which puts its output into the file 'usage'. Since the '&' was at the end of the line, these two programs were started together as a background job. After starting the job, the shell prints the job number in brackets (2 in this case) followed by the process number of each program started in the job. Then the shell immediates prompts for a new command, leaving the job running simultaneously.

As mentioned in section 1.8, foreground jobs become ***suspended*** by typing ↑Z which sends a STOP signal to the currently running foreground job. A background job can become suspended by using the ***stop*** command described below. When jobs are suspended they merely stop any further progress until started again, either in the foreground or the backgound. The shell notices when a job becomes stopped and reports this fact, much like it reports the termination of background jobs. For foreground jobs this looks like

```
% du > usage
↑Z
Stopped
%
```

'Stopped' message is typed by the shell when it notices that the *du* program stopped. For background jobs, using the *stop* command, it is

```
% sort usage &
[1] 2345
% stop %1
[1] + Stopped (signal)        sort usage
%
```

Suspending foreground jobs can be very useful when you need to temporarily change what you are doing (execute other commands) and then return to the suspended job. Also, foreground jobs can be suspended and then continued as background jobs using the *bg* command, allowing you to continue other work and stop waiting for the foreground job to finish. Thus

```
% du > usage
↑Z
Stopped
% bg
[1] du > usage &
%
```

starts 'du' in the foreground, stops it before it finishes, then continues it in the background allowing more foreground commands to be executed. This is especially helpful when a foreground job ends up taking longer than you expected and you wish you had started it in the backgound in the beginning.

All *job control* commands can take an argument that identifies a particular job. All job name arguments begin with the character '%', since some of the job control commands also accept process numbers (printed by the *ps* command.) The default job (when no argument is given) is called the *current* job and is identified by a '+' in the output of the *jobs* command, which shows you which jobs you have. When only one job is stopped or running in the background (the usual case) it is always the current job thus no argument is needed. If a job is stopped while running in the foreground it becomes the *current* job and the existing current job becomes the *previous* job – identified by a '–' in the output of *jobs*. When the current job terminates, the previous job becomes the current job. When given, the argument is either '%–' (indicating the previous job); '%#', where # is the job number; '%pref' where pref is some unique prefix of the command name and arguments of one of the jobs; or '%?' followed by some string found in only one of the jobs.

The *jobs* command types the table of jobs, giving the job number, commands and status ('Stopped' or 'Running') of each backgound or suspended job. With the '–l' option the process numbers are also typed.

```
% du > usage &
[1] 3398
% ls -s | sort -n > myfile &
[2] 3405
% mail bill
↑Z
Stopped
% jobs
[1] - Running          du > usage
[2]   Running          ls -s | sort -n > myfile
[3] + Stopped          mail bill
% fg %ls
ls -s | sort -n > myfile
% more myfile
```

The *fg* command runs a suspended or background job in the foreground. It is used to restart a previously suspended job or change a background job to run in the foreground (allowing signals or input from the terminal). In the above example we used *fg* to change the 'ls' job from the background to the foreground since we wanted to wait for it to finish before looking at its output file. The *bg* command runs a suspended job in the background. It is usually used after stopping the currently running foreground job with the STOP signal. The combination of the STOP signal and the *bg* command changes a foreground job into a background job. The *stop* command suspends a background job.

The *kill* command terminates a background or suspended job immediately. In addition to jobs, it may be given process numbers as arguments, as printed by *ps*. Thus, in the example above, the running *du* command could have been terminated by the command

```
% kill %1
[1] Terminated          du > usage
%
```

The *notify* command (not the variable mentioned earlier) indicates that the termination of a specific job should be reported at the time it finishes instead of waiting for the next prompt.

If a job running in the background tries to read input from the terminal it is automatically stopped. When such a job is then run in the foreground, input can be given to the job. If desired, the job can be run in the background again until it requests input again. This is illustrated in the following sequence where the 's' command in the text editor might take a long time.

```
% ed bigfile
120000
1,$s/thisword/thatword/
↑Z
Stopped
% bg
[1] ed bigfile &
%
 . . . some foreground commands
[1] Stopped (tty input)       ed bigfile
% fg
ed bigfile
w
120000
q
%
```

So after the 's' command was issued, the 'ed' job was stopped with ↑Z and then put in the background using *bg*. Some time later when the 's' command was finished, *ed* tried to read another command and was stopped because jobs in the backgound cannot read from the terminal. The *fg* command returned the 'ed' job to the foreground where it could once again accept commands from the terminal.

The command

```
stty tostop
```

causes all background jobs run on your terminal to stop when they are about to write output to the terminal. This prevents messages from background jobs from interrupting foreground job output and allows you to run a job in the background without losing terminal output. It also can be used for interactive programs that sometimes have long periods without interaction. Thus each time it outputs a prompt for more input it will stop before the prompt. It can then be run in the foreground using *fg*, more input can be given and, if necessary stopped and returned to the background. This *stty* command might be a good thing to put in your *.login* file if you do not like output from background jobs interrupting your work. It also can reduce the need for redirecting the output of background jobs if the output is not very big:

```
% stty tostop
% wc hugefile &
[1] 10387
% ed text
. . . some time later
q
[1] Stopped (tty output)       wc hugefile
% fg wc
wc hugefile
  13371  30123  302577
% stty –tostop
```

Thus after some time the 'wc' command, which counts the lines, words and characters in a file, had one line of output. When it tried to write this to the terminal it stopped. By restarting it in the foreground we allowed it to write on the terminal exactly when we were ready to look at its output. Programs which attempt to change the mode of the terminal will also block, whether or not *tostop* is set, when they are not in the foreground, as it would be very unpleasant to have a background job change the state of the terminal.

Since the *jobs* command only prints jobs started in the currently executing shell, it knows nothing about background jobs started in other login sessions or within shell files. The *ps* can be used in this case to find out about background jobs not started in the current shell.

2.7. Working Directories

As mentioned in section 1.6, the shell is always in a particular *working directory*. The 'change directory' command *chdir* (its short form *cd* may also be used) changes the working directory of the shell, that is, changes the directory you are located in.

It is useful to make a directory for each project you wish to work on and to place all files related to that project in that directory. The 'make directory' command, *mkdir*, creates a new directory. The *pwd* ('print working directory') command reports the absolute pathname of the working directory of the shell, that is, the directory you are located in. Thus in the example below:

```
% pwd
/usr/bill
% mkdir newpaper
% chdir newpaper
% pwd
/usr/bill/newpaper
%
```

the user has created and moved to the directory *newpaper*. where, for example, he might place a group of related files.

No matter where you have moved to in a directory hierarchy, you can return to your 'home' login directory by doing just

```
cd
```

with no arguments. The name '..' always means the directory above the current one in the hierarchy, thus

```
cd ..
```

changes the shell's working directory to the one directly above the current one. The name '..' can be used in any pathname, thus,

```
cd ../programs
```

means change to the directory 'programs' contained in the directory above the current one. If you have several directories for different projects under, say, your home directory, this shorthand notation permits you to switch easily between them.

The shell always remembers the pathname of its current working directory in the variable *cwd*. The shell can also be requested to remember the previous directory when you change to a new working directory. If the 'push directory' command *pushd* is used in place of the *cd* command, the shell saves the name of the current working directory on a *directory stack* before changing to the new one. You can see this list at any time by typing the 'directories' command *dirs*.

```
% pushd newpaper/references
~/newpaper/references  ~
% pushd /usr/lib/tmac
/usr/lib/tmac  ~/newpaper/references  ~
% dirs
/usr/lib/tmac  ~/newpaper/references  ~
% popd
~/newpaper/references  ~
% popd
~
%
```

The list is printed in a horizontal line, reading left to right, with a tilde (~) as shorthand for your home directory—in this case '/usr/bill'. The directory stack is printed whenever there is more than one entry on it and it changes. It is also printed by a *dirs* command. *Dirs* is usually faster and more informative than *pwd* since it shows the current working directory as well as any other directories remembered in the stack.

The *pushd* command with no argument alternates the current directory with the first directory in the list. The 'pop directory' *popd* command without an argument returns you to the directory you were in prior to the current one, discarding the previous current directory from the stack (forgetting it). Typing *popd* several times in a series takes you backward through the directories you had been in (changed to) by *pushd* command. There are other options to *pushd* and *popd* to manipulate the contents of the directory stack and to change to directories not at the top of the stack; see the *csh* manual page for details.

Since the shell remembers the working directory in which each job was started, it warns you when you might be confused by restarting a job in the foreground which has a different working directory than the current working directory of the shell. Thus if you start a background job, then change the shell's working directory and then cause the background job to run in the foreground, the shell warns you that the working directory of the currently running foreground job is different from that of the shell.

```
% dirs -l
/mnt/bill
% cd myproject
% dirs
~/myproject
% ed prog.c
1143
↑Z
Stopped
% cd ..
% ls
myproject
textfile
% fg
ed prog.c (wd: ~/myproject)
```

This way the shell warns you when there is an implied change of working directory, even though no cd command was issued. In the above example the 'ed' job was still in '/mnt/bill/project' even though the shell had changed to '/mnt/bill'. A similar warning is given when such a foreground job terminates or is suspended (using the STOP signal) since the return to the shell again implies a change of working directory.

```
% fg
ed prog.c (wd: ~/myproject)
 . . . after some editing
q
(wd now: ~)
%
```

These messages are sometimes confusing if you use programs that change their own working directories, since the shell only remembers which directory a job is started in, and assumes it stays there. The '-l' option of *jobs* will type the working directory of suspended or background jobs when it is different from the current working directory of the shell.

2.8. Useful built-in commands

We now give a few of the useful built-in commands of the shell describing how they are used.

The *alias* command described above is used to assign new aliases and to show the existing aliases. With no arguments it prints the current aliases. It may also be given only one argument such as

```
alias ls
```

to show the current alias for, e.g., 'ls'.

The *echo* command prints its arguments. It is often used in *shell scripts* or as an interactive command to see what filename expansions will produce.

The *history* command will show the contents of the history list. The numbers given with the history events can be used to reference previous events which are difficult to reference using the contextual mechanisms introduced above. There is also a shell variable called *prompt*. By placing a '!' character in its value the shell will there substitute the number of the current command in the history list. You can use this number to refer to this command in a history substitution. Thus you could

set prompt=′\! % ′

Note that the '!' character had to be *escaped* here even within '′' characters.

The *limit* command is used to restrict use of resources. With no arguments it prints the current limitations:

cputime	unlimited
filesize	unlimited
datasize	5616 kbytes
stacksize	512 kbytes
coredumpsize	unlimited

Limits can be set, e.g.:

limit coredumpsize 128k

Most reasonable units abbreviations will work; see the *csh* manual page for more details.

The *logout* command can be used to terminate a login shell which has *ignoreeof* set.

The *rehash* command causes the shell to recompute a table of where commands are located. This is necessary if you add a command to a directory in the current shell's search path and wish the shell to find it, since otherwise the hashing algorithm may tell the shell that the command wasn't in that directory when the hash table was computed.

The *repeat* command can be used to repeat a command several times. Thus to make 5 copies of the file *one* in the file *five* you could do

repeat 5 cat one >> five

The *setenv* command can be used to set variables in the environment. Thus

setenv TERM adm3a

will set the value of the environment variable TERM to 'adm3a'. A user program *printenv* exists which will print out the environment. It might then show:

```
% printenv
HOME=/usr/bill
SHELL=/bin/csh
PATH=:/usr/ucb:/bin:/usr/bin:/usr/local
TERM=adm3a
USER=bill
%
```

The *source* command can be used to force the current shell to read commands from a file. Thus

source .cshrc

can be used after editing in a change to the *.cshrc* file which you wish to take effect before the next time you login.

The *time* command can be used to cause a command to be timed no matter how much CPU time it takes. Thus

```
% time cp /etc/rc /usr/bill/rc
0.0u 0.1s 0:01 8% 2+1k 3+2io 1pf+0w
% time wc /etc/rc /usr/bill/rc
    52   178   1347 /etc/rc
    52   178   1347 /usr/bill/rc
   104   356   2694 total
0.1u 0.1s 0:00 13% 3+3k 5+3io 7pf+0w
%
```

indicates that the *cp* command used a negligible amount of user time (u) and about 1/10th of a system time (s); the elapsed time was 1 second (0:01), there was an average memory usage of 2k bytes of program space and 1k bytes of data space over the cpu time involved (2+1k); the program did three disk reads and two disk writes (3+2io), and took one page fault and was not swapped (1pf+0w). The word count command *wc* on the other hand used 0.1 seconds of user time and 0.1 seconds of system time in less than a second of elapsed time. The percentage '13%' indicates that over the period when it was active the command 'wc' used an average of 13 percent of the available CPU cycles of the machine.

The *unalias* and *unset* commands can be used to remove aliases and variable definitions from the shell, and *unsetenv* removes variables from the environment.

2.9. What else?

This concludes the basic discussion of the shell for terminal users. There are more features of the shell to be discussed here, and all features of the shell are discussed in its manual pages. One useful feature which is discussed later is the *foreach* built-in command which can be used to run the same command sequence with a number of different arguments.

If you intend to use UNIX a lot you you should look through the rest of this document and the shell manual pages to become familiar with the other facilities which are available to you.

3. Shell control structures and command scripts

3.1. Introduction

It is possible to place commands in files and to cause shells to be invoked to read and execute commands from these files, which are called ***shell scripts.*** We here detail those features of the shell useful to the writers of such scripts.

3.2. Make

It is important to first note what shell scripts are *not* useful for. There is a program called *make* which is very useful for maintaining a group of related files or performing sets of operations on related files. For instance a large program consisting of one or more files can have its dependencies described in a *makefile* which contains definitions of the commands used to create these different files when changes occur. Definitions of the means for printing listings, cleaning up the directory in which the files reside, and installing the resultant programs are easily, and most appropriately placed in this *makefile.* This format is superior and preferable to maintaining a group of shell procedures to maintain these files.

Similarly when working on a document a *makefile* may be created which defines how different versions of the document are to be created and which options of *nroff* or *troff* are appropriate.

3.3. Invocation and the argv variable

A *csh* command script may be interpreted by saying

% csh script ...

where *script* is the name of the file containing a group of *csh* commands and '...' is replaced by a sequence of arguments. The shell places these arguments in the variable *argv* and then begins to read commands from the script. These parameters are then available through the same mechanisms which are used to reference any other shell variables.

If you make the file 'script' executable by doing

chmod 755 script

and place a shell comment at the beginning of the shell script (i.e. begin the file with a '#' character) then a '/bin/csh' will automatically be invoked to execute 'script' when you type

script

If the file does not begin with a '#' then the standard shell '/bin/sh' will be used to execute it. This allows you to convert your older shell scripts to use *csh* at your convenience.

3.4. Variable substitution

After each input line is broken into words and history substitutions are done on it, the input line is parsed into distinct commands. Before each command is executed a mechanism know as *variable substitution* is done on these words. Keyed by the character '$' this substitution replaces the names of variables by their values. Thus

echo $argv

when placed in a command script would cause the current value of the variable *argv* to be echoed to the output of the shell script. It is an error for *argv* to be unset at this point.

A number of notations are provided for accessing components and attributes of variables. The notation

$?name

expands to '1' if name is *set* or to '0' if name is not *set.* It is the fundamental mechanism used for checking whether particular variables have been assigned values. All other forms of reference to undefined variables cause errors.

The notation

```
$#name
```

expands to the number of elements in the variable *name*. Thus

```
% set argv=(a b c)
% echo $?argv
1
% echo $#argv
3
% unset argv
% echo $?argv
0
% echo $argv
Undefined variable: argv.
%
```

It is also possible to access the components of a variable which has several values. Thus

```
$argv[1]
```

gives the first component of *argv* or in the example above 'a'. Similarly

```
$argv[$#argv]
```

would give 'c', and

```
$argv[1-2]
```

would give 'a b'. Other notations useful in shell scripts are

$*n*

where *n* is an integer as a shorthand for

$argv[*n*]

the *n th* parameter and

```
$*
```

which is a shorthand for

```
$argv
```

The form

```
$$
```

expands to the process number of the current shell. Since this process number is unique in the system it can be used in generation of unique temporary file names. The form

```
$<
```

is quite special and is replaced by the next line of input read from the shell's standard input (not the script it is reading). This is useful for writing shell scripts that are interactive, reading commands from the terminal, or even writing a shell script that acts as a filter, reading lines from its input file. Thus the sequence

```
echo 'yes or no?\c'
set a=($<)
```

would write out the prompt 'yes or no?' without a newline and then read the answer into the variable 'a'. In this case '$#a' would be '0' if either a blank line or end-of-file (↑D) was typed.

One minor difference between '$*n*' and '$argv[*n*]' should be noted here. The form '$argv[*n*]' will yield an error if *n* is not in the range '1–$#argv' while '$n' will never yield an out of range subscript error. This is for compatibility with the way older shells handled parameters.

Another important point is that it is never an error to give a subrange of the form 'n–'; if there are less than *n* components of the given variable then no words are substituted. A range of the form 'm–n' likewise returns an empty vector without giving an error when *m* exceeds the number of elements of the given variable, provided the subscript *n* is in range.

3.5. Expressions

In order for interesting shell scripts to be constructed it must be possible to evaluate expressions in the shell based on the values of variables. In fact, all the arithmetic operations of the language C are available in the shell with the same precedence that they have in C. In particular, the operations '==' and '!=' compare strings and the operators '&&' and '||' implement the boolean and/or operations. The special operators '=~' and '!~' are similar to '==' and '!=' except that the string on the right side can have pattern matching characters (like *, ? or []) and the test is whether the string on the left matches the pattern on the right.

The shell also allows file enquiries of the form

–? filename

where '?' is replace by a number of single characters. For instance the expression primitive

–e filename

tell whether the file 'filename' exists. Other primitives test for read, write and execute access to the file, whether it is a directory, or has non-zero length.

It is possible to test whether a command terminates normally, by a primitive of the form '{ command }' which returns true, i.e. '1' if the command succeeds exiting normally with exit status 0, or '0' if the command terminates abnormally or with exit status non-zero. If more detailed information about the execution status of a command is required, it can be executed and the variable '$status' examined in the next command. Since '$status' is set by every command, it is very transient. It can be saved if it is inconvenient to use it only in the single immediately following command.

For a full list of expression components available see the manual section for the shell.

3.6. Sample shell script

A sample shell script which makes use of the expression mechanism of the shell and some of its control structure follows:

```
% cat copyc
#
# Copyc copies those C programs in the specified list
# to the directory ~/backup if they differ from the files
# already in ~/backup
#
set noglob
foreach i ($argv)

        if ($i !~ *.c) continue  # not a .c file so do nothing

        if (! -r ~/backup/$i:t) then
                echo $i:t not in backup... not cp\'ed
                continue
        endif

        cmp -s $i ~/backup/$i:t # to set $status

        if ($status != 0) then
                echo new backup of $i
                cp $i ~/backup/$i:t
        endif
end
```

This script makes use of the *foreach* command, which causes the shell to execute the commands between the *foreach* and the matching *end* for each of the values given between '(' and ')' with the named variable, in this case 'i' set to successive values in the list. Within this loop we may use the command *break* to stop executing the loop and *continue* to prematurely terminate one iteration and begin the next. After the *foreach* loop the iteration variable (*i* in this case) has the value at the last iteration.

We set the variable *noglob* here to prevent filename expansion of the members of *argv*. This is a good idea, in general, if the arguments to a shell script are filenames which have already been expanded or if the arguments may contain filename expansion metacharacters. It is also possible to quote each use of a '$' variable expansion, but this is harder and less reliable.

The other control construct used here is a statement of the form

```
if ( expression ) then
        command
        ...
endif
```

The placement of the keywords here is **not** flexible due to the current implementation of the shell.†

†The following two formats are not currently acceptable to the shell:

```
if ( expression )               # Won't work!
then
        command
        ...
endif
```

and

```
if ( expression ) then command endif            # Won't work
```

The shell does have another form of the if statement of the form

if (expression) **command**

which can be written

if (expression) \
    command

Here we have escaped the newline for the sake of appearance. The command must not involve '|', '&' or ';' and must not be another control command. The second form requires the final '\' to **immediately** precede the end-of-line.

The more general *if* statements above also admit a sequence of *else-if* pairs followed by a single *else* and an *endif*, e.g.:

```
if ( expression ) then
      commands
else if (expression ) then
      commands
...

else
      commands
endif
```

Another important mechanism used in shell scripts is the ':' modifier. We can use the modifier ':r' here to extract a root of a filename or ':e' to extract the *extension*. Thus if the variable *i* has the value '/mnt/foo.bar' then

```
% echo $i $i:r $i:e
/mnt/foo.bar /mnt/foo bar
%
```

shows how the ':r' modifier strips off the trailing '.bar' and the the ':e' modifier leaves only the 'bar'. Other modifiers will take off the last component of a pathname leaving the head ':h' or all but the last component of a pathname leaving the tail ':t'. These modifiers are fully described in the *csh* manual pages in the programmers manual. It is also possible to use the *command substitution* mechanism described in the next major section to perform modifications on strings to then reenter the shells environment. Since each usage of this mechanism involves the creation of a new process, it is much more expensive to use than the ':' modification mechanism.# Finally, we note that the character '#' lexically introduces a shell comment in shell scripts (but not from the terminal). All subsequent characters on the input line after a '#' are discarded by the shell. This character can be quoted using ''' or '\' to place it in an argument word.

3.7. Other control structures

The shell also has control structures *while* and *switch* similar to those of C. These take the forms

---

#It is also important to note that the current implementation of the shell limits the number of ':' modifiers on a '$' substitution to 1. Thus

```
% echo $i $i:h:t
/a/b/c /a/b:t
%
```

does not do what one would expect.

```
while ( expression )
      commands
end
```

and

```
switch ( word )

case str1:
      commands
      breaksw

 ...

case strn:
      commands
      breaksw

default:
      commands
      breaksw

endsw
```

For details see the manual section for *csh.* C programmers should note that we use *breaksw* to exit from a *switch* while *break* exits a *while* or *foreach* loop. A common mistake to make in *csh* scripts is to use *break* rather than *breaksw* in switches.

Finally, *csh* allows a *goto* statement, with labels looking like they do in C, i.e.:

```
loop:
      commands
      goto loop
```

3.8. Supplying input to commands

Commands run from shell scripts receive by default the standard input of the shell which is running the script. This is different from previous shells running under UNIX. It allows shell scripts to fully participate in pipelines, but mandates extra notation for commands which are to take inline data.

Thus we need a metanotation for supplying inline data to commands in shell scripts. As an example, consider this script which runs the editor to delete leading blanks from the lines in each argument file

```
% cat deblank
# deblank — remove leading blanks
foreach i ($argv)
ed - $i << 'EOF'
1,$s/↑[ ]*//
w
q
'EOF'
end
%
```

The notation '<< 'EOF'' means that the standard input for the *ed* command is to come from the text in the shell script file up to the next line consisting of exactly ''EOF''. The fact that the 'EOF' is enclosed in '' characters, i.e. quoted, causes the shell to not perform variable substitution on the intervening lines. In general, if any part of the word following the '<<' which the shell uses to terminate the text to be given to the command is quoted then these substitutions will not be performed. In this case since we used the form '1,$' in our editor script we needed to insure that this '$' was not variable substituted. We could also have insured this by preceding the '$' here with a '\', i.e.:

    1,\$s/↑[ ]*//

but quoting the 'EOF' terminator is a more reliable way of achieving the same thing.

3.9. Catching interrupts

If our shell script creates temporary files, we may wish to catch interruptions of the shell script so that we can clean up these files. We can then do

    onintr label

where *label* is a label in our program. If an interrupt is received the shell will do a 'goto label' and we can remove the temporary files and then do an *exit* command (which is built in to the shell) to exit from the shell script. If we wish to exit with a non-zero status we can do

    exit(1)

e.g. to exit with status '1'.

3.10. What else?

There are other features of the shell useful to writers of shell procedures. The *verbose* and *echo* options and the related *-v* and *-x* command line options can be used to help trace the actions of the shell. The *-n* option causes the shell only to read commands and not to execute them and may sometimes be of use.

One other thing to note is that *csh* will not execute shell scripts which do not begin with the character '#', that is shell scripts that do not begin with a comment. Similarly, the '/bin/sh' on your system may well defer to 'csh' to interpret shell scripts which begin with '#'. This allows shell scripts for both shells to live in harmony.

There is also another quotation mechanism using '"' which allows only some of the expansion mechanisms we have so far discussed to occur on the quoted string and serves to make this string into a single word as ' ' does.

4. Other, less commonly used, shell features

4.1. Loops at the terminal; variables as vectors

It is occasionally useful to use the *foreach* control structure at the terminal to aid in performing a number of similar commands. For instance, there were at one point three shells in use on the Cory UNIX system at Cory Hall, '/bin/sh', '/bin/nsh', and '/bin/csh'. To count the number of persons using each shell one could have issued the commands

```
% grep –c csh$ /etc/passwd
27
% grep –c nsh$ /etc/passwd
128
% grep –c –v sh$ /etc/passwd
430
%
```

Since these commands are very similar we can use *foreach* to do this more easily.

```
% foreach i ('sh$' 'csh$' '–v sh$')
? grep –c $i /etc/passwd
? end
27
128
430
%
```

Note here that the shell prompts for input with '? ' when reading the body of the loop.

Very useful with loops are variables which contain lists of filenames or other words. You can, for example, do

```
% set a=(`ls`)
% echo $a
csh.n csh.rm
% ls
csh.n
csh.rm
% echo $#a
2
%
```

The *set* command here gave the variable *a* a list of all the filenames in the current directory as value. We can then iterate over these names to perform any chosen function.

The output of a command within '`' characters is converted by the shell to a list of words. You can also place the '`' quoted string within '"' characters to take each (non-empty) line as a component of the variable; preventing the lines from being split into words at blanks and tabs. A modifier ':x' exists which can be used later to expand each component of the variable into another variable splitting it into separate words at embedded blanks and tabs.

4.2. Braces { ... } in argument expansion

Another form of filename expansion, alluded to before involves the characters '{' and '}'. These characters specify that the contained strings, separated by ',' are to be consecutively substituted into the containing characters and the results expanded left to right. Thus

```
A{str1,str2,...strn}B
```

expands to

AstrlB Astr2B ... AstrnB

This expansion occurs before the other filename expansions, and may be applied recursively (i.e. nested). The results of each expanded string are sorted separately, left to right order being preserved. The resulting filenames are not required to exist if no other expansion mechanisms are used. This means that this mechanism can be used to generate arguments which are not filenames, but which have common parts.

A typical use of this would be

```
mkdir ~/{hdrs,retrofit,csh}
```

to make subdirectories 'hdrs', 'retrofit' and 'csh' in your home directory. This mechanism is most useful when the common prefix is longer than in this example, i.e.

```
chown root /usr/{ucb/{ex,edit},lib/{ex?.?*,how_ex}}
```

4.3. Command substitution

A command enclosed in '`' characters is replaced, just before filenames are expanded, by the output from that command. Thus it is possible to do

```
set pwd=`pwd`
```

to save the current directory in the variable *pwd* or to do

```
ex `grep -l TRACE *.c`
```

to run the editor *ex* supplying as arguments those files whose names end in '.c' which have the string 'TRACE' in them.*

4.4. Other details not covered here

In particular circumstances it may be necessary to know the exact nature and order of different substitutions performed by the shell. The exact meaning of certain combinations of quotations is also occasionally important. These are detailed fully in its manual section.

The shell has a number of command line option flags mostly of use in writing UNIX programs, and debugging shell scripts. See the shells manual section for a list of these options.

*Command expansion also occurs in input redirected with '<<' and within '"' quotations. Refer to the shell manual section for full details.

Appendix – Special characters

The following table lists the special characters of *csh* and the UNIX system, giving for each the section(s) in which it is discussed. A number of these characters also have special meaning in expressions. See the *csh* manual section for a complete list.

Syntactic metacharacters

;	2.4	separates commands to be executed sequentially
|	1.5	separates commands in a pipeline
( )	2.2,3.6	brackets expressions and variable values
&	2.5	follows commands to be executed without waiting for completion

Filename metacharacters

/	1.6	separates components of a file's pathname
?	1.6	expansion character matching any single character
*	1.6	expansion character matching any sequence of characters
[ ]	1.6	expansion sequence matching any single character from a set
~	1.6	used at the beginning of a filename to indicate home directories
{ }	4.2	used to specify groups of arguments with common parts

Quotation metacharacters

\	1.7	prevents meta-meaning of following single character
´	1.7	prevents meta-meaning of a group of characters
"	4.3	like ´, but allows variable and command expansion

Input/output metacharacters

<	1.5	indicates redirected input
>	1.3	indicates redirected output

Expansion/substitution metacharacters

$	3.4	indicates variable substitution
!	2.3	indicates history substitution
:	3.6	precedes substitution modifiers
↑	2.3	used in special forms of history substitution
`	4.3	indicates command substitution

Other metacharacters

#	1.3,3.6	begins scratch file names; indicates shell comments
–	1.2	prefixes option (flag) arguments to commands
%	2.6	prefixes job name specifications

Glossary

This glossary lists the most important terms introduced in the introduction to the shell and gives references to sections of the shell document for further information about them. References of the form 'pr (1)' indicate that the command *pr* is in the UNIX programmer's manual in section 1. You can get an online copy of its manual page by doing

```
man 1 pr
```

References of the form (2.5) indicate that more information can be found in section 2.5 of this manual.

. Your current directory has the name '.' as well as the name printed by the command *pwd;* see also *dirs*. The current directory '.' is usually the first *component* of the search path contained in the variable *path*, thus commands which are in '.' are found first (2.2). The character '.' is also used in separating *components* of filenames (1.6). The character '.' at the beginning of a *component* of a *pathname* is treated specially and not matched by the *filename expansion* metacharacters '?', '*', and '[' ']' pairs (1.6).

.. Each directory has a file '..' in it which is a reference to its parent directory. After changing into the directory with *chdir*, i.e.

```
chdir paper
```

you can return to the parent directory by doing

```
chdir ..
```

The current directory is printed by *pwd* (2.7).

a.out Compilers which create executable images create them, by default, in the file *a.out.* for historical reasons (2.3).

absolute pathname
A *pathname* which begins with a '/' is *absolute* since it specifies the *path* of directories from the beginning of the entire directory system – called the *root* directory. *Pathname*s which are not *absolute* are called *relative* (see definition of *relative pathname*) (1.6).

alias An *alias* specifies a shorter or different name for a UNIX command, or a transformation on a command to be performed in the shell. The shell has a command *alias* which establishes *aliases* and can print their current values. The command *unalias* is used to remove *aliases* (2.4).

argument Commands in UNIX receive a list of *argument* words. Thus the command

```
echo a b c
```

consists of the *command name* 'echo' and three *argument* words 'a', 'b' and 'c'. The set of *arguments* after the *command name* is said to be the *argument list* of the command (1.1).

argv The list of arguments to a command written in the shell language (a shell script or shell procedure) is stored in a variable called *argv* within the shell. This name is taken from the conventional name in the C programming language (3.4).

background Commands started without waiting for them to complete are called *background* commands (2.6).

base A filename is sometimes thought of as consisting of a *base* part, before any '.' character, and an *extension* – the part after the '.'. See *filename* and *extension* (1.6)

bg The *bg* command causes a *suspended* job to continue execution in the *background* (2.6).

bin — A directory containing binaries of programs and shell scripts to be executed is typically called a *bin* directory. The standard system *bin* directories are '/bin' containing the most heavily used commands and '/usr/bin' which contains most other user programs. Programs developed at UC Berkeley live in '/usr/ucb', while locally written programs live in '/usr/local'. Games are kept in the directory '/usr/games'. You can place binaries in any directory. If you wish to execute them often, the name of the directories should be a *component* of the variable *path*.

break — *Break* is a builtin command used to exit from loops within the control structure of the shell (3.7).

breaksw — The *breaksw* builtin command is used to exit from a *switch* control structure, like a *break* exits from loops (3.7).

builtin — A command executed directly by the shell is called a *builtin* command. Most commands in UNIX are not built into the shell, but rather exist as files in *bin* directories. These commands are accessible because the directories in which they reside are named in the *path* variable.

case — A *case* command is used as a label in a *switch* statement in the shell's control structure, similar to that of the language C. Details are given in the shell documentation 'csh(1)' (3.7).

cat — The *cat* program catenates a list of specified files on the *standard output*. It is usually used to look at the contents of a single file on the terminal, to 'cat a file' (1.8, 2.3).

cd — The *cd* command is used to change the *working directory*. With no arguments, *cd* changes your *working directory* to be your *home* directory (2.4, 2.7).

chdir — The *chdir* command is a synonym for *cd*. *Cd* is usually used because it is easier to type.

chsh — The *chsh* command is used to change the shell which you use on UNIX. By default, you use an different version of the shell which resides in '/bin/sh'. You can change your shell to '/bin/csh' by doing

    chsh your-login-name /bin/csh

Thus I would do

    chsh bill /bin/csh

It is only necessary to do this once. The next time you log in to UNIX after doing this command, you will be using *csh* rather than the shell in '/bin/sh' (1.9).

cmp — *Cmp* is a program which compares files. It is usually used on binary files, or to see if two files are identical (3.6). For comparing text files the program *diff*, described in 'diff (1)' is used.

command — A function performed by the system, either by the shell (a builtin *command*) or by a program residing in a file in a directory within the UNIX system, is called a *command* (1.1).

command name — When a command is issued, it consists of a *command name*, which is the first word of the command, followed by arguments. The convention on UNIX is that the first word of a command names the function to be performed (1.1).

command substitution — The replacement of a command enclosed in '\`' characters by the text output by that command is called *command substitution* (4.3).

component — A part of a *pathname* between '/' characters is called a *component* of that *pathname*. A variable which has multiple strings as value is said to have several *components*; each string is a *component* of the variable.

continue — A builtin command which causes execution of the enclosing *foreach* or *while* loop to cycle prematurely. Similar to the *continue* command in the programming language C (3.6).

control- — Certain special characters, called *control* characters, are produced by holding down the CONTROL key on your terminal and simultaneously pressing another character, much like the SHIFT key is used to produce upper case characters. Thus *control-* c is produced by holding down the CONTROL key while pressing the 'c' key. Usually UNIX prints an up-arrow (↑) followed by the corresponding letter when you type a *control* character (e.g. '↑C' for *control-* c (1.8).

core dump — When a program terminates abnormally, the system places an image of its current state in a file named 'core'. This *core dump* can be examined with the system debugger 'adb(1)' or 'sdb(1)' in order to determine what went wrong with the program (1.8). If the shell produces a message of the form

    Illegal instruction (core dumped)

(where 'Illegal instruction' is only one of several possible messages), you should report this to the author of the program or a system administrator, saving the 'core' file.

cp — The *cp* (copy) program is used to copy the contents of one file into another file. It is one of the most commonly used UNIX commands (1.6).

csh — The name of the shell program that this document describes.

.cshrc — The file *.cshrc* in your *home* directory is read by each shell as it begins execution. It is usually used to change the setting of the variable *path* and to set *alias* parameters which are to take effect globally (2.1).

cwd — The *cwd* variable in the shell holds the *absolute pathname* of the current *working directory*. It is changed by the shell whenever your current *working directory* changes and should not be changed otherwise (2.2).

date — The *date* command prints the current date and time (1.3).

debugging — *Debugging* is the process of correcting mistakes in programs and shell scripts. The shell has several options and variables which may be used to aid in shell *debugging* (4.4).

default: — The label *default:* is used within shell *switch* statements, as it is in the C language to label the code to be executed if none of the *case* labels matches the value switched on (3.7).

DELETE — The DELETE or RUBOUT key on the terminal normally causes an interrupt to be sent to the current job. Many users change the interrupt character to be ↑C.

detached — A command that continues running in the *background* after you logout is said to be *detached*.

diagnostic — An error message produced by a program is often referred to as a *diagnostic*. Most error messages are not written to the *standard output*, since that is often directed away from the terminal (1.3, 1.5). Error messsages are instead written to the *diagnostic output* which may be directed away from the terminal, but usually is not. Thus *diagnostics* will usually appear on the terminal (2.5).

directory — A structure which contains files. At any time you are in one particular *directory* whose names can be printed by the command *pwd*. The *chdir* command will change you to another *directory*, and make the files in that *directory* visible. The *directory* in which you are when you first login is your *home* directory (1.1, 2.7).

directory stack — The shell saves the names of previous *working directories* in the *directory stack* when you change your current *working directory* via the *pushd* command. The

directory stack can be printed by using the ***dirs*** command, which includes your current ***working directory*** as the first directory name on the left (2.7).

dirs — The ***dirs*** command prints the shell's ***directory stack*** (2.7).

du — The ***du*** command is a program (described in 'du(1)') which prints the number of disk blocks is all directories below and including your current ***working directory*** (2.6).

echo — The ***echo*** command prints its arguments (1.6, 3.6).

else — The ***else*** command is part of the 'if-then-else-endif' control command construct (3.6).

endif — If an ***if*** statement is ended with the word ***then***, all lines following the ***if*** up to a line starting with the word ***endif*** or ***else*** are executed if the condition between parentheses after the ***if*** is true (3.6).

EOF — An ***end-of-file*** is generated by the terminal by a control-d, and whenever a command reads to the end of a file which it has been given as input. Commands receiving input from a ***pipe*** receive an ***end-of-file*** when the command sending them input completes. Most commands terminate when they receive an ***end-of-file***. The shell has an option to ignore ***end-of-file*** from a terminal input which may help you keep from logging out accidentally by typing too many control-d's (1.1, 1.8, 3.8).

escape — A character '\' used to prevent the special meaning of a metacharacter is said to ***escape*** the character from its special meaning. Thus

```
echo \*
```

will echo the character '*' while just

```
echo *
```

will echo the names of the file in the current directory. In this example, \ ***escape*** s '*' (1.7). There is also a non-printing character called ***escape***, usually labelled ESC or ALTMODE on terminal keyboards. Some older UNIX systems use this character to indicate that output is to be ***suspended***. Most systems use control-s to stop the output and control-q to start it.

/etc/passwd — This file contains information about the accounts currently on the system. It consists of a line for each account with fields separated by ':' characters (1.8). You can look at this file by saying

```
cat /etc/passwd
```

The commands ***finger*** and ***grep*** are often used to search for information in this file. See 'finger(1)', 'passwd(5)', and 'grep(1)' for more details.

exit — The ***exit*** command is used to force termination of a shell script, and is built into the shell (3.9).

exit status — A command which discovers a problem may reflect this back to the command (such as a shell) which invoked (executed) it. It does this by returning a non-zero number as its ***exit status***, a status of zero being considered 'normal termination'. The ***exit*** command can be used to force a shell command script to give a non-zero ***exit status*** (3.6).

expansion — The replacement of strings in the shell input which contain metacharacters by other strings is referred to as the process of ***expansion***. Thus the replacement of the word '*' by a sorted list of files in the current directory is a 'filename expansion'. Similarly the replacement of the characters '!!' by the text of the last command is a 'history expansion'. ***Expansions*** are also referred to as ***substitutions*** (1.6, 3.4, 4.2).

expressions — *Expressions* are used in the shell to control the conditional structures used in the writing of shell scripts and in calculating values for these scripts. The operators available in shell *expressions* are those of the language C (3.5).

extension — Filenames often consist of a *base* name and an *extension* separated by the character '.'. By convention, groups of related files often share the same *root* name. Thus if 'prog.c' were a C program, then the object file for this program would be stored in 'prog.o'. Similarly a paper written with the '–me' nroff macro package might be stored in 'paper.me' while a formatted version of this paper might be kept in 'paper.out' and a list of spelling errors in 'paper.errs' (1.6).

fg — The *job control* command *fg* is used to run a *background* or *suspended* job in the *foreground* (1.8, 2.6).

filename — Each file in UNIX has a name consisting of up to 14 characters and not including the character '/' which is used in *pathname* building. Most *filenames* do not begin with the character '.', and contain only letters and digits with perhaps a '.' separating the *base* portion of the *filename* from an *extension* (1.6).

filename expansion — *Filename expansion* uses the metacharacters '*', '?' and '[' and ']' to provide a convenient mechanism for naming files. Using *filename expansion* it is easy to name all the files in the current directory, or all files which have a common *root* name. Other *filename expansion* mechanisms use the metacharacter '~' and allow files in other users' directories to be named easily (1.6, 4.2).

flag — Many UNIX commands accept arguments which are not the names of files or other users but are used to modify the action of the commands. These are referred to as *flag* options, and by convention consist of one or more letters preceded by the character '–' (1.2). Thus the *ls* (list files) command has an option '–s' to list the sizes of files. This is specified

```
ls –s
```

foreach — The *foreach* command is used in shell scripts and at the terminal to specify repetition of a sequence of commands while the value of a certain shell variable ranges through a specified list (3.6, 4.1).

foreground — When commands are executing in the normal way such that the shell is waiting for them to finish before prompting for another command they are said to be *foreground jobs* or *running in the foreground*. This is as opposed to *background*. *Foreground* jobs can be stopped by signals from the terminal caused by typing different control characters at the keyboard (1.8, 2.6).

goto — The shell has a command *goto* used in shell scripts to transfer control to a given label (3.7).

grep — The *grep* command searches through a list of argument files for a specified string. Thus

```
grep bill /etc/passwd
```

will print each line in the file */etc/passwd* which contains the string 'bill'. Actually, *grep* scans for *regular expressions* in the sense of the editors 'ed(1)' and 'ex(1)'. *Grep* stands for 'globally find *regular expression* and print' (2.4).

head — The *head* command prints the first few lines of one or more files. If you have a bunch of files containing text which you are wondering about it is sometimes useful to run *head* with these files as arguments. This will usually show enough of what is in these files to let you decide which you are interested in (1.5).
Head is also used to describe the part of a *pathname* before and including the last '/' character. The *tail* of a *pathname* is the part after the last '/'. The ':h' and ':t' modifiers allow the *head* or *tail* of a *pathname* stored in a shell variable

to be used (3.6).

history — The *history* mechanism of the shell allows previous commands to be repeated, possibly after modification to correct typing mistakes or to change the meaning of the command. The shell has a *history list* where these commands are kept, and a *history* variable which controls how large this list is (2.3).

home directory — Each user has a *home directory*, which is given in your entry in the password file, */etc/passwd*. This is the directory which you are placed in when you first login. The *cd* or *chdir* command with no arguments takes you back to this directory, whose name is recorded in the shell variable *home*. You can also access the *home directories* of other users in forming filenames using a *filename expansion* notation and the character '~' (1.6).

if — A conditional command within the shell, the *if* command is used in shell command scripts to make decisions about what course of action to take next (3.6).

ignoreeof — Normally, your shell will exit, printing 'logout' if you type a control-d at a prompt of '% '. This is the way you usually log off the system. You can *set* the *ignoreeof* variable if you wish in your *.login* file and then use the command *logout* to logout. This is useful if you sometimes accidentally type too many control-d characters, logging yourself off (2.2).

input — Many commands on UNIX take information from the terminal or from files which they then act on. This information is called *input*. Commands normally read for *input* from their *standard input* which is, by default, the terminal. This *standard input* can be redirected from a file using a shell metanotation with the character '<'. Many commands will also read from a file specified as argument. Commands placed in *pipelines* will read from the output of the previous command in the *pipeline*. The leftmost command in a *pipeline* reads from the terminal if you neither redirect its *input* nor give it a filename to use as *standard input*. Special mechanisms exist for supplying input to commands in shell scripts (1.5, 3.8).

interrupt — An *interrupt* is a signal to a program that is generated by hitting the RUBOUT or DELETE key (although users can and often do change the interrupt character, usually to ↑C). It causes most programs to stop execution. Certain programs, such as the shell and the editors, handle an *interrupt* in special ways, usually by stopping what they are doing and prompting for another command. While the shell is executing another command and waiting for it to finish, the shell does not listen to *interrupts.* The shell often wakes up when you hit *interrupt* because many commands die when they receive an *interrupt* (1.8, 3.9).

job — One or more commands typed on the same input line separated by '|' or ';' characters are run together and are called a *job*. Simple commands run by themselves without any '|' or ';' characters are the simplest *jobs. Jobs* are classified as *foreground*, *background*, or *suspended* (2.6).

job control — The builtin functions that control the execution of jobs are called *job control* commands. These are *bg, fg, stop, kill* (2.6).

job number — When each job is started it is assigned a small number called a *job number* which is printed next to the job in the output of the *jobs* command. This number, preceded by a '%' character, can be used as an argument to *job control* commands to indicate a specific job (2.6).

jobs — The *jobs* command prints a table showing jobs that are either running in the *background* or are *suspended* (2.6).

kill — A command which sends a signal to a job causing it to terminate (2.6).

.login — The file *.login* in your *home* directory is read by the shell each time you login to UNIX and the commands there are executed. There are a number of commands which are usefully placed here, especially *set* commands to the shell itself (2.1).

login shell — The shell that is started on your terminal when you login is called your *login shell*. It is different from other shells which you may run (e.g. on shell scripts) in that it reads the *.login* file before reading commands from the terminal and it reads the *.logout* file after you logout (2.1).

logout — The *logout* command causes a login shell to exit. Normally, a login shell will exit when you hit control-d generating an *end-of-file*, but if you have set *ignoreeof* in you *.login* file then this will not work and you must use *logout* to log off the UNIX system (2.8).

.logout — When you log off of UNIX the shell will execute commands from the file *.logout* in your *home* directory after it prints 'logout'.

lpr — The command *lpr* is the line printer daemon. The standard input of *lpr* spooled and printed on the UNIX line printer. You can also give *lpr* a list of filenames as arguments to be printed. It is most common to use *lpr* as the last component of a *pipeline* (2.3).

ls — The *ls* (list files) command is one of the most commonly used UNIX commands. With no argument filenames it prints the names of the files in the current directory. It has a number of useful *flag* arguments, and can also be given the names of directories as arguments, in which case it lists the names of the files in these directories (1.2).

mail — The *mail* program is used to send and receive messages from other UNIX users (1.1, 2.1).

make — The *make* command is used to maintain one or more related files and to organize functions to be performed on these files. In many ways *make* is easier to use, and more helpful than shell command scripts (3.2).

makefile — The file containing commands for *make* is called *makefile* (3.2).

manual — The *manual* often referred to is the 'UNIX programmer's manual'. It contains a number of sections and a description of each UNIX program. An online version of the *manual* is accessible through the *man* command. Its documentation can be obtained online via

```
man man
```

metacharacter — Many characters which are neither letters nor digits have special meaning either to the shell or to UNIX. These characters are called *metacharacters*. If it is necessary to place these characters in arguments to commands without them having their special meaning then they must be *quoted*. An example of a *metacharacter* is the character '>' which is used to indicate placement of output into a file. For the purposes of the *history* mechanism, most unquoted *metacharacters* form separate words (1.4). The appendix to this user's manual lists the *metacharacters* in groups by their function.

mkdir — The *mkdir* command is used to create a new directory.

modifier — Substitutions with the *history* mechanism, keyed by the character '!' or of variables using the metacharacter '$', are often subjected to modifications, indicated by placing the character ':' after the substitution and following this with the *modifier* itself. The *command substitution* mechanism can also be used to perform modification in a similar way, but this notation is less clear (3.6).

more — The program *more* writes a file on your terminal allowing you to control how much text is displayed at a time. *More* can move through the file screenful by screenful, line by line, search forward for a string, or start again at the beginning of the file. It is generally the easiest way of viewing a file (1.8).

noclobber The shell has a variable *noclobber* which may be set in the file *.login* to prevent accidental destruction of files by the '>' output redirection metasyntax of the shell (2.2, 2.5).

noglob The shell variable *noglob* is set to suppress the *filename expansion* of arguments containing the metacharacters '~', '*', '?', '[' and ']' (3.6).

notify The *notify* command tells the shell to report on the termination of a specific *background job* at the exact time it occurs as opposed to waiting until just before the next prompt to report the termination. The *notify* variable, if set, causes the shell to always report the termination of *background* jobs exactly when they occur (2.6).

onintr The *onintr* command is built into the shell and is used to control the action of a shell command script when an *interrupt* signal is received (3.9).

output Many commands in UNIX result in some lines of text which are called their *output*. This *output* is usually placed on what is known as the *standard output* which is normally connected to the user's terminal. The shell has a syntax using the metacharacter '>' for redirecting the *standard output* of a command to a file (1.3). Using the *pipe* mechanism and the metacharacter '|' it is also possible for the *standard output* of one command to become the *standard input* of another command (1.5). Certain commands such as the line printer daemon *p* do not place their results on the *standard output* but rather in more useful places such as on the line printer (2.3). Similarly the *write* command places its output on another user's terminal rather than its *standard output* (2.3). Commands also have a *diagnostic output* where they write their error messages. Normally these go to the terminal even if the *standard output* has been sent to a file or another command, but it is possible to direct error diagnostics along with *standard output* using a special metanotation (2.5).

pushd The *pushd* command, which means 'push directory', changes the shell's *working directory* and also remembers the current *working directory* before the change is made, allowing you to return to the same directory via the *popd* command later without retyping its name (2.7).

path The shell has a variable *path* which gives the names of the directories in which it searches for the commands which it is given. It always checks first to see if the command it is given is built into the shell. If it is, then it need not search for the command as it can do it internally. If the command is not builtin, then the shell searches for a file with the name given in each of the directories in the *path* variable, left to right. Since the normal definition of the *path* variable is

```
path (. /usr/ucb /bin /usr/bin)
```

the shell normally looks in the current directory, and then in the standard system directories '/usr/ucb', '/bin' and '/usr/bin' for the named command (2.2). If the command cannot be found the shell will print an error diagnostic. Scripts of shell commands will be executed using another shell to interpret them if they have 'execute' permission set. This is normally true because a command of the form

```
chmod 755 script
```

was executed to turn this execute permission on (3.3). If you add new commands to a directory in the *path*, you should issue the command *rehash* (2.2).

pathname A list of names, separated by '/' characters, forms a *pathname*. Each *component*, between successive '/' characters, names a directory in which the next *component* file resides. *Pathnames* which begin with the character '/' are interpreted relative to the *root* directory in the filesystem. Other *pathnames* are interpreted relative to the current directory as reported by *pwd*. The last component of a *pathname* may name a directory, but usually names a file.

pipeline — A group of commands which are connected together, the *standard output* of each connected to the *standard input* of the next, is called a *pipeline*. The *pipe* mechanism used to connect these commands is indicated by the shell metacharacter '|' (1.5, 2.3).

popd — The *popd* command changes the shell's *working directory* to the directory you most recently left using the *pushd* command. It returns to the directory without having to type its name, forgetting the name of the current *working directory* before doing so (2.7).

port — The part of a computer system to which each terminal is connected is called a *port*. Usually the system has a fixed number of *ports*, some of which are connected to telephone lines for dial-up access, and some of which are permanently wired directly to specific terminals.

pr — The *pr* command is used to prepare listings of the contents of files with headers giving the name of the file and the date and time at which the file was last modified (2.3).

printenv — The *printenv* command is used to print the current setting of variables in the environment (2.8).

process — An instance of a running program is called a *process* (2.6). UNIX assigns each *process* a unique number when it is started – called the *process number*. *Process numbers* can be used to stop individual *processes* using the *kill* or *stop* commands when the *processes* are part of a detached *background* job.

program — Usually synonymous with *command*; a binary file or shell command script which performs a useful function is often called a *program*.

programmer's manuals manual'u>(450u+1n) .br
Also referred to as the *manual*. See the glossary entry for 'manual'.

prompt — Many programs will print a *prompt* on the terminal when they expect input. Thus the editor 'ex(1)' will print a ':' when it expects input. The shell *prompts* for input with '% ' and occasionally with '? ' when reading commands from the terminal (1.1). The shell has a variable *prompt* which may be set to a different value to change the shell's main *prompt*. This is mostly used when debugging the shell (2.8).

ps — The *ps* command is used to show the processes you are currently running. Each process is shown with its unique process number, an indication of the terminal name it is attached to, an indication of the state of the process (whether it is running, stopped, awaiting some event (sleeping), and whether it is swapped out), and the amount of CPU time it has used so far. The command is identified by printing some of the words used when it was invoked (2.6). Shells, such as the *csh* you use to run the *ps* command, are not normally shown in the output.

pwd — The *pwd* command prints the full *pathname* of the current *working directory*. The *dirs* builtin command is usually a better and faster choice.

quit — The *quit* signal, generated by a control-\, is used to terminate programs which are behaving unreasonably. It normally produces a core image file (1.8).

quotation — The process by which metacharacters are prevented their special meaning, usually by using the character '´' in pairs, or by using the character '\', is referred to as *quotation* (1.7).

redirection — The routing of input or output from or to a file is known as *redirection* of input or output (1.3).

rehash — The *rehash* command tells the shell to rebuild its internal table of which commands are found in which directories in your *path*. This is necessary when a new program is installed in one of these directories (2.8).

relative pathname — A *pathname* which does not begin with a '/' is called a *relative pathname* since it is interpreted *relative* to the current *working directory*. The first *component* of such a *pathname* refers to some file or directory in the *working directory*, and subsequent *components* between '/' characters refer to directories below the *working directory*. *Pathnames* that are not *relative* are called *absolute pathnames* (1.6).

repeat — The *repeat* command iterates another command a specified number of times.

root — The directory that is at the top of the entire directory structure is called the *root* directory since it is the 'root' of the entire tree structure of directories. The name used in *pathnames* to indicate the *root* is '/'. *Pathnames* starting with '/' are said to be *absolute* since they start at the *root* directory. *Root* is also used as the part of a *pathname* that is left after removing the *extension*. See *filename* for a further explanation (1.6).

RUBOUT — The RUBOUT or DELETE key sends an interrupt to the current job. Most interactive commands return to their command level upon receipt of an interrupt, while non-interactive commands usually terminate, returning control to the shell. Users often change interrupt to be generated by ↑C rather than DELETE by using the *stty* command.

scratch file — Files whose names begin with a '#' are referred to as *scratch files*, since they are automatically removed by the system after a couple of days of non-use, or more frequently if disk space becomes tight (1.3).

script — Sequences of shell commands placed in a file are called shell command *scripts*. It is often possible to perform simple tasks using these *scripts* without writing a program in a language such as C, by using the shell to selectively run other programs (3.3, 3.10).

set — The builtin *set* command is used to assign new values to shell variables and to show the values of the current variables. Many shell variables have special meaning to the shell itself. Thus by using the *set* command the behavior of the shell can be affected (2.1).

setenv — Variables in the environment 'environ(5)' can be changed by using the *setenv* builtin command (2.8). The *printenv* command can be used to print the value of the variables in the environment.

shell — A *shell* is a command language interpreter. It is possible to write and run your own *shell*, as *shells* are no different than any other programs as far as the system is concerned. This manual deals with the details of one particular *shell*, called *csh*.

shell script — See *script* (3.3, 3.10).

signal — A *signal* in UNIX is a short message that is sent to a running program which causes something to happen to that process. *Signals* are sent either by typing special *control* characters on the keyboard or by using the *kill* or *stop* commands (1.8, 2.6).

sort — The *sort* program sorts a sequence of lines in ways that can be controlled by argument *flags* (1.5).

source — The *source* command causes the shell to read commands from a specified file. It is most useful for reading files such as *.cshrc* after changing them (2.8).

special character
: See ***metacharacters*** and the appendix to this manual.

standard
: We refer often to the ***standard input*** and ***standard output*** of commands. See ***input*** and ***output*** (1.3, 3.8).

status
: A command normally returns a ***status*** when it finishes. By convention a ***status*** of zero indicates that the command succeeded. Commands may return non-zero ***status*** to indicate that some abnormal event has occurred. The shell variable ***status*** is set to the ***status*** returned by the last command. It is most useful in shell commmand scripts (3.6).

stop
: The ***stop*** command causes a ***background*** job to become ***suspended*** (2.6).

string
: A sequential group of characters taken together is called a ***string***. ***Strings*** can contain any printable characters (2.2).

stty
: The ***stty*** program changes certain parameters inside UNIX which determine how your terminal is handled. See 'stty(1)' for a complete description (2.6).

substitution
: The shell implements a number of ***substitutions*** where sequences indicated by metacharacters are replaced by other sequences. Notable examples of this are history ***substitution*** keyed by the metacharacter '!' and variable ***substitution*** indicated by '$'. We also refer to ***substitutions*** as ***expansions*** (3.4).

suspended
: A job becomes ***suspended*** after a STOP signal is sent to it, either by typing a ***control***-z at the terminal (for ***foreground*** jobs) or by using the ***stop*** command (for ***background*** jobs). When ***suspended***, a job temporarily stops running until it is restarted by either the ***fg*** or ***bg*** command (2.6).

switch
: The ***switch*** command of the shell allows the shell to select one of a number of sequences of commands based on an argument string. It is similar to the ***switch*** statement in the language C (3.7).

termination
: When a command which is being executed finishes we say it undergoes ***termination*** or ***terminates***. Commands normally terminate when they read an ***end-of-file*** from their ***standard input***. It is also possible to terminate commands by sending them an ***interrupt*** or ***quit*** signal (1.8). The ***kill*** program terminates specified jobs (2.6).

then
: The ***then*** command is part of the shell's 'if-then-else-endif' control construct used in command scripts (3.6).

time
: The ***time*** command can be used to measure the amount of CPU and real time consumed by a specified command as well as the amount of disk i/o, memory utilized, and number of page faults and swaps taken by the command (2.1, 2.8).

tset
: The ***tset*** program is used to set standard erase and kill characters and to tell the system what kind of terminal you are using. It is often invoked in a ***.login*** file (2.1).

tty
: The word ***tty*** is a historical abbreviation for 'teletype' which is frequently used in UNIX to indicate the ***port*** to which a given terminal is connected. The ***tty*** command will print the name of the ***tty*** or ***port*** to which your terminal is presently connected.

unalias
: The ***unalias*** command removes aliases (2.8).

UNIX
: UNIX is an operating system on which ***csh*** runs. UNIX provides facilities which allow ***csh*** to invoke other programs such as editors and text formatters which you may wish to use.

unset
: The ***unset*** command removes the definitions of shell variables (2.2, 2.8).

variable expansion
: See ***variables*** and ***expansion*** (2.2, 3.4).

variables — *Variables* in *csh* hold one or more strings as value. The most common use of *variables* is in controlling the behavior of the shell. See *path*, *noclobber*, and *ignoreeof* for examples. *Variables* such as *argv* are also used in writing shell programs (shell command scripts) (2.2).

verbose — The *verbose* shell variable can be set to cause commands to be echoed after they are history expanded. This is often useful in debugging shell scripts. The *verbose* variable is set by the shell's *-v* command line option (3.10).

wc — The *wc* program calculates the number of characters, words, and lines in the files whose names are given as arguments (2.6).

while — The *while* builtin control construct is used in shell command scripts (3.7).

word — A sequence of characters which forms an argument to a command is called a *word*. Many characters which are neither letters, digits, '–', '.' nor '/' form *words* all by themselves even if they are not surrounded by blanks. Any sequence of characters may be made into a *word* by surrounding it with '" characters except for the characters '" and '!' which require special treatment (1.1). This process of placing special characters in *words* without their special meaning is called *quoting*.

working directory — At any given time you are in one particular directory, called your *working directory*. This directory's name is printed by the *pwd* command and the files listed by *ls* are the ones in this directory. You can change *working directories* using *chdir*.

write — The *write* command is used to communicate with other users who are logged in to UNIX.

An Introduction to Display Editing with Vi

William Joy

Revised for versions 3.5/2.13 by
Mark Horton

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, Ca. 94720

ABSTRACT

Vi (visual) is a display oriented interactive text editor. When using *vi* the screen of your terminal acts as a window into the file which you are editing. Changes which you make to the file are reflected in what you see.

Using *vi* you can insert new text any place in the file quite easily. Most of the commands to *vi* move the cursor around in the file. There are commands to move the cursor forward and backward in units of characters, words, sentences and paragraphs. A small set of operators, like **d** for delete and **c** for change, are combined with the motion commands to form operations such as delete word or change paragraph, in a simple and natural way. This regularity and the mnemonic assignment of commands to keys makes the editor command set easy to remember and to use.

Vi will work on a large number of display terminals, and new terminals are easily driven after editing a terminal description file. While it is advantageous to have an intelligent terminal which can locally insert and delete lines and characters from the display, the editor will function quite well on dumb terminals over slow phone lines. The editor makes allowance for the low bandwidth in these situations and uses smaller window sizes and different display updating algorithms to make best use of the limited speed available.

It is also possible to use the command set of *vi* on hardcopy terminals, storage tubes and "glass tty's" using a one line editing window; thus *vi's* command set is available on all terminals. The full command set of the more traditional, line oriented editor *ex* is available within *vi;* it is quite simple to switch between the two modes of editing.

September 15, 1986

An Introduction to Display Editing with Vi

William Joy

Revised for versions 3.5/2.13 by
Mark Horton

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, Ca. 94720

1. Getting started

This document provides a quick introduction to *vi*. (Pronounced *vee-eye*.) You should be running *vi* on a file you are familiar with while you are reading this. The first part of this document (sections 1 through 5) describes the basics of using *vi*. Some topics of special interest are presented in section 6, and some nitty-gritty details of how the editor functions are saved for section 7 to avoid cluttering the presentation here.

There is also a short appendix here, which gives for each character the special meanings which this character has in *vi*. Attached to this document should be a quick reference card. This card summarizes the commands of *vi* in a very compact format. You should have the card handy while you are learning *vi*.

1.1. Specifying terminal type

Before you can start *vi* you must tell the system what kind of terminal you are using. Here is a (necessarily incomplete) list of terminal type codes. If your terminal does not appear here, you should consult with one of the staff members on your system to find out the code for your terminal. If your terminal does not have a code, one can be assigned and a description for the terminal can be created.

Code	Full name	Type
2621	Hewlett-Packard 2621A/P	Intelligent
2645	Hewlett-Packard 264x	Intelligent
act4	Microterm ACT-IV	Dumb
act5	Microterm ACT-V	Dumb
adm3a	Lear Siegler ADM-3a	Dumb
adm31	Lear Siegler ADM-31	Intelligent
c100	Human Design Concept 100	Intelligent
dm1520	Datamedia 1520	Dumb
dm2500	Datamedia 2500	Intelligent
dm3025	Datamedia 3025	Intelligent
fox	Perkin-Elmer Fox	Dumb
h1500	Hazeltine 1500	Intelligent
h19	Heathkit h19	Intelligent
i100	Infoton 100	Intelligent
mime	Imitating a smart act4	Intelligent

The financial support of an IBM Graduate Fellowship and the National Science Foundation under grants MCS74-07644-A03 and MCS78-07291 is gratefully acknowledged.

t1061	Teleray 1061	Intelligent
vt52	Dec VT-52	Dumb

Suppose for example that you have a Hewlett-Packard HP2621A terminal. The code used by the system for this terminal is '2621'. In this case you can use one of the following commands to tell the system the type of your terminal:

% **setenv TERM** 2621

This command works with the shell *csh* on both version 6 and 7 systems. If you are using the standard version 7 shell then you should give the commands

$ TERM=2621
$ export TERM

If you want to arrange to have your terminal type set up automatically when you log in, you can use the *tset* program. If you dial in on a *mime*, but often use hardwired ports, a typical line for your *.login* file (if you use csh) would be

**setenv TERM `tset** - –d mime`

or for your *.profile* file (if you use sh)

**TERM=`tset** - –d mime`

Tset knows which terminals are hardwired to each port and needs only to be told that when you dial in you are probably on a *mime*. *Tset* is usually used to change the erase and kill characters, too.

1.2. Editing a file

After telling the system which kind of terminal you have, you should make a copy of a file you are familiar with, and run *vi* on this file, giving the command

% **vi** *name*

replacing *name* with the name of the copy file you just created. The screen should clear and the text of your file should appear on the screen. If something else happens refer to the footnote.‡

1.3. The editor's copy: the buffer

The editor does not directly modify the file which you are editing. Rather, the editor makes a copy of this file, in a place called the *buffer,* and remembers the file's name. You do not affect the contents of the file unless and until you write the changes you make back into the original file.

1.4. Notational conventions

In our examples, input which must be typed as is will be presented in **bold face**. Text which should be replaced with appropriate input will be given in *italics*. We will represent special characters in SMALL CAPITALS.

‡ If you gave the system an incorrect terminal type code then the editor may have just made a mess out of your screen. This happens when it sends control codes for one kind of terminal to some other kind of terminal. In this case hit the keys :q (colon and the q key) and then hit the RETURN key. This should get you back to the command level interpreter. Figure out what you did wrong (ask someone else if necessary) and try again.

Another thing which can go wrong is that you typed the wrong file name and the editor just printed an error diagnostic. In this case you should follow the above procedure for getting out of the editor, and try again this time spelling the file name correctly.

If the editor doesn't seem to respond to the commands which you type here, try sending an interrupt to it by hitting the DEL or RUB key on your terminal, and then hitting the :q command again followed by a carriage return.

1.5. Arrow keys

The editor command set is independent of the terminal you are using. On most terminals with cursor positioning keys, these keys will also work within the editor. If you don't have cursor positioning keys, or even if you do, you can use the **h j k** and **l** keys as cursor positioning keys (these are labelled with arrows on an *adm3a).**

(Particular note for the HP2621: on this terminal the function keys must be *shifted* (ick) to send to the machine, otherwise they only act locally. Unshifted use will leave the cursor positioned incorrectly.)

1.6. Special characters: ESC, CR and DEL

Several of these special characters are very important, so be sure to find them right now. Look on your keyboard for a key labelled ESC or ALT. It should be near the upper left corner of your terminal. Try hitting this key a few times. The editor will ring the bell to indicate that it is in a quiescent state.‡ Partially formed commands are cancelled by ESC, and when you insert text in the file you end the text insertion with ESC. This key is a fairly harmless one to hit, so you can just hit it if you don't know what is going on until the editor rings the bell.

The CR or RETURN key is important because it is used to terminate certain commands. It is usually at the right side of the keyboard, and is the same command used at the end of each shell command.

Another very useful key is the DEL or RUB key, which generates an interrupt, telling the editor to stop what it is doing. It is a forceful way of making the editor listen to you, or to return it to the quiescent state if you don't know or don't like what is going on. Try hitting the '/' key on your terminal. This key is used when you want to specify a string to be searched for. The cursor should now be positioned at the bottom line of the terminal after a '/' printed as a prompt. You can get the cursor back to the current position by hitting the DEL or RUB key; try this now.* From now on we will simply refer to hitting the DEL or RUB key as "sending an interrupt."**

The editor often echoes your commands on the last line of the terminal. If the cursor is on the first position of this last line, then the editor is performing a computation, such as computing a new position in the file after a search or running a command to reformat part of the buffer. When this is happening you can stop the editor by sending an interrupt.

1.7. Getting out of the editor

After you have worked with this introduction for a while, and you wish to do something else, you can give the command **ZZ** to the editor. This will write the contents of the editor's buffer back into the file you are editing, if you made any changes, and then quit from the editor. You can also end an editor session by giving the command **:q!**CR;† this is a dangerous but occasionally essential command which ends the editor session and discards all your changes. You need to know about this command in case you change the editor's copy of a file you wish only to look at. Be very careful not to give this command when you really want to save the changes you have made.

2. Moving around in the file

* As we will see later, *h* moves back to the left (like control-h which is a backspace), *j* moves down (in the same column), *k* moves up (in the same column), and *l* moves to the right.

‡ On smart terminals where it is possible, the editor will quietly flash the screen rather than ringing the bell.

* Backspacing over the '/' will also cancel the search.

** On some systems, this interruptibility comes at a price: you cannot type ahead when the editor is computing with the cursor on the bottom line.

† All commands which read from the last display line can also be terminated with a ESC as well as an CR.

2.1. Scrolling and paging

The editor has a number of commands for moving around in the file. The most useful of these is generated by hitting the control and D keys at the same time, a control-D or '^D'. We will use this two character notation for referring to these control keys from now on. You may have a key labelled '^' on your terminal. This key will be represented as '↑' in this document; '^' is exclusively used as part of the '^x' notation for control characters.‡

As you know now if you tried hitting **^D**, this command scrolls down in the file. The **D** thus stands for down. Many editor commands are mnemonic and this makes them much easier to remember. For instance the command to scroll up is **^U**. Many dumb terminals can't scroll up at all, in which case hitting **^U** clears the screen and refreshes it with a line which is farther back in the file at the top.

If you want to see more of the file below where you are, you can hit **^E** to expose one more line at the bottom of the screen, leaving the cursor where it is. ‡‡ The command **^Y** (which is hopelessly non-mnemonic, but next to **^U** on the keyboard) exposes one more line at the top of the screen.

There are other ways to move around in the file; the keys **^F** and **^B** ‡ move forward and backward a page, keeping a couple of lines of continuity between screens so that it is possible to read through a file using these rather than **^D** and **^U** if you wish.

Notice the difference between scrolling and paging. If you are trying to read the text in a file, hitting **^F** to move forward a page will leave you only a little context to look back at. Scrolling on the other hand leaves more context, and happens more smoothly. You can continue to read the text as scrolling is taking place.

2.2. Searching, goto, and previous context

Another way to position yourself in the file is by giving the editor a string to search for. Type the character / followed by a string of characters terminated by CR. The editor will position the cursor at the next occurrence of this string. Try hitting **n** to then go to the next occurrence of this string. The character **?** will search backwards from where you are, and is otherwise like /.†

If the search string you give the editor is not present in the file the editor will print a diagnostic on the last line of the screen, and the cursor will be returned to its initial position.

If you wish the search to match only at the beginning of a line, begin the search string with an **↑**. To match only at the end of a line, end the search string with a **$**. Thus **/↑search**CR will search for the word 'search' at the beginning of a line, and **/last$**CR searches for the word 'last' at the end of a line.*

The command **G**, when preceded by a number will position the cursor at that line in the file. Thus **1G** will move the cursor to the first line of the file. If you give **G** no count, then it moves to the end of the file.

If you are near the end of the file, and the last line is not at the bottom of the screen, the editor will place only the character '~' on each remaining line. This indicates that the last line in the file is on the screen; that is, the '~' lines are past the end of the file.

---

‡ If you don't have a '^' key on your terminal then there is probably a key labelled '↑'; in any case these characters are one and the same.

‡‡ Version 3 only.

‡ Not available in all v2 editors due to memory constraints.

† These searches will normally wrap around the end of the file, and thus find the string even if it is not on a line in the direction you search provided it is anywhere else in the file. You can disable this wraparound in scans by giving the command **:se nowrapscan**CR, or more briefly **:se nows**CR.

*Actually, the string you give to search for here can be a *regular expression* in the sense of the editors *ex*(1) and *ed*(1). If you don't wish to learn about this yet, you can disable this more general facility by doing **:se nomagic**CR; by putting this command in EXINIT in your environment, you can have this always be in effect (more about *EXINIT* later.)

You can find out the state of the file you are editing by typing a **^G**. The editor will show you the name of the file you are editing, the number of the current line, the number of lines in the buffer, and the percentage of the way through the buffer which you are. Try doing this now, and remember the number of the line you are on. Give a **G** command to get to the end and then another **G** command to get back where you were.

You can also get back to a previous position by using the command `` (two back quotes). This is often more convenient than **G** because it requires no advance preparation. Try giving a **G** or a search with **/** or **?** and then a `` to get back to where you were. If you accidentally hit **n** or any command which moves you far away from a context of interest, you can quickly get back by hitting ``.

2.3. Moving around on the screen

Now try just moving the cursor around on the screen. If your terminal has arrow keys (4 or 5 keys with arrows going in each direction) try them and convince yourself that they work. (On certain terminals using v2 editors, they won't.) If you don't have working arrow keys, you can always use **h**, **j**, **k**, and **l**. Experienced users of *vi* prefer these keys to arrow keys, because they are usually right underneath their fingers.

Hit the **+** key. Each time you do, notice that the cursor advances to the next line in the file, at the first non-white position on the line. The **–** key is like **+** but goes the other way.

These are very common keys for moving up and down lines in the file. Notice that if you go off the bottom or top with these keys then the screen will scroll down (and up if possible) to bring a line at a time into view. The RETURN key has the same effect as the **+** key.

Vi also has commands to take you to the top, middle and bottom of the screen. **H** will take you to the top (home) line on the screen. Try preceding it with a number as in **3H**. This will take you to the third line on the screen. Many *vi* commands take preceding numbers and do interesting things with them. Try **M**, which takes you to the middle line on the screen, and **L**, which takes you to the last line on the screen. **L** also takes counts, thus **5L** will take you to the fifth line from the bottom.

2.4. Moving within a line

Now try picking a word on some line on the screen, not the first word on the line. move the cursor using RETURN and **–** to be on the line where the word is. Try hitting the **w** key. This will advance the cursor to the next word on the line. Try hitting the **b** key to back up words in the line. Also try the **e** key which advances you to the end of the current word rather than to the beginning of the next word. Also try SPACE (the space bar) which moves right one character and the BS (backspace or **^H**) key which moves left one character. The key **h** works as **^H** does and is useful if you don't have a BS key. (Also, as noted just above, **l** will move to the right.)

If the line had punctuation in it you may have noticed that that the **w** and **b** keys stopped at each group of punctuation. You can also go back and forwards words without stopping at punctuation by using **W** and **B** rather than the lower case equivalents. Think of these as bigger words. Try these on a few lines with punctuation to see how they differ from the lower case **w** and **b**.

The word keys wrap around the end of line, rather than stopping at the end. Try moving to a word on a line below where you are by repeatedly hitting **w**.

2.5. Summary

SPACE	advance the cursor one position
^B	backwards to previous page
^D	scrolls down in the file
^E	exposes another line at the bottom (v3)
^F	forward to next page

^G	tell what is going on
^H	backspace the cursor
^N	next line, same column
^P	previous line, same column
^U	scrolls up in the file
^Y	exposes another line at the top (v3)
+	next line, at the beginning
–	previous line, at the beginning
/	scan for a following string forwards
?	scan backwards
B	back a word, ignoring punctuation
G	go to specified line, last default
H	home screen line
M	middle screen line
L	last screen line
W	forward a word, ignoring punctuation
b	back a word
e	end of current word
n	scan for next instance of **/** or **?** pattern
w	word after this word

2.6. View ‡

If you want to use the editor to look at a file, rather than to make changes, invoke it as *view* instead of *vi*. This will set the *readonly* option which will prevent you from accidently overwriting the file.

3. Making simple changes

3.1. Inserting

One of the most useful commands is the **i** (insert) command. After you type **i**, everything you type until you hit ESC is inserted into the file. Try this now; position yourself to some word in the file and try inserting text before this word. If you are on an dumb terminal it will seem, for a minute, that some of the characters in your line have been overwritten, but they will reappear when you hit ESC.

Now try finding a word which can, but does not, end in an 's'. Position yourself at this word and type **e** (move to end of word), then **a** for append and then 'sESC' to terminate the textual insert. This sequence of commands can be used to easily pluralize a word.

Try inserting and appending a few times to make sure you understand how this works; **i** placing text to the left of the cursor, **a** to the right.

It is often the case that you want to add new lines to the file you are editing, before or after some specific line in the file. Find a line where this makes sense and then give the command **o** to create a new line after the line you are on, or the command **O** to create a new line before the line you are on. After you create a new line in this way, text you type up to an ESC is inserted on the new line.

Many related editor commands are invoked by the same letter key and differ only in that one is given by a lower case key and the other is given by an upper case key. In these cases, the upper case key often differs from the lower case key in its sense of direction, with the upper case key working backward and/or up, while the lower case key moves forward and/or down.

‡ Not available in all v2 editors due to memory constraints.

Whenever you are typing in text, you can give many lines of input or just a few characters. To type in more than one line of text, hit a RETURN at the middle of your input. A new line will be created for text, and you can continue to type. If you are on a slow and dumb terminal the editor may choose to wait to redraw the tail of the screen, and will let you type over the existing screen lines. This avoids the lengthy delay which would occur if the editor attempted to keep the tail of the screen always up to date. The tail of the screen will be fixed up, and the missing lines will reappear, when you hit ESC.

While you are inserting new text, you can use the characters you normally use at the system command level (usually **^H** or #) to backspace over the last character which you typed, and the character which you use to kill input lines (usually @, **^X**, or **^U**) to erase the input you have typed on the current line.† The character **^W** will erase a whole word and leave you after the space after the previous word; it is useful for quickly backing up in an insert.

Notice that when you backspace during an insertion the characters you backspace over are not erased; the cursor moves backwards, and the characters remain on the display. This is often useful if you are planning to type in something similar. In any case the characters disappear when when you hit ESC; if you want to get rid of them immediately, hit an ESC and then **a** again.

Notice also that you can't erase characters which you didn't insert, and that you can't backspace around the end of a line. If you need to back up to the previous line to make a correction, just hit ESC and move the cursor back to the previous line. After making the correction you can return to where you were and use the insert or append command again.

3.2. Making small corrections

You can make small corrections in existing text quite easily. Find a single character which is wrong or just pick any character. Use the arrow keys to find the character, or get near the character with the word motion keys and then either backspace (hit the BS key or **^H** or even just **h**) or SPACE (using the space bar) until the cursor is on the character which is wrong. If the character is not needed then hit the **x** key; this deletes the character from the file. It is analogous to the way you **x** out characters when you make mistakes on a typewriter (except it's not as messy).

If the character is incorrect, you can replace it with the correct character by giving the command **r***c*, where *c* is replaced by the correct character. Finally if the character which is incorrect should be replaced by more than one character, give the command **s** which substitutes a string of characters, ending with ESC, for it. If there are a small number of characters which are wrong you can precede **s** with a count of the number of characters to be replaced. Counts are also useful with **x** to specify the number of characters to be deleted.

3.3. More corrections: operators

You already know almost enough to make changes at a higher level. All you need to know now is that the **d** key acts as a delete operator. Try the command **dw** to delete a word. Try hitting **.** a few times. Notice that this repeats the effect of the **dw**. The command **.** repeats the last command which made a change. You can remember it by analogy with an ellipsis '**...**'.

Now try **db**. This deletes a word backwards, namely the preceding word. Try **d**SPACE. This deletes a single character, and is equivalent to the **x** command.

Another very useful operator is **c** or change. The command **cw** thus changes the text of a single word. You follow it by the replacement text ending with an ESC. Find a word which you can change to another, and try this now. Notice that the end of the text to be changed was marked with the character '**$**' so that you can see this as you are typing in the new material.

† In fact, the character **^H** (backspace) always works to erase the last input character here, regardless of what your erase character is.

3.4. Operating on lines

It is often the case that you want to operate on lines. Find a line which you want to delete, and type **dd**, the **d** operator twice. This will delete the line. If you are on a dumb terminal, the editor may just erase the line on the screen, replacing it with a line with only an @ on it. This line does not correspond to any line in your file, but only acts as a place holder. It helps to avoid a lengthy redraw of the rest of the screen which would be necessary to close up the hole created by the deletion on a terminal without a delete line capability.

Try repeating the **c** operator twice; this will change a whole line, erasing its previous contents and replacing them with text you type up to an ESC.†

You can delete or change more than one line by preceding the **dd** or **cc** with a count, i.e. **5dd** deletes 5 lines. You can also give a command like **dL** to delete all the lines up to and including the last line on the screen, or **d3L** to delete through the third from the bottom line. Try some commands like this now.* Notice that the editor lets you know when you change a large number of lines so that you can see the extent of the change. The editor will also always tell you when a change you make affects text which you cannot see.

3.5. Undoing

Now suppose that the last change which you made was incorrect; you could use the insert, delete and append commands to put the correct material back. However, since it is often the case that we regret a change or make a change incorrectly, the editor provides a **u** (undo) command to reverse the last change which you made. Try this a few times, and give it twice in a row to notice that an **u** also undoes a **u.**

The undo command lets you reverse only a single change. After you make a number of changes to a line, you may decide that you would rather have the original state of the line back. The **U** command restores the current line to the state before you started changing it.

You can recover text which you delete, even if undo will not bring it back; see the section on recovering lost text below.

3.6. Summary

SPACE	advance the cursor one position
^H	backspace the cursor
^W	erase a word during an insert
erase	your erase (usually ^H or #), erases a character during an insert
kill	your kill (usually @, ^X, or ^U), kills the insert on this line
.	repeats the changing command
O	opens and inputs new lines, above the current
U	undoes the changes you made to the current line
a	appends text after the cursor
c	changes the object you specify to the following text
d	deletes the object you specify
i	inserts text before the cursor
o	opens and inputs new lines, below the current
u	undoes the last change

† The command S is a convenient synonym for for **cc**, by analogy with **s**. Think of S as a substitute on lines, while **s** is a substitute on characters.

* One subtle point here involves using the **/** search after a **d**. This will normally delete characters from the current position to the point of the match. If what is desired is to delete whole lines including the two points, give the pattern as **/pat/+0**, a line address.

4. Moving about; rearranging and duplicating text

4.1. Low level character motions

Now move the cursor to a line where there is a punctuation or a bracketing character such as a parenthesis or a comma or period. Try the command **f***x* where *x* is this character. This command finds the next *x* character to the right of the cursor in the current line. Try then hitting a **;**, which finds the next instance of the same character. By using the **f** command and then a sequence of **;**'s you can often get to a particular place in a line much faster than with a sequence of word motions or SPACEs. There is also a **F** command, which is like **f**, but searches backward. The **;** command repeats **F** also.

When you are operating on the text in a line it is often desirable to deal with the characters up to, but not including, the first instance of a character. Try **df***x* for some *x* now and notice that the *x* character is deleted. Undo this with **u** and then try **dt***x*; the **t** here stands for to, i.e. delete up to the next *x*, but not the *x*. The command **T** is the reverse of **t**.

When working with the text of a single line, an **↑** moves the cursor to the first non-white position on the line, and a **$** moves it to the end of the line. Thus **$a** will append new text at the end of the current line.

Your file may have tab (^I) characters in it. These characters are represented as a number of spaces expanding to a tab stop, where tab stops are every 8 positions.* When the cursor is at a tab, it sits on the last of the several spaces which represent that tab. Try moving the cursor back and forth over tabs so you understand how this works.

On rare occasions, your file may have nonprinting characters in it. These characters are displayed in the same way they are represented in this document, that is with a two character code, the first character of which is '^'. On the screen non-printing characters resemble a '^' character adjacent to another, but spacing or backspacing over the character will reveal that the two characters are, like the spaces representing a tab character, a single character.

The editor sometimes discards control characters, depending on the character and the setting of the *beautify* option, if you attempt to insert them in your file. You can get a control character in the file by beginning an insert and then typing a **^V** before the control character. The **^V** quotes the following character, causing it to be inserted directly into the file.

4.2. Higher level text objects

In working with a document it is often advantageous to work in terms of sentences, paragraphs, and sections. The operations **(** and **)** move to the beginning of the previous and next sentences respectively. Thus the command **d)** will delete the rest of the current sentence; likewise **d(** will delete the previous sentence if you are at the beginning of the current sentence, or the current sentence up to where you are if you are not at the beginning of the current sentence.

A sentence is defined to end at a '.', '!' or '?' which is followed by either the end of a line, or by two spaces. Any number of closing ')', ']', '"' and ''' characters may appear after the '.', '!' or '?' before the spaces or end of line.

The operations { and } move over paragraphs and the operations [[and]] move over sections.†

A paragraph begins after each empty line, and also at each of a set of paragraph macros, specified by the pairs of characters in the definition of the string valued option *paragraphs*. The default setting for this option defines the paragraph macros of the *-ms* and *-mm* macro packages,

---

* This is settable by a command of the form **:se ts=***x*CR, where *x* is 4 to set tabstops every four columns. This has effect on the screen representation within the editor.

† The [[and]] operations require the operation character to be doubled because they can move the cursor far from where it currently is. While it is easy to get back with the command **``**, these commands would still be frustrating if they were easy to hit accidentally.

i.e. the '.IP', '.LP', '.PP' and '.QP', '.P' and '.LI' macros.‡ Each paragraph boundary is also a sentence boundary. The sentence and paragraph commands can be given counts to operate over groups of sentences and paragraphs.

Sections in the editor begin after each macro in the *sections* option, normally '.NH', '.SH', '.H' and '.HU', and each line with a formfeed **^L** in the first column. Section boundaries are always line and paragraph boundaries also.

Try experimenting with the sentence and paragraph commands until you are sure how they work. If you have a large document, try looking through it using the section commands. The section commands interpret a preceding count as a different window size in which to redraw the screen at the new location, and this window size is the base size for newly drawn windows until another size is specified. This is very useful if you are on a slow terminal and are looking for a particular section. You can give the first section command a small count to then see each successive section heading in a small window.

4.3. Rearranging and duplicating text

The editor has a single unnamed buffer where the last deleted or changed away text is saved, and a set of named buffers **a**–**z** which you can use to save copies of text and to move text around in your file and between files.

The operator **y** yanks a copy of the object which follows into the unnamed buffer. If preceded by a buffer name, **"*x*y**, where *x* here is replaced by a letter **a**–**z**, it places the text in the named buffer. The text can then be put back in the file with the commands **p** and **P**; **p** puts the text after or below the cursor, while **P** puts the text before or above the cursor.

If the text which you yank forms a part of a line, or is an object such as a sentence which partially spans more than one line, then when you put the text back, it will be placed after the cursor (or before if you use **P**). If the yanked text forms whole lines, they will be put back as whole lines, without changing the current line. In this case, the put acts much like a **o** or **O** command.

Try the command **YP**. This makes a copy of the current line and leaves you on this copy, which is placed before the current line. The command **Y** is a convenient abbreviation for **yy**. The command **Yp** will also make a copy of the current line, and place it after the current line. You can give **Y** a count of lines to yank, and thus duplicate several lines; try **3YP**.

To move text within the buffer, you need to delete it in one place, and put it back in another. You can precede a delete operation by the name of a buffer in which the text is to be stored as in **"a5dd** deleting 5 lines into the named buffer *a*. You can then move the cursor to the eventual resting place of the these lines and do a **"ap** or **"aP** to put them back. In fact, you can switch and edit another file before you put the lines back, by giving a command of the form **:e** ***name***CR where *name* is the name of the other file you want to edit. You will have to write back the contents of the current editor buffer (or discard them) if you have made changes before the editor will let you switch to the other file. An ordinary delete command saves the text in the unnamed buffer, so that an ordinary put can move it elsewhere. However, the unnamed buffer is lost when you change files, so to move text from one file to another you should use an unnamed buffer.

4.4. Summary.

↑	first non-white on line
$	end of line
)	forward sentence
}	forward paragraph

‡ You can easily change or extend this set of macros by assigning a different string to the *paragraphs* option in your EXINIT. See section 6.2 for details. The '.bp' directive is also considered to start a paragraph.

]]	forward section
(	backward sentence
{	backward paragraph
[[	backward section
f*x*	find *x* forward in line
p	put text back, after cursor or below current line
y	yank operator, for copies and moves
t*x*	up to *x* forward, for operators
F*x*	f backward in line
P	put text back, before cursor or above current line
T*x*	t backward in line

5. High level commands

5.1. Writing, quitting, editing new files

So far we have seen how to enter *vi* and to write out our file using either **ZZ** or **:w**CR. The first exits from the editor, (writing if changes were made), the second writes and stays in the editor.

If you have changed the editor's copy of the file but do not wish to save your changes, either because you messed up the file or decided that the changes are not an improvement to the file, then you can give the command **:q!**CR to quit from the editor without writing the changes. You can also reedit the same file (starting over) by giving the command **:e!**CR. These commands should be used only rarely, and with caution, as it is not possible to recover the changes you have made after you discard them in this manner.

You can edit a different file without leaving the editor by giving the command **:e** *name*CR. If you have not written out your file before you try to do this, then the editor will tell you this, and delay editing the other file. You can then give the command **:w**CR to save your work and then the **:e** *name*CR command again, or carefully give the command **:e!** *name*CR, which edits the other file discarding the changes you have made to the current file. To have the editor automatically save changes, include ***set autowrite*** in your EXINIT, and use **:n** instead of **:e**.

5.2. Escaping to a shell

You can get to a shell to execute a single command by giving a *vi* command of the form **:!***cmd*CR. The system will run the single command *cmd* and when the command finishes, the editor will ask you to hit a RETURN to continue. When you have finished looking at the output on the screen, you should hit RETURN and the editor will clear the screen and redraw it. You can then continue editing. You can also give another **:** command when it asks you for a RETURN; in this case the screen will not be redrawn.

If you wish to execute more than one command in the shell, then you can give the command **:sh**CR. This will give you a new shell, and when you finish with the shell, ending it by typing a **^D**, the editor will clear the screen and continue.

On systems which support it, **^Z** will suspend the editor and return to the (top level) shell. When the editor is resumed, the screen will be redrawn.

5.3. Marking and returning

The command **``** returned to the previous place after a motion of the cursor by a command such as **/**, **?** or **G**. You can also mark lines in the file with single letter tags and return to these marks later by naming the tags. Try marking the current line with the command **m***x*, where you should pick some letter for *x*, say 'a'. Then move the cursor to a different line (any way you like) and hit **`a**. The cursor will return to the place which you marked. Marks last only until you edit another file.

When using operators such as **d** and referring to marked lines, it is often desirable to delete whole lines rather than deleting to the exact position in the line marked by **m**. In this case you can use the form *'x* rather than *`x*. Used without an operator, *'x* will move to the first non-white character of the marked line; similarly '' moves to the first non-white character of the line containing the previous context mark ``.

5.4. Adjusting the screen

If the screen image is messed up because of a transmission error to your terminal, or because some program other than the editor wrote output to your terminal, you can hit a **^L**, the ASCII form-feed character, to cause the screen to be refreshed.

On a dumb terminal, if there are @ lines in the middle of the screen as a result of line deletion, you may get rid of these lines by typing **^R** to cause the editor to retype the screen, closing up these holes.

Finally, if you wish to place a certain line on the screen at the top middle or bottom of the screen, you can position the cursor to that line, and then give a **z** command. You should follow the **z** command with a RETURN if you want the line to appear at the top of the window, a **.** if you want it at the center, or a **–** if you want it at the bottom. (**z.**, **z-**, and **z+** are not available on all v2 editors.)

6. Special topics

6.1. Editing on slow terminals

When you are on a slow terminal, it is important to limit the amount of output which is generated to your screen so that you will not suffer long delays, waiting for the screen to be refreshed. We have already pointed out how the editor optimizes the updating of the screen during insertions on dumb terminals to limit the delays, and how the editor erases lines to @ when they are deleted on dumb terminals.

The use of the slow terminal insertion mode is controlled by the ***slowopen*** option. You can force the editor to use this mode even on faster terminals by giving the command **:se slow**CR. If your system is sluggish this helps lessen the amount of output coming to your terminal. You can disable this option by **:se noslow**CR.

The editor can simulate an intelligent terminal on a dumb one. Try giving the command **:se redraw**CR. This simulation generates a great deal of output and is generally tolerable only on lightly loaded systems and fast terminals. You can disable this by giving the command **:se noredraw**CR.

The editor also makes editing more pleasant at low speed by starting editing in a small window, and letting the window expand as you edit. This works particularly well on intelligent terminals. The editor can expand the window easily when you insert in the middle of the screen on these terminals. If possible, try the editor on an intelligent terminal to see how this works.

You can control the size of the window which is redrawn each time the screen is cleared by giving window sizes as argument to the commands which cause large screen motions:

: / ? [[ ]] ` '

Thus if you are searching for a particular instance of a common string in a file you can precede the first search command by a small number, say 3, and the editor will draw three line windows around each instance of the string which it locates.

You can easily expand or contract the window, placing the current line as you choose, by giving a number on a **z** command, after the **z** and before the following RETURN, **.** or –. Thus the command **z5.** redraws the screen with the current line in the center of a five line window.†

† Note that the command **5z.** has an entirely different effect, placing line 5 in the center of a new window.

If the editor is redrawing or otherwise updating large portions of the display, you can interrupt this updating by hitting a DEL or RUB as usual. If you do this you may partially confuse the editor about what is displayed on the screen. You can still edit the text on the screen if you wish; clear up the confusion by hitting a **^L**; or move or search again, ignoring the current state of the display.

See section 7.8 on *open* mode for another way to use the *vi* command set on slow terminals.

6.2. Options, set, and editor startup files

The editor has a set of options, some of which have been mentioned above. The most useful options are given in the following table.

Name	Default	Description
autoindent	noai	Supply indentation automatically
autowrite	noaw	Automatic write before **:n**, **:ta**, **^↑**, **!**
ignorecase	noic	Ignore case in searching
lisp	nolisp	**(** { **)** } commands deal with S-expressions
list	nolist	Tabs print as ^I; end of lines marked with **$**
magic	nomagic	The characters . [and * are special in scans
number	nonu	Lines are displayed prefixed with line numbers
paragraphs	para=IPLPPPQPbpP LI	Macro names which start paragraphs
redraw	nore	Simulate a smart terminal on a dumb one
sections	sect=NHSHH HU	Macro names which start new sections
shiftwidth	sw=8	Shift distance for <, > and input **^D** and **^T**
showmatch	nosm	Show matching **(** or { as **)** or } is typed
slowopen	slow	Postpone display updates during inserts
term	dumb	The kind of terminal you are using.

The options are of three kinds: numeric options, string options, and toggle options. You can set numeric and string options by a statement of the form

set *opt*=*val*

and toggle options can be set or unset by statements of one of the forms

set *opt*
set no*opt*

These statements can be placed in your EXINIT in your environment, or given while you are running *vi* by preceding them with a **:** and following them with a CR.

You can get a list of all options which you have changed by the command **:set**CR, or the value of a single option by the command **:set** *opt***?**CR. A list of all possible options and their values is generated by **:set all**CR. Set can be abbreviated **se**. Multiple options can be placed on one line, e.g. **:se ai aw nu**CR.

Options set by the **set** command only last while you stay in the editor. It is common to want to have certain options set whenever you use the editor. This can be accomplished by creating a list of *ex* commands† which are to be run every time you start up *ex*, *edit*, or *vi*. A typical list includes a **set** command, and possibly a few **map** commands (on v3 editors). Since it is advisable to get these commands on one line, they can be separated with the | character, for example:

set ai aw terse|**map** @ dd|**map** # x

which sets the options *autoindent*, *autowrite*, *terse*, (the **set** command), makes @ delete a line, (the first **map**), and makes # delete a character, (the second **map**). (See section 6.9 for a description of the **map** command, which only works in version 3.) This string should be placed in the variable

† All commands which start with **:** are *ex* commands.

EXINIT in your environment. If you use *csh*, put this line in the file *.login* in your home directory:

setenv EXINIT 'set ai aw terse|map @ dd|map # x'

If you use the standard v7 shell, put these lines in the file *.profile* in your home directory:

EXINIT='set ai aw terse|map @ dd|map # x'
export EXINIT

On a version 6 system, the concept of environments is not present. In this case, put the line in the file *.exrc* in your home directory.

set ai aw terse|map @ dd|map # x

Of course, the particulars of the line would depend on which options you wanted to set.

6.3. Recovering lost lines

You might have a serious problem if you delete a number of lines and then regret that they were deleted. Despair not, the editor saves the last 9 deleted blocks of text in a set of numbered registers 1–9. You can get the *n*'th previous deleted text back in your file by the command "*n***p**. The " here says that a buffer name is to follow, *n* is the number of the buffer you wish to try (use the number 1 for now), and **p** is the put command, which puts text in the buffer after the cursor. If this doesn't bring back the text you wanted, hit **u** to undo this and then **.** (period) to repeat the put command. In general the **.** command will repeat the last change you made. As a special case, when the last command refers to a numbered text buffer, the **.** command increments the number of the buffer before repeating the command. Thus a sequence of the form

"1pu.u.u.

will, if repeated long enough, show you all the deleted text which has been saved for you. You can omit the **u** commands here to gather up all this text in the buffer, or stop after any **.** command to keep just the then recovered text. The command **P** can also be used rather than **p** to put the recovered text before rather than after the cursor.

6.4. Recovering lost files

If the system crashes, you can recover the work you were doing to within a few changes. You will normally receive mail when you next login giving you the name of the file which has been saved for you. You should then change to the directory where you were when the system crashed and give a command of the form:

% **vi –r** *name*

replacing *name* with the name of the file which you were editing. This will recover your work to a point near where you left off.†

You can get a listing of the files which are saved for you by giving the command:

% **vi –r**

If there is more than one instance of a particular file saved, the editor gives you the newest instance each time you recover it. You can thus get an older saved copy back by first recovering the newer copies.

For this feature to work, *vi* must be correctly installed by a super user on your system, and the *mail* program must exist to receive mail. The invocation "*vi -r*" will not always list all saved

† In rare cases, some of the lines of the file may be lost. The editor will give you the numbers of these lines and the text of the lines will be replaced by the string 'LOST'. These lines will almost always be among the last few which you changed. You can either choose to discard the changes which you made (if they are easy to remake) or to replace the few lost lines by hand.

files, but they can be recovered even if they are not listed.

6.5. Continuous text input

When you are typing in large amounts of text it is convenient to have lines broken near the right margin automatically. You can cause this to happen by giving the command **:se wm=10**CR. This causes all lines to be broken at a space at least 10 columns from the right hand edge of the screen.*

If the editor breaks an input line and you wish to put it back together you can tell it to join the lines with **J**. You can give **J** a count of the number of lines to be joined as in **3J** to join 3 lines. The editor supplies white space, if appropriate, at the juncture of the joined lines, and leaves the cursor at this white space. You can kill the white space with **x** if you don't want it.

6.6. Features for editing programs

The editor has a number of commands for editing programs. The thing that most distinguishes editing of programs from editing of text is the desirability of maintaining an indented structure to the body of the program. The editor has a *autoindent* facility for helping you generate correctly indented programs.

To enable this facility you can give the command **:se ai**CR. Now try opening a new line with **o** and type some characters on the line after a few tabs. If you now start another line, notice that the editor supplies white space at the beginning of the line to line it up with the previous line. You cannot backspace over this indentation, but you can use **^D** key to backtab over the supplied indentation.

Each time you type **^D** you back up one position, normally to an 8 column boundary. This amount is settable; the editor has an option called *shiftwidth* which you can set to change this value. Try giving the command **:se sw=4**CR and then experimenting with autoindent again.

For shifting lines in the program left and right, there are operators **<** and **>**. These shift the lines you specify right or left by one *shiftwidth*. Try **<<** and **>>** which shift one line left or right, and **<L** and **>L** shifting the rest of the display left and right.

If you have a complicated expression and wish to see how the parentheses match, put the cursor at a left or right parenthesis and hit **%**. This will show you the matching parenthesis. This works also for braces { and }, and brackets [and].

If you are editing C programs, you can use the [[and]] keys to advance or retreat to a line starting with a {, i.e. a function declaration at a time. When]] is used with an operator it stops after a line which starts with }; this is sometimes useful with **y**]].

6.7. Filtering portions of the buffer

You can run system commands over portions of the buffer using the operator **!**. You can use this to sort lines in the buffer, or to reformat portions of the buffer with a pretty-printer. Try typing in a list of random words, one per line and ending them with a blank line. Back up to the beginning of the list, and then give the command **!}sort**CR. This says to sort the next paragraph of material, and the blank line ends a paragraph.

6.8. Commands for editing LISP†

If you are editing a LISP program you should set the option *lisp* by doing **:se lisp**CR. This changes the **(** and **)** commands to move backward and forward over s-expressions. The **{** and **}** commands are like **(** and **)** but don't stop at atoms. These can be used to skip to the next list, or through a comment quickly.

---

* This feature is not available on some v2 editors. In v2 editors where it is available, the break can only occur to the right of the specified boundary instead of to the left.

† The LISP features are not available on some v2 editors due to memory constraints.

The *autoindent* option works differently for LISP, supplying indent to align at the first argument to the last open list. If there is no such argument then the indent is two spaces more than the last level.

There is another option which is useful for typing in LISP, the *showmatch* option. Try setting it with **:se sm**CR and then try typing a '(' some words and then a ')'. Notice that the cursor shows the position of the '(' which matches the ')' briefly. This happens only if the matching '(' is on the screen, and the cursor stays there for at most one second.

The editor also has an operator to realign existing lines as though they had been typed in with *lisp* and *autoindent* set. This is the **=** operator. Try the command **=%** at the beginning of a function. This will realign all the lines of the function declaration.

When you are editing LISP,, the [[and]] advance and retreat to lines beginning with a (, and are useful for dealing with entire function definitions.

6.9. Macros‡

Vi has a parameterless macro facility, which lets you set it up so that when you hit a single keystroke, the editor will act as though you had hit some longer sequence of keys. You can set this up if you find yourself typing the same sequence of commands repeatedly.

Briefly, there are two flavors of macros:

a) Ones where you put the macro body in a buffer register, say *x*. You can then type **@x** to invoke the macro. The @ may be followed by another @ to repeat the last macro.

b) You can use the *map* command from *vi* (typically in your *EXINIT*) with a command of the form:

:map *lhs rhs*CR

mapping *lhs* into *rhs*. There are restrictions: *lhs* should be one keystroke (either 1 character or one function key) since it must be entered within one second (unless *notimeout* is set, in which case you can type it as slowly as you wish, and *vi* will wait for you to finish it before it echoes anything). The *lhs* can be no longer than 10 characters, the *rhs* no longer than 100. To get a space, tab or newline into *lhs* or *rhs* you should escape them with a **^V**. (It may be necessary to double the **^V** if the map command is given inside *vi*, rather than in *ex.)* Spaces and tabs inside the *rhs* need not be escaped.

Thus to make the **q** key write and exit the editor, you can give the command

:map q :wq**^V^V**CR CR

which means that whenever you type **q**, it will be as though you had typed the four characters **:wq**CR. A **^V**'s is needed because without it the CR would end the **:** command, rather than becoming part of the *map* definition. There are two **^V**'s because from within *vi*, two **^V**'s must be typed to get one. The first CR is part of the *rhs*, the second terminates the : command.

Macros can be deleted with

unmap lhs

If the *lhs* of a macro is "#0" through "#9", this maps the particular function key instead of the 2 character "#" sequence. So that terminals without function keys can access such definitions, the form "#x" will mean function key *x* on all terminals (and need not be typed within one second.) The character "#" can be changed by using a macro in the usual way:

:map **^V^V^I #**

to use tab, for example. (This won't affect the *map* command, which still uses **#**, but just the invocation from visual mode.

‡ The macro feature is available only in version 3 editors.

The undo command reverses an entire macro call as a unit, if it made any changes.

Placing a '!' after the word **map** causes the mapping to apply to input mode, rather than command mode. Thus, to arrange for **^T** to be the same as 4 spaces in input mode, you can type:

:map **^T ^V**b̸b̸b̸b̸

where b̸ is a blank. The **^V** is necessary to prevent the blanks from being taken as white space between the *lhs* and *rhs*.

7. Word Abbreviations ‡‡

A feature similar to macros in input mode is word abbreviation. This allows you to type a short word and have it expanded into a longer word or words. The commands are **:abbreviate** and **:unabbreviate** (**:ab** and **:una**) and have the same syntax as **:map**. For example:

:ab eecs Electrical Engineering and Computer Sciences

causes the word 'eecs' to always be changed into the phrase 'Electrical Engineering and Computer Sciences'. Word abbreviation is different from macros in that only whole words are affected. If 'eecs' were typed as part of a larger word, it would be left alone. Also, the partial word is echoed as it is typed. There is no need for an abbreviation to be a single keystroke, as it should be with a macro.

7.1. Abbreviations

The editor has a number of short commands which abbreviate longer commands which we have introduced here. You can find these commands easily on the quick reference card. They often save a bit of typing and you can learn them as convenient.

8. Nitty-gritty details

8.1. Line representation in the display

The editor folds long logical lines onto many physical lines in the display. Commands which advance lines advance logical lines and will skip over all the segments of a line in one motion. The command | moves the cursor to a specific column, and may be useful for getting near the middle of a long line to split it in half. Try **80**| on a line which is more than 80 columns long.†

The editor only puts full lines on the display; if there is not enough room on the display to fit a logical line, the editor leaves the physical line empty, placing only an @ on the line as a place holder. When you delete lines on a dumb terminal, the editor will often just clear the lines to @ to save time (rather than rewriting the rest of the screen.) You can always maximize the information on the screen by giving the **^R** command.

If you wish, you can have the editor place line numbers before each line on the display. Give the command **:se nu**CR to enable this, and the command **:se nonu**CR to turn it off. You can have tabs represented as **^I** and the ends of lines indicated with '**$**' by giving the command **:se list**CR; **:se nolist**CR turns this off.

Finally, lines consisting of only the character '~' are displayed when the last line in the file is in the middle of the screen. These represent physical lines which are past the logical end of file.

8.2. Counts

Most *vi* commands will use a preceding count to affect their behavior in some way. The following table gives the common ways in which the counts are used:

---

‡‡ Version 3 only.

† You can make long lines very easily by using **J** to join together short lines.

new window size	**: / ? [[ ]] ` '**
scroll amount	**^D ^U**
line/column number	**z G |**
repeat effect	most of the rest

The editor maintains a notion of the current default window size. On terminals which run at speeds greater than 1200 baud the editor uses the full terminal screen. On terminals which are slower than 1200 baud (most dialup lines are in this group) the editor uses 8 lines as the default window size. At 1200 baud the default is 16 lines.

This size is the size used when the editor clears and refills the screen after a search or other motion moves far from the edge of the current window. The commands which take a new window size as count all often cause the screen to be redrawn. If you anticipate this, but do not need as large a window as you are currently using, you may wish to change the screen size by specifying the new size before these commands. In any case, the number of lines used on the screen will expand if you move off the top with a **-** or similar command or off the bottom with a command such as RETURN or **^D**. The window will revert to the last specified size the next time it is cleared and refilled.†

The scroll commands **^D** and **^U** likewise remember the amount of scroll last specified, using half the basic window size initially. The simple insert commands use a count to specify a repetition of the inserted text. Thus **10a+——**ESC will insert a grid-like string of text. A few commands also use a preceding count as a line or column number.

Except for a few commands which ignore any counts (such as **^R**), the rest of the editor commands use a count to indicate a simple repetition of their effect. Thus **5w** advances five words on the current line, while **5**RETURN advances five lines. A very useful instance of a count as a repetition is a count given to the **.** command, which repeats the last changing command. If you do **dw** and then **3.**, you will delete first one and then three words. You can then delete two more words with **2.**.

8.3. More file manipulation commands

The following table lists the file manipulation commands which you can use when you are in *vi*. All of these commands are followed by a CR or ESC. The most basic commands are **:w** and **:e**. A normal editing session on a single file will end with a **ZZ** command. If you are editing for a long period of time you can give **:w** commands occasionally after major amounts of editing, and then finish with a **ZZ**. When you edit more than one file, you can finish with one with a **:w** and start editing a new file by giving a **:e** command, or set *autowrite* and use **:n** <file>.

If you make changes to the editor's copy of a file, but do not wish to write them back, then you must give an **!** after the command you would otherwise use; this forces the editor to discard any changes you have made. Use this carefully.

The **:e** command can be given a **+** argument to start at the end of the file, or a **+*n*** argument to start at line *n*. In actuality, *n* may be any editor command not containing a space, usefully a scan like **+/*pat*** or **+?*pat***. In forming new names to the **e** command, you can use the character **%** which is replaced by the current file name, or the character **#** which is replaced by the alternate file name. The alternate file name is generally the last name you typed other than the current file. Thus if you try to do a **:e** and get a diagnostic that you haven't written the file, you can give a **:w** command and then a **:e #** command to redo the previous **:e**.

You can write part of the buffer to a file by finding out the lines that bound the range to be written using **^G**, and giving these numbers after the **:** and before the **w**, separated by **,**'s. You can also mark these lines with **m** and then use an address of the form **'*x*,'*y*** on the **w** command

† But not by a **^L** which just redraws the screen as it is.

:w	write back changes
:wq	write and quit
:x	write (if necessary) and quit (same as ZZ).
:e *name*	edit file *name*
:e!	reedit, discarding changes
:e + *name*	edit, starting at end
:e +*n*	edit, starting at line *n*
:e #	edit alternate file
:w *name*	write file *name*
:w! *name*	overwrite file *name*
:*x,y***w** *name*	write lines *x* through *y* to *name*
:r *name*	read file *name* into buffer
:r !*cmd*	read output of *cmd* into buffer
:n	edit next file in argument list
:n!	edit next file, discarding changes to current
:n *args*	specify new argument list
:ta *tag*	edit file containing tag *tag*, at *tag*

here.

You can read another file into the buffer after the current line by using the **:r** command. You can similarly read in the output from a command, just use **!***cmd* instead of a file name.

If you wish to edit a set of files in succession, you can give all the names on the command line, and then edit each one in turn using the command **:n**. It is also possible to respecify the list of files to be edited by giving the **:n** command a list of file names, or a pattern to be expanded as you would have given it on the initial *vi* command.

If you are editing large programs, you will find the **:ta** command very useful. It utilizes a data base of function names and their locations, which can be created by programs such as *ctags*, to quickly find a function whose name you give. If the **:ta** command will require the editor to switch files, then you must **:w** or abandon any changes before switching. You can repeat the **:ta** command without any arguments to look for the same tag again. (The tag feature is not available in some v2 editors.)

8.4. More about searching for strings

When you are searching for strings in the file with / and **?**, the editor normally places you at the next or previous occurrence of the string. If you are using an operator such as **d**, **c** or **y**, then you may well wish to affect lines up to the line before the line containing the pattern. You can give a search of the form */pat/–n* to refer to the *n*'th line before the next line containing *pat*, or you can use + instead of – to refer to the lines after the one containing *pat*. If you don't give a line offset, then the editor will affect characters up to the match place, rather than whole lines; thus use "+0" to affect to the line which matches.

You can have the editor ignore the case of words in the searches it does by giving the command **:se ic**CR. The command **:se noic**CR turns this off.

Strings given to searches may actually be regular expressions. If you do not want or need this facility, you should

    set nomagic

in your EXINIT. In this case, only the characters **↑** and **$** are special in patterns. The character \ is also then special (as it is most everywhere in the system), and may be used to get at the an extended pattern matching facility. It is also necessary to use a \ before a / in a forward scan or a **?** in a backward scan, in any case. The following table gives the extended forms when **magic** is set.

↑	at beginning of pattern, matches beginning of line
$	at end of pattern, matches end of line
.	matches any character
\\<	matches the beginning of a word
\\>	matches the end of a word
[*str*]	matches any single character in *str*
[↑*str*]	matches any single character not in *str*
[*x–y*]	matches any character between *x* and *y*
*	matches any number of the preceding pattern

If you use **nomagic** mode, then the **.** [and * primitives are given with a preceding \.

8.5. More about input mode

There are a number of characters which you can use to make corrections during input mode. These are summarized in the following table.

^H	deletes the last input character
^W	deletes the last input word, defined as by **b**
erase	your erase character, same as **^H**
kill	your kill character, deletes the input on this line
\	escapes a following **^H** and your erase and kill
ESC	ends an insertion
DEL	interrupts an insertion, terminating it abnormally
CR	starts a new line
^D	backtabs over *autoindent*
0^D	kills all the *autoindent*
↑^D	same as **0^D**, but restores indent next line
^V	quotes the next non-printing character into the file

The most usual way of making corrections to input is by typing **^H** to correct a single character, or by typing one or more **^W**'s to back over incorrect words. If you use # as your erase character in the normal system, it will work like **^H**.

Your system kill character, normally @, **^X** or **^U**, will erase all the input you have given on the current line. In general, you can neither erase input back around a line boundary nor can you erase characters which you did not insert with this insertion command. To make corrections on the previous line after a new line has been started you can hit ESC to end the insertion, move over and make the correction, and then return to where you were to continue. The command **A** which appends at the end of the current line is often useful for continuing.

If you wish to type in your erase or kill character (say # or @) then you must precede it with a \, just as you would do at the normal system command level. A more general way of typing non-printing characters into the file is to precede them with a **^V**. The **^V** echoes as a ↑ character on which the cursor rests. This indicates that the editor expects you to type a control character. In fact you may type any character and it will be inserted into the file at that point.*

If you are using *autoindent* you can backtab over the indent which it supplies by typing a **^D**. This backs up to a *shiftwidth* boundary. This only works immediately after the supplied

* This is not quite true. The implementation of the editor does not allow the NULL (^@) character to appear in files. Also the LF (linefeed or ^J) character is used by the editor to separate lines in the file, so it cannot appear in the middle of a line. You can insert any other character, however, if you wait for the editor to echo the ↑ before you type the character. In fact, the editor will treat a following letter as a request for the corresponding control character. This is the only way to type ^S or ^Q, since the system normally uses them to suspend and resume output and never gives them to the editor to process.

autoindent.

When you are using *autoindent* you may wish to place a label at the left margin of a line. The way to do this easily is to type ↑ and then **^D**. The editor will move the cursor to the left margin for one line, and restore the previous indent on the next. You can also type a **0** followed immediately by a **^D** if you wish to kill all the indent and not have it come back on the next line.

8.6. Upper case only terminals

If your terminal has only upper case, you can still use *vi* by using the normal system convention for typing on such a terminal. Characters which you normally type are converted to lower case, and you can type upper case letters by preceding them with a \. The characters { ˜ } | ` are not available on such terminals, but you can escape them as \(\↑ \) \! \´. These characters are represented on the display in the same way they are typed.‡ ‡

8.7. Vi and ex

Vi is actually one mode of editing within the editor *ex*. When you are running *vi* you can escape to the line oriented editor of *ex* by giving the command **Q**. All of the **:** commands which were introduced above are available in *ex*. Likewise, most *ex* commands can be invoked from *vi* using :. Just give them without the **:** and follow them with a CR.

In rare instances, an internal error may occur in *vi*. In this case you will get a diagnostic and be left in the command mode of *ex*. You can then save your work and quit if you wish by giving a command **x** after the **:** which *ex* prompts you with, or you can reenter *vi* by giving *ex* a *vi* command.

There are a number of things which you can do more easily in *ex* than in *vi*. Systematic changes in line oriented material are particularly easy. You can read the advanced editing documents for the editor *ed* to find out a lot more about this style of editing. Experienced users often mix their use of *ex* command mode and *vi* command mode to speed the work they are doing.

8.8. Open mode: vi on hardcopy terminals and "glass tty's" ‡

If you are on a hardcopy terminal or a terminal which does not have a cursor which can move off the bottom line, you can still use the command set of *vi*, but in a different mode. When you give a *vi* command, the editor will tell you that it is using *open* mode. This name comes from the *open* command in *ex*, which is used to get into the same mode.

The only difference between *visual* mode and *open* mode is the way in which the text is displayed.

In *open* mode the editor uses a single line window into the file, and moving backward and forward in the file causes new lines to be displayed, always below the current line. Two commands of *vi* work differently in *open:* **z** and **^R**. The **z** command does not take parameters, but rather draws a window of context around the current line and then returns you to the current line.

If you are on a hardcopy terminal, the **^R** command will retype the current line. On such terminals, the editor normally uses two lines to represent the current line. The first line is a copy of the line as you started to edit it, and you work on the line below this line. When you delete characters, the editor types a number of \'s to show you the characters which are deleted. The editor also reprints the current line soon after such changes so that you can see what the line looks like again.

It is sometimes useful to use this mode on very slow terminals which can support *vi* in the full screen mode. You can do this by entering *ex* and using an *open* command.

---

‡ The \ character you give will not echo until you type another key.

‡ Not available in all v2 editors due to memory constraints.

Acknowledgements

Bruce Englar encouraged the early development of this display editor. Peter Kessler helped bring sanity to version 2's command layout. Bill Joy wrote versions 1 and 2.0 through 2.7, and created the framework that users see in the present editor. Mark Horton added macros and other features and made the editor work on a large number of terminals and Unix systems.

Appendix: character functions

This appendix gives the uses the editor makes of each character. The characters are presented in their order in the ASCII character set: Control characters come first, then most special characters, then the digits, upper and then lower case characters.

For each character we tell a meaning it has as a command and any meaning it has during an insert. If it has only meaning as a command, then only this is discussed. Section numbers in parentheses indicate where the character is discussed; a 'f' after the section number means that the character is mentioned in a footnote.

^@	Not a command character. If typed as the first character of an insertion it is replaced with the last text inserted, and the insert terminates. Only 128 characters are saved from the last insert; if more characters were inserted the mechanism is not available. A **^@** cannot be part of the file due to the editor implementation (7.5f).
^A	Unused.
^B	Backward window. A count specifies repetition. Two lines of continuity are kept if possible (2.1, 6.1, 7.2).
^C	Unused.
^D	As a command, scrolls down a half-window of text. A count gives the number of (logical) lines to scroll, and is remembered for future **^D** and **^U** commands (2.1, 7.2). During an insert, backtabs over *autoindent* white space at the beginning of a line (6.6, 7.5); this white space cannot be backspaced over.
^E	Exposes one more line below the current screen in the file, leaving the cursor where it is if possible. (Version 3 only.)
^F	Forward window. A count specifies repetition. Two lines of continuity are kept if possible (2.1, 6.1, 7.2).
^G	Equivalent to **:f**CR, printing the current file, whether it has been modified, the current line number and the number of lines in the file, and the percentage of the way through the file that you are.
^H (BS)	Same as **left arrow**. (See **h**). During an insert, eliminates the last input character, backing over it but not erasing it; it remains so you can see what you typed if you wish to type something only slightly different (3.1, 7.5).
^I (TAB)	Not a command character. When inserted it prints as some number of spaces. When the cursor is at a tab character it rests at the last of the spaces which represent the tab. The spacing of tabstops is controlled by the *tabstop* option (4.1, 6.6).
^J (LF)	Same as **down arrow** (see **j**).
^K	Unused.
^L	The ASCII formfeed character, this causes the screen to be cleared and redrawn. This is useful after a transmission error, if characters typed by a program other than the editor scramble the screen, or after output is stopped by an interrupt (5.4, 7.2f).
^M (CR)	A carriage return advances to the next line, at the first non-white position in the line. Given a count, it advances that many lines (2.3). During an insert, a CR causes the insert to continue onto another line (3.1).
^N	Same as **down arrow** (see **j**).
^O	Unused.
^P	Same as **up arrow** (see **k**).
^Q	Not a command character. In input mode, **^Q** quotes the next character, the same as **^V**, except that some teletype drivers will eat the **^Q** so that the editor

never sees it.

^R Redraws the current screen, eliminating logical lines not corresponding to physical lines (lines with only a single @ character on them). On hardcopy terminals in *open* mode, retypes the current line (5.4, 7.2, 7.8).

^S Unused. Some teletype drivers use **^S** to suspend output until **^Q**is

^T Not a command character. During an insert, with *autoindent* set and at the beginning of the line, inserts *shiftwidth* white space.

^U Scrolls the screen up, inverting **^D** which scrolls down. Counts work as they do for **^D**, and the previous scroll amount is common to both. On a dumb terminal, **^U** will often necessitate clearing and redrawing the screen further back in the file (2.1, 7.2).

^V Not a command character. In input mode, quotes the next character so that it is possible to insert non-printing and special characters into the file (4.2, 7.5).

^W Not a command character. During an insert, backs up as **b** would in command mode; the deleted characters remain on the display (see **^H**) (7.5).

^X Unused.

^Y Exposes one more line above the current screen, leaving the cursor where it is if possible. (No mnemonic value for this key; however, it is next to **^U** which scrolls up a bunch.) (Version 3 only.)

^Z If supported by the Unix system, stops the editor, exiting to the top level shell. Same as **:stop**CR. Otherwise, unused.

^[(ESC) Cancels a partially formed command, such as a **z** when no following character has yet been given; terminates inputs on the last line (read by commands such as **:** / and **?**); ends insertions of new text into the buffer. If an ESC is given when quiescent in command state, the editor rings the bell or flashes the screen. You can thus hit ESC if you don't know what is happening till the editor rings the bell. If you don't know if you are in insert mode you can type ESC**a**, and then material to be input; the material will be inserted correctly whether or not you were in insert mode when you started (1.5, 3.1, 7.5).

^\ Unused.

^] Searches for the word which is after the cursor as a tag. Equivalent to typing **:ta**, this word, and then a CR. Mnemonically, this command is "go right to" (7.3).

^↑ Equivalent to **:e** #CR, returning to the previous position in the last edited file, or editing a file which you specified if you got a 'No write since last change diagnostic' and do not want to have to type the file name again (7.3). (You have to do a **:w** before ^↑ will work in this case. If you do not wish to write the file you should do **:e!** #CR instead.)

^_ Unused. Reserved as the command character for the Tektronix 4025 and 4027 terminal.

SPACE Same as **right arrow** (see **l**).

! An operator, which processes lines from the buffer with reformatting commands. Follow **!** with the object to be processed, and then the command name terminated by CR. Doubling **!** and preceding it by a count causes count lines to be filtered; otherwise the count is passed on to the object after the **!**. Thus **2!**}*fmt*CR reformats the next two paragraphs by running them through the program *fmt*. If you are working on LISP, the command **!%***grind*CR,* given at the beginning of a

*Both *fmt* and *grind* are Berkeley programs and may not be present at all installations.

function, will run the text of the function through the LISP grinder (6.7, 7.3). To read a file or the output of a command into the buffer use **:r** (7.3). To simply execute a command use **:!** (7.3).

" Precedes a named buffer specification. There are named buffers **1–9** used for saving deleted text and named buffers **a–z** into which you can place text (4.3, 6.3)

The macro character which, when followed by a number, will substitute for a function key on terminals without function keys (6.9). In input mode, if this is your erase character, it will delete the last character you typed in input mode, and must be preceded with a **** to insert it, since it normally backs over the last input character you gave.

$** Moves to the end of the current line. If you **:se list**CR, then the end of each line will be shown by printing a **$ after the end of the displayed text in the line. Given a count, advances to the count'th following end of line; thus **2$** advances to the end of the following line.

% Moves to the parenthesis or brace { } which balances the parenthesis or brace at the current cursor position.

& A synonym for **:&**CR, by analogy with the *ex* **&** command.

´ When followed by a ´ returns to the previous context at the beginning of a line. The previous context is set whenever the current line is moved in a non-relative way. When followed by a letter **a–z**, returns to the line which was marked with this letter with a **m** command, at the first non-white character in the line. (2.2, 5.3). When used with an operator such as **d**, the operation takes place over complete lines; if you use `, the operation takes place from the exact marked place to the current cursor position within the line.

(Retreats to the beginning of a sentence, or to the beginning of a LISP s-expression if the *lisp* option is set. A sentence ends at a **. !** or **?** which is followed by either the end of a line or by two spaces. Any number of closing **)] "** and **´** characters may appear after the **. !** or **?**, and before the spaces or end of line. Sentences also begin at paragraph and section boundaries (see **{** and **[[** below). A count advances that many sentences (4.2, 6.8).

) Advances to the beginning of a sentence. A count repeats the effect. See **(** above for the definition of a sentence (4.2, 6.8).

***** Unused.

+ Same as CR when used as a command.

, Reverse of the last **f F t** or **T** command, looking the other way in the current line. Especially useful after hitting too many **;** characters. A count repeats the search.

– Retreats to the previous line at the first non-white character. This is the inverse of **+** and RETURN. If the line moved to is not on the screen, the screen is scrolled, or cleared and redrawn if this is not possible. If a large amount of scrolling would be required the screen is also cleared and redrawn, with the current line at the center (2.3).

. Repeats the last command which changed the buffer. Especially useful when deleting words or lines; you can delete some words/lines and then hit **.** to delete more and more words/lines. Given a count, it passes it on to the command being repeated. Thus after a **2dw**, **3.** deletes three words (3.3, 6.3, 7.2, 7.4).

/ Reads a string from the last line on the screen, and scans forward for the next occurrence of this string. The normal input editing sequences may be used during the input on the bottom line; an returns to command state without ever searching. The search begins when you hit CR to terminate the pattern; the cursor moves to the beginning of the last line to indicate that the search is in progress; the search may then be terminated with a DEL or RUB, or by backspacing when at the beginning of the bottom line, returning the cursor to its initial position. Searches normally wrap end-around to find a string anywhere in the buffer.

When used with an operator the enclosed region is normally affected. By mentioning an offset from the line matched by the pattern you can force whole lines to be affected. To do this give a pattern with a closing a closing / and then an offset $+n$ or $-n$.

To include the character / in the search string, you must escape it with a preceding \. A ↑ at the beginning of the pattern forces the match to occur at the beginning of a line only; this speeds the search. A **$** at the end of the pattern forces the match to occur at the end of a line only. More extended pattern matching is available, see section 7.4; unless you set **nomagic** in your *.exrc* file you will have to preceed the characters **.** **[** ***** and **~** in the search pattern with a \ to get them to work as you would naively expect (1.5, 2,2, 6.1, 7.2, 7.4).

0 Moves to the first character on the current line. Also used, in forming numbers, after an initial **1–9**.

1–9 Used to form numeric arguments to commands (2.3, 7.2).

: A prefix to a set of commands for file and option manipulation and escapes to the system. Input is given on the bottom line and terminated with an CR, and the command then executed. You can return to where you were by hitting DEL or RUB if you hit **:** accidentally (see primarily 6.2 and 7.3).

; Repeats the last single character find which used **f F t** or **T**. A count iterates the basic scan (4.1).

< An operator which shifts lines left one *shiftwidth*, normally 8 spaces. Like all operators, affects lines when repeated, as in **<<**. Counts are passed through to the basic object, thus **3<<** shifts three lines (6.6, 7.2).

= Reindents line for LISP, as though they were typed in with *lisp* and *autoindent* set (6.8).

> An operator which shifts lines right one *shiftwidth*, normally 8 spaces. Affects lines when repeated as in **>>**. Counts repeat the basic object (6.6, 7.2).

? Scans backwards, the opposite of /. See the / description above for details on scanning (2.2, 6.1, 7.4).

@ A macro character (6.9). If this is your kill character, you must escape it with a \ to type it in during input mode, as it normally backs over the input you have given on the current line (3.1, 3.4, 7.5).

A Appends at the end of line, a synonym for **$a** (7.2).

B Backs up a word, where words are composed of non-blank sequences, placing the cursor at the beginning of the word. A count repeats the effect (2.4).

C Changes the rest of the text on the current line; a synonym for **c$**.

D Deletes the rest of the text on the current line; a synonym for **d$**.

E Moves forward to the end of a word, defined as blanks and non-blanks, like **B** and **W**. A count repeats the effect.

F Finds a single following character, backwards in the current line. A count repeats this search that many times (4.1).

G Goes to the line number given as preceding argument, or the end of the file if no preceding count is given. The screen is redrawn with the new current line in the center if necessary (7.2).

H — **Home arrow**. Homes the cursor to the top line on the screen. If a count is given, then the cursor is moved to the count'th line on the screen. In any case the cursor is moved to the first non-white character on the line. If used as the target of an operator, full lines are affected (2.3, 3.2).

I — Inserts at the beginning of a line; a synonym for **↑i**.

J — Joins together lines, supplying appropriate white space: one space between words, two spaces after a **.**, and no spaces at all if the first character of the joined on line is **)**. A count causes that many lines to be joined rather than the default two (6.5, 7.1f).

K — Unused.

L — Moves the cursor to the first non-white character of the last line on the screen. With a count, to the first non-white of the count'th line from the bottom. Operators affect whole lines when used with **L** (2.3).

M — Moves the cursor to the middle line on the screen, at the first non-white position on the line (2.3).

N — Scans for the next match of the last pattern given to **/** or **?**, but in the reverse direction; this is the reverse of **n**.

O — Opens a new line above the current line and inputs text there up to an ESC. A count can be used on dumb terminals to specify a number of lines to be opened; this is generally obsolete, as the *slowopen* option works better (3.1).

P — Puts the last deleted text back before/above the cursor. The text goes back as whole lines above the cursor if it was deleted as whole lines. Otherwise the text is inserted between the characters before and at the cursor. May be preceded by a named buffer specification "*x* to retrieve the contents of the buffer; buffers **1–9** contain deleted material, buffers **a–z** are available for general use (6.3).

Q — Quits from *vi* to *ex* command mode. In this mode, whole lines form commands, ending with a RETURN. You can give all the **:** commands; the editor supplies the **:** as a prompt (7.7).

R — Replaces characters on the screen with characters you type (overlay fashion). Terminates with an ESC.

S — Changes whole lines, a synonym for **cc**. A count substitutes for that many lines. The lines are saved in the numeric buffers, and erased on the screen before the substitution begins.

T — Takes a single following character, locates the character before the cursor in the current line, and places the cursor just after that character. A count repeats the effect. Most useful with operators such as **d** (4.1).

U — Restores the current line to its state before you started changing it (3.5).

V — Unused.

W — Moves forward to the beginning of a word in the current line, where words are defined as sequences of blank/non-blank characters. A count repeats the effect (2.4).

X — Deletes the character before the cursor. A count repeats the effect, but only characters on the current line are deleted.

Y — Yanks a copy of the current line into the unnamed buffer, to be put back by a later **p** or **P**; a very useful synonym for **yy**. A count yanks that many lines. May be preceded by a buffer name to put lines in that buffer (7.4).

ZZ — Exits the editor. (Same as **:x**CR.) If any changes have been made, the buffer is written out to the current file. Then the editor quits.

[[Backs up to the previous section boundary. A section begins at each macro in the *sections* option, normally a '.NH' or '.SH' and also at lines which which start with a formfeed **^L**. Lines beginning with **{** also stop **[[**; this makes it useful for looking backwards, a function at a time, in C programs. If the option *lisp* is set, stops at each **(** at the beginning of a line, and is thus useful for moving backwards at the top level LISP objects. (4.2, 6.1, 6.6, 7.2).

\ Unused.

]] Forward to a section boundary, see **[[** for a definition (4.2, 6.1, 6.6, 7.2).

↑ Moves to the first non-white position on the current line (4.4).

_ Unused.

\` When followed by a **\`** returns to the previous context. The previous context is set whenever the current line is moved in a non-relative way. When followed by a letter **a**–**z**, returns to the position which was marked with this letter with a **m** command. When used with an operator such as **d**, the operation takes place from the exact marked place to the current position within the line; if you use **´**, the operation takes place over complete lines (2.2, 5.3).

a Appends arbitrary text after the current cursor position; the insert can continue onto multiple lines by using RETURN within the insert. A count causes the inserted text to be replicated, but only if the inserted text is all on one line. The insertion terminates with an ESC (3.1, 7.2).

b Backs up to the beginning of a word in the current line. A word is a sequence of alphanumerics, or a sequence of special characters. A count repeats the effect (2.4).

c An operator which changes the following object, replacing it with the following input text up to an ESC. If more than part of a single line is affected, the text which is changed away is saved in the numeric named buffers. If only part of the current line is affected, then the last character to be changed away is marked with a **$**. A count causes that many objects to be affected, thus both **3c)** and **c3)** change the following three sentences (7.4).

d An operator which deletes the following object. If more than part of a line is affected, the text is saved in the numeric buffers. A count causes that many objects to be affected; thus **3dw** is the same as **d3w** (3.3, 3.4, 4.1, 7.4).

e Advances to the end of the next word, defined as for **b** and **w**. A count repeats the effect (2.4, 3.1).

f Finds the first instance of the next character following the cursor on the current line. A count repeats the find (4.1).

g Unused.

Arrow keys **h**, **j**, **k**, **l**, and **H**.

h **Left arrow**. Moves the cursor one character to the left. Like the other arrow keys, either **h**, the **left arrow** key, or one of the synonyms (**^H**) has the same effect. On v2 editors, arrow keys on certain kinds of terminals (those which send escape sequences, such as vt52, c100, or hp) cannot be used. A count repeats the effect (3.1, 7.5).

i Inserts text before the cursor, otherwise like **a** (7.2).

j **Down arrow**. Moves the cursor one line down in the same column. If the position does not exist, *vi* comes as close as possible to the same column. Synonyms include **^J** (linefeed) and **^N**.

k **Up arrow**. Moves the cursor one line up. **^P** is a synonym.

l — **Right arrow**. Moves the cursor one character to the right. SPACE is a synonym.

m — Marks the current position of the cursor in the mark register which is specified by the next character **a–z**. Return to this position or use with an operator using ` or ′ (5.3).

n — Repeats the last / or **?** scanning commands (2.2).

o — Opens new lines below the current line; otherwise like **O** (3.1).

p — Puts text after/below the cursor; otherwise like **P** (6.3).

q — Unused.

r — Replaces the single character at the cursor with a single character you type. The new character may be a RETURN; this is the easiest way to split lines. A count replaces each of the following count characters with the single character given; see **R** above which is the more usually useful iteration of **r** (3.2).

s — Changes the single character under the cursor to the text which follows up to an ESC; given a count, that many characters from the current line are changed. The last character to be changed is marked with **$** as in **c** (3.2).

t — Advances the cursor upto the character before the next character typed. Most useful with operators such as **d** and **c** to delete the characters up to a following character. You can use **.** to delete more if this doesn't delete enough the first time (4.1).

u — Undoes the last change made to the current buffer. If repeated, will alternate between these two states, thus is its own inverse. When used after an insert which inserted text on more than one line, the lines are saved in the numeric named buffers (3.5).

v — Unused.

w — Advances to the beginning of the next word, as defined by **b** (2.4).

x — Deletes the single character under the cursor. With a count deletes deletes that many characters forward from the cursor position, but only on the current line (6.5).

y — An operator, yanks the following object into the unnamed temporary buffer. If preceded by a named buffer specification, "*x*, the text is placed in that buffer also. Text can be recovered by a later **p** or **P** (7.4).

z — Redraws the screen with the current line placed as specified by the following character: RETURN specifies the top of the screen, **.** the center of the screen, and **–** at the bottom of the screen. A count may be given after the **z** and before the following character to specify the new screen size for the redraw. A count before the **z** gives the number of the line to place in the center of the screen instead of the default current line. (5.4)

{ — Retreats to the beginning of the beginning of the preceding paragraph. A paragraph begins at each macro in the *paragraphs* option, normally '.IP', '.LP', '.PP', '.QP' and '.bp'. A paragraph also begins after a completely empty line, and at each section boundary (see [[above) (4.2, 6.8, 7.6).

| — Places the cursor on the character in the column specified by the count (7.1, 7.2).

} — Advances to the beginning of the next paragraph. See { for the definition of paragraph (4.2, 6.8, 7.6).

~ — Unused.

^? (DEL) — Interrupts the editor, returning it to command accepting state (1.5, 7.5)

Vi Command & Function Reference

Alan P.W. Hewett

Revised for version 2.12 by Mark Horton

1. Author's Disclaimer

This document does not claim to be 100% complete. There are a few commands listed in the original document that I was unable to test either because I do not speak **lisp**, because they required programs we don't have, or because I wasn't able to make them work. In these cases I left the command out. The commands listed in this document have been tried and are known to work. It is expected that prospective users of this document will read it once to get the flavor of everything that **vi** can do and then use it as a reference document. Experimentation is recommended. If you don't understand a command, try it and see what happens.

[Note: In revising this document, I have attempted to make it completely reflect version 2.12 of **vi**. It does not attempt to document the VAX version (version 3), but with one or two exceptions (wrapmargin, arrow keys) everything said about 2.12 should apply to 3.1. *Mark Horton*]

2. Notation

[**option**] is used to denote optional parts of a command. Many **vi** commands have an optional count. [**cnt**] means that an optional number may precede the command to multiply or iterate the command. {**variable item**} is used to denote parts of the command which must appear, but can take a number of different values. <**character** [**-character**]> means that the character or one of the characters in the range described between the two angle brackets is to be typed. For example <**esc**> means the **escape** key is to be typed. <**a-z**> means that a lower case letter is to be typed. ^<**character**> means that the character is to be typed as a **control** character, that is, with the <**cntl**> key held down while simultaneously typing the specified character. In this document control characters will be denoted using the ***upper case*** character, but ^<uppercase chr> and ^<lowercase chr> are equivalent. That is, for example, <^**D**> is equal to <^**d**>. The most common character abbreviations used in this list are as follows:

<esc>	escape, octal 033
<cr>	carriage return, ^M, octal 015
<lf>	linefeed ^J, octal 012
<nl>	newline, ^J, octal 012 (same as linefeed)
<bs>	backspace, ^H, octal 010
<tab>	tab, ^I, octal 011
<bell>	bell, ^G, octal 07
<ff>	formfeed, ^L, octal 014
<sp>	space, octal 040
<del>	delete, octal 0177

[cnt]B
: Move the cursor to the preceding word that is separated from the current word by a "white space" (<sp>,<tab>, or <nl>).

[cnt]e
: Move the cursor to the end of the current word or the end of the "cnt"'th word hence. Mnemonic: **e**nd-of-word

[cnt]E
: Move the cursor to the end of the current word which is delimited by "white space" (<sp>,<tab>, or <nl>).

[line number]G
: Move the cursor to the line specified. Of particular use are the sequences "1G" and "G", which move the cursor to the beginning and the end of the file respectively. Mnemonic: **G**o-to

NOTE: The next four commands (^D, ^U, ^F, ^B) are not true motion commands, in that they cannot be used as the object of commands such as delete or change.

[cnt]^D
: Move the cursor down in the file by "cnt" lines (or the last "cnt" if a new count isn't given. The initial default is half a page.) The screen is simultaneously scrolled up. Mnemonic: **D**own

[cnt]^U
: Move the cursor up in the file by "cnt" lines. The screen is simultaneously scrolled down. Mnemonic: **U**p

[cnt]^F
: Move the cursor to the next page. A count moves that many pages. Two lines of the previous page are kept on the screen for continuity if possible. Mnemonic: **F**orward-a-page

[cnt]^B
: Move the cursor to the previous page. Two lines of the current page are kept if possible. Mnemonic: **B**ackup-a-page

[cnt](
: Move the cursor to the beginning of the next sentence. A sentence is defined as ending with a ".", "!", or "?" followed by two spaces or a <nl>.

[cnt])
: Move the cursor backwards to the beginning of a sentence.

[cnt]}
: Move the cursor to the beginning of the next paragraph. This command works best inside **nroff** documents. It understands two sets of **nroff** macros, **–ms** and **–mm**, for which the commands ".IP", ".LP", ".PP", ".QP", "P", as well as the nroff command ".bp" are considered to be paragraph delimiters. A blank line also delimits a paragraph. The **nroff** macros that it accepts as paragraph delimiters is adjustable. See **paragraphs** under the **Set Commands** section.

[cnt]{
: Move the cursor backwards to the beginning of a paragraph.

]]
: Move the cursor to the next "section", where a section is defined by two sets of **nroff** macros, **–ms** and **–mm**, in which ".NH", ".SH", and ".H" delimit a section. A line beginning with a <ff><nl> sequence, or a line beginning with a "{" are also considered to be section delimiters. The last option makes it useful for finding the beginnings of C functions. The **nroff** macros that are used for section delimiters can be adjusted. See **sections** under the **Set Commands** section.

[[
: Move the cursor backwards to the beginning of a section.

%
: Move the cursor to the matching parenthesis or brace. This is very useful in C or lisp code. If the cursor is sitting on a ( ) { or } the cursor is moved to the matching character at the other end of the section. If the cursor is not sitting on a brace or a parenthesis, **vi** searches forward until it finds one and then jumps to the match mate.

[cnt]H
: If there is no count move the cursor to the top left position on the screen. If there is a count, then move the cursor to the beginning of the line "cnt" lines from the top of the screen. Mnemonic: **H**ome

4.1. Entry and Exit

To enter **vi** on a particular *file*, type

vi *file*

The file will be read in and the cursor will be placed at the beginning of the first line. The first screenfull of the file will be displayed on the terminal.

To get out of the editor, type

ZZ

If you are in some special mode, such as input mode or the middle of a multi-keystroke command, it may be necessary to type <esc> first.

4.2. Cursor and Page Motion

NOTE: The arrow keys (see the next four commands) on certain kinds of terminals will not work with the PDP-11 version of vi. The control versions or the hjkl versions will work on any terminal. Experienced users prefer the hjkl keys because they are always right under their fingers. Beginners often prefer the arrow keys, since they do not require memorization of which hjkl key is which. The mnemonic value of hjkl is clear from looking at the keyboard of an adm3a.

[cnt]<bs> or [cnt]h or [cnt]←
: Move the cursor to the left one character. Cursor stops at the left margin of the page. If cnt is given, these commands move that many spaces.

[cnt]^N or [cnt]j or [cnt]↓ or [cnt]<lf>
: Move down one line. Moving off the screen scrolls the window to force a new line onto the screen. Mnemonic: **N**ext

[cnt]^P or [cnt]k or [cnt]↑
: Move up one line. Moving off the top of the screen forces new text onto the screen. Mnemonic: **P**revious

[cnt]<sp> or [cnt]l or [cnt]→
: Move to the right one character. Cursor will not go beyond the end of the line.

[cnt]-
: Move the cursor up the screen to the beginning of the next line. Scroll if necessary.

[cnt]+ or [cnt]<cr>
: Move the cursor down the screen to the beginning of the next line. Scroll up if necessary.

[cnt]$
: Move the cursor to the end of the line. If there is a count, move to the end of the line "cnt" lines forward in the file.

^
: Move the cursor to the beginning of the first word on the line.

0
: Move the cursor to the left margin of the current line.

[cnt]|
: Move the cursor to the column specified by the count. The default is column zero.

[cnt]w
: Move the cursor to the beginning of the next word. If there is a count, then move forward that many words and position the cursor at the beginning of the word. Mnemonic: next-**w**ord

[cnt]W
: Move the cursor to the beginning of the next word which follows a "white space" (<sp>,<tab>, or <nl>). Ignore other punctuation.

[cnt]b
: Move the cursor to the preceding word. Mnemonic: **b**ackup-word

n	Repeat the last /[string]/ or ?[string]? search. Mnemonic: **n**ext occurrence.
N	Repeat the last /[string]/ or ?[string]? search, but in the reverse direction.
:g/[string]/[editor command]<nl>	Using the : syntax it is possible to do global searches ala the standard UNIX "ed" editor.

4.4. Text Insertion

The following commands allow for the insertion of text. All multicharacter text insertions are terminated with an <esc> character. The last change can always be **undone** by typing a **u**. The text insert in insertion mode can contain newlines.

a{text}<esc>	Insert text immediately following the cursor position. Mnemonic: **a**ppend
A{text}<esc>	Insert text at the end of the current line. Mnemonic: **A**ppend
i{text}<esc>	Insert text immediately preceding the cursor position. Mnemonic: **i**nsert
I{text}<esc>	Insert text at the beginning of the current line.
o{text}<esc>	Insert a new line after the line on which the cursor appears and insert text there. Mnemonic: **o**pen new line
O{text}<esc>	Insert a new line preceding the line on which the cursor appears and insert text there.

4.5. Text Deletion

The following commands allow the user to delete text in various ways. All changes can always be **undone** by typing the **u** command.

[cnt]x	Delete the character or characters starting at the cursor position.
[cnt]X	Delete the character or characters starting at the character preceding the cursor position.
D	Deletes the remainder of the line starting at the cursor. Mnemonic: **D**elete the rest of line
[cnt]d{motion}	Deletes one or more occurrences of the specified motion. Any motion from sections 4.1 and 4.2 can be used here. The d can be stuttered (e.g. [cnt]dd) to delete cnt lines.

4.6. Text Replacement

The following commands allow the user to simultaneously delete and insert new text. All such actions can be **undone** by typing **u** following the command.

r<chr>	Replaces the character at the current cursor position with <chr>. This is a one character replacement. No <esc> is required for termination. Mnemonic: **r**eplace character
R{text}<esc>	Starts overlaying the characters on the screen with whatever you type. It does not stop until an <esc> is typed.
[cnt]s{text}<esc>	Substitute for "cnt" characters beginning at the current cursor position. A "$" will appear at the position in the text where the "cnt"'th character appears so you will know how much you are erasing. Mnemonic: **s**ubstitute
[cnt]S{text}<esc>	Substitute for the entire current line (or lines). If no count is given, a "$" appears at the end of the current line. If a count of more than 1 is given, all the lines to be replaced are deleted before the insertion begins.
[cnt]c{motion}{text}<esc>	Change the specified "motion" by replacing it with the insertion text. A "$" will

[cnt]L — If there is no count move the cursor to the beginning of the last line on the screen. If there is a count, then move the cursor to the beginning of the line "cnt" lines from the bottom of the screen. Mnemonic: **L**ast

M — Move the cursor to the beginning of the middle line on the screen. Mnemonic: **M**iddle

m<a-z> — This command does not move the cursor, but it **marks** the place in the file and the character "<a-z>" becomes the label for referring to this location in the file. See the next two commands. Mnemonic: **m**ark **NOTE:** The mark command is not a motion, and cannot be used as the target of commands such as delete.

'<a-z> — Move the cursor to the beginning of the line that is marked with the label "<a-z>".

`<a-z> — Move the cursor to the exact position on the line that was marked with with the label "<a-z>".

'' — Move the cursor back to the beginning of the line where it was before the last "non-relative" move. A "non-relative" move is something such as a search or a jump to a specific line in the file, rather than moving the cursor or scrolling the screen.

`` — Move the cursor back to the exact spot on the line where it was located before the last "non-relative" move.

4.3. Searches

The following commands allow you to search for items in a file.

[cnt]f{chr}

Search forward on the line for the next or "cnt"'th occurrence of the character "chr". The cursor is placed **at** the character of interest. Mnemonic: **f**ind character

[cnt]F{chr}

Search backwards on the line for the next or "cnt"'th occurrence of the character "chr". The cursor is placed **at** the character of interest.

[cnt]t{chr}

Search forward on the line for the next or "cnt"'th occurrence of the character "chr". The cursor is placed **just preceding** the character of interest. Mnemonic: move cursor up **t**o character

[cnt]T{chr}

Search backwards on the line for the next or "cnt"'th occurrence of the character "chr". The cursor is placed **just preceding** the character of interest.

[cnt]; — Repeat the last "f", "F", "t" or "T" command.

[cnt], — Repeat the last "f", "F", "t" or "T" command, but in the opposite search direction. This is useful if you overshoot.

[cnt]/[string]/<nl>
Search forward for the next occurrence of "string". Wrap around at the end of the file does occur. The final </> is not required.

[cnt]?[string]?<nl>
Search backwards for the next occurrence of "string". If a count is specified, the count becomes the new window size. Wrap around at the beginning of the file does occur. The final <?> is not required.

4.8. Miscellaneous Commands

Vi has a number of miscellaneous commands that are very useful. They are:

ZZ — This is the normal way to exit from vi. If any changes have been made, the file is written out. Then you are returned to the shell.

^L — Redraw the current screen. This is useful if someone "write"s you while you are in "vi" or if for any reason garbage gets onto the screen.

^R — On dumb terminals, those not having the "delete line" function (the vt100 is such a terminal), **vi** saves redrawing the screen when you delete a line by just marking the line with an "@" at the beginning and blanking the line. If you want to actually get rid of the lines marked with "@" and see what the page looks like, typing a ^R will do this.

. — "Dot" is a particularly useful command. It repeats the last text modifying command. Therefore you can type a command once and then to another place and repeat it by just typing ".".

u — Perhaps the most important command in the editor, u undoes the last command that changed the buffer. Mnemonic: **u**ndo

U — Undo all the text modifying commands performed on the current line since the last time you moved onto it.

[cnt]J — Join the current line and the following line. The <nl> is deleted and the two lines joined, usually with a space between the end of the first line and the beginning of what was the second line. If the first line ended with a "period", then two spaces are inserted. A count joins the next cnt lines. Mnemonic: Join lines

Q — Switch to **ex** editing mode. In this mode **vi** will behave very much like **ed**. The editor in this mode will operate on single lines normally and will not attempt to keep the "window" up to date. Once in this mode it is also possible to switch to the **open** mode of editing. By entering the command **[line number]open<nl>** you enter this mode. It is similar to the normal visual mode except the window is only **one** line long. Mnemonic: **Q**uit visual mode

^] — An abbreviation for a tag command. The cursor should be positioned at the beginning of a word. That word is taken as a tag name, and the tag with that name is found as if it had been typed in a :tag command.

[cnt]!{motion}{UNIX cmd}<nl> — Any UNIX filter (e.g. command that reads the standard input and outputs something to the standard output) can be sent a section of the current file and have the output of the command replace the original text. Useful examples are programs like **cb**, **sort**, and **nroff**. For instance, using **sort** it would be possible to sort a section of the current file into a new list. Using **!!** means take a line or lines starting at the line the cursor is currently on and pass them to the UNIX command. **NOTE:** To just escape to the shell for one command, use :!{cmd}<nl>, see section 5.

z{cnt}<nl> — This resets the current window size to "cnt" lines and redraws the screen.

4.9. Special Insert Characters

There are some characters that have special meanings during insert modes. They are:

^V — During inserts, typing a ^V allows you to quote control characters into the file. Any character typed after the ^V will be inserted into the file.

[^]^D or [0]^D — <^D> without any argument backs up one **shiftwidth**. This is necessary to remove indentation that was inserted by the **autoindent** feature. ^<^D> temporarily removes all the autoindentation, thus placing the cursor at the left margin. On the next line, the previous indent level will be restored. This is appear at the end of the last item that is being deleted unless the deletion involves whole lines. Motion's can be any motion from sections 4.1 or 4.2. Stuttering the c (e.g. [cnt]cc) changes cnt lines.

4.7. Moving Text

Vi provides a number of ways of moving chunks of text around. There are nine buffers into which each piece of text which is deleted or "yanked" is put in addition to the "undo" buffer. The most recent deletion or yank is in the "undo" buffer and also usually in buffer 1, the next most recent in buffer 2, and so forth. Each new deletion pushes down all the older deletions. Deletions older than 9 disappear. There is also a set of named registers, a-z, into which text can optionally be placed. If any delete or replacement type command is preceded by **"<a-z>**, that named buffer will contain the text deleted after the command is executed. For example, **"a3dd** will delete three lines starting at the current line and put them in buffer **"a.*** There are two more basic commands and some variations useful in getting and putting text into a file.

["<a-z>][cnt]y{motion}

Yank the specified item or "cnt" items and put in the "undo" buffer or the specified buffer. The variety of "items" that can be yanked is the same as those that can be deleted with the "d" command or changed with the "c" command. In the same way that "dd" means delete the current line and "cc" means replace the current line, "yy" means yank the current line.

["<a-z>][cnt]Y
Yank the current line or the "cnt" lines starting from the current line. If no buffer is specified, they will go into the "undo" buffer, like any delete would. It is equivalent to "yy". Mnemonic: **Y**ank

["<a-z>]p
Put "undo" buffer or the specified buffer down **after** the cursor. If whole lines were yanked or deleted into the buffer, then they will be put down on the line following the line the cursor is on. If something else was deleted, like a word or sentence, then it will be inserted immediately following the cursor. Mnemonic: **p**ut buffer

It should be noted that text in the named buffers remains there when you start editing a new file with the **:e file<esc>** command. Since this is so, it is possible to copy or delete text from one file and carry it over to another file in the buffers. However, the undo buffer and the ability to undo are lost when changing files.

["<a-z>]P
Put "undo" buffer or the specified buffer down **before** the cursor. If whole lines where yanked or deleted into the buffer, then they will be put down on the line preceding the line the cursor is on. If something else was deleted, like a word or sentence, then it will be inserted immediately preceding the cursor.

[cnt]>{motion}
The shift operator will right shift all the text from the line on which the cursor is located to the line where the **motion** is located. The text is shifted by one **shiftwidth**. (See section 6.) >> means right shift the current line or lines.

[cnt]<{motion}
The shift operator will left shift all the text from the line on which the cursor is located to the line where the **item** is located. The text is shifted by one **shiftwidth**. (See section 6.) << means left shift the current line or lines. Once the line has reached the left margin it is not further affected.

[cnt]={motion}
Prettyprints the indicated area according to **lisp** conventions. The area should be a lisp s-expression.

* Referring to an upper case letter as a buffer name (A-Z) is the same as referring to the lower case letter, except that text placed in such a buffer is appended to it instead of replacing it.

tag filename vi-search-command

If **vi** finds the tag you specified in the **:ta** command, it stops editing the current file if necessary and if the current file is up to date on the disk and switches to the file specified and uses the search pattern specified to find the "tagged" item of interest. This is particularly useful when editing multi-file C programs such as the operating system. There is a program called **ctags** which will generate an appropriate **tags** file for C and f77 programs so that by saying **:ta function<nl>** you will be switched to that function. It could also be useful when editing multi-file documents, though the **tags** file would have to be generated manually.

6. Special Arrangements for Startup

Vi takes the value of **$TERM** and looks up the characteristics of that terminal in the file **/etc/termcap**. If you don't know **vi**'s name for the terminal you are working on, look in **/etc/termcap**.

When **vi** starts, it attempts to read the variable EXINIT from your environment.* If that exists, it takes the values in it as the default values for certain of its internal constants. See the section on "Set Values" for further details. If EXINIT doesn't exist you will get all the normal defaults.

Should you inadvertently hang up the phone while inside **vi**, or should the computer crash, all may not be lost. Upon returning to the system, type:

vi –r file

This will normally recover the file. If there is more than one temporary file for a specific file name, **vi** recovers the newest one. You can get an older version by recovering the file more than once. The command "vi -r" without a file name gives you the list of files that were saved in the last system crash (but *not* the file just saved when the phone was hung up).

7. Set Commands

Vi has a number of internal variables and switches which can be set to achieve special affects. These options come in three forms, those that are switches, which toggle from off to on and back, those that require a numeric value, and those that require an alphanumeric string value. The toggle options are set by a command of the form:

:set option<nl>

and turned off with the command:

:set nooption<nl>

Commands requiring a value are set with a command of the form:

:set option=value<nl>

To display the value of a specific option type:

:set option?<nl>

To display only those that you have changed type:

:set<nl>

and to display the long table of all the settable parameters and their current values type:

* On version 6 systems Instead of EXINIT, put the startup commands in the file .exrc in your home directory.

useful for putting "labels" at the left margin. 0<^D> says remove all autoindents and stay that way. Thus the cursor moves to the left margin and stays there on successive lines until <tab>'s are typed. As with the <tab>, the <^D> is only effective before any other "non-autoindent" controlling characters are typed. Mnemonic: Delete a shiftwidth

^W — If the cursor is sitting on a word, <^W> moves the cursor back to the beginning of the word, thus erasing the word from the insert. Mnemonic: erase Word

<bs> — The backspace always serves as an erase during insert modes in addition to your normal "erase" character. To insert a <bs> into your file, use the <^V> to quote it.

5. : Commands

Typing a ":" during command mode causes **vi** to put the cursor at the bottom on the screen in preparation for a command. In the ":" mode, **vi** can be given most **ed** commands. It is also from this mode that you exit from **vi** or switch to different files. All commands of this variety are terminated by a <nl>, <cr>, or <esc>.

:w[!] [file] — Causes **vi** to write out the current text to the disk. It is written to the file you are editing unless "file" is supplied. If "file" is supplied, the write is directed to that file instead. If that file already exists, **vi** will not perform the write unless the "!" is supplied indicating you *really* want to destroy the older copy of the file.

:q[!] — Causes **vi** to exit. If you have modified the file you are looking at currently and haven't written it out, **vi** will refuse to exit unless the "!" is supplied.

:e[!] [+[cmd]] [file]

Start editing a new file called "file" or start editing the current file over again. The command ":e!" says "ignore the changes I've made to this file and start over from the beginning". It is useful if you really mess up the file. The optional "+" says instead of starting at the beginning, start at the "end", or, if "cmd" is supplied, execute "cmd" first. Useful cases of this are where cmd is "n" (any integer) which starts at line number n, and "/text", which searches for "text" and starts at the line where it is found.

^^ — Switch back to the place you were before your last tag command. If your last tag command stayed within the file, ^^ returns to that tag. If you have no recent tag command, it will return to the same place in the previous file that it was showing when you switched to the current file.

:n[!] — Start editing the next file in the argument list. Since **vi** can be called with multiple file names, the ":n" command tells it to stop work on the current file and switch to the next file. If the current file was modifies, it has to be written out before the ":n" will work or else the "!" must be supplied, which says discard the changes I made to the current file.

:n[!] file [file file ...]

Replace the current argument list with a new list of files and start editing the first file in this new list.

:r file — Read in a copy of "file" on the line after the cursor.

:r !cmd — Execute the "cmd" and take its output and put it into the file after the current line.

:!cmd — Execute any UNIX shell command.

:ta[!] tag — **Vi** looks in the file named **tags** in the current directory. **Tags** is a file of lines in the format:

open
: Default: open Type: toggle
When set, prevents entering open or visual modes from ex or edit. Not of interest from vi.

optimize opt
: Default: opt Type: toggle
Basically of use only when using the **ex** capabilities. This option prevents automatic <cr>s from taking place, and speeds up output of indented lines, at the expense of losing typeahead on some versions of UNIX.

paragraphs para
: Default: para=IPLPPPQPP bp Type: string
Each pair of characters in the string indicate **nroff** macros which are to be treated as the beginning of a paragraph for the { and } commands. The default string is for the **-ms** and **-mm** macros. To indicate one letter **nroff** macros, such as **.P** or **.H**, quote a space in for the second character position. For example:

:set paragraphs=P\ bp<nl>

would cause **vi** to consider **.P** and **.bp** as paragraph delimiters.

prompt
: Default: prompt Type: toggle
In **ex** command mode the prompt character **:** will be printed when **ex** is waiting for a command. This is not of interest from vi.

redraw
: Default: noredraw Type: toggle
On dumb terminals, force the screen to always be up to date, by sending great amounts of output. Useful only at high speeds.

report
: Default: report=5 Type: numeric
This sets the threshold for the number of lines modified. When more than this number of lines are modified, removed, or yanked, **vi** will report the number of lines changed at the bottom of the screen.

scroll
: Default: scroll={1/2 window} Type: numeric
This is the number of lines that the screen scrolls up or down when using the <^U> and <^D> commands.

sections
: Default: sections=SHNHH HU Type: string
Each two character pair of this string specify **nroff** macro names which are to be treated as the beginning of a section by the]] and [[commands. The default string is for the **-ms** and **-mm** macros. To enter one letter **nroff** macros, use a quoted space as the second character. See **paragraphs** for a fuller explanation.

shell sh
: Default: sh=from environment SHELL or /bin/sh Type: string
This is the name of the **sh** to be used for "escaped" commands.

shiftwidth sw
: Default: sw=8 Type: numeric
This is the number of spaces that a <^T> or <^D> will move over for indenting, and the amount < and > shift by.

showmatch sm
: Default: nosm Type: toggle
When a) or } is typed, show the matching (or { by moving the cursor to it for one second if it is on the current screen.

slowopen slow
: Default: terminal dependent Type: toggle
On terminals that are slow and unintelligent, this option prevents the updating of the screen some of the time to improve speed.

tabstop ts
: Default: ts=8 Type: numeric
<tab>s are expanded to boundaries that are multiples of this value.

taglength tl
: Default: tl=0 Type: numeric
If nonzero, tag names are only significant to this many characters.

:set all<nl>

Most of the options have a long form and an abbreviation. Both are listed in the following table as well as the normal default value.

To arrange to have values other than the default used every time you enter **vi**, place the appropriate **set** command in EXINIT in your environment, e.g.

```
EXINIT='set ai aw terse sh=/bin/csh'
export EXINIT
```

or

```
setenv EXINIT 'set ai aw terse sh=/bin/csh'
```

for **sh** and **csh**, respectively. These are usually placed in your .profile or .login. If you are running a system without environments (such as version 6) you can place the set command in the file .exrc in your home directory.

autoindent ai — Default: noai Type: toggle
When in autoindent mode, vi helps you indent code by starting each line in the same column as the preceding line. Tabbing to the right with <tab> or <^T> will move this boundary to the right, and it can be moved to the left with <^D>.

autoprint ap — Default: ap Type: toggle
Causes the current line to be printed after each ex text modifying command. This is not of much interest in the normal **vi** visual mode.

autowrite aw — Default: noaw type: toggle
Autowrite causes an automatic write to be done if there are unsaved changes before certain commands which change files or otherwise interact with the outside world. These commands are :!, :tag, :next, :rewind, ^^, and ^].

beautify bf — Default: nobf Type: toggle
Causes all control characters except <tab>, <nl>, and <ff> to be discarded.

directory dir — Default: dir=/tmp Type: string
This is the directory in which **vi** puts its temporary file.

errorbells eb — Default: noeb Type: toggle
Error messages are preceded by a <bell>.

hardtabs ht — Default: hardtabs=8 Type: numeric
This option contains the value of hardware tabs in your terminal, or of software tabs expanded by the Unix system.

ignorecase ic — Default: noic Type: toggle
All upper case characters are mapped to lower case in regular expression matching.

lisp — Default: nolisp Type: toggle
Autoindent for **lisp** code. The commands ( ) [[and]] are modified appropriately to affect s-expressions and functions.

list — Default: nolist Type: toggle
All printed lines have the <tab> and <nl> characters displayed visually.

magic — Default: magic Type: toggle
Enable the metacharacters for matching. These include **. * < > [string] [^string]** and **[<chr>-<chr>]**.

number nu — Default: nonu Type: toggle
Each line is displayed with its line number.

term — Default: (from environment **TERM**, else dumb) Type: string
This is the terminal and controls the visual displays. It cannot be changed when in "visual" mode, you have to Q to command mode, type a set term command, and do "vi." to get back into visual. Or exit vi, fix $TERM, and reenter. The definitions that drive a particular terminal type are found in the file **/etc/termcap**.

terse — Default: terse Type: toggle
When set, the error diagnostics are short.

warn — Default: warn Type: toggle
The user is warned if she/he tries to escape to the shell without writing out the current changes.

window — Default: window={8 at 600 baud or less, 16 at 1200 baud, and screen size – 1 at 2400 baud or more} Type: numeric
This is the number of lines in the window whenever **vi** must redraw an entire screen. It is useful to make this size smaller if you are on a slow line.

w300, w1200, w9600
These set window, but only within the corresponding speed ranges. They are useful in an EXINIT to fine tune window sizes. For example,

    set w300=4 w1200=12

causes a 4 lines window at speed up to 600 baud, a 12 line window at 1200 baud, and a full screen (the default) at over 1200 baud.

wrapscan ws — Default: ws Type: toggle
Searches will wrap around the end of the file when is option is set. When it is off, the search will terminate when it reaches the end or the beginning of the file.

wrapmargin wm — Default: wm=0 Type: numeric
Vi will automatically insert a <nl> when it finds a natural break point (usually a <sp> between words) that occurs within "wm" spaces of the right margin. Therefore with "wm=0" the option is off. Setting it to 10 would mean that any time you are within 10 spaces of the right margin **vi** would be looking for a <sp> or <tab> which it could replace with a <nl>. This is convenient for people who forget to look at the screen while they type. (In version 3, wrapmargin behaves more like nroff, in that the boundary specified by the distance from the right edge of the screen is taken as the rightmost edge of the area where a break is allowed, instead of the leftmost edge.)

writeany wa — Default: nowa Type: toggle
Vi normally makes a number of checks before it writes out a file. This prevents the user from inadvertently destroying a file. When the "writeany" option is enabled, **vi** no longer makes these checks.

A Tutorial Introduction to the UNIX Text Editor

Brian W. Kernighan

ABSTRACT

Almost all text input on the UNIX† operating system is done with the text-editor ***ed***. This memorandum is a tutorial guide to help beginners get started with text editing.

Although it does not cover everything, it does discuss enough for most users' day-to-day needs. This includes printing, appending, changing, deleting, moving and inserting entire lines of text; reading and writing files; context searching and line addressing; the substitute command; the global commands; and the use of special characters for advanced editing.

September 12, 1986

a
Now is the time
for all good men
to come to the aid of their party.

The only way to stop appending is to type a line that contains only a period. The "." is used to tell *ed* that you have finished appending. (Even experienced users forget that terminating "." sometimes. If *ed* seems to be ignoring you, type an extra line with just "." on it. You may then find you've added some garbage lines to your text, which you'll have to take out later.)

After the append command has been done, the buffer will contain the three lines

Now is the time
for all good men
to come to the aid of their party.

The "**a**" and "." aren't there, because they are not text.

To add more text to what you already have, just issue another **a** command, and continue typing.

Error Messages – "?"

If at any time you make an error in the commands you type to *ed*, it will tell you by typing

?

This is about as cryptic as it can be, but with practice, you can usually figure out how you goofed.

Writing text out as a file – the Write command "w"

It's likely that you'll want to save your text for later use. To write out the contents of the buffer onto a file, use the *write* command

w

followed by the filename you want to write on. This will copy the buffer's contents onto the specified file (destroying any previous information on the file). To save the text on a file named **junk**, for example, type

w junk

Leave a space between **w** and the file name. *Ed* will respond by printing the number of characters it wrote out. In this case, *ed* would respond with

68

(Remember that blanks and the return character at the end of each line are included in the character count.) Writing a file just makes a copy of the text – the buffer's contents are not disturbed, so you can go on adding lines to it. This is an important point. *Ed* at all times works on a copy of a file, not the file itself. No change in the contents of a file takes place until you give a **w** command. (Writing out the text onto a file from time to time as it is being created is a good idea, since if the system crashes or if you make some horrible mistake, you will lose all the text in the buffer but any text that was written onto a file is relatively safe.)

Leaving ed – the Quit command "q"

To terminate a session with *ed*, save the text you're working on by writing it onto a file using the **w** command, and then type the command

q

which stands for *quit*. The system will respond with the prompt character ($ or %). At this point your buffer vanishes, with all its text, which is why you want to write it out before quitting.†

Exercise 1:

Enter *ed* and create some text using

a
. . . text . . .
.

Write it out using **w**. Then leave *ed* with the **q** command, and print the file, to see that everything worked. (To print a file, say

pr filename

or

cat filename

in response to the prompt character. Try both.)

Reading text from a file – the Edit command "e"

A common way to get text into the buffer is to read it from a file in the file system. This is what you do to edit text that you saved with the **w** command in a previous session. The *edit* command **e** fetches the entire contents of a file into the buffer. So if you had saved the three lines "Now is the time", etc., with a **w** command in an earlier session, the *ed* command

e junk

would fetch the entire contents of the file **junk** into the buffer, and respond

† Actually, *ed* will print **?** if you try to quit without writing. At that point, write if you want; if not, another q will get you out regardless.

A Tutorial Introduction to the UNIX Text Editor

Brian W. Kernighan

Introduction

Ed is a "text editor", that is, an interactive program for creating and modifying "text", using directions provided by a user at a terminal. The text is often a document like this one, or a program or perhaps data for a program.

This introduction is meant to simplify learning *ed*. The recommended way to learn *ed* is to read this document, simultaneously using *ed* to follow the examples, then to read the description in section I of the *UNIX Programmer's Manual,* all the while experimenting with *ed.* (Solicitation of advice from experienced users is also useful.)

Do the exercises! They cover material not completely discussed in the actual text. An appendix summarizes the commands.

Disclaimer

This is an introduction and a tutorial. For this reason, no attempt is made to cover more than a part of the facilities that *ed* offers (although this fraction includes the most useful and frequently used parts). When you have mastered the Tutorial, try *Advanced Editing on UNIX.* Also, there is not enough space to explain basic UNIX procedures. We will assume that you know how to log on to UNIX, and that you have at least a vague understanding of what a file is. For more on that, read *UNIX for Beginners.*

You must also know what character to type as the end-of-line on your particular terminal. This character is the RETURN key on most terminals. Throughout, we will refer to this character, whatever it is, as RETURN.

Getting Started

We'll assume that you have logged in to your system and it has just printed the prompt character, usually either a **$** or a **%**. The easiest way to get *ed* is to type

ed (followed by a return)

You are now ready to go – *ed* is waiting for you to tell it what to do.

Creating Text – the Append command "a"

As your first problem, suppose you want to create some text starting from scratch. Perhaps you are typing the very first draft of a paper; clearly it will have to start somewhere, and undergo modifications later. This section will show how to get some text in, just to get started. Later we'll talk about how to change it.

When *ed* is first started, it is rather like working with a blank piece of paper – there is no text or information present. This must be supplied by the person using *ed;* it is usually done by typing in the text, or by reading it into *ed* from a file. We will start by typing in some text, and return shortly to how to read files.

First a bit of terminology. In *ed* jargon, the text being worked on is said to be "kept in a buffer." Think of the buffer as a work space, if you like, or simply as the information that you are going to be editing. In effect the buffer is like the piece of paper, on which we will write things, then change some of them, and finally file the whole thing away for another day.

The user tells *ed* what to do to his text by typing instructions called "commands." Most commands consist of a single letter, which must be typed in lower case. Each command is typed on a separate line. (Sometimes the command is preceded by information about what line or lines of text are to be affected – we will discuss these shortly.) *Ed* makes no response to most commands – there is no prompting or typing of messages like "ready". (This silence is preferred by experienced users, but sometimes a hangup for beginners.)

The first command is ***append,*** written as the letter

a

all by itself. It means "append (or add) text lines to the buffer, as I type them in." Appending is rather like writing fresh material on a piece of paper.

So to enter lines of text into the buffer, just type an **a** followed by a RETURN, followed by the lines of text you want, like this:

To print the *last* line of the buffer, you could use

$,$p

but *ed* lets you abbreviate this to

$p

You can print any single line by typing the line number followed by a **p**. Thus

1p

produces the response

Now is the time

which is the first line of the buffer.

In fact, *ed* lets you abbreviate even further: you can print any single line by typing *just* the line number – no need to type the letter **p**. So if you say

$

ed will print the last line of the buffer.

You can also use **$** in combinations like

$–1,$p

which prints the last two lines of the buffer. This helps when you want to see how far you got in typing.

Exercise 3:

As before, create some text using the **a** command and experiment with the **p** command. You will find, for example, that you can't print line 0 or a line beyond the end of the buffer, and that attempts to print a buffer in reverse order by saying

3,1p

don't work.

The current line – "Dot" or "."

Suppose your buffer still contains the six lines as above, that you have just typed

1,3p

and *ed* has printed the three lines for you. Try typing just

p (no line numbers)

This will print

to come to the aid of their party.

which is the third line of the buffer. In fact it is the last (most recent) line that you have done anything with. (You just printed it!) You can repeat this **p** command without line numbers, and it will continue to print line 3.

The reason is that *ed* maintains a record of the last line that you did anything to (in this case, line 3, which you just printed) so that it can be used instead of an explicit line number. This most recent line is referred to by the shorthand symbol

(pronounced "dot").

Dot is a line number in the same way that **$** is; it means exactly "the current line", or loosely, "the line you most recently did something to." You can use it in several ways – one possibility is to say

,$p

This will print all the lines from (including) the current line to the end of the buffer. In our example these are lines 3 through 6.

Some commands change the value of dot, while others do not. The **p** command sets dot to the number of the last line printed; the last command will set both

and **$** to 6.

Dot is most useful when used in combinations like this one:

+1 (or equivalently, **+1p**)

This means "print the next line" and is a handy way to step slowly through a buffer. You can also say

–1 (or **–1p**)

which means "print the line *before* the current line." This enables you to go backwards if you wish. Another useful one is something like

–3,–1p

which prints the previous three lines.

Don't forget that all of these change the value of dot. You can find out what dot is at any time by typing

=

Ed will respond by printing the value of dot.

Let's summarize some things about the **p** command and dot. Essentially **p** can be preceded by 0, 1, or 2 line numbers. If there is no line number given, it prints the "current line", the line that dot refers to. If there is one line number given (with or without the letter **p**), it prints that line (and dot is set there); and if there are two line numbers, it prints all the lines in that range (and sets dot to the last line printed.) If two line numbers are specified the first can't be bigger than the second (see Exercise 2.)

Typing a single return will cause printing of the next line – it's equivalent to **.+1p**. Try it. Try typing a **–**; you will find that it's equivalent

68

which is the number of characters in **junk**. *If anything was already in the buffer, it is deleted first.*

If you use the **e** command to read a file into the buffer, then you need not use a file name after a subsequent **w** command; *ed* remembers the last file name used in an **e** command, and **w** will write on this file. Thus a good way to operate is

```
ed
e file
[editing session]
w
q
```

This way, you can simply say **w** from time to time, and be secure in the knowledge that if you got the file name right at the beginning, you are writing into the proper file each time.

You can find out at any time what file name *ed* is remembering by typing the *file* command **f**. In this example, if you typed

```
f
```

ed would reply

```
junk
```

Reading text from a file – the Read command "r"

Sometimes you want to read a file into the buffer without destroying anything that is already there. This is done by the *read* command **r**. The command

```
r junk
```

will read the file **junk** into the buffer; it adds it to the end of whatever is already in the buffer. So if you do a read after an edit:

```
e junk
r junk
```

the buffer will contain *two* copies of the text (six lines).

```
Now is the time
for all good men
to come to the aid of their party.
Now is the time
for all good men
to come to the aid of their party.
```

Like the **w** and **e** commands, **r** prints the number of characters read in, after the reading operation is complete.

Generally speaking, **r** is much less used than **e**.

Exercise 2:

Experiment with the **e** command – try reading and printing various files. You may get an error **?name**, where **name** is the name of a file; this means that the file doesn't exist, typically because you spelled the file name wrong, or perhaps that you are not allowed to read or write it. Try alternately reading and appending to see that they work similarly. Verify that

```
ed filename
```

is exactly equivalent to

```
ed
e filename
```

What does

```
f filename
```

do?

Printing the contents of the buffer – the Print command "p"

To *print* or list the contents of the buffer (or parts of it) on the terminal, use the print command

```
p
```

The way this is done is as follows. Specify the lines where you want printing to begin and where you want it to end, separated by a comma, and followed by the letter **p**. Thus to print the first two lines of the buffer, for example, (that is, lines 1 through 2) say

```
1,2p  (starting line=1, ending line=2 p)
```

Ed will respond with

```
Now is the time
for all good men
```

Suppose you want to print *all* the lines in the buffer. You could use **1,3p** as above if you knew there were exactly 3 lines in the buffer. But in general, you don't know how many there are, so what do you use for the ending line number? *Ed* provides a shorthand symbol for "line number of last line in buffer" – the dollar sign **$**. Use it this way:

```
1,$p
```

This will print *all* the lines in the buffer (line 1 to last line.) If you want to stop the printing before it is finished, push the DEL or Delete key; *ed* will type

```
?
```

and wait for the next command.

right. If it didn't, you can try again. (Notice that there is a **p** on the same line as the **s** command. With few exceptions, **p** can follow any command; no other multi-command lines are legal.)

It's also legal to say

    s/ . . . //

which means "change the first string of characters to "*nothing*", i.e., remove them. This is useful for deleting extra words in a line or removing extra letters from words. For instance, if you had

    Nowxx is the time

you can say

    s/xx//p

to get

    Now is the time

Notice that // (two adjacent slashes) means "no characters", not a blank. There *is* a difference! (See below for another meaning of //.)

Exercise 5:

Experiment with the substitute command. See what happens if you substitute for some word on a line with several occurrences of that word. For example, do this:

    a
    the other side of the coin

    s/the/on the/p

You will get

    on the other side of the coin

A substitute command changes only the first occurrence of the first string. You can change all occurrences by adding a **g** (for "global") to the **s** command, like this:

    s/ . . . / . . . /gp

Try other characters instead of slashes to delimit the two sets of characters in the **s** command – anything should work except blanks or tabs.

(If you get funny results using any of the characters

    ^ $ [* \ &

read the section on "Special Characters".)

Context searching – "/ . . . /"

With the substitute command mastered, you can move on to another highly important idea of *ed* – context searching.

Suppose you have the original three line text in the buffer:

    Now is the time
    for all good men
    to come to the aid of their party.

Suppose you want to find the line that contains *their* so you can change it to *the*. Now with only three lines in the buffer, it's pretty easy to keep track of what line the word *their* is on. But if the buffer contained several hundred lines, and you'd been making changes, deleting and rearranging lines, and so on, you would no longer really know what this line number would be. Context searching is simply a method of specifying the desired line, regardless of what its number is, by specifying some context on it.

The way to say "search for a line that contains this particular string of characters" is to type

    /*string of characters we want to find*/

For example, the *ed* command

    /their/

is a context search which is sufficient to find the desired line – it will locate the next occurrence of the characters between slashes ("their"). It also sets dot to that line and prints the line for verification:

    to come to the aid of their party.

"Next occurrence" means that *ed* starts looking for the string at line **.+1**, searches to the end of the buffer, then continues at line 1 and searches to line dot. (That is, the search "wraps around" from $ to 1.) It scans all the lines in the buffer until it either finds the desired line or gets back to dot again. If the given string of characters can't be found in any line, *ed* types the error message

    ?

Otherwise it prints the line it found.

You can do both the search for the desired line *and* a substitution all at once, like this:

    /their/s/their/the/p

which will yield

    to come to the aid of the party.

There were three parts to that last command: context search for the desired line, make the substitution, print the line.

The expression **/their/** is a context search expression. In their simplest form, all context search expressions are like this – a string of characters surrounded by slashes. Context searches are interchangeable with line numbers, so they can be used by themselves to find and print a desired line, or as line numbers for some other command, like **s**. They were used both ways in the examples above.

to **.−1p**.

Deleting lines: the "d" command

Suppose you want to get rid of the three extra lines in the buffer. This is done by the *delete* command

d

Except that **d** deletes lines instead of printing them, its action is similar to that of **p**. The lines to be deleted are specified for **d** exactly as they are for **p**:

starting line, ending line d

Thus the command

4,$d

deletes lines 4 through the end. There are now three lines left, as you can check by using

1,$p

And notice that **$** now is line 3! Dot is set to the next line after the last line deleted, unless the last line deleted is the last line in the buffer. In that case, dot is set to **$**.

Exercise 4:

Experiment with **a**, **e**, **r**, **w**, **p** and **d** until you are sure that you know what they do, and until you understand how dot, **$**, and line numbers are used.

If you are adventurous, try using line numbers with **a**, **r** and **w** as well. You will find that **a** will append lines *after* the line number that you specify (rather than after dot); that **r** reads a file in *after* the line number you specify (not necessarily at the end of the buffer); and that **w** will write out exactly the lines you specify, not necessarily the whole buffer. These variations are sometimes handy. For instance you can insert a file at the beginning of a buffer by saying

0r filename

and you can enter lines at the beginning of the buffer by saying

0a
. . . *text* . . .

Notice that **.w** is *very* different from

w

Modifying text: the Substitute command "s"

We are now ready to try one of the most important of all commands – the substitute command

s

This is the command that is used to change individual words or letters within a line or group of lines. It is what you use, for example, for correcting spelling mistakes and typing errors.

Suppose that by a typing error, line 1 says

Now is th time

– the *e* has been left off *the*. You can use **s** to fix this up as follows:

1s/th/the/

This says: "in line 1, substitute for the characters *th* the characters *the*." To verify that it works (*ed* will not print the result automatically) say

p

and get

Now is the time

which is what you wanted. Notice that dot must have been set to the line where the substitution took place, since the **p** command printed that line. Dot is always set this way with the **s** command.

The general way to use the substitute command is

starting-line, ending-line s/*change this*/*to this*/

Whatever string of characters is between the first pair of slashes is replaced by whatever is between the second pair, in *all* the lines between *starting-line* and *ending-line*. Only the first occurrence on each line is changed, however. If you want to change *every* occurrence, see Exercise 5. The rules for line numbers are the same as those for **p**, except that dot is set to the last line changed. (But there is a trap for the unwary: if no substitution took place, dot is *not* changed. This causes an error **?** as a warning.)

Thus you can say

1,$s/speling/spelling/

and correct the first spelling mistake on each line in the text. (This is useful for people who are consistent misspellers!)

If no line numbers are given, the **s** command assumes we mean "make the substitution on line dot", so it changes things only on the current line. This leads to the very common sequence

s/something/something else/p

which makes some correction on the current line, and then prints it, to make sure it worked out you typed in.

"Insert" is similar to append – for instance

```
/string/i
. . . type the lines to be inserted here . . .
```

will insert the given text *before* the next line that contains "string". The text between **i** and **.** is *inserted before* the specified line. If no line number is specified dot is used. Dot is set to the last line inserted.

Exercise 7:

"Change" is rather like a combination of delete followed by insert. Experiment to verify that

start, end d
i
. . . *text* . . .

is almost the same as

start, end c
. . . *text* . . .

These are not *precisely* the same if line **$** gets deleted. Check this out. What is dot?

Experiment with **a** and **i**, to see that they are similar, but not the same. You will observe that

line-number a
. . . *text* . . .

appends *after* the given line, while

line-number i
. . . *text* . . .

inserts *before* it. Observe that if no line number is given, **i** inserts before line dot, while **a** appends after line dot.

Moving text around: the "m" command

The move command **m** is used for cutting and pasting – it lets you move a group of lines from one place to another in the buffer. Suppose you want to put the first three lines of the buffer at the end instead. You could do it by saying:

1,3w temp
$r temp
1,3d

(Do you see why?) but you can do it a lot easier with the **m** command:

1,3m$

The general case is

start line, end line **m** *after this line*

Notice that there is a third line to be specified – the place where the moved stuff gets put. Of course the lines to be moved can be specified by context searches; if you had

First paragraph
. . .
end of first paragraph.
Second paragraph
. . .
end of second paragraph.

you could reverse the two paragraphs like this:

/Second/,/end of second/m/First/–1

Notice the **–1**: the moved text goes *after* the line mentioned. Dot gets set to the last line moved.

The global commands "g" and "v"

The *global* command **g** is used to execute one or more *ed* commands on all those lines in the buffer that match some specified string. For example

```
g/peling/p
```

prints all lines that contain **peling**. More usefully,

```
g/peling/s//pelling/gp
```

makes the substitution everywhere on the line, then prints each corrected line. Compare this to

```
1,$s/peling/pelling/gp
```

which only prints the last line substituted. Another subtle difference is that the **g** command does not give a **?** if **peling** is not found where the **s** command will.

There may be several commands (including **a**, **c**, **i**, **r**, **w**, but not **g**); in that case, every line except the last must end with a backslash \:

```
g/xxx/-1s/abc/def/B
+2s/ghi/jkl/B
-2,p
```

makes changes in the lines before and after each line that contains **xxx**, then prints all three lines.

The **v** command is the same as **g**, except that the commands are executed on every line that does *not* match the string following **v**:

```
v/ /d
```

deletes every line that does not contain a blank.

Special Characters

You may have noticed that things just don't work right when you used some characters like , *****, **$**, and others in context searches and the substi-

Suppose the buffer contains the three familiar lines

> **Now is the time**
> **for all good men**
> **to come to the aid of their party.**

Then the *ed* line numbers

> /Now/+1
> /good/
> /party/-1

are all context search expressions, and they all refer to the same line (line 2). To make a change in line 2, you could say

> /Now/+1s/good/bad/

or

> /good/s/good/bad/

or

> /party/-1s/good/bad/

The choice is dictated only by convenience. You could print all three lines by, for instance

> /Now/,/party/p

or

> /Now/,/Now/+2p

or by any number of similar combinations. The first one of these might be better if you don't know how many lines are involved. (Of course, if there were only three lines in the buffer, you'd use

> **1,$**p

but not if there were several hundred.)

The basic rule is: a context search expression is *the same as* a line number, so it can be used wherever a line number is needed.

Exercise 6:

Experiment with context searching. Try a body of text with several occurrences of the same string of characters, and scan through it using the same context search.

Try using context searches as line numbers for the substitute, print and delete commands. (They can also be used with **r**, **w**, and **a**.)

Try context searching using **?text?** instead of **/text/**. This scans lines in the buffer in reverse order rather than normal. This is sometimes useful if you go too far while looking for some string of characters – it's an easy way to back up.

(If you get funny results with any of the characters

> ^ **$** [* \ &

read the section on "Special Characters".)

Ed provides a shorthand for repeating a context search for the same string. For example, the *ed* line number

> **/string/**

will find the next occurrence of **string**. It often happens that this is not the desired line, so the search must be repeated. This can be done by typing merely

> //

This shorthand stands for "the most recently used context search expression." It can also be used as the first string of the substitute command, as in

> **/string1/s//string2/**

which will find the next occurrence of **string1** and replace it by **string2**. This can save a lot of typing. Similarly

> ??

means "scan backwards for the same expression."

Change and Insert – "c" and "i"

This section discusses the *change* command

> c

which is used to change or replace a group of one or more lines, and the *insert* command

> i

which is used for inserting a group of one or more lines.

"Change", written as

> c

is used to replace a number of lines with different lines, which are typed in at the terminal. For example, to change lines **.+1** through **$** to something else, type

> .+1,**$c**
> . . . *type the lines of text you want here* . . .

The lines you type between the **c** command and the **.** will take the place of the original lines between start line and end line. This is most useful in replacing a line or several lines which have errors in them.

If only one line is specified in the **c** command, then just that line is replaced. (You can type in as many replacement lines as you like.) Notice the use of **.** to end the input – this works just like the **.** in the append command and must appear by itself on a new line. If no line number is given, line dot is replaced. The value of dot is set to the last line to produce

the end of the world is at hand

Observe this expression carefully, for it illustrates how to take advantage of *ed* to save typing. The string **/world/** found the desired line; the shorthand **//** found the same word in the line; and the **&** saves you from typing it again.

The **&** is a special character only within the replacement text of a substitute command, and has no special meaning elsewhere. You can turn off the special meaning of **&** by preceding it with a ****:

s/ampersand/\&/

will convert the word "ampersand" into the literal symbol **&** in the current line.

Summary of Commands and Line Numbers

The general form of *ed* commands is the command name, perhaps preceded by one or two line numbers, and, in the case of **e**, **r**, and **w**, followed by a file name. Only one command is allowed per line, but a **p** command may follow any other command (except for **e**, **r**, **w** and **q**).

a: Append, that is, add lines to the buffer (at line dot, unless a different line is specified). Appending continues until

is typed on a new line. Dot is set to the last line appended.

c: Change the specified lines to the new text which follows. The new lines are terminated by a , as with **a**. If no lines are specified, replace line dot. Dot is set to last line changed.

d: Delete the lines specified. If none are specified, delete line dot. Dot is set to the first undeleted line, unless **$** is deleted, in which case dot is set to **$**.

e: Edit new file. Any previous contents of the buffer are thrown away, so issue a **w** beforehand.

f: Print remembered filename. If a name follows **f** the remembered name will be set to it.

g: The command

g/---/commands

will execute the commands on those lines that contain **---**, which can be any context search expression.

i: Insert lines before specified line (or dot) until a

is typed on a new line. Dot is set to last line inserted.

m: Move lines specified to after the line named after **m**. Dot is set to the last line moved.

p: Print specified lines. If none specified, print line dot. A single line number is equivalent to *line-number* **p**. A single return prints **.+1**, the next line.

q: Quit *ed*. Wipes out all text in buffer if you give it twice in a row without first giving a **w** command.

r: Read a file into buffer (at end unless specified elsewhere.) Dot set to last line read.

s: The command

s/string1/string2/

substitutes the characters **string1** into **string2** in the specified lines. If no lines are specified, make the substitution in line dot. Dot is set to last line in which a substitution took place, which means that if no substitution took place, dot is not changed. **s** changes only the first occurrence of **string1** on a line; to change all of them, type a **g** after the final slash.

v: The command

v/---/commands

executes **commands** on those lines that *do not* contain **---**.

w: Write out buffer onto a file. Dot is not changed.

.=: Print value of dot. (**=** by itself prints the value of **$**.)

!: The line

!command-line

causes **command-line** to be executed as a UNIX command.

/-----/: Context search. Search for next line which contains this string of characters. Print it. Dot is set to the line where string was found. Search starts at **.+1**, wraps around from **$** to 1, and continues to dot, if necessary.

?-----?: Context search in reverse direction. Start search at **.–1**, scan to **1**, wrap around to **$**.

tute command. The reason is rather complex, although the cure is simple. Basically, ***ed*** treats these characters as special, with special meanings. For instance, ***in a context search or the first string of the substitute command only,***

means "any character," not a period, so

/xy/

means "a line with an **x**, ***any character,*** and a **y**," ***not*** just "a line with an **x**, a period, and a **y**." A complete list of the special characters that can cause trouble is the following:

^ $ [* \

Warning: The backslash character \ is special to *ed*. For safety's sake, avoid it where possible. If you have to use one of the special characters in a substitute command, you can turn off its magic meaning temporarily by preceding it with the backslash. Thus

s/*/backslash dot star/

will change \.* into "backslash dot star".

Here is a hurried synopsis of the other special characters. First, the circumflex ^ signifies the beginning of a line. Thus

/^string/

finds **string** only if it is at the beginning of a line: it will find

string

but not

the string...

The dollar-sign **$** is just the opposite of the circumflex; it means the end of a line:

/string$/

will only find an occurrence of **string** that is at the end of some line. This implies, of course, that

/^string$/

will find only a line that contains just **string**, and

/^$/

finds a line containing exactly one character.

The character **.**, as we mentioned above, matches anything;

/xy/

matches any of

x+y
x–y
x y
xy

This is useful in conjunction with *, which is a repetition character; **a*** is a shorthand for "any number of **a**'s," so **.*** matches any number of anythings. This is used like this:

s/*/stuff/

which changes an entire line, or

s/*,//

which deletes all characters in the line up to and including the last comma. (Since **.*** finds the longest possible match, this goes up to the last comma.)

[is used with] to form "character classes"; for example,

/[0123456789]/

matches any single digit – any one of the characters inside the braces will cause a match. This can be abbreviated to **[0–9]**.

Finally, the & is another shorthand character – it is used only on the right-hand part of a substitute command where it means "whatever was matched on the left-hand side". It is used to save typing. Suppose the current line contained

Now is the time

and you wanted to put parentheses around it. You could just retype the line, but this is tedious. Or you could say

s/^/(/
s/$/)/

using your knowledge of ^ and **$**. But the easiest way uses the &:

s/*/(&)/

This says "match the whole line, and replace it by itself surrounded by parentheses." The & can be used several times in a line; consider using

s/*/&? &!!/

to produce

Now is the time? Now is the time!!

You don't have to match the whole line, of course: if the buffer contains

the end of the world

you could type

/world/s//& is at hand/

Contents

Edit: A Tutorial

Ricki Blau

James Joyce

Computing Services
University of California
Berkeley, California 94720

ABSTRACT

This narrative introduction to the use of the text editor ***edit*** assumes no prior familiarity with computers or with text editing. Its aim is to lead the beginning UNIX† user through the fundamental steps of writing and revising a file of text. Edit, a version of the text editor *ex,* was designed to provide an informative environment for new and casual users.

We welcome comments and suggestions about this tutorial and the UNIX documentation in general.

September 1981

†UNIX is a trademark of Bell Laboratories.

Session 1

Making contact with UNIX

To use the editor you must first make contact with the computer by logging in to UNIX. We'll quickly review the standard UNIX login procedure for the two ways you can make contact: on a terminal that is directly linked to the computer, or over a telephone line where the computer answers your call.

Directly-linked terminals

Turn on your terminal and press the RETURN key. You are now ready to login.

Dial-up terminals

If your terminal connects with the computer over a telephone line, turn on the terminal, dial the system access number, and, when you hear a high-pitched tone, place the receiver of the telephone in the acoustic coupler. You are now ready to login.

Logging in

The message inviting you to login is:

:login:

Type your login name, which identifies you to UNIX, on the same line as the login message, and press RETURN. If the terminal you are using has both upper and lower case, **be sure you enter your login name in lower case;** otherwise UNIX assumes your terminal has only upper case and will not recognize lower case letters you may type. UNIX types ":login:" and you reply with your login name, for example "susan":

:login: **susan** *(and press the RETURN key)*

(In the examples, input you would type appears in **bold face** to distinguish it from the responses from UNIX.)

UNIX will next respond with a request for a password as an additional precaution to prevent unauthorized people from using your account. The password will not appear when you type it, to prevent others from seeing it. The message is:

Password: *(type your password and press RETURN)*

If any of the information you gave during the login sequence was mistyped or incorrect, UNIX will respond with

Login incorrect.
:login:

in which case you should start the login process anew. Assuming that you have successfully logged in, UNIX will print the message of the day and eventually will present you with a % at the beginning of a fresh line. The % is the UNIX prompt symbol which tells you that UNIX is ready to accept a command.

Asking for *edit*

You are ready to tell UNIX that you want to work with edit, the text editor. Now is a convenient time to choose a name for the file of text you are about to create. To begin your editing session, type **edit** followed by a space and then the filename you have selected; for example, "text". When you have completed the command, press the RETURN key and wait for edit's response:

Introduction

Text editing using a terminal connected to a computer allows you to create, modify, and print text easily. A *text editor* is a program that assists you as you create and modify text. The text editor you will learn here is named *edit.* Creating text using edit is as easy as typing it on an electric typewriter. Modifying text involves telling the text editor what you want to add, change, or delete. You can review your text by typing a command to print the file contents as they were entered by you. Another program, a text formatter, rearranges your text for you into "finished form." This document does not discuss the use of a text formatter.

These lessons assume no prior familiarity with computers or with text editing. They consist of a series of text editing sessions which lead you through the fundamental steps of creating and revising text. After scanning each lesson and before beginning the next, you should practice the examples at a terminal to get a feeling for the actual process of text editing. If you set aside some time for experimentation, you will soon become familiar with using the computer to write and modify text. In addition to the actual use of the text editor, other features of UNIX will be very important to your work. You can begin to learn about these other features by reading "Communicating with UNIX" or one of the other tutorials that provide a general introduction to the system. You will be ready to proceed with this lesson as soon as you are familiar with (1) your terminal and its special keys, (2) the login procedure, (3) and the ways of correcting typing errors. Let's first define some terms:

program
: A set of instructions, given to the computer, describing the sequence of steps the computer performs in order to accomplish a specific task. The tasks must be specific, such as balancing your checkbook or editing your text. A general task, such as working for world peace, is something we can do, but not something we can write programs to do.

UNIX
: UNIX is a special type of program, called an operating system, that supervises the machinery and all other programs comprising the total computer system.

edit
: *edit* is the name of the UNIX text editor you will be learning to use, and is a program that aids you in writing or revising text. Edit was designed for beginning users, and is a simplified version of an editor named *ex.*

file
: Each UNIX account is allotted space for the permanent storage of information, such as programs, data or text. A file is a logical unit of data, for example, an essay, a program, or a chapter from a book, which is stored on a computer system. Once you create a file, it is kept until you instruct the system to remove it. You may create a file during one UNIX session, end the session, and return to use it at a later time. Files contain anything you choose to write and store in them. The sizes of files vary to suit your needs; one file might hold only a single number, yet another might contain a very long document or program. The only way to save information from one session to the next is to store it in a file, which you will learn in Session 1.

filename
: Filenames are used to distinguish one file from another, serving the same purpose as the labels of manila folders in a file cabinet. In order to write or access information in a file, you use the name of that file in a UNIX command, and the system will automatically locate the file.

disk
: Files are stored on an input/output device called a disk, which looks something like a stack of phonograph records. Each surface is coated with a material similar to the coating on magnetic recording tape, and information is recorded on it.

buffer
: A temporary work space, made available to the user for the duration of a session of text editing and used for creating and modifying the text file. We can think of the buffer as a blackboard that is erased after each class, where each session with the editor is a class.

> **:add**
> add: Not an editor command
> :

When you receive a diagnostic message, check what you typed in order to determine what part of your command confused edit. The message above means that edit was unable to recognize your mistyped command and, therefore, did not execute it. Instead, a new ":" appeared to let you know that edit is again ready to execute a command.

Text input mode

By giving the command "append" (or using the abbreviation "a"), you entered *text input mode,* also known as *append mode.* When you enter text input mode, edit stops sending you a prompt. You will not receive any prompts or error messages while in text input mode. You can enter pretty much anything you want on the lines. The lines are transmitted one by one to the buffer and held there during the editing session. You may append as much text as you want, and *when you wish to stop entering text lines you should type a period as the only character on the line and press the RETURN key.* When you type the period and press RETURN, you signal that you want to stop appending text, and edit responds by allowing you to exit text input mode and reenter command mode. Edit will again prompt you for a command by printing ":".

Leaving append mode does not destroy the text in the buffer. You have to leave append mode to do any of the other kinds of editing, such as changing, adding, or printing text. If you type a period as the first character and type any other character on the same line, edit will believe you want to remain in append mode and will not let you out. As this can be very frustrating, be sure to type **only** the period and the RETURN key.

This is a good place to learn an important lesson about computers and text: a blank space is a character as far as a computer is concerned. If you so much as type a period followed by a blank (that is, type a period and then the space bar on the keyboard), you will remain in append mode with the last line of text being:

> .

Let's say that the lines of text you enter are (try to type **exactly** what you see, including "thiss"):

> **This is some sample text.**
> **And thiss is some more text.**
> **Text editing is strange, but nice.**
> **.**

The last line is the period followed by a RETURN that gets you out of append mode.

Making corrections

If you have read a general introduction to UNIX, such as "Communicating with UNIX", you will recall that it is possible to erase individual letters that you have typed. This is done by typing the designated erase character as many times as there are characters you want to erase.

The usual erase character is the backspace (control-H), and you can correct typing errors in the line you are typing by holding down the CTRL key and typing the "H" key. If you try typing control-H you will notice that the terminal backspaces in the line you are on. You can backspace over your error, and then type what you want to be the rest of the line.

If you make a bad start in a line and would like to begin again, you can either backspace to the beginning of the line or you can use the at-sign "@" to erase everything on the line:

> **Text edtiing is strange, but@**
> **Text editing is strange, but nice.**

When you type the at-sign (@), you erase the entire line typed so far and are given a fresh line to type on. You may immediately begin to retype the line. This, unfortunately, does not help after

% **edit text** ***(followed by a RETURN)***
"text" No such file or directory
:

If you typed the command correctly, you will now be in communication with edit. Edit has set aside a buffer for use as a temporary working space during your current editing session. It also checked to see if the file you named, "text", already existed. It was unable to find such a file, since "text" is a new file we are about to create. Edit confirms this with the line:

"text" No such file or directory

On the next line appears edit's prompt ":", announcing that you are in *command mode* and edit expects a command from you. You may now begin to create the new file.

The "Command not found" message

If you misspelled edit by typing, say, "editor", your request would be handled as follows:

% **editor**
editor: Command not found
%

Your mistake in calling edit "editor" was treated by UNIX as a request for a program named "editor". Since there is no program named "editor", UNIX reported that the program was "not found". A new % indicates that UNIX is ready for another command, and you may then enter the correct command.

A summary

Your exchange with UNIX as you logged in and made contact with edit should look something like this:

:login: **susan**
Password:
... A Message of General Interest ...
% **edit text**
"text" No such file or directory
:

Entering text

You may now begin entering text into the buffer. This is done by *appending* (or adding) text to whatever is currently in the buffer. Since there is nothing in the buffer at the moment, you are appending text to nothing; in effect, since you are adding text to nothing you are creating text. Most edit commands have two forms: a word that suggests what the command does, and a shorter abbreviation of that word. Either form may be used. Many beginners find the full command names easier to remember at first, but once you are familiar with editing you may prefer to type the shorter abbreviations. The command to input text is "append", and it may be abbreviated "a". Type **append** and press the RETURN key.

% **edit text**
: **append**

Messages from *edit*

If you make a mistake in entering a command and type something that edit does not recognize, edit will respond with a message intended to help you diagnose your error. For example, if you misspell the command to input text by typing, perhaps, "add" instead of "append" or "a", you will receive this message:

Session 2

Login with UNIX as in the first session:

:login: **susan** *(carriage return)*
Password: *(give password and carriage return)*
... A Message of General Interest ...
%

When you indicate you want to edit, you can specify the name of the file you worked on last time. This will start edit working, and it will fetch the contents of the file into the buffer, so that you can resume editing the same file. When edit has copied the file into the buffer, it will repeat its name and tell you the number of lines and characters it contains. Thus,

% edit text
"text" 3 lines, 90 characters
:

means you asked edit to fetch the file named "text" for editing, causing it to copy the 90 characters of text into the buffer. Edit awaits your further instructions, and indicates this by its prompt character, the colon (:). In this session, we will append more text to our file, print the contents of the buffer, and learn to change the text of a line.

Adding more text to the file

If you want to add more to the end of your text you may do so by using the append command to enter text input mode. When "append" is the first command of your editing session, the lines you enter are placed at the end of the buffer. Here we'll use the abbreviation for the append command, "a":

:**a**
This is text added in Session 2.
It doesn't mean much here, but
it does illustrate the editor.
.

You may recall that once you enter append mode using the "a" (or "append") command, you need to type a line containing only a period (.) to exit append mode.

Interrupt

Should you press the RUB key (sometimes labelled DELETE) while working with edit, it will send this message to you:

Interrupt
:

Any command that edit might be executing is terminated by rub or delete, causing edit to prompt you for a new command. If you are appending text at the time, you will exit from append mode and be expected to give another command. The line of text you were typing when the append command was interrupted will not be entered into the buffer.

Making corrections

If while typing the line you hit an incorrect key, recall that you may delete the incorrect character or cancel the entire line of input by erasing in the usual way. Refer either to the last few pages of Session 1 or to "Communicating with UNIX" if you need to review the procedures for making a correction. The most important idea to remember is that erasing a character or cancelling a line must be done before you press the RETURN key.

you type the line and press RETURN. To make corrections in lines that have been completed, it is necessary to use the editing commands covered in the next session and those that follow.

Writing text to disk

You are now ready to edit the text. The simplest kind of editing is to write it to disk as a file for safekeeping after the session is over. This is the only way to save information from one session to the next, since the editor's buffer is temporary and will last only until the end of the editing session. Learning how to write a file to disk is second in importance only to entering the text. To write the contents of the buffer to a disk file, use the command "write" (or its abbreviation "w"):

    : **write**

Edit will copy the contents of the buffer to a disk file. If the file does not yet exist, a new file will be created automatically and the presence of a "[New file]" will be noted. The newly-created file will be given the name specified when you entered the editor, in this case "text". To confirm that the disk file has been successfully written, edit will repeat the filename and give the number of lines and the total number of characters in the file. The buffer remains unchanged by the "write" command. All of the lines that were written to disk will still be in the buffer, should you want to modify or add to them.

Edit must have a filename to use before it can write a file. If you forgot to indicate the name of the file when you began the editing session, edit will print

    No current filename

in response to your write command. If this happens, you can specify the filename in a new write command:

    : **write text**

After the "write" (or "w"), type a space and then the name of the file.

Signing off

We have done enough for this first lesson on using the UNIX text editor, and are ready to quit the session with edit. To do this we type "quit" (or "q") and press RETURN:

    : **write**
    "text" [New file] 3 lines, 90 characters
    : **quit**
    %

The % is from UNIX to tell you that your session with edit is over and you may command UNIX further. Since we want to end the entire session at the terminal, we also need to exit from UNIX. In response to the UNIX prompt of " % " type the command

    % **logout**

This will end your session with UNIX, and will ready the terminal for the next user. It is always important to type **logout** at the end of a session to make absolutely sure no one could accidentally stumble into your abandoned session and thus gain access to your files, tempting even the most honest of souls.

This is the end of the first session on UNIX text editing.

The current line

Edit keeps track of the line in the buffer where it is located at all times during an editing session. In general, the line that has been most recently printed, entered, or changed is the current location in the buffer. The editor is prepared to make changes at the current location in the buffer, unless you direct it to another location.

In particular, when you bring a file into the buffer, you will be located at the last line in the file, where the editor left off copying the lines from the file to the buffer. If your first editing command is "append", the lines you enter are added to the end of the file, after the current line — the last line in the file.

You can refer to your current location in the buffer by the symbol period (.) usually known by the name "dot". If you type "." and carriage return you will be instructing edit to print the current line:

```
:.
And thiss is some more text.
```

If you want to know the number of the current line, you can type **.=** and press RETURN, and edit will respond with the line number:

```
:.=
2
```

If you type the number of any line and press RETURN, edit will position you at that line and print its contents:

```
:2
And thiss is some more text.
```

You should experiment with these commands to gain experience in using them to make changes.

Numbering lines (nu)

The **number (nu)** command is similar to print, giving both the number and the text of each printed line. To see the number and the text of the current line type

```
:nu
         2  And thiss is some more text.
```

Note that the shortest abbreviation for the number command is "nu" (and not "n", which is used for a different command). You may specify a range of lines to be listed by the number command in the same way that lines are specified for print. For example, **1,$nu** lists all lines in the buffer with their corresponding line numbers.

Substitute command (s)

Now that you have found the misspelled word, you can change it from "thiss" to "this". As far as edit is concerned, changing things is a matter of substituting one thing for another. As *a* stood for ***append,*** so *s* stands for ***substitute***. We will use the abbreviation "s" to reduce the chance of mistyping the substitute command. This command will instruct edit to make the change:

2s/thiss/this/

We first indicate the line to be changed, line 2, and then type an "s" to indicate we want edit to make a substitution. Inside the first set of slashes are the characters that we want to change, followed by the characters to replace them, and then a closing slash mark. To summarize:

2s/ *what is to be changed* / *what to change it to* /

If edit finds an exact match of the characters to be changed it will make the change **only** in the first occurrence of the characters. If it does not find the characters to be changed, it will respond:

Listing what's in the buffer (p)

Having appended text to what you wrote in Session 1, you might want to see all the lines in the buffer. To print the contents of the buffer, type the command:

: **1,$p**

The "1"† stands for line 1 of the buffer, the "**$**" is a special symbol designating the last line of the buffer, and "p" (or **print**) is the command to print from line 1 to the end of the buffer. The command "1,$p" gives you:

This is some sample text.
And thiss is some more text.
Text editing is strange, but nice.
This is text added in Session 2.
It doesn't mean much here, but
it does illustrate the editor.

Occasionally, you may accidentally type a character that can't be printed, which can be done by striking a key while the CTRL key is pressed. In printing lines, edit uses a special notation to show the existence of non-printing characters. Suppose you had introduced the non-printing character "control-A" into the word "illustrate" by accidently pressing the CTRL key while typing "a". This can happen on many terminals because the CTRL key and the "A" key are beside each other. If your finger presses between the two keys, control-A results. When asked to print the contents of the buffer, edit would display

it does illustr^Ate the editor.

To represent the control-A, edit shows "^A". The sequence "^" followed by a capital letter stands for the one character entered by holding down the CTRL key and typing the letter which appears after the "^". We'll soon discuss the commands that can be used to correct this typing error.

In looking over the text we see that "this" is typed as "thiss" in the second line, a deliberate error so we can learn to make corrections. Let's correct the spelling.

Finding things in the buffer

In order to change something in the buffer we first need to find it. We can find "thiss" in the text we have entered by looking at a listing of the lines. Physically speaking, we search the lines of text looking for "thiss" and stop searching when we have found it. The way to tell edit to search for something is to type it inside slash marks:

: **/thiss/**

By typing **/thiss/** and pressing RETURN, you instruct edit to search for "thiss". If you ask edit to look for a pattern of characters which it cannot find in the buffer, it will respond "Pattern not found". When edit finds the characters "thiss", it will print the line of text for your inspection:

And thiss is some more text.

Edit is now positioned in the buffer at the line it just printed, ready to make a change in the line.

†The numeral "one" is the top left-most key, and should not be confused with the letter "el".

:z

If no starting line number is given for the **z** command, printing will start at the "current" line, in this case the last line printed. Viewing lines in the buffer one screen full at a time is known as *paging*. Paging can also be used to print a section of text on a hard-copy terminal.

Saving the modified text

This seems to be a good place to pause in our work, and so we should end the second session. If you (in haste) type "q" to quit the session your dialogue with edit will be:

:**q**
No write since last change (:quit! overrides)
:

This is edit's warning that you have not written the modified contents of the buffer to disk. You run the risk of losing the work you did during the editing session since you typed the latest write command. Because in this lesson we have not written to disk at all, everything we have done would have been lost if edit had obeyed the **q** command. If you did not want to save the work done during this editing session, you would have to type "q!" or ("quit!") to confirm that you indeed wanted to end the session immediately, leaving the file as it was after the most recent "write" command. However, since you want to save what you have edited, you need to type:

:**w**
"text" 6 lines, 171 characters

and then follow with the commands to quit and logout:

:**q**
% logout

and hang up the phone or turn off the terminal when UNIX asks for a name. Terminals connected to the port selector will stop after the logout command, and pressing keys on the keyboard will do nothing.

This is the end of the second session on UNIX text editing.

Substitute pattern match failed

indicating that your instructions could not be carried out. When edit does find the characters that you want to change, it will make the substitution and automatically print the changed line, so that you can check that the correct substitution was made. In the example,

: **2s/thiss/this/**
And this is some more text.

line 2 (and line 2 only) will be searched for the characters "thiss", and when the first exact match is found, "thiss" will be changed to "this". Strictly speaking, it was not necessary above to specify the number of the line to be changed. In

: **s/thiss/this/**

edit will assume that we mean to change the line where we are currently located ("."). In this case, the command without a line number would have produced the same result because we were already located at the line we wished to change.

For another illustration of the substitute command, let us choose the line:

Text editing is strange, but nice.

You can make this line a bit more positive by taking out the characters "strange, but " so the line reads:

Text editing is nice.

A command that will first position edit at the desired line and then make the substitution is:

: **/strange/s/strange, but //**

What we have done here is combine our search with our substitution. Such combinations are perfectly legal, and speed up editing quite a bit once you get used to them. That is, you do not necessarily have to use line numbers to identify a line to edit. Instead, you may identify the line you want to change by asking edit to search for a specified pattern of letters that occurs in that line. The parts of the above command are:

/strange/	tells edit to find the characters "strange" in the text
s	tells edit to make a substitution
/strange, but //	substitutes nothing at all for the characters "strange, but "

You should note the space after "but" in "/strange, but /". If you do not indicate that the space is to be taken out, your line will read:

Text editing is nice.

which looks a little funny because of the extra space between "is" and "nice". Again, we realize from this that a blank space is a real character to a computer, and in editing text we need to be aware of spaces within a line just as we would be aware of an "a" or a "4".

Another way to list what's in the buffer (z)

Although the print command is useful for looking at specific lines in the buffer, other commands may be more convenient for viewing large sections of text. You can ask to see a screen full of text at a time by using the command **z**. If you type

: **1z**

edit will start with line 1 and continue printing lines, stopping either when the screen of your terminal is full or when the last line in the buffer has been printed. If you want to read the next segment of text, type the command

This is some sample text.
It doesn't mean much here, but
it does illustrate the editor.
And this is some more text.
Text editing is nice.
This is text added in Session 2.

You can restore the original order by typing:

:**4,$m1**

or, combining context searching and the move command:

:**/And this is some/,/This is text/m/This is some sample/**

(Do not type both examples here!) The problem with combining context searching with the move command is that your chance of making a typing error in such a long command is greater than if you type line numbers.

Copying lines (copy)

The **copy** command is used to make a second copy of specified lines, leaving the original lines where they were. Copy has the same format as the move command, for example:

:**2,5copy $**

makes a copy of lines 2 through 5, placing the added lines after the buffer's end ($). Experiment with the copy command so that you can become familiar with how it works. Note that the shortest abbreviation for copy is **co** (and not the letter "c", which has another meaning).

Deleting lines (d)

Suppose you want to delete the line

This is text added in Session 2.

from the buffer. If you know the number of the line to be deleted, you can type that number followed by **delete** or **d**. This example deletes line 4, which is "This is text added in Session 2." if you typed the commands suggested so far.

:**4d**
It doesn't mean much here, but

Here "4" is the number of the line to be deleted, and "delete" or "d" is the command to delete the line. After executing the delete command, edit prints the line that has become the current line (".").

If you do not happen to know the line number you can search for the line and then delete it using this sequence of commands:

:**/added in Session 2./**
This is text added in Session 2.
:**d**
It doesn't mean much here, but

The "/added in Session 2./" asks edit to locate and print the line containing the indicated text, starting its search at the current line and moving line by line until it finds the text. Once you are sure that you have correctly specified the line you want to delete, you can enter the delete (d) command. In this case it is not necessary to specify a line number before the "d". If no line number is given, edit deletes the current line ("."), that is, the line found by our search. After the deletion, your buffer should contain:

Session 3

Bringing text into the buffer (e)

Login to UNIX and make contact with edit. You should try to login without looking at the notes, but if you must then by all means do.

Did you remember to give the name of the file you wanted to edit? That is, did you type

% edit text

or simply

% edit

Both ways get you in contact with edit, but the first way will bring a copy of the file named "text" into the buffer. If you did forget to tell edit the name of your file, you can get it into the buffer by typing:

:**e text**
"text" 6 lines, 171 characters

The command **edit,** which may be abbreviated **e**, tells edit that you want to erase anything that might already be in the buffer and bring a copy of the file "text" into the buffer for editing. You may also use the edit (e) command to change files in the middle of an editing session, or to give edit the name of a new file that you want to create. Because the edit command clears the buffer, you will receive a warning if you try to edit a new file without having saved a copy of the old file. This gives you a chance to write the contents of the buffer to disk before editing the next file.

Moving text in the buffer (m)

Edit allows you to move lines of text from one location in the buffer to another by means of the **move** (**m**) command. The first two examples are for illustration only, though after you have read this Session you are welcome to return to them for practice. The command

:**2,4m$**

directs edit to move lines 2, 3, and 4 to the end of the buffer ($). The format for the move command is that you specify the first line to be moved, the last line to be moved, the move command "m", and the line after which the moved text is to be placed. So,

:**1,3m6**

would instruct edit to move lines 1 through 3 (inclusive) to a location after line 6 in the buffer. To move only one line, say, line 4, to a location in the buffer after line 5, the command would be "4m5".

Let's move some text using the command:

:**5,$m1**
2 lines moved
it does illustrate the editor.

After executing a command that moves more than one line of the buffer, edit tells how many lines were affected by the move and prints the last moved line for your inspection. If you want to see more than just the last line, you can then use the print (p), z, or number (nu) command to view more text. The buffer should now contain:

```
:u
2 more lines in file after undo
And this is some more text.
```

Here again, edit informs you if the command affects more than one line, and prints the text of the line which is now "dot" (the current line).

More about the dot (.) and buffer end ($)

The function assumed by the symbol dot depends on its context. It can be used:

1. to exit from append mode; we type dot (and only a dot) on a line and press RETURN;
2. to refer to the line we are at in the buffer.

Dot can also be combined with the equal sign to get the number of the line currently being edited:

```
:.=
```

If we type ".=" we are asking for the number of the line, and if we type "." we are asking for the text of the line.

In this editing session and the last, we used the dollar sign to indicate the end of the buffer in commands such as print, copy, and move. The dollar sign as a command asks edit to print the last line in the buffer. If the dollar sign is combined with the equal sign ($=) edit will print the line number corresponding to the last line in the buffer.

"." and "$", then, represent line numbers. Whenever appropriate, these symbols can be used in place of line numbers in commands. For example

```
:.,$d
```

instructs edit to delete all lines from the current line (.) to the end of the buffer.

Moving around in the buffer (+ and –)

When you are editing you often want to go back and re-read a previous line. You could specify a context search for a line you want to read if you remember some of its text, but if you simply want to see what was written a few, say 3, lines ago, you can type

```
-3p
```

This tells edit to move back to a position 3 lines before the current line (.) and print that line. You can move forward in the buffer similarly:

```
+2p
```

instructs edit to print the line that is 2 ahead of your current position.

You may use "+" and "–" in any command where edit accepts line numbers. Line numbers specified with "+" or "–" can be combined to print a range of lines. The command

```
:-1,+2copy$
```

makes a copy of 4 lines: the current line, the line before it, and the two after it. The copied lines will be placed after the last line in the buffer ($), and the original lines referred to by "–1" and "+2" remain where they are.

Try typing only "–"; you will move back one line just as if you had typed "–1p". Typing the command "+" works similarly. You might also try typing a few plus or minus signs in a row (such as "+++") to see edit's response. Typing RETURN alone on a line is the equivalent of typing "+1p"; it will move you one line ahead in the buffer and print that line.

If you are at the last line of the buffer and try to move further ahead, perhaps by typing a "+" or a carriage return alone on the line, edit will remind you that you are at the end of the buffer:

> This is some sample text.
> And this is some more text.
> Text editing is nice.
> It doesn't mean much here, but
> it does illustrate the editor.
> And this is some more text.
> Text editing is nice.
> This is text added in Session 2.
> It doesn't mean much here, but

To delete both lines 2 and 3:

> And this is some more text.
> Text editing is nice.

you type

> :**2,3d**
> 2 lines deleted

which specifies the range of lines from 2 to 3, and the operation on those lines — "d" for delete. If you delete more than one line you will receive a message telling you the number of lines deleted, as indicated in the example above.

The previous example assumes that you know the line numbers for the lines to be deleted. If you do not you might combine the search command with the delete command:

> :**/And this is some/,/Text editing is nice./d**

A word or two of caution

In using the search function to locate lines to be deleted you should be **absolutely sure** the characters you give as the basis for the search will take edit to the line you want deleted. Edit will search for the first occurrence of the characters starting from where you last edited – that is, from the line you see printed if you type dot (.).

A search based on too few characters may result in the wrong lines being deleted, which edit will do as easily as if you had meant it. For this reason, it is usually safer to specify the search and then delete in two separate steps, at least until you become familiar enough with using the editor that you understand how best to specify searches. For a beginner it is not a bad idea to double-check each command before pressing RETURN to send the command on its way.

Undo (u) to the rescue

The **undo (u)** command has the ability to reverse the effects of the last command that changed the buffer. To undo the previous command, type "u" or "undo". Undo can rescue the contents of the buffer from many an unfortunate mistake. However, its powers are not unlimited, so it is still wise to be reasonably careful about the commands you give.

It is possible to undo only commands which have the power to change the buffer — for example, delete, append, move, copy, substitute, and even undo itself. The commands write (w) and edit (e), which interact with disk files, cannot be undone, nor can commands that do not change the buffer, such as print. Most importantly, the **only** command that can be reversed by undo is the last "undo-able" command you typed. You can use control-H and @ to change commands while you are typing them, and undo to reverse the effect of the commands after you have typed them and pressed RETURN.

To illustrate, let's issue an undo command. Recall that the last buffer-changing command we gave deleted the lines formerly numbered 2 and 3. Typing undo at this moment will reverse the effects of the deletion, causing those two lines to be replaced in the buffer.

Session 4

This lesson covers several topics, starting with commands that apply throughout the buffer, characters with special meanings, and how to issue UNIX commands while in the editor. The next topics deal with files: more on reading and writing, and methods of recovering files lost in a crash. The final section suggests sources of further information.

Making commands global (g)

One disadvantage to the commands we have used for searching or substituting is that if you have a number of instances of a word to change it appears that you have to type the command repeatedly, once for each time the change needs to be made. Edit, however, provides a way to make commands apply to the entire contents of the buffer – the **global (g)** command.

To print all lines containing a certain sequence of characters (say, "text") the command is:

:**g/text/p**

The "g" instructs edit to make a global search for all lines in the buffer containing the characters "text". The "p" prints the lines found.

To issue a global command, start by typing a "g" and then a search pattern identifying the lines to be affected. Then, on the same line, type the command to be executed for the identified lines. Global substitutions are frequently useful. For example, to change all instances of the word "text" to the word "material" the command would be a combination of the global search and the substitute command:

:**g/text/s/text/material/g**

Note the "g" at the end of the global command, which instructs edit to change each and every instance of "text" to "material". If you do not type the "g" at the end of the command only the *first* instance of "text" *in each line* will be changed (the normal result of the substitute command). The "g" at the end of the command is independent of the "g" at the beginning. You may give a command such as:

:**5s/text/material/g**

to change every instance of "text" in line 5 alone. Further, neither command will change "text" to "material" if "Text" begins with a capital rather than a lower-case *t*.

Edit does not automatically print the lines modified by a global command. If you want the lines to be printed, type a "p" at the end of the global command:

:**g/text/s/text/material/gp**

You should be careful about using the global command in combination with any other – in essence, be sure of what you are telling edit to do to the entire buffer. For example,

:**g/ /d**
72 less lines in file after global

will delete every line containing a blank anywhere in it. This could adversely affect your document, since most lines have spaces between words and thus would be deleted. After executing the global command, edit will print a warning if the command added or deleted more than one line. Fortunately, the undo command can reverse the effects of a global command. You should experiment with the global command on a small file of text to see what it can do for you.

More about searching and substituting

In using slashes to identify a character string that we want to search for or change, we have always specified the exact characters. There is a less tedious way to repeat the same string of characters. To change "text" to "texts" we may type either

At end-of-file

or

Not that many lines in buffer

Similarly, if you try to move to a position before the first line, edit will print one of these messages:

Nonzero address required on this command

or

Negative address – first buffer line is 1

The number associated with a buffer line is the line's "address", in that it can be used to locate the line.

Changing lines (c)

You can also delete certain lines and insert new text in their place. This can be accomplished easily with the **change (c)** command. The change command instructs edit to delete specified lines and then switch to text input mode to accept the text that will replace them. Let's say you want to change the first two lines in the buffer:

This is some sample text.
And this is some more text.

to read

This text was created with the UNIX text editor.

To do so, you type:

: **1,2c**
2 lines changed
This text was created with the UNIX text editor.
.
:

In the command **1,2c** we specify that we want to change the range of lines beginning with 1 and ending with 2 by giving line numbers as with the print command. These lines will be deleted. After you type RETURN to end the change command, edit notifies you if more than one line will be changed and places you in text input mode. Any text typed on the following lines will be inserted into the position where lines were deleted by the change command. **You will remain in text input mode until you exit in the usual way, by typing a period alone on a line.** Note that the number of lines added to the buffer need not be the same as the number of lines deleted.

This is the end of the third session on text editing with UNIX.

:s/\$/dollar/

looks for the character "$" in the current line and replaces it by the word "dollar". Were it not for the backslash, the "$" would have represented "the end of the line" in your search rather than the character "$". The backslash retains its special significance unless it is preceded by another backslash.

Issuing UNIX commands from the editor

After creating several files with the editor, you may want to delete files no longer useful to you or ask for a list of your files. Removing and listing files are not functions of the editor, and so they require the use of UNIX system commands (also referred to as "shell" commands, as "shell" is the name of the program that processes UNIX commands). You do not need to quit the editor to execute a UNIX command as long as you indicate that it is to be sent to the shell for execution. To use the UNIX command **rm** to remove the file named "junk" type:

:**!rm junk**
!
:

The exclamation mark (!) indicates that the rest of the line is to be processed as a shell command. If the buffer contents have not been written since the last change, a warning will be printed before the command is executed:

[No write since last change]

The editor prints a "!" when the command is completed. The tutorial "Communicating with UNIX" describes useful features of the system, of which the editor is only one part.

Filenames and file manipulation

Throughout each editing session, edit keeps track of the name of the file being edited as the *current filename*. Edit remembers as the current filename the name given when you entered the editor. The current filename changes whenever the edit (e) command is used to specify a new file. Once edit has recorded a current filename, it inserts that name into any command where a filename has been omitted. If a write command does not specify a file, edit, as we have seen, supplies the current filename. If you are editing a file named "draft3" having 283 lines in it, you can have the editor write onto a different file by including its name in the write command:

:**w chapter3**
"chapter3" [new file] 283 lines, 8698 characters

The current filename remembered by the editor *will not be changed as a result of the write command.* Thus, if the next write command does not specify a name, edit will write onto the current file ("draft3") and not onto the file "chapter3".

The file (f) command

To ask for the current filename, type **file** (or **f**). In response, the editor provides current information about the buffer, including the filename, your current position, the number of lines in the buffer, and the percent of the distance through the file your current location is.

:**f**
"text" [Modified] line 3 of 4 --75%--

If the contents of the buffer have changed since the last time the file was written, the editor will tell you that the file has been "[Modified]". After you save the changes by writing onto a disk file, the buffer will no longer be considered modified:

: /text/s/text/texts/

as we have done in the past, or a somewhat abbreviated command:

: /text/s//texts/

In this example, the characters to be changed are not specified – there are no characters, not even a space, between the two slash marks that indicate what is to be changed. This lack of characters between the slashes is taken by the editor to mean "use the characters we last searched for as the characters to be changed."

Similarly, the last context search may be repeated by typing a pair of slashes with nothing between them:

: /does/
It doesn't mean much here, but
: //
it does illustrate the editor.

(You should note that the search command found the characters "does" in the word "doesn't" in the first search request.) Because no characters are specified for the second search, the editor scans the buffer for the next occurrence of the characters "does".

Edit normally searches forward through the buffer, wrapping around from the end of the buffer to the beginning, until the specified character string is found. If you want to search in the reverse direction, use question marks (?) instead of slashes to surround the characters you are searching for.

It is also possible to repeat the last substitution without having to retype the entire command. An ampersand (&) used as a command repeats the most recent substitute command, using the same search and replacement patterns. After altering the current line by typing

: s/text/texts/

you type

: /text/&

or simply

: //&

to make the same change on the next line in the buffer containing the characters "text".

Special characters

Two characters have special meanings when used in specifying searches: "$" and "^". "$" is taken by the editor to mean "end of the line" and is used to identify strings that occur at the end of a line.

: g/text.$/s//material./p

tells the editor to search for all lines ending in "text." (and nothing else, not even a blank space), to change each final "text." to "material.", and print the changed lines.

The symbol "^" indicates the beginning of a line. Thus,

: s/^/1. /

instructs the editor to insert "1." and a space at the beginning of the current line.

The characters "$" and "^" have special meanings only in the context of searching. At other times, they are ordinary characters. If you ever need to search for a character that has a special meaning, you must indicate that the character is to lose temporarily its special significance by typing another special character, the backslash (\), before it.

If this is not possible and you cannot find someone to help you, enter the command

:preserve

and wait for the reply,

File preserved.

If you do not receive this reply, seek help immediately. Do not simply leave the editor. If you do, the buffer will be lost, and you may not be able to save your file. If the reply is "File preserved." you can leave the editor (or logout) to remedy the situation. After a preserve, you can use the recover command once the problem has been corrected, or the **–r** option of the edit command if you leave the editor and want to return.

If you make an undesirable change to the buffer and type a write command before discovering your mistake, the modified version will replace any previous version of the file. Should you ever lose a good version of a document in this way, do not panic and leave the editor. As long as you stay in the editor, the contents of the buffer remain accessible. Depending on the nature of the problem, it may be possible to restore the buffer to a more complete state with the undo command. After fixing the damaged buffer, you can again write the file to disk.

Further reading and other information

Edit is an editor designed for beginning and casual users. It is actually a version of a more powerful editor called *ex*. These lessons are intended to introduce you to the editor and its more commonly-used commands. We have not covered all of the editor's commands, but a selection of commands that should be sufficient to accomplish most of your editing tasks. You can find out more about the editor in the *Ex Reference Manual,* which is applicable to both *ex* and *edit.* The manual is available from the Computing Services Library, 218 Evans Hall. One way to become familiar with the manual is to begin by reading the description of commands that you already know.

Using *ex*

As you become more experienced with using the editor, you may still find that edit continues to meet your needs. However, should you become interested in using ex, it is easy to switch. To begin an editing session with ex, use the name **ex** in your command instead of **edit.**

Edit commands work the same way in ex, but the editing environment is somewhat different. You should be aware of a few differences that exist between the two versions of the editor. In edit, only the characters "^", "$", and "\" have special meanings in searching the buffer or indicating characters to be changed by a substitute command. Several additional characters have special meanings in ex, as described in the *Ex Reference Manual.* Another feature of the edit environment prevents users from accidently entering two alternative modes of editing, *open* and *visual,* in which the editor behaves quite differently from normal command mode. If you are using ex and the editor behaves strangely, you may have accidently entered open mode by typing "o". Type the ESC key and then a "Q" to get out of open or visual mode and back into the regular editor command mode. The document *An Introduction to Display Editing with Vi* provides a full discussion of visual mode.

: **w**
"text" 4 lines, 88 characters
: **f**
"text" line 3 of 4 --75%--

Reading additional files (r)

The **read (r)** command allows you to add the contents of a file to the buffer at a specified location, essentially copying new lines between two existing lines. To use it, specify the line after which the new text will be placed, the **read (r)** command, and then the name of the file. If you have a file named "example", the command

: **$r example**
"example" 18 lines, 473 characters

reads the file "example" and adds it to the buffer after the last line. The current filename is not changed by the read command.

Writing parts of the buffer

The **write (w)** command can write all or part of the buffer to a file you specify. We are already familiar with writing the entire contents of the buffer to a disk file. To write only part of the buffer onto a file, indicate the beginning and ending lines before the write command, for example

: **45,$w ending**

Here all lines from 45 through the end of the buffer are written onto the file named *ending*. The lines remain in the buffer as part of the document you are editing, and you may continue to edit the entire buffer. Your original file is unaffected by your command to write part of the buffer to another file. Edit still remembers whether you have saved changes to the buffer in your original file or not.

Recovering files

Although it does not happen very often, there are times UNIX stops working because of some malfunction. This situation is known as a *crash*. Under most circumstances, edit's crash recovery feature is able to save work to within a few lines of changes before a crash (or an accidental phone hang up). If you lose the contents of an editing buffer in a system crash, you will normally receive mail when you login that gives the name of the recovered file. To recover the file, enter the editor and type the command **recover (rec)**, followed by the name of the lost file. For example, to recover the buffer for an edit session involving the file "chap6", the command is:

: **recover chap6**

Recover is sometimes unable to save the entire buffer successfully, so always check the contents of the saved buffer carefully before writing it back onto the original file. For best results, write the buffer to a new file temporarily so you can examine it without risk to the original file. Unfortunately, you cannot use the recover command to retrieve a file you removed using the shell command **rm**.

Other recovery techniques

If something goes wrong when you are using the editor, it may be possible to save your work by using the command **preserve (pre)**, which saves the buffer as if the system had crashed. If you are writing a file and you get the message "Quota exceeded", you have tried to use more disk storage than is allotted to your account. ***Proceed with caution*** because it is likely that only a part of the editor's buffer is now present in the file you tried to write. In this case you should use the shell escape from the editor (!) to remove some files you don't need and try to write the file again.

Index

3. Basics

To run **vi** the shell variable **TERM** must be defined and exported to your environment. How you do this depends on which shell you are using. You can tell which shell you have by the character it prompts you for commands with. The Bourne shell prompts with '$', and the C shell prompts with '%'. For these examples, we will suppose that you are using an HP 2621 terminal, whose termcap name is "2621".

3.1. Bourne Shell

To manually set your terminal type to 2621 you would type:

```
TERM=2621
export TERM
```

There are various ways of having this automatically or semi-automatically done when you log in. Suppose you usually dial in on a 2621. You want to tell this to the machine, but still have it work when you use a hardwired terminal. The recommended way, if you have the **tset** program, is to use the sequence

```
tset -s -d 2621 > tset$$
. tset$$
rm tset$$
```

in your .login (for csh) or the same thing using '.' instead of 'source' in your .profile (for sh). The above line says that if you are dialing in you are on a 2621, but if you are on a hardwired terminal it figures out your terminal type from an on-line list.

3.2. The C Shell

To manually set your terminal type to 2621 you would type:

```
setenv TERM 2621
```

There are various ways of having this automatically or semi-automatically done when you log in. Suppose you usually dial in on a 2621. You want to tell this to the machine, but still have it work when you use a hardwired terminal. The recommended way, if you have the **tset** program, is to use the sequence

```
tset -s -d 2621 > tset$$
source tset$$
rm tset$$
```

in your .login.* The above line says that if you are dialing in you are on a 2621, but if you are on a hardwired terminal it figures out your terminal type from an on-line list.

4. Normal Commands

Vi is a visual editor with a window on the file. What you see on the screen is **vi**'s current notion of what your file will contain, (at this point in the file), when it is written out. Most commands do not cause any change in the screen until the complete command is typed. Should you get confused while typing a command, you can abort the command by typing an <del> character. You will know you are back to command level when you hear a <bell>. Usually typing an <esc> will produce the same result. When **vi** gets an improperly formatted command it rings the <bell>. Following are the **vi** commands broken down by function.

* On a version 6 system without environments, the invocation of tset is simpler, just add the line "tset -d 2621" to your .login or .profile.

Ex Reference Manual
Version 3.5/2.13 – September, 1980

William Joy

Revised for versions 3.5/2.13 by
Mark Horton

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, Ca. 94720

ABSTRACT

Ex a line oriented text editor, which supports both command and display oriented editing. This reference manual describes the command oriented part of *ex;* the display editing features of *ex* are described in *An Introduction to Display Editing with Vi.* Other documents about the editor include the introduction *Edit: A tutorial*, the *Ex/edit Command Summary*, and a *Vi Quick Reference* card.

September 15, 1986

Ex Reference Manual
Version 3.5/2.13 – September, 1980

William Joy

Revised for versions 3.5/2.13 by
Mark Horton

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, Ca. 94720

1. Starting ex

Each instance of the editor has a set of options, which can be set to tailor it to your liking. The command *edit* invokes a version of *ex* designed for more casual or beginning users by changing the default settings of some of these options. To simplify the description which follows we assume the default settings of the options.

When invoked, *ex* determines the terminal type from the TERM variable in the environment. It there is a TERMCAP variable in the environment, and the type of the terminal described there matches the TERM variable, then that description is used. Also if the TERMCAP variable contains a pathname (beginning with a /) then the editor will seek the description of the terminal in that file (rather than the default /etc/termcap.) If there is a variable EXINIT in the environment, then the editor will execute the commands in that variable, otherwise if there is a file *.exrc* in your HOME directory *ex* reads commands from that file, simulating a *source* command. Option setting commands placed in EXINIT or *.exrc* will be executed before each editor session.

A command to enter *ex* has the following prototype:†

ex [–] [**–v**] [**–t** *tag*] [**–r**] [**–l**] [**–w***n*] [**–x**] [**–R**] [+*command*] name ...

The most common case edits a single file with no options, i.e.:

ex name

The – command line option option suppresses all interactive-user feedback and is useful in processing editor scripts in command files. The **–v** option is equivalent to using *vi* rather than *ex*. The **–t** option is equivalent to an initial *tag* command, editing the file containing the *tag* and positioning the editor at its definition. The **–r** option is used in recovering after an editor or system crash, retrieving the last saved version of the named file or, if no file is specified, typing a list of saved files. The **–l** option sets up for editing LISP, setting the *showmatch* and *lisp* options. The **–w** option sets the default window size to *n*, and is useful on dialups to start in small windows. The **–x** option causes *ex* to prompt for a *key*, which is used to encrypt and decrypt the contents of the file, which should already be encrypted using the same key, see *crypt*(1). The **–R** option sets the *readonly* option at the start. ‡ *Name* arguments indicate files to be edited. An argument of the form +*command* indicates that the editor should begin by executing the specified command. If *command* is omitted, then it defaults to "$", positioning the editor at the last line of the first file

The financial support of an IBM Graduate Fellowship and the National Science Foundation under grants MCS74-07644-A03 and MCS78-07291 is gratefully acknowledged.

† Brackets '[' ']' surround optional parameters here.

‡ Not available in all v2 editors due to memory constraints.

initially. Other useful commands here are scanning patterns of the form "/pat" or line numbers, e.g. "+100" starting at line 100.

2. File manipulation

2.1. Current file

Ex is normally editing the contents of a single file, whose name is recorded in the *current* file name. *Ex* performs all editing actions in a buffer (actually a temporary file) into which the text of the file is initially read. Changes made to the buffer have no effect on the file being edited unless and until the buffer contents are written out to the file with a *write* command. After the buffer contents are written, the previous contents of the written file are no longer accessible. When a file is edited, its name becomes the current file name, and its contents are read into the buffer.

The current file is almost always considered to be *edited.* This means that the contents of the buffer are logically connected with the current file name, so that writing the current buffer contents onto that file, even if it exists, is a reasonable action. If the current file is not *edited* then *ex* will not normally write on it if it already exists.*

2.2. Alternate file

Each time a new value is given to the current file name, the previous current file name is saved as the *alternate* file name. Similarly if a file is mentioned but does not become the current file, it is saved as the alternate file name.

2.3. Filename expansion

Filenames within the editor may be specified using the normal shell expansion conventions. In addition, the character '%' in filenames is replaced by the *current* file name and the character '#' by the *alternate* file name.†

2.4. Multiple files and named buffers

If more than one file is given on the command line, then the first file is edited as described above. The remaining arguments are placed with the first file in the *argument list.* The current argument list may be displayed with the *args* command. The next file in the argument list may be edited with the *next* command. The argument list may also be respecified by specifying a list of names to the *next* command. These names are expanded, the resulting list of names becomes the new argument list, and *ex* edits the first file on the list.

For saving blocks of text while editing, and especially when editing more than one file, *ex* has a group of named buffers. These are similar to the normal buffer, except that only a limited number of operations are available on them. The buffers have names *a* through *z*.‡

2.5. Read only

It is possible to use *ex* in *read only* mode to look at files that you have no intention of modifying. This mode protects you from accidently overwriting the file. Read only mode is on when the *readonly* option is set. It can be turned on with the **-R** command line option, by the *view* command line invocation, or by setting the *readonly* option. It can be cleared by setting *noreadonly*. It is possible to write, even while in read only mode, by indicating that you really know what you are doing. You can write to a different file, or can use the ! form of write, even while in read only mode.

---

* The *file* command will say "[Not edited]" if the current file is not considered edited.

† This makes it easy to deal alternately with two files and eliminates the need for retyping the name supplied on an *edit* command after a *No write since last change* diagnostic is received.

‡ It is also possible to refer to *A* through *Z;* the upper case buffers are the same as the lower but commands append to named buffers rather than replacing if upper case names are used.

3. Exceptional Conditions

3.1. Errors and interrupts

When errors occur *ex* (optionally) rings the terminal bell and, in any case, prints an error diagnostic. If the primary input is from a file, editor processing will terminate. If an interrupt signal is received, *ex* prints "Interrupt" and returns to its command level. If the primary input is a file, then *ex* will exit when this occurs.

3.2. Recovering from hangups and crashes

If a hangup signal is received and the buffer has been modified since it was last written out, or if the system crashes, either the editor (in the first case) or the system (after it reboots in the second) will attempt to preserve the buffer. The next time you log in you should be able to recover the work you were doing, losing at most a few lines of changes from the last point before the hangup or editor crash. To recover a file you can use the **–r** option. If you were editing the file *resume,* then you should change to the directory where you were when the crash occurred, giving the command

ex –r *resume*

After checking that the retrieved file is indeed ok, you can *write* it over the previous contents of that file.

You will normally get mail from the system telling you when a file has been saved after a crash. The command

ex –r

will print a list of the files which have been saved for you. (In the case of a hangup, the file will not appear in the list, although it can be recovered.)

4. Editing modes

Ex has five distinct modes. The primary mode is *command* mode. Commands are entered in command mode when a ':' prompt is present, and are executed each time a complete line is sent. In *text input* mode *ex* gathers input lines and places them in the file. The *append, insert,* and *change* commands use text input mode. No prompt is printed when you are in text input mode. This mode is left by typing a '.' alone at the beginning of a line, and *command* mode resumes.

The last three modes are *open* and *visual* modes, entered by the commands of the same name, and, within open and visual modes *text insertion* mode. *Open* and *visual* modes allow local editing operations to be performed on the text in the file. The *open* command displays one line at a time on any terminal while *visual* works on CRT terminals with random positioning cursors, using the screen as a (single) window for file editing changes. These modes are described (only) in *An Introduction to Display Editing with Vi.*

5. Command structure

Most command names are English words, and initial prefixes of the words are acceptable abbreviations. The ambiguity of abbreviations is resolved in favor of the more commonly used commands.*

5.1. Command parameters

Most commands accept prefix addresses specifying the lines in the file upon which they are to have effect. The forms of these addresses will be discussed below. A number of commands also may take a trailing *count* specifying the number of lines to be involved in the command.† Thus

* As an example, the command *substitute* can be abbreviated 's' while the shortest available abbreviation for the *set* command is 'se'.

† Counts are rounded down if necessary.

the command "10p" will print the tenth line in the buffer while "delete 5" will delete five lines from the buffer, starting with the current line.

Some commands take other information or parameters, this information always being given after the command name.‡

5.2. Command variants

A number of commands have two distinct variants. The variant form of the command is invoked by placing an '!' immediately after the command name. Some of the default variants may be controlled by options; in this case, the '!' serves to toggle the default.

5.3. Flags after commands

The characters '#', 'p' and 'l' may be placed after many commands.** In this case, the command abbreviated by these characters is executed after the command completes. Since *ex* normally prints the new current line after each change, 'p' is rarely necessary. Any number of '+' or '–' characters may also be given with these flags. If they appear, the specified offset is applied to the current line value before the printing command is executed.

5.4. Comments

It is possible to give editor commands which are ignored. This is useful when making complex editor scripts for which comments are desired. The comment character is the double quote: ". Any command line beginning with " is ignored. Comments beginning with " may also be placed at the ends of commands, except in cases where they could be confused as part of text (shell escapes and the substitute and map commands).

5.5. Multiple commands per line

More than one command may be placed on a line by separating each pair of commands by a '|' character. However the *global* commands, comments, and the shell escape '!' must be the last command on a line, as they are not terminated by a '|'.

5.6. Reporting large changes

Most commands which change the contents of the editor buffer give feedback if the scope of the change exceeds a threshold given by the *report* option. This feedback helps to detect undesirably large changes so that they may be quickly and easily reversed with an *undo*. After commands with more global effect such as *global* or *visual*, you will be informed if the net change in the number of lines in the buffer during this command exceeds this threshold.

6. Command addressing

6.1. Addressing primitives

.	The current line. Most commands leave the current line as the last line which they affect. The default address for most commands is the current line, thus '.' is rarely used alone as an address.
n	The *n*th line in the editor's buffer, lines being numbered sequentially from 1.
$	The last line in the buffer.
%	An abbreviation for "1,$", the entire buffer.
+n –n	An offset relative to the current buffer line.†

‡ Examples would be option names in a *set* command i.e. "set number", a file name in an *edit* command, a regular expression in a *substitute* command, or a target address for a *copy* command, i.e. "1,5 copy 25".

** A 'p' or 'l' must be preceded by a blank or tab except in the single special case 'dp'.

† The forms '.+3' '+3' and '+++' are all equivalent; if the current line is line 100 they all address line 103.

/pat/ ?pat? Scan forward and backward respectively for a line containing *pat*, a regular expression (as defined below). The scans normally wrap around the end of the buffer. If all that is desired is to print the next line containing *pat*, then the trailing / or ? may be omitted. If *pat* is omitted or explicitly empty, then the last regular expression specified is located.‡

'' 'x Before each non-relative motion of the current line '.', the previous current line is marked with a tag, subsequently referred to as ''''. This makes it easy to refer or return to this previous context. Marks may also be established by the *mark* command, using single lower case letters *x* and the marked lines referred to as '*'x*'.

6.2. Combining addressing primitives

Addresses to commands consist of a series of addressing primitives, separated by ',' or ';'. Such address lists are evaluated left-to-right. When addresses are separated by ';' the current line '.' is set to the value of the previous addressing expression before the next address is interpreted. If more addresses are given than the command requires, then all but the last one or two are ignored. If the command takes two addresses, the first addressed line must precede the second in the buffer.†

7. Command descriptions

The following form is a prototype for all *ex* commands:

address **command** *! parameters count flags*

All parts are optional; the degenerate case is the empty command which prints the next line in the file. For sanity with use from within *visual* mode, *ex* ignores a ":" preceding any command.

In the following command descriptions, the default addresses are shown in parentheses, which are *not,* however, part of the command.

abbreviate *word rhs* abbr: **ab**

Add the named abbreviation to the current list. When in input mode in visual, if *word* is typed as a complete word, it will be changed to *rhs*.

(.) **append** abbr: **a**
text
.

Reads the input text and places it after the specified line. After the command, '.' addresses the last line input or the specified line if no lines were input. If address '0' is given, text is placed at the beginning of the buffer.

a!
text
.

The variant flag to *append* toggles the setting for the *autoindent* option during the input of *text.*

‡ The forms \/ and \? scan using the last regular expression used in a scan; after a substitute // and ?? would scan using the substitute's regular expression.

† Null address specifications are permitted in a list of addresses, the default in this case is the current line '.'; thus ',100' is equivalent to '.,100'. It is an error to give a prefix address to a command which expects none.

args

The members of the argument list are printed, with the current argument delimited by '[' and ']'.

(. , .) **change** *count*   abbr: **c**
text
.

Replaces the specified lines with the input *text*. The current line becomes the last line input; if no lines were input it is left as for a *delete.*

c!
text
.

The variant toggles *autoindent* during the *change.*

(. , .) **copy** *addr flags*   abbr: **co**

A *copy* of the specified lines is placed after *addr,* which may be '0'. The current line '.' addresses the last line of the copy. The command *t* is a synonym for *copy.*

(. , .) **delete** *buffer count flags*   abbr: **d**

Removes the specified lines from the buffer. The line after the last line deleted becomes the current line; if the lines deleted were originally at the end, the new last line becomes the current line. If a named *buffer* is specified by giving a letter, then the specified lines are saved in that buffer, or appended to it if an upper case letter is used.

edit *file*   abbr: **e**
ex *file*

Used to begin an editing session on a new file. The editor first checks to see if the buffer has been modified since the last *write* command was issued. If it has been, a warning is issued and the command is aborted. The command otherwise deletes the entire contents of the editor buffer, makes the named file the current file and prints the new filename. After insuring that this file is sensible† the editor reads the file into its buffer.

If the read of the file completes without error, the number of lines and characters read is typed. If there were any non-ASCII characters in the file they are stripped of their non-ASCII high bits, and any null characters in the file are discarded. If none of these errors occurred, the file is considered *edited.* If the last line of the input file is missing the trailing newline character, it will be supplied and a complaint will be issued. This command leaves the current line '.' at the last line read.‡

e! *file*

The variant form suppresses the complaint about modifications having been made and not written from the editor buffer, thus discarding all changes which have been made before editing the new file.

e +*n* *file*

Causes the editor to begin at line *n* rather than at the last line; *n* may also be an editor command containing no spaces, e.g.: "+/pat".

---

† I.e., that it is not a binary file such as a directory, a block or character special file other than */dev/tty,* a terminal, or a binary or executable file (as indicated by the first word).
‡ If executed from within *open* or *visual,* the current line is initially the first line of the file.

file abbr: **f**

Prints the current file name, whether it has been '[Modified]' since the last *write* command, whether it is *read only*, the current line, the number of lines in the buffer, and the percentage of the way through the buffer of the current line.*

file *file*

The current file name is changed to *file* which is considered '[Not edited]'.

(1 , $) **global** */pat/ cmds* abbr: **g**

First marks each line among those specified which matches the given regular expression. Then the given command list is executed with '.' initially set to each marked line.

The command list consists of the remaining commands on the current input line and may continue to multiple lines by ending all but the last such line with a '\'. If *cmds* (and possibly the trailing / delimiter) is omitted, each line matching *pat* is printed. *Append, insert,* and *change* commands and associated input are permitted; the '.' terminating input may be omitted if it would be on the last line of the command list. *Open* and *visual* commands are permitted in the command list and take input from the terminal.

The *global* command itself may not appear in *cmds*. The *undo* command is also not permitted there, as *undo* instead can be used to reverse the entire *global* command. The options *autoprint* and *autoindent* are inhibited during a *global*, (and possibly the trailing / delimiter) and the value of the *report* option is temporarily infinite, in deference to a *report* for the entire global. Finally, the context mark '´´' is set to the value of '.' before the global command begins and is not changed during a global command, except perhaps by an *open* or *visual* within the *global*.

g! */pat/ cmds* abbr: **v**

The variant form of *global* runs *cmds* at each line not matching *pat*.

(.) **insert** abbr: **i**
text
.

Places the given text before the specified line. The current line is left at the last line input; if there were none input it is left at the line before the addressed line. This command differs from *append* only in the placement of text.

i!
text
.

The variant toggles *autoindent* during the *insert*.

(. , .+1) **join** *count flags* abbr: **j**

Places the text from a specified range of lines together on one line. White space is adjusted at each junction to provide at least one blank character, two if there was a '.' at the end of the line, or none if the first following character is a ')'. If there is already white space at the end of the line, then the white space at the start of the next line will be discarded.

* In the rare case that the current file is '[Not edited]' this is noted also; in this case you have to use the form **w!** to write to the file, since the editor is not sure that a **write** will not destroy a file unrelated to the current contents of the buffer.

j!

The variant causes a simpler *join* with no white space processing; the characters in the lines are simply concatenated.

(.) k *x*

The *k* command is a synonym for *mark*. It does not require a blank or tab before the following letter.

(. , .) list *count flags*

Prints the specified lines in a more unambiguous way: tabs are printed as '^I' and the end of each line is marked with a trailing '$'. The current line is left at the last line printed.

map *lhs rhs*

The *map* command is used to define macros for use in *visual* mode. *Lhs* should be a single character, or the sequence "#n", for n a digit, referring to function key *n*. When this character or function key is typed in *visual* mode, it will be as though the corresponding *rhs* had been typed. On terminals without function keys, you can type "#n". See section 6.9 of the "Introduction to Display Editing with Vi" for more details.

(.) mark *x*

Gives the specified line mark *x*, a single lower case letter. The *x* must be preceded by a blank or a tab. The addressing form ''x' then addresses this line. The current line is not affected by this command.

(. , .) move *addr* abbr: **m**

The *move* command repositions the specified lines to be after *addr*. The first of the moved lines becomes the current line.

next abbr: **n**

The next file from the command line argument list is edited.

n!

The variant suppresses warnings about the modifications to the buffer not having been written out, discarding (irretrievably) any changes which may have been made.

n *filelist*
n +*command filelist*

The specified *filelist* is expanded and the resulting list replaces the current argument list; the first file in the new list is then edited. If *command* is given (it must contain no spaces), then it is executed after editing the first such file.

(. , .) number *count flags* abbr: **#** or **nu**

Prints each specified line preceded by its buffer line number. The current line is left at the last line printed.

(.) open *flags* abbr: **o**
(.) open /*pat*/ *flags*

Enters intraline editing *open* mode at each addressed line. If *pat* is given, then the cursor will be placed initially at the beginning of the string matched by the pattern. To exit this mode use Q. See *An Introduction to Display Editing with Vi* for more details.
‡

‡ Not available in all v2 editors due to memory constraints.

preserve

The current editor buffer is saved as though the system had just crashed. This command is for use only in emergencies when a *write* command has resulted in an error and you don't know how to save your work. After a *preserve* you should seek help.

(. , .) **print** *count* abbr: **p** or **P**

Prints the specified lines with non-printing characters printed as control characters '^*x*'; delete (octal 177) is represented as '^?'. The current line is left at the last line printed.

(.) **put** *buffer* abbr: **pu**

Puts back previously *deleted* or *yanked* lines. Normally used with *delete* to effect movement of lines, or with *yank* to effect duplication of lines. If no *buffer* is specified, then the last *deleted* or *yanked* text is restored.* By using a named buffer, text may be restored that was saved there at any previous time.

quit abbr: **q**

Causes *ex* to terminate. No automatic write of the editor buffer to a file is performed. However, *ex* issues a warning message if the file has changed since the last *write* command was issued, and does not *quit*.† Normally, you will wish to save your changes, and you should give a *write* command; if you wish to discard them, use the **q!** command variant.

q!

Quits from the editor, discarding changes to the buffer without complaint.

(.) **read** *file* abbr: **r**

Places a copy of the text of the given file in the editing buffer after the specified line. If no *file* is given the current file name is used. The current file name is not changed unless there is none in which case *file* becomes the current name. The sensibility restrictions for the *edit* command apply here also. If the file buffer is empty and there is no current name then *ex* treats this as an *edit* command.

Address '0' is legal for this command and causes the file to be read at the beginning of the buffer. Statistics are given as for the *edit* command when the *read* successfully terminates. After a *read* the current line is the last line read.‡

(.) **read** !*command*

Reads the output of the command *command* into the buffer after the specified line. This is not a variant form of the command, rather a read specifying a *command* rather than a *filename;* a blank or tab before the ! is mandatory.

recover *file*

Recovers *file* from the system save area. Used after a accidental hangup of the phone** or a system crash** or *preserve* command. Except when you use *preserve* you will be notified by mail when a file is saved.

---

* But no modifying commands may intervene between the *delete* or *yank* and the *put,* nor may lines be moved between files without using a named buffer.

† *Ex* will also issue a diagnostic if there are more files in the argument list.

‡ Within *open* and *visual* the current line is set to the first line read rather than the last.

** The system saves a copy of the file you were editing only if you have made changes to the file.

rewind abbr: **rew**

The argument list is rewound, and the first file in the list is edited.

rew!

Rewinds the argument list discarding any changes made to the current buffer.

set *parameter*

With no arguments, prints those options whose values have been changed from their defaults; with parameter *all* it prints all of the option values.

Giving an option name followed by a '?' causes the current value of that option to be printed. The '?' is unnecessary unless the option is Boolean valued. Boolean options are given values either by the form 'set *option*' to turn them on or 'set no*option*' to turn them off; string and numeric options are assigned via the form 'set *option*=value'.

More than one parameter may be given to *set* ; they are interpreted left-to-right.

shell abbr: **sh**

A new shell is created. When it terminates, editing resumes.

source *file* abbr: **so**

Reads and executes commands from the specified file. *Source* commands may be nested.

(. , .) **substitute** */pat/repl/ options count flags* abbr: **s**

On each specified line, the first instance of pattern *pat* is replaced by replacement pattern *repl*. If the *global* indicator option character 'g' appears, then all instances are substituted; if the *confirm* indication character 'c' appears, then before each substitution the line to be substituted is typed with the string to be substituted marked with '↑' characters. By typing an 'y' one can cause the substitution to be performed, any other input causes no change to take place. After a *substitute* the current line is the last line substituted.

Lines may be split by substituting new-line characters into them. The newline in *repl* must be escaped by preceding it with a '\'. Other metacharacters available in *pat* and *repl* are described below.

stop

Suspends the editor, returning control to the top level shell. If *autowrite* is set and there are unsaved changes, a write is done first unless the form **stop**! is used. This commands is only available where supported by the teletype driver and operating system.

(. , .) **substitute** *options count flags* abbr: **s**

If *pat* and *repl* are omitted, then the last substitution is repeated. This is a synonym for the **&** command.

(. , .) **t** *addr flags*

The *t* command is a synonym for *copy*.

ta *tag*

The focus of editing switches to the location of *tag,* switching to a different line in the current file where it is defined, or if necessary to another file.‡

‡ If you have modified the current file before giving a *tag* command, you must write it out; giving another *tag* command, specifying no *tag* will reuse the previous tag.

The tags file is normally created by a program such as *ctags,* and consists of a number of lines with three fields separated by blanks or tabs. The first field gives the name of the tag, the second the name of the file where the tag resides, and the third gives an addressing form which can be used by the editor to find the tag; this field is usually a contextual scan using '*/pat/*' to be immune to minor changes in the file. Such scans are always performed as if *nomagic* was set.

The tag names in the tags file must be sorted alphabetically. ‡

unabbreviate *word* abbr: **una**

Delete *word* from the list of abbreviations.

undo abbr: **u**

Reverses the changes made in the buffer by the last buffer editing command. Note that *global* commands are considered a single command for the purpose of *undo* (as are *open* and *visual.)* Also, the commands *write* and *edit* which interact with the file system cannot be undone. *Undo* is its own inverse.

Undo always marks the previous value of the current line '.' as ''''. After an *undo* the current line is the first line restored or the line before the first line deleted if no lines were restored. For commands with more global effect such as *global* and *visual* the current line regains it's pre-command value after an *undo.*

unmap *lhs*

The macro expansion associated by *map* for *lhs* is removed.

(1 , $) **v** */pat/ cmds*

A synonym for the *global* command variant **g!**, running the specified *cmds* on each line which does not match *pat.*

version abbr: **ve**

Prints the current version number of the editor as well as the date the editor was last changed.

(.) **visual** *type count flags* abbr: **vi**

Enters visual mode at the specified line. *Type* is optional and may be '–' , '↑' or '.' as in the *z* command to specify the placement of the specified line on the screen. By default, if *type* is omitted, the specified line is placed as the first on the screen. A *count* specifies an initial window size; the default is the value of the option *window.* See the document *An Introduction to Display Editing with Vi* for more details. To exit this mode, type Q.

visual file
visual +*n* file

From visual mode, this command is the same as edit.

(1 , $) **write** *file* abbr: **w**

Writes changes made back to *file*, printing the number of lines and characters written. Normally *file* is omitted and the text goes back where it came from. If a *file* is specified, then text will be written to that file.* If the file does not exist it is created. The current file name is changed only if there is no current file name; the current line is never changed.

‡ Not available in all v2 editors due to memory constraints.

* The editor writes to a file only if it is the current file and is *edited*, if the file does not exist, or if the file is actually a teletype, */dev/tty, /dev/null.* Otherwise, you must give the variant form **w!** to force the write.

If an error occurs while writing the current and *edited* file, the editor considers that there has been "No write since last change" even if the buffer had not previously been modified.

(1 , $) **write>>** *file* abbr: **w>>**

Writes the buffer contents at the end of an existing file.

w! *name*

Overrides the checking of the normal *write* command, and will write to any file which the system permits.

(1 , $) **w** **!***command*

Writes the specified lines into *command.* Note the difference between **w!** which overrides checks and **w !** which writes to a command.

wq *name*

Like a *write* and then a *quit* command.

wq! *name*

The variant overrides checking on the sensibility of the *write* command, as **w!** does.

xit *name*

If any changes have been made and not written, writes the buffer out. Then, in any case, quits.

(. , .) **yank** *buffer count* abbr: **ya**

Places the specified lines in the named *buffer,* for later retrieval via *put.* If no buffer name is specified, the lines go to a more volatile place; see the *put* command description.

(.+1) **z** *count*

Print the next *count* lines, default *window.*

(.) **z** *type count*

Prints a window of text with the specified line at the top. If *type* is '–' the line is placed at the bottom; a '.' causes the line to be placed in the center.* A count gives the number of lines to be displayed rather than double the number specified by the *scroll* option. On a CRT the screen is cleared before display begins unless a count which is less than the screen size is given. The current line is left at the last line printed.

! *command*

The remainder of the line after the '!' character is sent to a shell to be executed. Within the text of *command* the characters '%' and '#' are expanded as in filenames and the character '!' is replaced with the text of the previous command. Thus, in particular, '!!' repeats the last such shell escape. If any such expansion is performed, the expanded line will be echoed. The current line is unchanged by this command.

If there has been "[No write]" of the buffer contents since the last change to the editing buffer, then a diagnostic will be printed before the command is executed as a warning. A single '!' is printed when the command completes.

* Forms 'z=' and 'z↑' also exist; 'z=' places the current line in the center, surrounds it with lines of '–' characters and leaves the current line at this line. The form 'z↑' prints the window before 'z–' would. The characters '+', '↑' and '–' may be repeated for cumulative effect. On some v2 editors, no *type* may be given.

(*addr* , *addr*) **!** ***command***

Takes the specified address range and supplies it as standard input to ***command;*** the resulting output then replaces the input lines.

(**$**) **=**

Prints the line number of the addressed line. The current line is unchanged.

(. , .) $>$ ***count flags***
(. , .) $<$ ***count flags***

Perform intelligent shifting on the specified lines; $<$ shifts left and $>$ shift right. The quantity of shift is determined by the ***shiftwidth*** option and the repetition of the specification character. Only white space (blanks and tabs) is shifted; no non-white characters are discarded in a left-shift. The current line becomes the last line which changed due to the shifting.

^D

An end-of-file from a terminal input scrolls through the file. The ***scroll*** option specifies the size of the scroll, normally a half screen of text.

(.+1 , .+1)
(.+1 , .+1) |

An address alone causes the addressed lines to be printed. A blank line prints the next line in the file.

(. , .) & ***options count flags***

Repeats the previous ***substitute*** command.

(. , .) ~ ***options count flags***

Replaces the previous regular expression with the previous replacement pattern from a substitution.

8. Regular expressions and substitute replacement patterns

8.1. Regular expressions

A regular expression specifies a set of strings of characters. A member of this set of strings is said to be *matched* by the regular expression. *Ex* remembers two previous regular expressions: the previous regular expression used in a ***substitute*** command and the previous regular expression used elsewhere (referred to as the previous ***scanning*** regular expression.) The previous regular expression can always be referred to by a null *re*, e.g. '//' or '??'.

8.2. Magic and nomagic

The regular expressions allowed by ***ex*** are constructed in one of two ways depending on the setting of the ***magic*** option. The ***ex*** and ***vi*** default setting of ***magic*** gives quick access to a powerful set of regular expression metacharacters. The disadvantage of ***magic*** is that the user must remember that these metacharacters are ***magic*** and precede them with the character '\' to use them as "ordinary" characters. With ***nomagic,*** the default for ***edit,*** regular expressions are much simpler, there being only two metacharacters. The power of the other metacharacters is still available by preceding the (now) ordinary character with a '\'. Note that '\' is thus always a metacharacter.

The remainder of the discussion of regular expressions assumes that that the setting of this option is ***magic.***†

† To discern what is true with ***nomagic*** it suffices to remember that the only special characters in this case will be '↑' at the beginning of a regular expression, '$' at the end of a regular expression, and '\'. With ***nomagic*** the

8.3. Basic regular expression summary

The following basic constructs are used to construct ***magic*** mode regular expressions.

char — An ordinary character matches itself. The characters '↑' at the beginning of a line, '$' at the end of line, '*' as any character other than the first, '.', '\', '[', and '~' are not ordinary characters and must be escaped (preceded) by '\' to be treated as such.

↑ — At the beginning of a pattern forces the match to succeed only at the beginning of a line.

$ — At the end of a regular expression forces the match to succeed only at the end of the line.

. — Matches any single character except the new-line character.

\< — Forces the match to occur only at the beginning of a "variable" or "word"; that is, either at the beginning of a line, or just before a letter, digit, or underline and after a character not one of these.

\> — Similar to '\<', but matching the end of a "variable" or "word", i.e. either the end of the line or before character which is neither a letter, nor a digit, nor the underline character.

[*string*] — Matches any (single) character in the class defined by ***string***. Most characters in ***string*** define themselves. A pair of characters separated by '–' in ***string*** defines the set of characters collating between the specified lower and upper bounds, thus '[a–z]' as a regular expression matches any (single) lower-case letter. If the first character of ***string*** is an '↑' then the construct matches those characters which it otherwise would not; thus '[↑a–z]' matches anything but a lower-case letter (and of course a newline). To place any of the characters '↑', '[', or '–' in ***string*** you must escape them with a preceding '\'.

8.4. Combining regular expression primitives

The concatenation of two regular expressions matches the leftmost and then longest string which can be divided with the first piece matching the first regular expression and the second piece matching the second. Any of the (single character matching) regular expressions mentioned above may be followed by the character '*' to form a regular expression which matches any number of adjacent occurrences (including 0) of characters matched by the regular expression it follows.

The character '~' may be used in a regular expression, and matches the text which defined the replacement part of the last ***substitute*** command. A regular expression may be enclosed between the sequences '\(' and '\)' with side effects in the ***substitute*** replacement patterns.

8.5. Substitute replacement patterns

The basic metacharacters for the replacement pattern are '&' and '~'; these are given as '\&' and '\~' when ***nomagic*** is set. Each instance of '&' is replaced by the characters which the regular expression matched. The metacharacter '~' stands, in the replacement pattern, for the defining text of the previous replacement pattern.

Other metasequences possible in the replacement pattern are always introduced by the escaping character '\'. The sequence '*n*' is replaced by the text matched by the *n*-th regular subexpression enclosed between '\(' and '\)'.† The sequences '\u' and '\l' cause the immediately following character in the replacement to be converted to upper- or lower-case respectively if this character is a letter. The sequences '\U' and '\L' turn such conversion on, either until '\E' or '\e' is encountered, or until the end of the replacement pattern.

---

characters '~' and '&' also lose their special meanings related to the replacement pattern of a substitute.

† When nested, parenthesized subexpressions are present, *n* is determined by counting occurrences of '\(' starting from the left.

9. Option descriptions

autoindent, ai default: noai

Can be used to ease the preparation of structured program text. At the beginning of each *append*, *change* or *insert* command or when a new line is *opened* or created by an *append*, *change*, *insert*, or *substitute* operation within *open* or *visual* mode, *ex* looks at the line being appended after, the first line changed or the line inserted before and calculates the amount of white space at the start of the line. It then aligns the cursor at the level of indentation so determined.

If the user then types lines of text in, they will continue to be justified at the displayed indenting level. If more white space is typed at the beginning of a line, the following line will start aligned with the first non-white character of the previous line. To back the cursor up to the preceding tab stop one can hit **^D**. The tab stops going backwards are defined at multiples of the *shiftwidth* option. You *cannot* backspace over the indent, except by sending an end-of-file with a **^D**.

Specially processed in this mode is a line with no characters added to it, which turns into a completely blank line (the white space provided for the *autoindent* is discarded.) Also specially processed in this mode are lines beginning with an '↑' and immediately followed by a **^D**. This causes the input to be repositioned at the beginning of the line, but retaining the previous indent for the next line. Similarly, a '0' followed by a **^D** repositions at the beginning but without retaining the previous indent.

Autoindent doesn't happen in *global* commands or when the input is not a terminal.

autoprint, ap default: ap

Causes the current line to be printed after each *delete*, *copy*, *join*, *move*, *substitute*, *t*, *undo* or shift command. This has the same effect as supplying a trailing 'p' to each such command. *Autoprint* is suppressed in globals, and only applies to the last of many commands on a line.

autowrite, aw default: noaw

Causes the contents of the buffer to be written to the current file if you have modified it and give a *next, rewind, stop, tag,* or *!* command, or a **^↑** (switch files) or ^] (tag goto) command in *visual*. Note, that the *edit* and *ex* commands do **not** autowrite. In each case, there is an equivalent way of switching when autowrite is set to avoid the *autowrite* (*edit* for *next*, *rewind!* for .I rewind , *stop!* for *stop*, *tag!* for *tag*, *shell* for *!*, and **:e #** and a **:ta!** command from within *visual).*

beautify, bf default: nobeautify

Causes all control characters except tab, newline and form-feed to be discarded from the input. A complaint is registered the first time a backspace character is discarded. *Beautify* does not apply to command input.

directory, dir default: dir=/tmp

Specifies the directory in which *ex* places its buffer file. If this directory in not writable, then the editor will exit abruptly when it fails to be able to create its buffer there.

edcompatible default: noedcompatible

Causes the presence of absence of **g** and **c** suffixes on substitute commands to be remembered, and to be toggled by repeating the suffices. The suffix **r** makes the substitution be as in the ~ command, instead of like *&*. ‡‡

‡‡ Version 3 only.

errorbells, eb default: noeb

Error messages are preceded by a bell.* If possible the editor always places the error message in a standout mode of the terminal (such as inverse video) instead of ringing the bell.

hardtabs, ht default: ht=8

Gives the boundaries on which terminal hardware tabs are set (or on which the system expands tabs).

ignorecase, ic default: noic

All upper case characters in the text are mapped to lower case in regular expression matching. In addition, all upper case characters in regular expressions are mapped to lower case except in character class specifications.

lisp default: nolisp

Autoindent indents appropriately for *lisp* code, and the ( ) { } [[and]] commands in *open* and *visual* are modified to have meaning for *lisp*.

list default: nolist

All printed lines will be displayed (more) unambiguously, showing tabs and end-of-lines as in the *list* command.

magic default: magic for *ex* and *vi*†

If *nomagic* is set, the number of regular expression metacharacters is greatly reduced, with only '↑' and '$' having special effects. In addition the metacharacters '~' and '&' of the replacement pattern are treated as normal characters. All the normal metacharacters may be made *magic* when *nomagic* is set by preceding them with a '\'.

mesg default: mesg

Causes write permission to be turned off to the terminal while you are in visual mode, if *nomesg* is set. ‡‡

number, nu default: nonumber

Causes all output lines to be printed with their line numbers. In addition each input line will be prompted for by supplying the line number it will have.

open default: open

If *noopen*, the commands *open* and *visual* are not permitted. This is set for *edit* to prevent confusion resulting from accidental entry to open or visual mode.

optimize, opt default: optimize

Throughput of text is expedited by setting the terminal to not do automatic carriage returns when printing more than one (logical) line of output, greatly speeding output on terminals without addressable cursors when text with leading white space is printed.

paragraphs, para default: para=IPLPPPQPP LIbp

Specifies the paragraphs for the { and } operations in *open* and *visual*. The pairs of characters in the option's value are the names of the macros which start paragraphs.

* Bell ringing in *open* and *visual* on errors is not suppressed by setting *noeb*.

† *Nomagic* for *edit*.

‡‡ Version 3 only.

prompt default: prompt

Command mode input is prompted for with a ':'.

redraw default: noredraw

The editor simulates (using great amounts of output), an intelligent terminal on a dumb terminal (e.g. during insertions in *visual* the characters to the right of the cursor position are refreshed as each input character is typed.) Useful only at very high speed.

remap default: remap

If on, macros are repeatedly tried until they are unchanged. ‡‡ For example, if **o** is mapped to **O**, and **O** is mapped to **I**, then if *remap* is set, **o** will map to **I**, but if *noremap* is set, it will map to **O**.

report default: report=5†

Specifies a threshold for feedback from commands. Any command which modifies more than the specified number of lines will provide feedback as to the scope of its changes. For commands such as *global*, *open*, *undo*, and *visual* which have potentially more far reaching scope, the net change in the number of lines in the buffer is presented at the end of the command, subject to this same threshold. Thus notification is suppressed during a *global* command on the individual commands performed.

scroll default: scroll=½ window

Determines the number of logical lines scrolled when an end-of-file is received from a terminal input in command mode, and the number of lines printed by a command mode *z* command (double the value of *scroll*).

sections default: sections=SHNHH HU

Specifies the section macros for the [[and]] operations in *open* and *visual*. The pairs of characters in the options's value are the names of the macros which start paragraphs.

shell, sh default: sh=/bin/sh

Gives the path name of the shell forked for the shell escape command '!', and by the *shell* command. The default is taken from SHELL in the environment, if present.

shiftwidth, sw default: sw=8

Gives the width a software tab stop, used in reverse tabbing with **^D** when using *autoindent* to append text, and by the shift commands.

showmatch, sm default: nosm

In *open* and *visual* mode, when a **)** or **}** is typed, move the cursor to the matching **(** or **{** for one second if this matching character is on the screen. Extremely useful with *lisp*.

slowopen, slow terminal dependent

Affects the display algorithm used in *visual* mode, holding off display updating during input of new text to improve throughput when the terminal in use is both slow and unintelligent. See *An Introduction to Display Editing with Vi* for more details.

‡‡ Version 3 only.
† 2 for *edit*.

tabstop, ts default: ts=8

The editor expands tabs in the input file to be on *tabstop* boundaries for the purposes of display.

taglength, tl default: tl=0

Tags are not significant beyond this many characters. A value of zero (the default) means that all characters are significant.

tags default: tags=tags /usr/lib/tags

A path of files to be used as tag files for the *tag* command. ‡‡ A requested tag is searched for in the specified files, sequentially. By default (even in version 2) files called **tags** are searched for in the current directory and in /usr/lib (a master file for the entire system.)

term from environment TERM

The terminal type of the output device.

terse default: noterse

Shorter error diagnostics are produced for the experienced user.

warn default: warn

Warn if there has been '[No write since last change]' before a '!' command escape.

window default: window=speed dependent

The number of lines in a text window in the *visual* command. The default is 8 at slow speeds (600 baud or less), 16 at medium speed (1200 baud), and the full screen (minus one line) at higher speeds.

w300, w1200, w9600

These are not true options but set **window** only if the speed is slow (300), medium (1200), or high (9600), respectively. They are suitable for an EXINIT and make it easy to change the 8/16/full screen rule.

wrapscan, ws default: ws

Searches using the regular expressions in addressing will wrap around past the end of the file.

wrapmargin, wm default: wm=0

Defines a margin for automatic wrapover of text during input in *open* and *visual* modes. See *An Introduction to Text Editing with Vi* for details.

writeany, wa default: nowa

Inhibit the checks normally made before *write* commands, allowing a write to any file which the system protection mechanism will allow.

10. Limitations

Editor limits that the user is likely to encounter are as follows: 1024 characters per line, 256 characters per global command list, 128 characters per file name, 128 characters in the previous inserted and deleted text in *open* or *visual*, 100 characters in a shell escape command, 63 characters in a string valued option, and 30 characters in a tag name, and a limit of 250000 lines in the file is silently enforced.

‡‡ Version 3 only.

The *visual* implementation limits the number of macros defined with map to 32, and the total number of characters in macros to be less than 512.

Acknowledgments. Chuck Haley contributed greatly to the early development of *ex*. Bruce Englar encouraged the redesign which led to *ex* version 1. Bill Joy wrote versions 1 and 2.0 through 2.7, and created the framework that users see in the present editor. Mark Horton added macros and other features and made the editor work on a large number of terminals and Unix systems.

Ex changes – Version 3.1 to 3.5

This update describes the new features and changes which have been made in converting from version 3.1 to 3.5 of *ex*. Each change is marked with the first version where it appeared.

Update to Ex Reference Manual

Command line options

3.4 A new command called *view* has been created. *View* is just like *vi* but it sets *readonly*.

3.4 The encryption code from the v7 editor is now part of *ex*. You can invoke *ex* with the –x option and it will ask for a key, as *ed*. The *ed x* command (to enter encryption mode from within the editor) is not available. This feature may not be available in all instances of *ex* due to memory limitations.

Commands

3.4 Provisions to handle the new process stopping features of the Berkeley TTY driver have been added. A new command, *stop*, takes you out of the editor cleanly and efficiently, returning you to the shell. Resuming the editor puts you back in command or visual mode, as appropriate. If *autowrite* is set and there are outstanding changes, a write is done first unless you say "stop!".

3.4 A

:vi <file>

command from visual mode is now treated the same as a

:edit <file> or :ex <file>

command. The meaning of the *vi* command from *ex* command mode is not affected.

3.3 A new command mode command *xit* (abbreviated *x*) has been added. This is the same as *wq* but will not bother to write if there have been no changes to the file.

Options

3.4 A read only mode now lets you guarantee you won't clobber your file by accident. You can set the on/off option *readonly* (*ro*), and writes will fail unless you use an ! after the write. Commands such as *x, ZZ,* the autowrite option, and in general anything that writes is affected. This option is turned on if you invoke ex with the –R flag.

3.4 The *wrapmargin* option is now usable. The way it works has been completely revamped. Now if you go past the margin (even in the middle of a word) the entire word is erased and rewritten on the next line. This changes the semantics of the number given to wrapmargin. 0 still means off. Any other number is still a distance from the right edge of the screen, but this location is now the right edge of the area where wraps can take place, instead of the left edge. Wrapmargin now behaves much like fill/nojustify mode in *nroff*.

3.3 The options *w300, w1200,* and *w9600* can be set. They are synonyms for *window,* but only apply at 300, 1200, or 9600 baud, respectively. Thus you can specify you want a 12 line window at 300 baud and a 23 line window at 1200 baud in your EXINIT with

:set w300=12 w1200=23

3.3 The new option *timeout* (default on) causes macros to time out after one second. Turn it off and they will wait forever. This is useful if you want multi character macros, but if your terminal sends escape sequences for arrow keys, it will be necessary to hit escape twice to get a beep.

3.3 The new option *remap* (default on) causes the editor to attempt to map the result of a macro mapping again until the mapping fails. This makes it possible, say, to map q to # and #1 to something else and get q1 mapped to something else. Turning it off makes it possible to map ^L to l and map ^R to ^L without having ^R map to l.

3.3 The new (string) valued option *tags* allows you to specify a list of tag files, similar to the "path" variable of csh. The files are separated by spaces (which are entered preceded by a backslash) and are searched left to right. The default value is "tags /usr/lib/tags", which has the same effect as before. It is recommended that "tags" always be the first entry. On Ernie CoVax, /usr/lib/tags contains entries for the system defined library procedures from section 3 of the manual.

Environment enquiries

3.4 The editor now adopts the convention that a null string in the environment is the same as not being set. This applies to TERM, TERMCAP, and EXINIT.

Vi Tutorial Update

Deleted features

3.3 The "q" command from visual no longer works at all. You must use "Q" to get to ex command mode. The "q" command was deleted because of user complaints about hitting it by accident too often.

3.5 The provisions for changing the window size with a numeric prefix argument to certain visual commands have been deleted. The correct way to change the window size is to use the z command, for example z5<cr> to change the window to 5 lines.

3.3 The option "mapinput" is dead. It has been replaced by a much more powerful mechanism: ":map!".

Change in default option settings

3.3 The default window sizes have been changed. At 300 baud the window is now 8 lines (it was 1/2 the screen size). At 1200 baud the window is now 16 lines (it was 2/3 the screen size, which was usually also 16 for a typical 24 line CRT). At 9600 baud the window is still the full screen size. Any baud rate less than 1200 behaves like 300, any over 1200 like 9600. This change makes *vi* more usable on a large screen at slow speeds.

Vi commands

3.3 The command "ZZ" from vi is the same as ":x<cr>". This is the recommended way to leave the editor. Z must be typed twice to avoid hitting it accidently.

3.4 The command ^Z is the same as ":stop<cr>". Note that if you have an arrow key that sends ^Z the stop function will take priority over the arrow function. If you have your "susp" character set to something besides ^Z, that key will be honored as well.

3.3 It is now possible from visual to string several search expressions together separated by semicolons the same as command mode. For example, you can say

```
/foo/;/bar
```

from visual and it will move to the first "bar" after the next "foo". This also works within one line.

3.3 ^R is now the same as ^L on terminals where the right arrow key sends ^L (This includes the Televideo 912/920 and the ADM 31 terminals.)

3.4 The visual page motion commands ^F and ^B now treat any preceding counts as number of pages to move, instead of changes to the window size. That is, 2^F moves forward 2 pages.

Macros

3.3 The "mapinput" mechanism of version 3.1 has been replaced by a more powerful mechanism. An "!" can follow the word "**map**" in the *map* command. **Map!**'ed macros only apply during input mode, while **map**'ed macros only apply during command mode. Using "**map**" or "**map!**" by itself produces a listing of macros in the corresponding mode.

3.4 A word abbreviation mode is now available. You can define abbreviations with the *abbreviate* command

:abbr foo find outer otter

which maps "foo" to "find outer otter". Abbreviations can be turned off with the *unabbreviate* command. The syntax of these commands is identical to the *map* and *unmap* commands, except that the ! forms do not exist. Abbreviations are considered when in visual input mode only, and only affect whole words typed in, using the conservative definition. (Thus "foobar" will not be mapped as it would using "map!") Abbreviate and unabbreviate can be abbreviated to "ab" and "una", respectively.

Ex changes – Version 2.0 to 3.1

This update describes the new features and changes which have been made in converting from version 2.0 to 3.1 of *ex*. Each change is marked with the first version where it appeared. Versions 2.1 through 2.7 were implemented by Bill Joy; Mark Horton produced versions 2.8, 2.9 and 3.1 and is maintaining the current version.

Update to Ex Reference Manual

Command line options

2.1 Invoking *ex* via

    % ex -l

now sets the *lisp* and *showmatch* options. This is suitable for invocations from within *lisp*(1). If you don't like *showmatch* you can still use "ex -l" to get *lisp* set, just put the command "set noshowmatch" in your *.exrc* file.

3.1 Invoking *ex* with an argument **–wn** sets the value of the *window* option before starting; this is particularly suitable when invoking *vi*, thus

    % vi -w5 ex2.0-3.1

edits the file with a 5 line initial window.

2.9 The text after a + on the command line is no longer limited to being a line number, but can be any single command. This generality is also available within the editor on *edit* and *next* commands (but no blanks are allowed in such commands.) A very useful form of this option is exemplified by

    % vi +/main more.c

Command addressing

2.9 The address form % is short for "1,$".

Commands

2.2 The editor now ignores a ":" in front of commands, so you can say ":wq" even in command mode.

2.8 The *global* command now does something sensible when you say

    g/pat/

printing all lines containing *pat;* before this printed the first line after each line containing *pat*. The trailing / may be omitted here.

3.1 New commands *map* and *unmap* have been added which are used with macros in *visual* mode. These are described below.

3.1 The *next* command now admits an argument of the form "+command" as described above.

3.1 The *substitute* command, given no arguments, now repeats the previous *substitute,* just as "&" does. This is easier to type.

2.8 The substititute command "s/str", omitting the delimiter on the regular expression, now deletes "str"; previously this was an error.

2.9 During pattern searches of a *tag* command, the editor uses *nomagic* mode; previously a funny, undocumented mode of searching was used.

3.1 The editor requires that the tag names in the *tags* file be sorted.

2.3 The command *P* is a synonym for *print.*

2.9 The default starting address for *z* is **.+1**. If *z* is followed by a number, then this number is remembered by setting the *scroll* option.

2.9 A command consisting of only two addresses, e.g. "1,10" now causes all the lines to be printed, rather than just the last line.

Options

2.8 *Autowrite* (which can be abbreviated *aw*) is an on/off option, off by default. If you set this option, then the editor will perform *write* commands if the current file is modified and you give a *next,* ^^ (in *visual), !* or *tag* commands, (and noticeably not before *edit* commands.) Note that there is an equivalent way to do the command with autowrite set without the write in each case: *edit, :e #, shell* and *tag!* do not *autowrite.*

3.1 A new option *edcompatible* causes the presence or absence of *g* and *c* suffixes on *substitute* commands to be remembered, and to be toggled by repeating the suffices. The suffix *r* makes the substitution be as in the ~ command instead of like *&*.

2.8 There is a new *hardtabs* option, which is numeric and defaults to 8. Changing this to, say, 4, tells *ex* that either you system expands tabs to every 4 spaces, or your terminal has hardware tabs set every 4 spaces.

3.1 There is a new boolean option *mapinput* which is described with the macro facility for *visual* below.

2.9 Whether *ex* prompts for commands now depends only on the setting of the *prompt* variable so that you can say "set prompt" inside *script*(1) and get *ex* to prompt.

Environment enquiries

3.1 *Ex* will now execute initial commands from the EXINIT environment variable rather than *.exrc* if it find such a variable.

2.9 *Ex* will read the terminal description from the TERMCAP environment variable if the description there is the one for the TERM in the environment. TERMCAP may still be a pathname (starting with a /; in that case this will be used as the termcap file rather than /etc/termcap, and the terminal description will be sought there.)

Vi Tutorial Update

Change in default option settings.

3.1 The default setting for the *magic* option is now *magic*. Thus the characters

. [* ~

are special in scanning patterns in *vi*. You should

set nomagic

in your *.exrc* if you don't use these regularly. This makes *vi* default like *ex*. In a related change, *beautify* is no longer the default for *vi*.

Line wrap around

2.4 The *w W b B e* and *E* operations in visual now wrap around line boundaries. Thus a sequence of enough **w** commands will get to any word after the current position in the file, and b's will back up to any previous position. Thus these are more like the sentence operations *(* and *)*. (You still can't back around line boundaries during inserts however.)

2.3 The / and *?* searches now find the next or previous instance of the searched for string. Previously, they would not find strings on the current line. Thus you can move to the right on the current line by typing "/pref<ESC>" where "pref" is a prefix of the word you wish to move to, and delete to a following string "str" by doing "d/str<ESC>", whether it is on the same or a succeeding line. (Previously the command "d/pat/" deleted lines through the next line containing "pat". This can be accomplished now by the somewhat unusual command "d/pat/0", which is short for "d/pat/+0". The point is that whole lines are affected if the search patter only specifies a line, and using address arithmetic makes the pattern only specify a line.)

3.1 Arrow keys on terminals that send more than 1 character now work. Home up keys are supported as are the four directions. (Note that the HP 2621 will turn on function key labels, and even then you have to hold shift down to use the arrow keys. To avoid turning on the labels, and to give up the function keys, use terminal type 2621nl instead of 2621.)

Macros

3.1 A parameterless macro facility is included from visual. This facility lets you say that when you type a particular key, you really mean some longer sequence of keys. It is useful when you find yourself typing the same sequence of commands repeatedly.

Briefly, there are two flavors of macros:

a) Put the macro body in a buffer register, say x. Then type @x to invoke it. @ may be followed by another @ to repeat the last macro. This allows macros up to 512 chars.

b) Use the map command from command mode (typically in the *.exrc* file) as follows:

map *lhs rhs*

where *lhs* will be mapped to *rhs*. There are restrictions: *lhs* should be 1-keystroke (either 1 char or 1 function key) since it must be entered within 1 second. The *lhs* can be no longer than 10 chars, the *rhs* no longer than 100. To get space, tab, "|", or newline into *lhs* or *rhs,* escape them with ctrl V. (It may be necessary to escape the ctrl V with ctrl V if the map command is given from visual mode.) Spaces and tabs inside the *rhs* need not be escaped.

For example, to make the Q key write and exit the editor, you can do

:map Q :wq^VCR

which means that whenever you type 'Q', it will be as though you had typed the four characters :wqCR. The control V is needed because without it the return would end the colon command.

For 1 shot macros it is best to put the macro in a buffer register and map a key to '@r', since this will allow the macro to be edited.

Macros can be deleted with

unmap lhs

If the lhs of a macro is "#0" through "#9", this maps the particular function key instead of the 2 char # sequence, if the terminal has function keys. For terminals without function keys, the sequence #*x* means function key *x*, for any digit *x*. As a special case, on terminals without function keys, the #x sequence need not be typed within one second. The character # can be changed by using a macro in the usual way:

map ^V^I #

to use tab, for example. (This won't affect the map command, which still uses #, but just the invocation from visual mode.) The undo command will undo an entire macro call as a unit.

3.1 New commands in visual: ^Y and ^E. These scroll the screen up and down 1 line, respectively. They can be given counts, controlling the number of lines the screen is scrolled. They differ from ^U and ^D in that the cursor stays over the same line in the buffer it was over before rather than staying in the same place on the screen. (^Y on a dumb terminal with a full screen will redraw the screen moving the cursor up a few lines.) If you're looking for mnemonic value in the names, try this: Y is right next to U and E is right next to D.

Miscellaneous

3.1 In visual: '&' is a synonym for ':&<cr>'.

2.2 In input mode in open and visual ^V (like tenex) is now equivalent to ^Q (which is reminiscent of ITS) superquoting the next character.

2.8 The j, k, and l keys now move the cursor down, up, and right, respectively, in visual mode, as they used to do (and always did on some terminals). This is to avoid the creeping of these keys into the map descriptions of terminals and to compensate for the lack of arrow keys on some terminals.

2.5 The $ command now sets the column for future cursor motions to effective infinity. Thus a '$' followed by up/down cursor motions moves at the right margin of each line.

2.9 The way window sizes and scrolling commands are based on the options window and scroll has been rearranged. All command mode scrolling commands (z and ctrl D) are based on scroll: ^D moves scroll lines, z moves scroll*2 lines. Everything in visual (^D, ^U, ^F, ^B, z, window sizes in general) are based on the window option. The defaults are arranged so that everything seems as before, but on hardcopy terminals at 300 baud the default for scroll is 11 instead of 6.

Ex/Edit Command Summary (Version 2.0)

Ex and *edit* are text editors, used for creating and modifying files of text on the UNIX computer system. *Edit* is a variant of *ex* with features designed to make it less complicated to learn and use. In terms of command syntax and effect the editors are essentially identical, and this command summary applies to both.

The summary is meant as a quick reference for users already acquainted with *edit* or *ex*. Fuller explanations of the editors are available in the documents *Edit: A Tutorial* (a self-teaching introduction) and the *Ex Reference Manual* (the comprehensive reference source for both *edit* and *ex*). Both of these writeups are available in the Computing Services Library.

In the examples included with the summary, commands and text entered by the user are printed in **boldface** to distinguish them from responses printed by the computer.

The Editor Buffer

In order to perform its tasks the editor sets aside a temporary work space, called a *buffer*, separate from the user's permanent file. Before starting to work on an existing file the editor makes a copy of it in the buffer, leaving the original untouched. All editing changes are made to the buffer copy, which must then be written back to the permanent file in order to update the old version. The buffer disappears at the end of the editing session.

Editing: Command and Text Input Modes

During an editing session there are two usual modes of operation: *command* mode and *text input* mode. (This disregards, for the moment, *open* and *visual* modes, discussed below.) In command mode, the editor issues a colon prompt (:) to show that it is ready to accept and execute a command. In text input mode, on the other hand, there is no prompt and the editor merely accepts text to be added to the buffer. Text input mode is initiated by the commands *append*, *insert*, and *change*, and is terminated by typing a period as the first and only character on a line.

Line Numbers and Command Syntax

The editor keeps track of lines of text in the buffer by numbering them consecutively starting with 1 and renumbering as lines are added or deleted. At any given time the editor is positioned at one of these lines; this position is called the *current line*. Generally, commands that change the contents of the buffer print the new current line at the end of their execution.

Most commands can be preceded by one or two line-number addresses which indicate the lines to be affected. If one number is given the command operates on that line only; if two, on an inclusive range of lines. Commands that can take line-number prefixes also assume default prefixes if none are given. The default assumed by each command is designed to make it convenient to use in many instances without any line-number prefix. For the most part, a command used without a prefix operates on the current line, though exceptions to this rule should be noted. The *print* command by itself, for instance, causes one line, the current line, to be printed at the terminal.

The summary shows the number of line addresses that can be prefixed to each command as well as the defaults assumed if they are omitted. For example, *(.,.)* means that up to 2 line-numbers may be given, and that if none is given the command operates on the current line. (In the address prefix notation, "." stands for the current line and "$" stands for the last line of the buffer.) If no such notation appears, no line-number prefix may be used.

Some commands take trailing information; only the more important instances of this are mentioned in the summary.

Open and Visual Modes

Besides command and text input modes, *ex* and *edit* provide on some CRT terminals other modes of editing, *open* and *visual*. In these modes the cursor can be moved to individual words or characters in a line. The commands then given are very different from the standard editor commands; most do not appear on the screen when typed. *An Introduction to Display Editing with Vi* provides a full discussion.

Special Characters

Some characters take on special meanings when used in context searches and in patterns given to the *substitute* command. For *edit*, these are "^" and "$", meaning the beginning and end of a line, respectively. *Ex* has the following additional special characters:

. & * [ ] ~

To use one of the special characters as its simple graphic representation rather than with its special meaning, precede it by a backslash (\). The backslash always has a special meaning.

Name	Abbr	Description	Examples
(.)append	a	Begins text input mode, adding lines to the buffer after the line specified. Appending continues until "." is typed alone at the beginning of a new line, followed by a carriage return. *0a* places lines at the beginning of the buffer.	**:a** **Three lines of text** **are added to the buffer** **after the current line.** **.** :
(.,.)change	c	Deletes indicated line(s) and initiates text input mode to replace them with new text which follows. New text is terminated the same way as with *append*.	**:5,6c** **Lines 5 and 6 are** **deleted and replaced by** **these three lines.** **.** :
(.,.)copy *addr*	co	Places a copy of the specified lines after the line indicated by *addr*. The example places a copy of lines 8 through 12, inclusive, after line 25.	**:8,12co 25** Last line copied is printed :
(.,.)delete	d	Removes lines from the buffer and prints the current line after the deletion.	**:13,15d** New current line is printed :
edit *file* edit! *file*	e e!	Clears the editor buffer and then copies into it the named *file*, which becomes the current file. This is a way of shifting to a different file without leaving the editor. The editor issues a warning message if this command is used before saving changes made to the file already in the buffer; using the form e! overrides this protective mechanism.	**:e ch10** No write since last change **:e! ch10** "ch10" 3 lines, 62 characters :
file *name*	f	If followed by a *name*, renames the current file to *name*. If used without *name*, prints the name of the current file.	**:f ch9** "ch9" [Modified] 3 lines ... **:f** "ch9" [Modified] 3 lines ... :
(1,$)global (1,$)global!	g g! or v	global/*pattern*/*commands* Searches the entire buffer (unless a smaller range is specified by line-number prefixes) and executes *commands* on every line with an expression matching *pattern*. The second form, abbreviated either **g!** or **v**, executes *commands* on lines that *do not* contain the expression *pattern*.	**:g/nonsense/d** :
(.)insert	i	Inserts new lines of text immediately before the specified line. Differs from *append* only in that text is placed before, rather than after, the indicated line. In other words, **1i** has the same effect as **0a**.	**:1i** **These lines of text will** **be added prior to line 1.** **.** :
(.,.+1)join	j	Join lines together, adjusting white space (spaces and tabs) as necessary.	**:2,5j** Resulting line is printed :

Name	Abbr	Description	Examples
(.,.)list	l	Prints lines in a more unambiguous way than the *print* command does. The end of a line, for example, is marked with a "$", and tabs printed as "^I".	:9l This is line 9$:
(.,.)move *addr*	m	Moves the specified lines to a position after the line indicated by *addr*.	:12,15m 25 New current line is printed :
(.,.)number	nu	Prints each line preceded by its buffer line number.	:nu 10 This is line 10 :
(.)open	o	Too involved to discuss here, but if you enter open mode accidentally, press the ESC key followed by **q** to get back into normal editor command mode. *Edit* is designed to prevent accidental use of the open command.	
preserve	pre	Saves a copy of the current buffer contents as though the system had just crashed. This is for use in an emergency when a *write* command has failed and you don't know how else to save your work.†	:preserve File preserved. :
(.,.)print	p	Prints the text of line(s).	:+2,+3p The second and third lines after the current line :
quit quit!	q q!	Ends the editing session. You will receive a warning if you have changed the buffer since last writing its contents to the file. In this event you must either type w to write, or type q! to exit from the editor without saving your changes.	:q No write since last change :q! %
(.)read *file*	r	Places a copy of *file* in the buffer after the specified line. Address 0 is permissible and causes the copy of *file* to be placed at the beginning of the buffer. The *read* command does not erase any text already in the buffer. If no line number is specified, *file* is placed after the current line.	:0r newfile "newfile" 5 lines, 86 characters :
recover *file*	rec	Retrieves a copy of the editor buffer after a system crash, editor crash, phone line disconnection, or *preserve* command.	
(.,.)substitute	s	**substitute/*pattern*/*replacement*/** **substitute/*pattern*/*replacement*/gc** Replaces the first occurrence of *pattern* on a line with *replacement*. Including a **g** after the command changes all occurrences of *pattern* on the line. The **c** option allows the user to confirm each substitution before it is made; see the manual for details.	:3p Line 3 contains a misstake :s/misstake/mistake/ Line 3 contains a mistake :

† Seek assistance from a consultant as soon as possible after saving a file with the *preserve* command, because the file is saved on system storage space for only one week.

Name	Abbr	Description	Examples
undo	**u**	Reverses the changes made in the buffer by the last buffer-editing command. Note that this example contains a notification about the number of lines affected.	**:1,15d** 15 lines deleted new line number 1 is printed **:u** 15 more lines in file ... old line number 1 is printed :
(1,$)**write** *file* (1,$)**write!** *file*	**w** **w!**	Copies data from the buffer onto a permanent file. If no *file* is named, the current filename is used. The file is automatically created if it does not yet exist. A response containing the number of lines and characters in the file indicates that the write has been completed successfully. The editor's built-in protections against overwriting existing files will in some circumstances inhibit a write. The form **w!** forces the write, confirming that an existing file is to be overwritten.	**:w** "file7" 64 lines, 1122 characters **:w file8** "file8" File exists ... **:w! file8** "file8" 64 lines, 1122 characters :
(.)**z** *count*	**z**	Prints a screen full of text starting with the line indicated; or, if *count* is specified, prints that number of lines. Variants of the *z* command are described in the manual.	
!*command*		Executes the remainder of the line after ! as a UNIX command. The buffer is unchanged by this, and control is returned to the editor when the execution of *command* is complete.	**:!date** Fri Jun 9 12:15:11 PDT 1978 ! :
control-d		Prints the next *scroll* of text, normally half of a screen. See the manual for details of the *scroll* option.	
(.+1)<cr>		An address alone followed by a carriage return causes the line to be printed. A carriage return by itself prints the line following the current line.	:<cr> the line after the current line :
/*pattern*/		Searches for the next line in which *pattern* occurs and prints it.	**:/This pattern/** This pattern next occurs here. :
//		Repeats the most recent search.	**://** This pattern also occurs here. :
?*pattern*?		Searches in the reverse direction for *pattern*.	
??		Repeats the most recent search, moving in the reverse direction through the buffer.	

Ex differences – version 1.1 to 2.0

This sheet summarizes the differences between the old version 1.1 of *ex* and the new version 2.0. The new *ex* is available as the standard *ex* on the VAX on the 5th floor of Evans, and as a new and experimental version in /usr/new on the Cory 11/70. It will soon be available in /usr/new on the Computer Center and Ingres Machines. Send problems over the Berkeley network to "vax:bill".

Changes to existing features

Options.

The options *editany, edited, fork, hush, printall* and *sticky* have been deleted because of lack of use. The *notify* option has been renamed *report.*

The *home* option will soon be superseded by the environment feature of version 7 UNIX and has been deleted. Similarly the *mode* option is superseded by the *umask* of version 7 and has also been deleted.

The *visualmessage* option has been deleted; use "mesg n" at the system command level to inhibit interconsole messages.

The *iul* option is replaced by a more general mechanism which allows portions of the buffer to be processed through specified commands; you can get *iul* processing on lines 1 to 100 of a file by doing "1,100!iul". This replaces the lines 1 to 100 by the output of an *iul* command, giving the command these lines as input.

Invocation

The options **–o, –n** and **–p** have been deleted.

Filename formation

The alternate filename is now represented as '#' rather than '`', since '`' is a shell metacharacter. The editor now uses a shell to expand filenames containing shell metacharacters. If you use *csh,* then you can use all the shell metasyntax in forming new filenames, including home directory references with '~' and variables you define in *.cshrc* using '$'.

Character representation

Control characters are now represented as '^*x*'; thus a control X is printed as '^X'; the delete character is represented '^?'.

Command changes

There have been major changes to open/visual (incompatible ones are described below).

It is no longer possible to discard changes by repeating the *quit* command twice. You must use the variant form **quit!** to get out of the editor discarding changes. Similarly the variant forms **e!** and **next!** must be used to edit a new file or the next file without saving changes you have made.†

A new form of the '!' shell escape replaces the *expand* and *tabulate* commands. Thus the command "1,10expand" of the old version is replaced by "1,10!expand" in the new. Note also that the command abbreviation **ta** no longer refers to the *tabulate* command, which has been deleted, but rather refers to the new *tag* command.

† Less useful are **rewind!** and **recover!**.

The format of the *args* command has been changed; the files are no longer numbered, rather the entire argument list is always printed with the current file name enclosed by '[' and ']'.

The format of the *file* command output has been changed; the editor says '[Not edited]' in the rare case that this is true rather than saying '[Edited]'. The command also gives the percentage of the way into the buffer that the current line is.

The format of the *set* command has been improved; "set all" now prints in a three column format. The commands "set %", "set !" and "set `" have been deleted. The command "set" now prints in a one line format rather than down the screen.

The commands *echo, expand, help, reset, sync, tabulate* and *xpand* have been deleted.

Changes to open and visual

A large number of changes have been made to open and visual; we summarize only the most noticeable ones here. See the attached reference card for more information, and (even if you know how to use visual already) you should look at *An Introduction to Text Editing with Vi*. We do not discuss any of the new commands in visual here.†

The delete line command is now **dd** rather than \\ (\\ no longer works!.) In fact, **d** and other operators can now operate on lines; thus **dL** deletes to the last line on the screen. The shift commands < and > are now operators, thus << and >> now have the effect that < and > used to have.

The command **v** has been deleted; only its synonym **z** remains. The **K** operation has been moved to **m**; **K** has no meaning in the new version. The **^S** operation has been deleted, but **^G** does a sync, and also prints some information. The **^W** operation has been deleted (use **B**). The #, @ and **^X** operations have been deleted. To delete to the beginning of the line use **d0**; the commands and **x** and **X** are similar to #.

During inputs, **^W** backs up like **b** rather than **B**.

Terminal support has been vastly improved; the editor will now drive most any display terminal, using all terminal features such as cursor addressing, clear to end of line, insert and delete line and insert and delete character. To help performance on slow terminals some options are now set based on the intelligence and speed of the terminal; in particular, the default window size is 1/2 a full screen at 300 baud, or 2/3 of a full screen at 1200 baud.

† It is now possible to edit with the focus of the editing being *visual* using a command *vi* rather than *ex* on the command line, and using a new **:** command from within visual to run command mode commands.

Advanced Editing on UNIX

Brian W. Kernighan

ABSTRACT

This paper is meant to help secretaries, typists and programmers to make effective use of the UNIX† facilities for preparing and editing text. It provides explanations and examples of

- special characters, line addressing and global commands in the editor **ed**;
- commands for "cut and paste" operations on files and parts of files, including the **mv**, **cp**, **cat** and **rm** commands, and the **r**, **w**, **m** and **t** commands of the editor;
- editing scripts and editor-based programs like **grep** and **sed**.

Although the treatment is aimed at non-programmers, new users with any background should find helpful hints on how to get their jobs done more easily.

November 14, 1986

† UNIX is a trademark of Bell Laboratories.

Advanced Editing on UNIX

Brian W. Kernighan

1. INTRODUCTION

Although UNIX provides remarkably effective tools for text editing, that by itself is no guarantee that everyone will automatically make the most effective use of them. In particular, people who are not computer specialists — typists, secretaries, casual users — often use the system less effectively than they might.

This document is intended as a sequel to *A Tutorial Introduction to the UNIX Text Editor* [1], providing explanations and examples of how to edit with less effort. (You should also be familiar with the material in *UNIX For Beginners* [2].) Further information on all commands discussed here can be found in *The UNIX Programmer's Manual* [3].

Examples are based on observations of users and the difficulties they encounter. Topics covered include special characters in searches and substitute commands, line addressing, the global commands, and line moving and copying. There are also brief discussions of effective use of related tools, like those for file manipulation, and those based on **ed**, like **grep** and **sed**.

A word of caution. There is only one way to learn to use something, and that is to *use* it. Reading a description is no substitute for trying something. A paper like this one should give you ideas about what to try, but until you actually try something, you will not learn it.

2. SPECIAL CHARACTERS

The editor **ed** is the primary interface to the system for many people, so it is worthwhile to know how to get the most out of **ed** for the least effort.

The next few sections will discuss shortcuts and labor-saving devices. Not all of these will be instantly useful to any one person, of course, but a few will be, and the others should give you ideas to store away for future use. And as always, until you try these things, they will remain theoretical knowledge, not something you have confidence in.

The List command 'l'

ed provides two commands for printing the contents of the lines you're editing. Most people are familiar with **p**, in combinations like

```
1,$p
```

to print all the lines you're editing, or

```
s/abc/def/p
```

to change 'abc' to 'def' on the current line. Less familiar is the *list* command **l** (the letter '*l*'), which gives slightly more information than **p**. In particular, **l** makes visible characters that are normally invisible, such as tabs and backspaces. If you list a line that contains some of these, **l** will print each tab as $\geqslant$ and each backspace as $\lessdot$. This makes it much easier to correct the sort of typing mistake that inserts extra spaces adjacent to tabs, or inserts a backspace followed by a space.

The **l** command also 'folds' long lines for printing — any line that exceeds 72 characters is printed on multiple lines; each printed line except the last is terminated by a backslash \, so you can tell it was folded. This is useful for printing long lines on short terminals.

Occasionally the **l** command will print in a line a string of numbers preceded by a backslash, such as \07 or \16. These combinations are used to make visible characters that normally don't print, like form feed or vertical tab or bell. Each such combination is a single character. When you see such characters, be wary — they may have surprising meanings when printed on some terminals. Often their presence means that your finger slipped while you were typing; you almost never want them.

The Substitute Command 's'

Most of the next few sections will be taken up with a discussion of the substitute command **s**. Since this is the command for changing the contents of individual lines, it probably has the most complexity of any **ed** command, and the most potential for effective use.

As the simplest place to begin, recall the meaning of a trailing **g** after a substitute command. With s/this/that/

and

    s/this/that/g

the first one replaces the *first* 'this' on the line with 'that'. If there is more than one 'this' on the line, the second form with the trailing **g** changes *all* of them.

Either form of the **s** command can be followed by **p** or **l** to 'print' or 'list' (as described in the previous section) the contents of the line:

    s/this/that/p
    s/this/that/l
    s/this/that/gp
    s/this/that/gl

are all legal, and mean slightly different things. Make sure you know what the differences are.

Of course, any **s** command can be preceded by one or two 'line numbers' to specify that the substitution is to take place on a group of lines. Thus

    1,$s/mispell/misspell/

changes the *first* occurrence of 'mispell' to 'misspell' on every line of the file. But

    1,$s/mispell/misspell/g

changes *every* occurrence in every line (and this is more likely to be what you wanted in this particular case).

You should also notice that if you add a **p** or **l** to the end of any of these substitute commands, only the last line that got changed will be printed, not all the lines. We will talk later about how to print all the lines that were modified.

The Undo Command 'u'

Occasionally you will make a substitution in a line, only to realize too late that it was a ghastly mistake. The 'undo' command **u** lets you 'undo' the last substitution: the last line that was substituted can be restored to its previous state by typing the command

    u

The Metacharacter '.'

As you have undoubtedly noticed when you use **ed**, certain characters have unexpected meanings when they occur in the left side of a substitute command, or in a search for a particular line. In the next several sections, we will talk about these special characters, which are often called 'metacharacters'.

The first one is the period '.'. On the left side of a substitute command, or in a search with '/.../', '.' stands for *any* single character. Thus the search

    /x.y/

finds any line where 'x' and 'y' occur separated by a single character, as in

    x+y
    x-y
    x□y
    x.y

and so on. (We will use □ to stand for a space whenever we need to make it visible.)

Since '.' matches a single character, that gives you a way to deal with funny characters printed by **l**. Suppose you have a line that, when printed with the **l** command, appears as

    .... th\07is ....

and you want to get rid of the \07 (which represents the bell character, by the way).

The most obvious solution is to try

    s/\07//

but this will fail. (Try it.) The brute force solution, which most people would now take, is to retype the entire line. This is guaranteed, and is actually quite a reasonable tactic if the line in question isn't too big, but for a very long line, retyping is a bore. This is where the metacharacter '.' comes in handy. Since '\07' really represents a single character, if we say

    s/th.is/this/

the job is done. The '.' matches the mysterious character between the 'h' and the 'i', *whatever it is.*

Bear in mind that since '.' matches any single character, the command

    s/./,/

converts the first character on a line into a ',', which very often is not what you intended.

As is true of many characters in **ed**, the '.' has several meanings, depending on its context. This line shows all three:

The first '.' is a line number, the number of the line we are editing, which is called 'line dot'. (We will discuss line dot more in Section 3.) The second '.' is a metacharacter that matches any single character on that line. The third '.' is the only one that really is an honest literal period. On the *right* side of a substitution, '.' is not special. If you apply this command to the line

Now is the time.

the result will be

which is probably not what you intended.

The Backslash '\'

Since a period means 'any character', the question naturally arises of what to do when you really want a period. For example, how do you convert the line

Now is the time.

into

Now is the time?

The backslash '\' does the job. A backslash turns off any special meaning that the next character might have; in particular, '\.' converts the '.' from a 'match anything' into a period, so you can use it to replace the period in

Now is the time.

like this:

```
s/\./?/
```

The pair of characters '\.' is considered by **ed** to be a single real period.

The backslash can also be used when searching for lines that contain a special character. Suppose you are looking for a line that contains

The search

```
/.PP/
```

isn't adequate, for it will find a line like

THE APPLICATION OF ...

because the '.' matches the letter 'A'. But if you say

```
/\.PP/
```

you will find only lines that contain '.PP'.

The backslash can also be used to turn off special meanings for characters other than '.'. For example, consider finding a line that contains a backslash. The search

```
/\/
```

won't work, because the '\' isn't a literal '\', but instead means that the second '/' no longer delimits the search. But by preceding a backslash with another one, you can search for a literal backslash. Thus

```
/\\/
```

does work. Similarly, you can search for a forward slash '/' with

```
/\//
```

The backslash turns off the meaning of the immediately following '/' so that it doesn't terminate the /.../ construction prematurely.

As an exercise, before reading further, find two substitute commands each of which will convert the line

```
\x\.\y
```

into the line

```
\x\y
```

Here are several solutions; verify that each works as advertised.

```
s/\\\.//
s/x../x/
s/..y/y/
```

A couple of miscellaneous notes about backslashes and special characters. First, you can use any character to delimit the pieces of an **s** command: there is nothing sacred about slashes. (But you must use slashes for context searching.) For instance, in a line that contains a lot of slashes already, like

//exec //sys.fort.go // etc...

you could use a colon as the delimiter — to delete all the slashes, type

```
s:/::g
```

Second, if # and @ are your character erase and line kill characters, you have to type \# and \@; this is true whether you're talking to **ed** or any other program.

When you are adding text with **a** or **i** or **c**, backslash is not special, and you should only put in one backslash for each one you really want.

The Dollar Sign '$'

The next metacharacter, the '$', stands for 'the end of the line'. As its most obvious use, suppose you have the line

Now is the

and you wish to add the word 'time' to the end. Use the **$** like this:

```
s/$/□time/
```

to get

Now is the time

Notice that a space is needed before 'time' in the substitute command, or you will get

Now is thetime

As another example, replace the second comma in the following line with a period without altering the first:

Now is the time, for all good men,

The command needed is

s/,$/./

The **$** sign here provides context to make specific which comma we mean. Without it, of course, the **s** command would operate on the first comma to produce

Now is the time. for all good men,

As another example, to convert

Now is the time.

into

Now is the time?

as we did earlier, we can use

s/.$/?/

Like '.', the '$' has multiple meanings depending on context. In the line

$s/$/$/

the first '$' refers to the last line of the file, the second refers to the end of that line, and the third is a literal dollar sign, to be added to that line.

The Circumflex '^'

The circumflex (or hat or caret) '^' stands for the beginning of the line. For example, suppose you are looking for a line that begins with 'the'. If you simply say

/the/

you will in all likelihood find several lines that contain 'the' in the middle before arriving at the one you want. But with

/^the/

you narrow the context, and thus arrive at the desired one more easily.

The other use of '^' is of course to enable you to insert something at the beginning of a line:

s/^/□/

places a space at the beginning of the current line.

Metacharacters can be combined. To search for a line that contains *only* the characters

you can use the command

/^\.PP$/

The Star '*'

Suppose you have a line that looks like this:

text x   y *text*

where *text* stands for lots of text, and there are some indeterminate number of spaces between the **x** and the **y**. Suppose the job is to replace all the spaces between **x** and **y** by a single space. The line is too long to retype, and there are too many spaces to count. What now?

This is where the metacharacter '*' comes in handy. A character followed by a star stands for as many consecutive occurrences of that character as possible. To refer to all the spaces at once, say

s/x□□*y/x□y/

The construction '□*' means 'as many spaces as possible'. Thus 'x□*y' means 'an x, as many spaces as possible, then a y'.

The star can be used with any character, not just space. If the original example was instead

text x————y *text*

then all '–' signs can be replaced by a single space with the command

s/x–*y/x□y/

Finally, suppose that the line was

text x..................y *text*

Can you see what trap lies in wait for the unwary? If you blindly type

s/x.*y/x□y/

what will happen? The answer, naturally, is that it depends. If there are no other x's or y's on the line, then everything works, but it's blind luck, not good management. Remember that '.' matches *any* single character? Then '.*' matches as many single characters as possible, and unless you're careful, it can eat up a lot more of the line than you expected. If the line was, for example, like this:

text x *text* x................y *text* y *text*

then saying

s/x.*y/x□y/

will take everything from the *first* 'x' to the *last* 'y', which, in this example, is undoubtedly more than you wanted.

The solution, of course, is to turn off the special meaning of '.' with '\.':

s/x\.*y/x□y/

Now everything works, for '\.*' means 'as many *periods* as possible'.

There are times when the pattern '.*' is exactly what you want. For example, to change

Now is the time for all good men ....

into

Now is the time.

use '.*' to eat up everything after the 'for':

s/□for.*/./

There are a couple of additional pitfalls associated with '*' that you should be aware of. Most notable is the fact that 'as many as possible' means *zero* or more. The fact that zero is a legitimate possibility is sometimes rather surprising. For example, if our line contained

text xy *text* x   y *text*

and we said

s/x□□*y/x□y/

the *first* 'xy' matches this pattern, for it consists of an 'x', zero spaces, and a 'y'. The result is that the substitute acts on the first 'xy', and does not touch the later one that actually contains some intervening spaces.

The way around this, if it matters, is to specify a pattern like

/x□□□*y/

which says 'an x, a space, then as many more spaces as possible, then a y', in other words, one or more spaces.

The other startling behavior of '*' is again related to the fact that zero is a legitimate number of occurrences of something followed by a star. The command

s/x*/y/g

when applied to the line

abcdef

produces

yaybycydyeyfy

which is almost certainly not what was intended. The reason for this behavior is that zero is a legal number of matches, and there are no x's at the beginning of the line (so that gets converted into a 'y'), nor between the 'a' and the 'b' (so that gets converted into a 'y'), nor ... and so on. Make sure you really want zero matches; if not, in this case write

s/xx*/y/g

'xx*' is one or more x's.

The Brackets '[ ]'

Suppose that you want to delete any numbers that appear at the beginning of all lines of a file. You might first think of trying a series of commands like

1,$s/^1*//
1,$s/^2*//
1,$s/^3*//

and so on, but this is clearly going to take forever if the numbers are at all long. Unless you want to repeat the commands over and over until finally all numbers are gone, you must get all the digits on one pass. This is the purpose of the brackets [and].

The construction

[0123456789]

matches any single digit — the whole thing is called a 'character class'. With a character class, the job is easy. The pattern '[0123456789]*' matches zero or more digits (an entire number), so

1,$s/^[0123456789]*//

deletes all digits from the beginning of all lines.

Any characters can appear within a character class, and just to confuse the issue there are essentially no special characters inside the brackets; even the backslash doesn't have a special meaning. To search for special characters, for example, you can say

/[.\$^[]/

Within [...], the '[' is not special. To get a ']' into a character class, make it the first character.

It's a nuisance to have to spell out the digits, so you can abbreviate them as [0–9]; similarly, [a–z] stands for the lower case letters, and [A–Z] for upper case.

As a final frill on character classes, you can specify a class that means 'none of the following characters'. This is done by beginning the class with a '^':

[^0–9]

stands for 'any character *except* a digit'. Thus you might find the first line that doesn't begin with a tab or space by a search like

/^[^(space)(tab)]/

Within a character class, the circumflex has a special meaning only if it occurs at the beginning. Just to convince yourself, verify that

/^[^^]/

finds a line that doesn't begin with a circumflex.

The Ampersand '&'

The ampersand '&' is used primarily to save typing. Suppose you have the line

Now is the time

and you want to make it

Now is the best time

Of course you can always say

s/the/the best/

but it seems silly to have to repeat the 'the'. The '&' is used to eliminate the repetition. On the *right* side of a substitute, the ampersand means 'whatever was just matched', so you can say

s/the/& best/

and the '&' will stand for 'the'. Of course this isn't much of a saving if the thing matched is just 'the', but if it is something truly long or awful, or if it is something like '.*' which matches a lot of text, you can save some tedious typing. There is also much less chance of making a typing error in the replacement text. For example, to parenthesize a line, regardless of its length,

s/.*/(&)/

The ampersand can occur more than once on the right side:

s/the/& best and & worst/

makes

Now is the best and the worst time

and

s/.*/&? &!!/

converts the original line into

Now is the time? Now is the time!!

To get a literal ampersand, naturally the backslash is used to turn off the special meaning:

s/ampersand/\&/

converts the word into the symbol. Notice that '&' is not special on the left side of a substitute, only on the *right* side.

Substituting Newlines

ed provides a facility for splitting a single line into two or more shorter lines by 'substituting in a newline'. As the simplest example, suppose a line has gotten unmanageably long because of editing (or merely because it was unwisely typed). If it looks like

text xy *text*

you can break it between the 'x' and the 'y' like this:

```
s/xy/x\
y/
```

This is actually a single command, although it is typed on two lines. Bearing in mind that '\' turns off special meanings, it seems relatively intuitive that a '\' at the end of a line would make the newline there no longer special.

You can in fact make a single line into several lines with this same mechanism. As a large example, consider underlining the word 'very' in a long line by splitting 'very' onto a separate line, and preceding it by the **roff** or **nroff** formatting command '.ul'.

text a very big *text*

The command

```
s/□very□/\
very\
/
```

converts the line into four shorter lines, preceding the word 'very' by the line '.ul', and eliminating the spaces around the 'very', all at the same time.

When a newline is substituted in, dot is left pointing at the last line created.

Joining Lines

Lines may also be joined together, but this is done with the **j** command instead of **s**. Given the lines

Now is
□the time

and supposing that dot is set to the first of them, then the command

j

joins them together. No blanks are added, which is why we carefully showed a blank at the beginning of the second line.

All by itself, a **j** command joins line dot to line dot+1, but any contiguous set of lines can be joined. Just specify the starting and ending line numbers. For example,

1,$jp

joins all the lines into one big one and prints it. (More on line numbers in Section 3.)

Rearranging a Line with \(... \)

(This section should be skipped on first reading.) Recall that '&' is a shorthand that stands for whatever was matched by the left side of an **s** command. In much the same way you can capture separate pieces of what was matched; the only difference is that you have to specify on the left side just what pieces you're interested in.

Suppose, for instance, that you have a file of lines that consist of names in the form

```
Smith, A. B.
Jones, C.
```

and so on, and you want the initials to precede the name, as in

```
A. B. Smith
C. Jones
```

It is possible to do this with a series of editing commands, but it is tedious and error-prone. (It is instructive to figure out how it is done, though.)

The alternative is to 'tag' the pieces of the pattern (in this case, the last name, and the initials), and then rearrange the pieces. On the left side of a substitution, if part of the pattern is enclosed between \(and \), whatever matched that part is remembered, and available for use on the right side. On the right side, the symbol '\1' refers to whatever matched the first \(...\) pair, '\2' to the second \(...\), and so on.

The command

```
1,$s/^\([^,]*\),*\(.*\)/\2\1/
```

although hard to read, does the job. The first \(...\) matches the last name, which is any string up to the comma; this is referred to on the right side with '\1'. The second \(...\) is whatever follows the comma and any spaces, and is referred to as '\2'.

Of course, with any editing sequence this complicated, it's foolhardy to simply run it and hope. The global commands **g** and **v** discussed in section 4 provide a way for you to print exactly those lines which were affected by the substitute command, and thus verify that it did what you wanted in all cases.

3. LINE ADDRESSING IN THE EDITOR

The next general area we will discuss is that of line addressing in **ed**, that is, how you specify what lines are to be affected by editing commands. We have already used constructions like

```
1,$s/x/y/
```

to specify a change on all lines. And most users are long since familiar with using a single newline (or return) to print the next line, and with

```
/thing/
```

to find a line that contains 'thing'. Less familiar, surprisingly enough, is the use of

```
?thing?
```

to scan *backwards* for the previous occurrence of 'thing'. This is especially handy when you realize that the thing you want to operate on is back up the page from where you are currently editing.

The slash and question mark are the only characters you can use to delimit a context search, though you can use essentially any character in a substitute command.

Address Arithmetic

The next step is to combine the line numbers like '.', '$', '/.../' and '?...?' with '+' and '−'. Thus

```
$−1
```

is a command to print the next to last line of the current file (that is, one line before line '$'). For example, to recall how far you got in a previous editing session,

```
$−5,$p
```

prints the last six lines. (Be sure you understand why it's six, not five.) If there aren't six, of course, you'll get an error message.

As another example,

prints from three lines before where you are now (at line dot) to three lines after, thus giving you a bit of context. By the way, the '+' can be omitted:

is absolutely identical in meaning.

Another area in which you can save typing effort in specifying lines is to use '−' and '+' as line numbers by themselves.

```
−
```

by itself is a command to move back up one line in the file. In fact, you can string several minus signs together to move back up that many lines:

```
−−−
```

moves up three lines, as does '−3'. Thus

```
−3,+3p
```

is also identical to the examples above.

Since '−' is shorter than '.−1', constructions like

```
−,.s/bad/good/
```

are useful. This changes 'bad' to 'good' on the previous line and on the current line.

'+' and '–' can be used in combination with searches using '/.../' and '?...?', and with '**$**'. The search

/thing/—

finds the line containing 'thing', and positions you two lines before it.

Repeated Searches

Suppose you ask for the search

/horrible thing/

and when the line is printed you discover that it isn't the horrible thing that you wanted, so it is necessary to repeat the search again. You don't have to re-type the search, for the construction

//

is a shorthand for 'the previous thing that was searched for', whatever it was. This can be repeated as many times as necessary. You can also go backwards:

??

searches for the same thing, but in the reverse direction.

Not only can you repeat the search, but you can use '//' as the left side of a substitute command, to mean 'the most recent pattern'.

```
/horrible thing/
 .... ed prints line with 'horrible thing' ...
s//good/p
```

To go backwards and change a line, say

??s//good/

Of course, you can still use the '&' on the right hand side of a substitute to stand for whatever got matched:

//s//&B&/p

finds the next occurrence of whatever you searched for last, replaces it by two copies of itself, then prints the line just to verify that it worked.

Default Line Numbers and the Value of Dot

One of the most effective ways to speed up your editing is always to know what lines will be affected by a command if you don't specify the lines it is to act on, and on what line you will be positioned (i.e., the value of dot) when a command finishes. If you can edit without specifying unnecessary line numbers, you can save a lot of typing.

As the most obvious example, if you issue a search command like

/thing/

you are left pointing at the next line that contains 'thing'. Then no address is required with commands like **s** to make a substitution on that line, or **p** to print it, or **l** to list it, or **d** to delete it, or **a** to append text after it, or **c** to change it, or **i** to insert text before it.

What happens if there was no 'thing'? Then you are left right where you were — dot is unchanged. This is also true if you were sitting on the only 'thing' when you issued the command. The same rules hold for searches that use '?...?'; the only difference is the direction in which you search.

The delete command **d** leaves dot pointing at the line that followed the last deleted line. When line '**$**' gets deleted, however, dot points at the *new* line '**$**'.

The line-changing commands **a**, **c** and **i** by default all affect the current line — if you give no line number with them, **a** appends text after the current line, **c** changes the current line, and **i** inserts text before the current line.

a, **c**, and **i** behave identically in one respect — when you stop appending, changing or inserting, dot points at the last line entered. This is exactly what you want for typing and editing on the fly. For example, you can say

```
a
 ... text ...
 ... botch ...            (minor error)
s/botch/correct/          (fix botched line)
a
 ... more text ...
```

without specifying any line number for the substitute command or for the second append command. Or you can say

```
a
 ... text ...
 ... horrible botch ...   (major error)
c                         (replace entire line)
 ... fixed up line ...
```

You should experiment to determine what happens if you add *no* lines with **a**, **c** or **i**.

The **r** command will read a file into the text being edited, either at the end if you give no address, or after the specified line if you do. In either case, dot points at the last line read in. Remember that you can even say **0r** to read a file in at the beginning of the text. (You can also say **0a** or **1i** to start adding text at the beginning.)

The **w** command writes out the entire file. If you precede the command by one line number, that line is written, while if you precede it by two line numbers, that range of lines is written. The **w** command does *not* change dot: the current line remains the same, regardless of what lines are written. This is true even if you say something like

```
/^\.AB/,/^\.AE/w abstract
```

which involves a context search.

Since the **w** command is so easy to use, you should save what you are editing regularly as you go along just in case the system crashes, or in case you do something foolish, like clobbering what you're editing.

The least intuitive behavior, in a sense, is that of the **s** command. The rule is simple — you are left sitting on the last line that got changed. If there were no changes, then dot is unchanged.

To illustrate, suppose that there are three lines in the buffer, and you are sitting on the middle one:

```
x1
x2
x3
```

Then the command

```
-,+s/x/y/p
```

prints the third line, which is the last one changed. But if the three lines had been

```
x1
y2
y3
```

and the same command had been issued while dot pointed at the second line, then the result would be to change and print only the first line, and that is where dot would be set.

Semicolon ';'

Searches with '/.../' and '?...?' start at the current line and move forward or backward respectively until they either find the pattern or get back to the current line. Sometimes this is not what is wanted. Suppose, for example, that the buffer contains lines like this:

```
.
.
ab
.
.
.
bc
.
.
```

Starting at line **1**, one would expect that the command

```
/a/,/b/p
```

prints all the lines from the 'ab' to the 'bc' inclusive. Actually this is not what happens. *Both* searches (for 'a' and for 'b') start from the same point, and thus they both find the line that contains 'ab'. The result is to print a single line. Worse, if there had been a line with a 'b' in it before the 'ab' line, then the print command would be in error, since the second line number would be less than the first, and it is illegal to try to print lines in reverse order.

This is because the comma separator for line numbers doesn't set dot as each address is processed; each search starts from the same place. In **ed**, the semicolon ';' can be used just like comma, with the single difference that use of a semicolon forces dot to be set at that point as the line numbers are being evaluated. In effect, the semicolon 'moves' dot. Thus in our example above, the command

```
/a/;/b/p
```

prints the range of lines from 'ab' to 'bc', because after the 'a' is found, dot is set to that line, and then 'b' is searched for, starting beyond that line.

This property is most often useful in a very simple situation. Suppose you want to find the *second* occurrence of 'thing'. You could say

```
/thing/
//
```

but this prints the first occurrence as well as the second, and is a nuisance when you know very well that it is only the second one you're interested in. The solution is to say

```
/thing/;//
```

This says to find the first occurrence of 'thing', set dot to that line, then find the second and print only that.

Closely related is searching for the second previous occurrence of something, as in

```
?something?;??
```

Printing the third or fourth or ... in either direction is left as an exercise.

Finally, bear in mind that if you want to find the first occurrence of something in a file, starting at an arbitrary place within the file, it is not sufficient to say

```
1;/thing/
```

because this fails if 'thing' occurs on line **1**. But it is possible to say

```
0;/thing/
```

(one of the few places where 0 is a legal line number), for this starts the search at line 1.

Interrupting the Editor

As a final note on what dot gets set to, you should be aware that if you hit the interrupt or delete or rubout or break key while **ed** is doing a command, things are put back together again and your state is restored as much as possible to what it was before the command began. Naturally, some changes are irrevocable — if you are reading or writing a file or making substitutions or deleting lines, these will be stopped in some clean but unpredictable state in the middle (which is why it is not usually wise to stop them). Dot may or may not be changed.

Printing is more clear cut. Dot is not changed until the printing is done. Thus if you print until you see an interesting line, then hit delete, you are *not* sitting on that line or even near it. Dot is left where it was when the **p** command was started.

4. GLOBAL COMMANDS

The global commands **g** and **v** are used to perform one or more editing commands on all lines that either contain (**g**) or don't contain (**v**) a specified pattern.

As the simplest example, the command

```
g/UNIX/p
```

prints all lines that contain the word 'UNIX'. The pattern that goes between the slashes can be anything that could be used in a line search or in a substitute command; exactly the same rules and limitations apply.

As another example, then,

```
g/^\./p
```

prints all the formatting commands in a file (lines that begin with '.').

The **v** command is identical to **g**, except that it operates on those line that do *not* contain an occurrence of the pattern. (Don't look too hard for mnemonic significance to the letter 'v'.) So

```
v/^\./p
```

prints all the lines that don't begin with '.' — the actual text lines.

The command that follows **g** or **v** can be anything:

```
g/^\./d
```

deletes all lines that begin with '.', and

```
g/^$/d
```

deletes all empty lines.

Probably the most useful command that can follow a global is the substitute command, for this can be used to make a change and print each affected line for verification. For example, we could change the word 'Unix' to 'UNIX' everywhere, and verify that it really worked, with

```
g/Unix/s//UNIX/gp
```

Notice that we used '//' in the substitute command to mean 'the previous pattern', in this case, 'Unix'. The **p** command is done on every line that matches the pattern, not just those on which a substitution took place.

The global command operates by making two passes over the file. On the first pass, all lines that match the pattern are marked. On the second pass, each marked line in turn is examined, dot is set to that line, and the command executed. This means that it is possible for the command that follows a **g** or **v** to use addresses, set dot, and so on, quite freely.

```
g/^\.PP/+
```

prints the line that follows each '.PP' command (the signal for a new paragraph in some formatting packages). Remember that '+' means 'one line past dot'. And

```
g/topic/?^\.SH?1
```

searches for each line that contains 'topic', scans backwards until it finds a line that begins '.SH' (a section heading) and prints the line that follows that, thus showing the section headings under which 'topic' is mentioned. Finally,

```
g/^\.EQ/+,/^\.EN/-p
```

prints all the lines that lie between lines beginning with '.EQ' and '.EN' formatting commands.

The **g** and **v** commands can also be preceded by line numbers, in which case the lines searched are only those in the range specified.

Multi-line Global Commands

It is possible to do more than one command under the control of a global command, although the syntax for expressing the operation is not especially natural or pleasant. As an example, suppose the task is to change 'x' to 'y' and 'a' to 'b' on all lines that contain 'thing'. Then

```
g/thing/s/x/y/\
s/a/b/
```

is sufficient. The '\' signals the **g** command that the set of commands continues on the next line; it terminates on the first line that does not end with '\'. (As a minor blemish, you can't use a substitute command to insert a newline within a **g** command.)

You should watch out for this problem: the command

```
g/x/s//y/\
s/a/b/
```

does *not* work as you expect. The remembered pattern is the last pattern that was actually executed, so sometimes it will be 'x' (as expected), and sometimes it will be 'a' (not expected). You must spell it out, like this:

```
g/x/s/x/y/\
s/a/b/
```

It is also possible to execute **a**, **c** and **i** commands under a global command; as with other multi-line constructions, all that is needed is to add a '\' at the end of each line except the last. Thus to add a '.nf' and '.sp' command before each '.EQ' line, type

```
g/^\.EQ/i\
```

There is no need for a final line containing a '.' to terminate the **i** command, unless there are further commands being done under the global. On the other hand, it does no harm to put it in either.

5. CUT AND PASTE WITH UNIX COMMANDS

One editing area in which non-programmers seem not very confident is in what might be called 'cut and paste' operations — changing the name of a file, making a copy of a file somewhere else, moving a few lines from one place to another in a file, inserting one file in the middle of another, splitting a file into pieces, and splicing two or more files together.

Yet most of these operations are actually quite easy, if you keep your wits about you and go cautiously. The next several sections talk about cut and paste. We will begin with the UNIX commands for moving entire files around, then discuss **ed** commands for operating on pieces of files.

Changing the Name of a File

You have a file named 'memo' and you want it to be called 'paper' instead. How is it done?

The UNIX program that renames files is called **mv** (for 'move'); it 'moves' the file from one name to another, like this:

```
mv memo paper
```

That's all there is to it: **mv** from the old name to the new name.

```
mv oldname newname
```

Warning: if there is already a file around with the new name, its present contents will be silently clobbered by the information from the other file. The one exception is that you can't move a file to itself —

```
mv x x
```

is illegal.

Making a Copy of a File

Sometimes what you want is a copy of a file — an entirely fresh version. This might be because you want to work on a file, and yet save a copy in case something gets fouled up, or just because you're paranoid.

In any case, the way to do it is with the **cp** command. (**cp** stands for 'copy'; the system is big on short command names, which are appreciated by heavy users, but sometimes a strain for novices.) Suppose you have a file called 'good' and you want to save a copy before you make some dramatic editing changes. Choose a name — 'savegood' might be acceptable — then type

```
cp good savegood
```

This copies 'good' onto 'savegood', and you now have two identical copies of the file 'good'. (If 'savegood' previously contained something, it gets overwritten.)

Now if you decide at some time that you want to get back to the original state of 'good', you can say

```
mv savegood good
```

(if you're not interested in 'savegood' any more), or

```
cp savegood good
```

if you still want to retain a safe copy.

In summary, **mv** just renames a file; **cp** makes a duplicate copy. Both of them clobber the 'target' file if it already exists, so you had better be sure that's what you want to do *before* you do it.

Removing a File

If you decide you are really done with a file forever, you can remove it with the **rm** command:

```
rm savegood
```

throws away (irrevocably) the file called 'savegood'.

Putting Two or More Files Together

The next step is the familiar one of collecting two or more files into one big one. This will be needed, for example, when the author of a paper decides that several sections need to be combined into one. There are several ways to do it, of which the cleanest, once you get used to it, is a program called **cat**. (Not *all* programs have two-letter names.) **cat** is short for 'concatenate', which is exactly what we want to do.

Suppose the job is to combine the files 'file1' and 'file2' into a single file called 'bigfile'. If you say

```
cat file
```

the contents of 'file' will get printed on your terminal. If you say

```
cat file1 file2
```

the contents of 'file1' and then the contents of 'file2' will *both* be printed on your terminal, in that order. So **cat** combines the files, all right, but it's not much help to print them on the terminal — we want them in 'bigfile'.

Fortunately, there is a way. You can tell the system that instead of printing on your terminal, you want the same information put in a file. The way to do it is to add to the command line the character > and the name of the file where you want the output to go. Then you can say

```
cat file1 file2 >bigfile
```

and the job is done. (As with **cp** and **mv**, you're putting something into 'bigfile', and anything that was already there is destroyed.)

This ability to 'capture' the output of a program is one of the most useful aspects of the system. Fortunately it's not limited to the **cat** program — you can use it with *any* program that prints on your terminal. We'll see some more uses for it in a moment.

Naturally, you can combine several files, not just two:

```
cat file1 file2 file3 ... >bigfile
```

collects a whole bunch.

Question: is there any difference between

```
cp good savegood
```

and

```
cat good >savegood
```

Answer: for most purposes, no. You might reasonably ask why there are two programs in that case, since **cat** is obviously all you need. The answer is that **cp** will do some other things as well, which you can investigate for yourself by reading the manual. For now we'll stick to simple usages.

Adding Something to the End of a File

Sometimes you want to add one file to the end of another. We have enough building blocks now that you can do it; in fact before reading further it would be valuable if you figured out how. To be specific, how would you use **cp**, **mv** and/or **cat** to add the file 'good1' to the end of the file 'good'?

You could try

```
cat good good1 >temp
mv temp good
```

which is probably most direct. You should also understand why

```
cat good good1 >good
```

doesn't work. (Don't practice with a good 'good'!)

The easy way is to use a variant of >, called >>. In fact, >> is identical to > except that instead of clobbering the old file, it simply tacks stuff on at the end. Thus you could say

```
cat good1 >>good
```

and 'good1' is added to the end of 'good'. (And if 'good' didn't exist, this makes a copy of 'good1' called 'good'.)

6. CUT AND PASTE WITH THE EDITOR

Now we move on to manipulating pieces of files — individual lines or groups of lines. This is another area where new users seem unsure of themselves.

Filenames

The first step is to ensure that you know the **ed** commands for reading and writing files. Of course you can't go very far without knowing **r** and **w**. Equally useful, but less well known, is the 'edit' command **e**. Within **ed**, the command

```
e newfile
```

says 'I want to edit a new file called *newfile*, without leaving the editor.' The **e** command discards whatever you're currently working on and starts over on *newfile*. It's exactly the same as if you had quit with the **q** command, then re-entered **ed** with a new file name, except that if you have a pattern remembered, then a command like // will still work.

If you enter **ed** with the command

```
ed file
```

ed remembers the name of the file, and any subsequent **e**, **r** or **w** commands that don't contain a filename will refer to this remembered file. Thus

```
ed file1
 ... (editing) ...
w        (writes back in file1)
e file2  (edit new file, without leaving editor)
 ... (editing on file2) ...
w        (writes back on file2)
```

(and so on) does a series of edits on various files without ever leaving **ed** and without typing the name of any file more than once. (As an aside, if you examine the sequence of commands here, you can see why many UNIX systems use **e** as a synonym for **ed**.)

You can find out the remembered file name at any time with the **f** command; just type **f** without a file name. You can also change the name of the remembered file name with **f**; a useful sequence is

```
ed precious
f junk
 ... (editing) ...
```

which gets a copy of a precious file, then uses **f** to guarantee that a careless **w** command won't clobber the original.

Inserting One File into Another

Suppose you have a file called 'memo', and you want the file called 'table' to be inserted just after the reference to Table 1. That is, in 'memo' somewhere is a line that says

Table 1 shows that ...

and the data contained in 'table' has to go there, probably so it will be formatted properly by **nroff** or **troff**. Now what?

This one is easy. Edit 'memo', find 'Table 1', and add the file 'table' right there:

ed memo
/Table 1/
Table 1 shows that ... [response from ed]

The critical line is the last one. As we said earlier, the **r** command reads a file; here you asked for it to be read in right after line dot. An **r** command without any address adds lines at the end, so it is the same as **$r**.

Writing out Part of a File

The other side of the coin is writing out part of the document you're editing. For example, maybe you want to split out into a separate file that table from the previous example, so it can be formatted and tested separately. Suppose that in the file being edited we have

...[lots of stuff]

which is the way a table is set up for the **tbl** program. To isolate the table in a separate file called 'table', first find the start of the table (the '.TS' line), then write out the interesting part:

```
/^\.TS/
```

and the job is done. If you are confident, you can do it all at once with

```
/^\.TS/;/^\.TE/w table
```

The point is that the **w** command can write out a group of lines, instead of the whole file. In fact, you can write out a single line if you like; just give one line number instead of two. For example, if you have just typed a horribly complicated line and you know that it (or something like it) is going to be needed later, then save it — don't re-type it. In the editor, say

```
a
a
a
```

This last example is worth studying, to be sure you appreciate what's going on.

Moving Lines Around

Suppose you want to move a paragraph from its present position in a paper to the end. How would you do it? As a concrete example, suppose each paragraph in the paper begins with the formatting command '.PP'. Think about it and write down the details before reading on.

The brute force way (not necessarily bad) is to write the paragraph onto a temporary file, delete it from its current position, then read in the temporary file at the end. Assuming that you are sitting on the '.PP' command that begins the paragraph, this is the sequence of commands:

```
$r temp
```

That is, from where you are now ('.') until one line before the next '.PP' ('/^\.PP/–') write onto 'temp'. Then delete the same lines. Finally, read 'temp' at the end.

As we said, that's the brute force way. The easier way (often) is to use the *move* command **m** that **ed** provides — it lets you do the whole set of operations at one crack, without any temporary file.

The **m** command is like many other **ed** commands in that it takes up to two line numbers in front that tell what lines are to be affected. It is also *followed* by a line number that tells where the lines are to go. Thus

```
line1, line2 m line3
```

says to move all the lines between 'line1' and 'line2' after 'line3'. Naturally, any of 'line1' etc., can be patterns between slashes, **$** signs, or other ways to specify lines.

Suppose again that you're sitting at the first line of the paragraph. Then you can say

That's all.

As another example of a frequent operation, you can reverse the order of two adjacent lines by moving the first one to after the second. Suppose that you are positioned at the first. Then

```
m+
```

does it. It says to move line dot to after one line after line dot. If you are positioned on the second line,

```
m—
```

does the interchange.

As you can see, the **m** command is more succinct and direct than writing, deleting and re-reading. When is brute force better anyway? This is a matter of personal taste — do what you have most confidence in. The main difficulty with the **m** command is that if you use patterns to specify both the lines you are moving and the target, you have to take care that you specify them properly, or you may well not move the lines you thought you did. The result of a botched **m** command can be a ghastly mess. Doing the job a step at a time makes it easier for you to verify at each step that you accomplished what you wanted to. It's also a good idea to issue a **w** command before doing anything complicated; then if you goof, it's easy to back up to where you were.

Marks

ed provides a facility for marking a line with a particular name so you can later reference it by name regardless of its actual line number. This can be handy for moving lines, and for keeping track of them as they move. The *mark* command is **k**; the command

```
kx
```

marks the current line with the name 'x'. If a line number precedes the **k**, that line is marked. (The mark name must be a single lower case letter.) Now you can refer to the marked line with the address

```
'x
```

Marks are most useful for moving things around. Find the first line of the block to be moved, and mark it with *'a*. Then find the last line and mark it with *'b*. Now position yourself at the place where the stuff is to go and say

```
'a,'bm.
```

Bear in mind that only one line can have a particular mark name associated with it at any given time.

Copying Lines

We mentioned earlier the idea of saving a line that was hard to type or used often, so as to cut down on typing time. Of course this could be more than one line; then the saving is presumably even greater.

ed provides another command, called **t** (for 'transfer') for making a copy of a group of one or more lines at any point. This is often easier than writing and reading.

The **t** command is identical to the **m** command, except that instead of moving lines it simply duplicates them at the place you named. Thus

```
1,$t$
```

duplicates the entire contents that you are editing. A more common use for **t** is for creating a series of lines that differ only slightly. For example, you can say

```
a
.......
.
t.          (make a copy)
s/x/y/      (change it a bit)
t.          (make third copy)
s/y/z/      (change it a bit)
```

and so on.

The Temporary Escape '!'

Sometimes it is convenient to be able to temporarily escape from the editor to do some other UNIX command, perhaps one of the file copy or move commands discussed in section 5, without leaving the editor. The 'escape' command **!** provides a way to do this.

If you say

```
!any UNIX command
```

your current editing state is suspended, and the UNIX command you asked for is executed. When the command finishes, **ed** will signal you by printing another **!**; at that point you can resume editing.

You can really do *any* UNIX command, including another **ed**. (This is quite common, in fact.) In this case, you can even do another **!**.

7. SUPPORTING TOOLS

There are several tools and techniques that go along with the editor, all of which are relatively easy once you know how **ed** works, because they are all based on the editor. In this section we will give some fairly cursory examples of these tools, more to indicate their existence than to provide a complete tutorial. More information on each can be found in [3].

Grep

Sometimes you want to find all occurrences of some word or pattern in a set of files, to edit them or perhaps just to verify their presence or absence. It may be possible to edit each file separately and look for the pattern of interest, but if there are many files this can get very tedious, and if the files are really big, it may be impossible because of limits in **ed**.

The program **grep** was invented to get around these limitations. The search patterns that we have described in the paper are often called 'regular expressions', and 'grep' stands for

```
g/re/p
```

That describes exactly what **grep** does — it prints every line in a set of files that contains a particular pattern. Thus

```
grep 'thing' file1 file2 file3 ...
```

finds 'thing' wherever it occurs in any of the files 'file1', 'file2', etc. **grep** also indicates the file in which the line was found, so you can later edit it if you like.

The pattern represented by 'thing' can be any pattern you can use in the editor, since **grep** and **ed** use exactly the same mechanism for pattern searching. It is wisest always to enclose the pattern in the single quotes '...' if it contains any non-alphabetic characters, since many such characters also mean something special to the UNIX command interpreter (the 'shell'). If you don't quote them, the command interpreter will try to interpret them before **grep** gets a chance.

There is also a way to find lines that *don't* contain a pattern:

```
grep -v 'thing' file1 file2 ...
```

finds all lines that don't contains 'thing'. The **-v** must occur in the position shown. Given **grep** and **grep -v**, it is possible to do things like selecting all lines that contain some combination of patterns. For example, to get all lines that contain 'x' but not 'y':

```
grep x file... | grep -v y
```

(The notation | is a 'pipe', which causes the output of the first command to be used as input to the second command; see [2].)

Editing Scripts

If a fairly complicated set of editing operations is to be done on a whole set of files, the easiest thing to do is to make up a 'script', i.e., a file that contains the operations you want to perform, then apply this script to each file in turn.

For example, suppose you want to change every 'Unix' to 'UNIX' and every 'Gcos' to 'GCOS' in a large number of files. Then put into the file 'script' the lines

```
g/Unix/s//UNIX/g
g/Gcos/s//GCOS/g
w
q
```

Now you can say

```
ed file1 <script
ed file2 <script
...
```

This causes **ed** to take its commands from the prepared script. Notice that the whole job has to be planned in advance.

And of course by using the UNIX command interpreter, you can cycle through a set of files automatically, with varying degrees of ease.

Sed

sed ('stream editor') is a version of the editor with restricted capabilities but which is capable of processing unlimited amounts of input. Basically **sed** copies its input to its output, applying one or more editing commands to each line of input.

As an example, suppose that we want to do the 'Unix' to 'UNIX' part of the example given above, but without rewriting the files. Then the command

```
sed 's/Unix/UNIX/g' file1 file2 ...
```

applies the command 's/Unix/UNIX/g' to all lines from 'file1', 'file2', etc., and copies all lines to the output. The advantage of using **sed** in such a case is that it can be used with input too large for **ed** to handle. All the output can be collected in one place, either in a file or perhaps piped into another program.

If the editing transformation is so complicated that more than one editing command is needed, commands can be supplied from a file, or on the command line, with a slightly more complex syntax. To take commands from a file, for example,

```
sed -f cmdfile input-files...
```

sed has further capabilities, including conditional testing and branching, which we cannot go into here.

Acknowledgement

I am grateful to Ted Dolotta for his careful reading and valuable suggestions.

References

[1] Brian W. Kernighan, *A Tutorial Introduction to the UNIX Text Editor,* Bell Laboratories internal memorandum.

[2] Brian W. Kernighan, *UNIX For Beginners,* Bell Laboratories internal memorandum.

[3] Ken L. Thompson and Dennis M. Ritchie, *The UNIX Programmer's Manual.* Bell Laboratories.

MAIL REFERENCE MANUAL

Kurt Shoens

Revised by

Craig Leres

Version 2.18

September 12, 1986

1. Introduction

Mail provides a simple and friendly environment for sending and receiving mail. It divides incoming mail into its constituent messages and allows the user to deal with them in any order. In addition, it provides a set of *ed*-like commands for manipulating messages and sending mail. *Mail* offers the user simple editing capabilities to ease the composition of outgoing messages, as well as providing the ability to define and send to names which address groups of users. Finally, *Mail* is able to send and receive messages across such networks as the ARPANET, UUCP, and Berkeley network.

This document describes how to use the *Mail* program to send and receive messages. The reader is not assumed to be familiar with other message handling systems, but should be familiar with the UNIX[1] shell, the text editor, and some of the common UNIX commands. "The UNIX Programmer's Manual," "An Introduction to Csh," and "Text Editing with Ex and Vi" can be consulted for more information on these topics.

Here is how messages are handled: the mail system accepts incoming *messages* for you from other people and collects them in a file, called your *system mailbox*. When you login, the system notifies you if there are any messages waiting in your system mailbox. If you are a *csh* user, you will be notified when new mail arrives if you inform the shell of the location of your mailbox. On version 7 systems, your system mailbox is located in the directory /usr/spool/mail in a file with your login name. If your login name is "sam," then you can make *csh* notify you of new mail by including the following line in your .cshrc file:

set mail=/usr/spool/mail/sam

When you read your mail using *Mail*, it reads your system mailbox and separates that file into the individual messages that have been sent to you. You can then read, reply to, delete, or save these messages. Each message is marked with its author and the date they sent it.

[1] UNIX is a trademark of Bell Laboratories.

2. Common usage

The *Mail* command has two distinct usages, according to whether one wants to send or receive mail. Sending mail is simple: to send a message to a user whose login name is, say, "root," use the shell command:

% Mail root

then type your message. When you reach the end of the message, type an EOT (control–d) at the beginning of a line, which will cause *Mail* to echo "EOT" and return you to the Shell. When the user you sent mail to next logs in, he will receive the message:

You have mail.

to alert him to the existence of your message.

If, while you are composing the message you decide that you do not wish to send it after all, you can abort the letter with a RUBOUT. Typing a single RUBOUT causes *Mail* to print

(Interrupt – one more to kill letter)

Typing a second RUBOUT causes *Mail* to save your partial letter on the file "dead.letter" in your home directory and abort the letter. Once you have sent mail to someone, there is no way to undo the act, so be careful.

The message your recipient reads will consist of the message you typed, preceded by a line telling who sent the message (your login name) and the date and time it was sent.

If you want to send the same message to several other people, you can list their login names on the command line. Thus,

```
% Mail sam bob john
Tuition fees are due next Friday. Don't forget!!
<Control–d>
EOT
%
```

will send the reminder to sam, bob, and john.

If, when you log in, you see the message,

You have mail.

you can read the mail by typing simply:

% Mail

Mail will respond by typing its version number and date and then listing the messages you have waiting. Then it will type a prompt and await your command. The messages are assigned numbers starting with 1 – you refer to the messages with these numbers. *Mail* keeps tack of which messages are *new* (have been sent since you last read your mail) and *read* (have been read by you). New messages have an **N** next to them in the header listing and old, but unread messages have a **U** next to them. *Mail* keeps track of new/old and read/unread messages by putting a header field called "Status" into your messages.

To look at a specific message, use the **type** command, which may be abbreviated to simply **t**. For example, if you had the following messages:

```
N 1 root   Wed Sep 21 09:21 "Tuition fees"
N 2 sam    Tue Sep 20 22:55
```

you could examine the first message by giving the command:

type 1

which might cause *Mail* to respond with, for example:

```
Message 1:
From root Wed Sep 21 09:21:45 1978
Subject: Tuition fees
```

Status: R

Tuition fees are due next Wednesday. Don't forget!!

Many *Mail* commands that operate on messages take a message number as an argument like the **type** command. For these commands, there is a notion of a current message. When you enter the *Mail* program, the current message is initially the first one. Thus, you can often omit the message number and use, for example,

t

to type the current message. As a further shorthand, you can type a message by simply giving its message number. Hence,

1

would type the first message.

Frequently, it is useful to read the messages in your mailbox in order, one after another. You can read the next message in *Mail* by simply typing a newline. As a special case, you can type a newline as your first command to *Mail* to type the first message.

If, after typing a message, you wish to immediately send a reply, you can do so with the **reply** command. **Reply**, like **type**, takes a message number as an argument. *Mail* then begins a message addressed to the user who sent you the message. You may then type in your letter in reply, followed by a <control-d> at the beginning of a line, as before. *Mail* will type EOT, then type the ampersand prompt to indicate its readiness to accept another command. In our example, if, after typing the first message, you wished to reply to it, you might give the command:

reply

Mail responds by typing:

To: root
Subject: Re: Tuition fees

and waiting for you to enter your letter. You are now in the message collection mode described at the beginning of this section and *Mail* will gather up your message up to a control–d. Note that it copies the subject header from the original message. This is useful in that correspondence about a particular matter will tend to retain the same subject heading, making it easy to recognize. If there are other header fields in the message, the information found will also be used. For example, if the letter had a "To:" header listing several recipients, *Mail* would arrange to send your replay to the same people as well. Similarly, if the original message contained a "Cc:" (carbon copies to) field, *Mail* would send your reply to *those* users, too. *Mail* is careful, though, not too send the message to ***you***, even if you appear in the "To:" or "Cc:" field, unless you ask to be included explicitly. See section 4 for more details.

After typing in your letter, the dialog with *Mail* might look like the following:

reply
To: root
Subject: Tuition fees

Thanks for the reminder
EOT
&

The **reply** command is especially useful for sustaining extended conversations over the message system, with other "listening" users receiving copies of the conversation. The **reply** command can be abbreviated to **r**.

Sometimes you will receive a message that has been sent to several people and wish to reply *only* to the person who sent it. **Reply** with a capital **R** replies to a message, but sends a copy to the sender only.

If you wish, while reading your mail, to send a message to someone, but not as a reply to one of your messages, you can send the message directly with the **mail** command, which takes as arguments the names of the recipients you wish to send to. For example, to send a message to "frank," you would do:

mail frank
This is to confirm our meeting next Friday at 4.
EOT
&

The **mail** command can be abbreviated to **m**.

Normally, each message you receive is saved in the file *mbox* in your login directory at the time you leave *Mail.* Often, however, you will not want to save a particular message you have received because it is only of passing interest. To avoid saving a message in *mbox* you can delete it using the **delete** command. In our example,

delete 1

will prevent *Mail* from saving message 1 (from root) in *mbox*. In addition to not saving deleted messages, *Mail* will not let you type them, either. The effect is to make the message disappear altogether, along with its number. The **delete** command can be abbreviated to simply **d**.

Many features of *Mail* can be tailored to your liking with the **set** command. The **set** command has two forms, depending on whether you are setting a *binary* option or a *valued* option. Binary options are either on or off. For example, the "ask" option informs *Mail* that each time you send a message, you want it to prompt you for a subject header, to be included in the message. To set the "ask" option, you would type

set ask

Another useful *Mail* option is "hold." Unless told otherwise, *Mail* moves the messages from your system mailbox to the file *mbox* in your home directory when you leave *Mail.* If you want *Mail* to keep your letters in the system mailbox instead, you can set the "hold" option.

Valued options are values which *Mail* uses to adapt to your tastes. For example, the "SHELL" option tells *Mail* which shell you like to use, and is specified by

set SHELL=/bin/csh

for example. Note that no spaces are allowed in "SHELL=/bin/csh." A complete list of the *Mail* options appears in section 5.

Another important valued option is "crt." If you use a fast video terminal, you will find that when you print long messages, they fly by too quickly for you to read them. With the "crt" option, you can make *Mail* print any message larger than a given number of lines by sending it through the paging program *more*. For example, most CRT users should do:

set crt=24

to paginate messages that will not fit on their screens. *More* prints a screenful of information, then types --MORE--. Type a space to see the next screenful.

Another adaptation to user needs that *Mail* provides is that of *aliases.* An alias is simply a name which stands for one or more real user names. *Mail* sent to an alias is really sent to the list of real users associated with it. For example, an alias can be defined for the members of a project, so that you can send mail to the whole project by sending mail to just a single name. The **alias** command in *Mail* defines an alias. Suppose that the users in a project are named Sam, Sally, Steve, and Susan. To define an alias called "project" for them, you would use the *Mail* command:

alias project sam sally steve susan

The **alias** command can also be used to provide a convenient name for someone whose user name is inconvenient. For example, if a user named "Bob Anderson" had the login name "anderson,"" you might want to use:

alias bob anderson

so that you could send mail to the shorter name, "bob."

While the **alias** and **set** commands allow you to customize *Mail*, they have the drawback that they must be retyped each time you enter *Mail*. To make them more convenient to use, *Mail* always looks for two files when it is invoked. It first reads a system wide file "/usr/lib/Mail.rc," then a user specific file, ".mailrc," which is found in the user's home directory. The system wide file is maintained by the system administrator and contains **set** commands that are applicable to all users of the system. The ".mailrc" file is usually used by each user to set options the way he likes and define individual aliases. For example, my .mailrc file looks like this:

set ask nosave SHELL=/bin/csh

As you can see, it is possible to set many options in the same **set** command. The "nosave" option is described in section 5.

Mail aliasing is implemented at the system-wide level by the mail delivery system *sendmail*. These aliases are stored in the file /usr/lib/aliases and are accessible to all users of the system. The lines in /usr/lib/aliases are of the form:

alias: $name_1$, $name_2$, $name_3$

where *alias* is the mailing list name and the $name_i$ are the members of the list. Long lists can be continued onto the next line by starting the next line with a space or tab. Remember that you must execute the shell command *newaliases* after editing /usr/lib/aliases since the delivery system uses an indexed file created by *newaliases*.

We have seen that *Mail* can be invoked with command line arguments which are people to send the message to, or with no arguments to read mail. Specifying the **-f** flag on the command line causes *Mail* to read messages from a file other than your system mailbox. For example, if you have a collection of messages in the file "letters" you can use *Mail* to read them with:

% Mail -f letters

You can use all the *Mail* commands described in this document to examine, modify, or delete messages from your "letters" file, which will be rewritten when you leave *Mail* with the **quit** command described below.

Since mail that you read is saved in the file *mbox* in your home directory by default, you can read *mbox* in your home directory by using simply

% Mail -f

Normally, messages that you examine using the **type** command are saved in the file "mbox" in your home directory if you leave *Mail* with the **quit** command described below. If you wish to retain a message in your system mailbox you can use the **preserve** command to tell *Mail* to leave it there. The **preserve** command accepts a list of message numbers, just like **type** and may be abbreviated to **pre**.

Messages in your system mailbox that you do not examine are normally retained in your system mailbox automatically. If you wish to have such a message saved in *mbox* without reading it, you may use the **mbox** command to have them so saved. For example,

mbox 2

in our example would cause the second message (from sam) to be saved in *mbox* when the **quit** command is executed. **Mbox** is also the way to direct messages to your *mbox* file if you have set the "hold" option described above. **Mbox** can be abbreviated to **mb**.

When you have perused all the messages of interest, you can leave *Mail* with the **quit** command, which saves the messages you have typed but not deleted in the file *mbox* in your login directory. Deleted messages are discarded irretrievably, and messages left untouched are preserved in your system mailbox so that you will see them the next time you type:

% Mail

The **quit** command can be abbreviated to simply **q**.

If you wish for some reason to leave *Mail* quickly without altering either your system mailbox or *mbox*, you can type the **x** command (short for **exit**), which will immediately return you to the Shell without changing anything.

If, instead, you want to execute a Shell command without leaving *Mail*, you can type the command preceded by an exclamation point, just as in the text editor. Thus, for instance:

!date

will print the current date without leaving *Mail.*

Finally, the **help** command is available to print out a brief summary of the *Mail* commands, using only the single character command abbreviations.

3. Maintaining folders

Mail includes a simple facility for maintaining groups of messages together in folders. This section describes this facility.

To use the folder facility, you must tell *Mail* where you wish to keep your folders. Each folder of messages will be a single file. For convenience, all of your folders are kept in a single directory of your choosing. To tell *Mail* where your folder directory is, put a line of the form

set folder=letters

in your *.mailrc* file. If, as in the example above, your folder directory does not begin with a '/,' *Mail* will assume that your folder directory is to be found starting from your home directory. Thus, if your home directory is **/usr/person** the above example told *Mail* to find your folder directory in **/usr/person/letters**.

Anywhere a file name is expected, you can use a folder name, preceded with '+.' For example, to put a message into a folder with the **save** command, you can use:

save +classwork

to save the current message in the *classwork* folder. If the *classwork* folder does not yet exist, it will be created. Note that messages which are saved with the **save** command are automatically removed from your system mailbox.

In order to make a copy of a message in a folder without causing that message to be removed from your system mailbox, use the **copy** command, which is identical in all other respects to the **save** command. For example,

copy +classwork

copies the current message into the *classwork* folder and leaves a copy in your system mailbox.

The **folder** command can be used to direct *Mail* to the contents of a different folder. For example,

folder +classwork

directs *Mail* to read the contents of the *classwork* folder. All of the commands that you can use on your system mailbox are also applicable to folders, including **type**, **delete**, and **reply**. To inquire which folder you are currently editing, use simply:

folder

To list your current set of folders, use the **folders** command.

To start *Mail* reading one of your folders, you can use the **–f** option described in section 2. For example:

% Mail –f +classwork

will cause *Mail* to read your *classwork* folder without looking at your system mailbox.

4. More about sending mail

4.1. Tilde escapes

While typing in a message to be sent to others, it is often useful to be able to invoke the text editor on the partial message, print the message, execute a shell command, or do some other auxiliary function. *Mail* provides these capabilities through *tilde escapes,* which consist of a tilde (~) at the beginning of a line, followed by a single character which indicates the function to be performed. For example, to print the text of the message so far, use:

~p

which will print a line of dashes, the recipients of your message, and the text of the message so far. Since *Mail* requires two consecutive RUBOUT's to abort a letter, you can use a single RUBOUT to abort the output of ~p or any other ~ escape without killing your letter.

If you are dissatisfied with the message as it stands, you can invoke the text editor on it using the escape

~e

which causes the message to be copied into a temporary file and an instance of the editor to be spawned. After modifying the message to your satisfaction, write it out and quit the editor. *Mail* will respond by typing

(continue)

after which you may continue typing text which will be appended to your message, or type <control-d> to end the message. A standard text editor is provided by *Mail.* You can override this default by setting the valued option "EDITOR" to something else. For example, you might prefer:

set EDITOR=/usr/ucb/ex

Many systems offer a screen editor as an alternative to the standard text editor, such as the *vi* editor from UC Berkeley. To use the screen, or *visual* editor, on your current message, you can use the escape,

~v

~v works like ~e, except that the screen editor is invoked instead. A default screen editor is defined by *Mail.* If it does not suit you, you can set the valued option "VISUAL" to the path name of a different editor.

It is often useful to be able to include the contents of some file in your message; the escape

~r filename

is provided for this purpose, and causes the named file to be appended to your current message. *Mail* complains if the file doesn't exist or can't be read. If the read is successful, the number of lines and characters appended to your message is printed, after which you may continue appending text. The filename may contain shell metacharacters like * and ? which are expanded according to the conventions of your shell.

As a special case of ~r, the escape

~d

reads in the file "dead.letter" in your home directory. This is often useful since *Mail* copies the text of your message there when you abort a message with RUBOUT.

To save the current text of your message on a file you may use the

~w filename

escape. *Mail* will print out the number of lines and characters written to the file, after which you may continue appending text to your message. Shell metacharacters may be used in the filename, as in ~r and are expanded with the conventions of your shell.

If you are sending mail from within *Mail's* command mode you can read a message sent to you into the message you are constructing with the escape:

~m 4

which will read message 4 into the current message, shifted right by one tab stop. You can name any non-deleted message, or list of messages. Messages can also be forwarded without shifting by a tab stop with ~f. This is the usual way to forward a message.

If, in the process of composing a message, you decide to add additional people to the list of message recipients, you can do so with the escape

~t name1 name2 ...

You may name as few or many additional recipients as you wish. Note that the users originally on the recipient list will still receive the message; you cannot remove someone from the recipient list with ~t.

If you wish, you can associate a subject with your message by using the escape

~s Arbitrary string of text

which replaces any previous subject with "Arbitrary string of text." The subject, if given, is sent near the top of the message prefixed with "Subject:" You can see what the message will look like by using ~p.

For political reasons, one occasionally prefers to list certain people as recipients of carbon copies of a message rather than direct recipients. The escape

~c name1 name2 ...

adds the named people to the "Cc:" list, similar to ~t. Again, you can execute ~p to see what the message will look like.

The recipients of the message together constitute the "To:" field, the subject the "Subject:" field, and the carbon copies the "Cc:" field. If you wish to edit these in ways impossible with the ~t, ~s, and ~c escapes, you can use the escape

~h

which prints "To:" followed by the current list of recipients and leaves the cursor (or printhead) at the end of the line. If you type in ordinary characters, they are appended to the end of the current list of recipients. You can also use your erase character to erase back into the list of recipients, or your kill character to erase them altogether. Thus, for example, if your erase and kill characters are the standard # and @ symbols,

~h
To: root kurt####bill

would change the initial recipients "root kurt" to "root bill." When you type a newline, *Mail* advances to the "Subject:" field, where the same rules apply. Another newline brings you to the "Cc:" field, which may be edited in the same fashion. Another newline leaves you appending text to the end of your message. You can use ~p to print the current text of the header fields and the body of the message.

To effect a temporary escape to the shell, the escape

~!command

is used, which executes *command* and returns you to mailing mode without altering the text of your message. If you wish, instead, to filter the body of your message through a shell command, then you can use

~|command

which pipes your message through the command and uses the output as the new text of your message. If the command produces no output, *Mail* assumes that something is amiss and retains the old version of your message. A frequently-used filter is the command *fmt*, designed to format outgoing mail.

To effect a temporary escape to *Mail* command mode instead, you can use the

~:*Mail command*

escape. This is especially useful for retyping the message you are replying to, using, for example:

~:t

It is also useful for setting options and modifying aliases.

If you wish (for some reason) to send a message that contains a line beginning with a tilde, you must double it. Thus, for example,

~~This line begins with a tilde.

sends the line

~This line begins with a tilde.

Finally, the escape

~?

prints out a brief summary of the available tilde escapes.

On some terminals (particularly ones with no lower case) tilde's are difficult to type. *Mail* allows you to change the escape character with the "escape" option. For example, I set

set escape=]

and use a right bracket instead of a tilde. If I ever need to send a line beginning with right bracket, I double it, just as for ~. Changing the escape character removes the special meaning of ~.

4.2. Network access

This section describes how to send mail to people on other machines. Recall that sending to a plain login name sends mail to that person on your machine. If your machine is directly (or sometimes, even, indirectly) connected to the Arpanet, you can send messages to people on the Arpanet using a name of the form

name@host

where *name* is the login name of the person you're trying to reach and *host* is the name of the machine where he logs in on the Arpanet.

If your recipient logs in on a machine connected to yours by UUCP (the Bell Laboratories supplied network that communicates over telephone lines), sending mail to him is a bit more complicated. You must know the list of machines through which your message must travel to arrive at his site. So, if his machine is directly connected to yours, you can send mail to him using the syntax:

host!name

where, again, *host* is the name of his machine and *name* is his login name. If your message must go through an intermediate machine first, you must use the syntax:

intermediate!host!name

and so on. It is actually a feature of UUCP that the map of all the systems in the network is not known anywhere (except where people decide to write it down for convenience). Talk to your system administrator about the machines connected to your site.

If you want to send a message to a recipient on the Berkeley network (Berknet), you use the syntax:

host:name

where *host* is his machine name and *name* is his login name. Unlike UUCP, you need not know the names of the intermediate machines.

When you use the **reply** command to respond to a letter, there is a problem of figuring out the names of the users in the "To:" and "Cc:" lists *relative to the current machine*. If the original letter was sent to you by someone on the local machine, then this problem does not exist, but if the message came from a remote machine, the problem must be dealt with. *Mail* uses a heuristic to build the correct name for each user relative to the local machine. So, when you **reply** to remote mail, the names in the "To:" and "Cc:" lists may change somewhat.

4.3. Special recipients

As described previously, you can send mail to either user names or **alias** names. It is also possible to send messages directly to files or to programs, using special conventions. If a recipient name has a '/' in it or begins with a '+', it is assumed to be the path name of a file into which to send the message. If the file already exists, the message is appended to the end of the file. If you want to name a file in your current directory (ie, one for which a '/' would not usually be needed) you can precede the name with './' So, to send mail to the file "memo" in the current directory, you can give the command:

% Mail ./memo

If the name begins with a '+,' it is expanded into the full path name of the folder name in your folder directory. This ability to send mail to files can be used for a variety of purposes, such as maintaining a journal and keeping a record of mail sent to a certain group of users. The second example can be done automatically by including the full pathname of the record file in the **alias** command for the group. Using our previous **alias** example, you might give the command:

alias project sam sally steve susan /usr/project/mail_record

Then, all mail sent to "project" would be saved on the file "/usr/project/mail_record" as well as being sent to the members of the project. This file can be examined using *Mail -f*.

It is sometimes useful to send mail directly to a program, for example one might write a project billboard program and want to access it using *Mail*. To send messages to the billboard program, one can send mail to the special name '|billboard' for example. *Mail* treats recipient names that begin with a '|' as a program to send the mail to. An **alias** can be set up to reference a '|' prefaced name if desired. *Caveats*: the shell treats '|' specially, so it must be quoted on the command line. Also, the '| program' must be presented as a single argument to mail. The safest course is to surround the entire name with double quotes. This also applies to usage in the **alias** command. For example, if we wanted to alias 'rmsgs' to 'rmsgs –s' we would need to say:

alias rmsgs "| rmsgs -s"

5. Additional features

This section describes some additional commands of use for reading your mail, setting options, and handling lists of messages.

5.1. Message lists

Several *Mail* commands accept a list of messages as an argument. Along with **type** and **delete**, described in section 2, there is the **from** command, which prints the message headers associated with the message list passed to it. The **from** command is particularly useful in conjunction with some of the message list features described below.

A *message list* consists of a list of message numbers, ranges, and names, separated by spaces or tabs. Message numbers may be either decimal numbers, which directly specify messages, or one of the special characters "↑" "." or "$" to specify the first relevant, current, or last relevant message, respectively. *Relevant* here means, for most commands "not deleted" and "deleted" for the **undelete** command.

A range of messages consists of two message numbers (of the form described in the previous paragraph) separated by a dash. Thus, to print the first four messages, use

    type 1–4

and to print all the messages from the current message to the last message, use

    type .–$

A *name* is a user name. The user names given in the message list are collected together and each message selected by other means is checked to make sure it was sent by one of the named users. If the message consists entirely of user names, then every message sent by one those users that is *relevant* (in the sense described earlier) is selected. Thus, to print every message sent to you by "root," do

    type root

As a shorthand notation, you can specify simply "*" to get every *relevant* (same sense) message. Thus,

    type *

prints all undeleted messages,

    delete *

deletes all undeleted messages, and

    undelete *

undeletes all deleted messages.

You can search for the presence of a word in subject lines with /. For example, to print the headers of all messages that contain the word "PASCAL," do:

    from /pascal

Note that subject searching ignores upper/lower case differences.

5.2. List of commands

This section describes all the *Mail* commands available when receiving mail.

! Used to preface a command to be executed by the shell.

– The – command goes to the previous message and prints it. The – command may be given a decimal number *n* as an argument, in which case the *n*th previous message is gone to and printed.

Print
Like **print**, but also print out ignored header fields. See also **print** and **ignore**.

Reply
Note the capital R in the name. Frame a reply to a one or more messages. The reply (or replies if you are using this on multiple messages) will be sent ONLY to the person who sent you the message (respectively, the set of people who sent the messages you are replying to). You can add people using the **~t** and **~c** tilde escapes. The subject in your reply is formed by prefacing the subject in the original message with "Re:" unless it already began thus. If the original message included a "reply-to" header field, the reply will go *only* to the recipient named by "reply-to." You type in your message using the same conventions available to you through the **mail** command. The **Reply** command is especially useful for replying to messages that were sent to enormous distribution groups when you really just want to send a message to the originator. Use it often.

Type
Identical to the **Print** command.

alias Define a name to stand for a set of other names. This is used when you want to send messages to a certain group of people and want to avoid retyping their names. For example

alias project john sue willie kathryn

creates an alias *project* which expands to the four people John, Sue, Willie, and Kathryn.

alternates
If you have accounts on several machines, you may find it convenient to use the /usr/lib/aliases on all the machines except one to direct your mail to a single account. The **alternates** command is used to inform *Mail* that each of these other addresses is really *you*. *Alternates* takes a list of user names and remembers that they are all actually you. When you **reply** to messages that were sent to one of these alternate names, *Mail* will not bother to send a copy of the message to this other address (which would simply be directed back to you by the alias mechanism). If *alternates* is given no argument, it lists the current set of alternate names. **Alternates** is usually used in the .mailrc file.

chdir
The **chdir** command allows you to change your current directory. **Chdir** takes a single argument, which is taken to be the pathname of the directory to change to. If no argument is given, **chdir** changes to your home directory.

copy The **copy** command does the same thing that **save** does, except that it does not mark the messages it is used on for deletion when you quit.

delete
Deletes a list of messages. Deleted messages can be reclaimed with the **undelete** command.

dt The **dt** command deletes the current message and prints the next message. It is useful for quickly reading and disposing of mail.

edit To edit individual messages using the text editor, the **edit** command is provided. The **edit** command takes a list of messages as described under the **type** command and processes each by writing it into the file Message*x* where *x* is the message number being edited and executing the text editor on it. When you have edited the message to your satisfaction, write the message out and quit, upon which *Mail* will read the message back and remove the file. **Edit** may be abbreviated to **e**.

else Marks the end of the then-part of an **if** statement and the beginning of the part to take effect if the condition of the **if** statement is false.

endifMarks the end of an **if** statement.

exit Leave *Mail* without updating the system mailbox or the file your were reading. Thus, if you accidentally delete several messages, you can use **exit** to avoid scrambling your mailbox.

file The same as **folder**.

folders
List the names of the folders in your folder directory.

folder

The **folder** command switches to a new mail file or folder. With no arguments, it tells you which file you are currently reading. If you give it an argument, it will write out changes (such as deletions) you have made in the current file and read the new file. Some special conventions are recognized for the name:

Name	Meaning
#	Previous file read
%	Your system mailbox
%name	*Name*'s system mailbox
&	Your ~/mbox file
+folder	A file in your folder directory

from The **from** command takes a list of messages and prints out the header lines for each one; hence

from joe

is the easy way to display all the message headers from "joe."

headers

When you start up *Mail* to read your mail, it lists the message headers that you have. These headers tell you who each message is from, when they were sent, how many lines and characters each message is, and the "Subject:" header field of each message, if present. In addition, *Mail* tags the message header of each message that has been the object of the **preserve** command with a "P." Messages that have been **saved** or **written** are flagged with a "*." Finally, **deleted** messages are not printed at all. If you wish to reprint the current list of message headers, you can do so with the **headers** command. The **headers** command (and thus the initial header listing) only lists the first so many message headers. The number of headers listed depends on the speed of your terminal. This can be overridden by specifying the number of headers you want with the *window* option. *Mail* maintains a notion of the current "window" into your messages for the purposes of printing headers. Use the **z** command to move forward and back a window. You can move *Mail's* notion of the current window directly to a particular message by using, for example,

headers 40

to move *Mail's* attention to the messages around message 40. The **headers** command can be abbreviated to **h**.

help Print a brief and usually out of date help message about the commands in *Mail*. Refer to this manual instead.

hold Arrange to hold a list of messages in the system mailbox, instead of moving them to the file *mbox* in your home directory. If you set the binary option *hold*, this will happen by default.

if Commands in your ".mailrc" file can be executed conditionally depending on whether you are sending or receiving mail with the **if** command. For example, you can do:

```
if receive
        commands...
endif
```

An **else** form is also available:

```
if send
        commands...
else
        commands...
endif
```

Note that the only allowed conditions are **receive** and **send**.

ignore
Add the list of header fields named to the *ignore list*. Header fields in the ignore list are not printed on your terminal when you print a message. This allows you to suppress printing of certain machine-generated header fields, such as *Via* which are not usually of interest. The **Type** and **Print** commands can be used to print a message in its entirety, including ignored fields. If **ignore** is executed with no arguments, it lists the current set of ignored fields.

list List the vaild *Mail* commands.

mail Send mail to one or more people. If you have the *ask* option set, *Mail* will prompt you for a subject to your message. Then you can type in your message, using tilde escapes as described in section 4 to edit, print, or modify your message. To signal your satisfaction with the message and send it, type control-d at the beginning of a line, or a . alone on a line if you set the option *dot*. To abort the message, type two interrupt characters (RUBOUT by default) in a row or use the **~q** escape.

mbox
Indicate that a list of messages be sent to *mbox* in your home directory when you quit. This is the default action for messages if you do *not* have the *hold* option set.

next The **next** command goes to the next message and types it. If given a message list, **next** goes to the first such message and types it. Thus,

    next root

goes to the next message sent by "root" and types it. The **next** command can be abbreviated to simply a newline, which means that one can go to and type a message by simply giving its message number or one of the magic characters "↑" "." or "$". Thus,

    .

prints the current message and

    4

prints message 4, as described previously.

preserve
Same as **hold**. Cause a list of messages to be held in your system mailbox when you quit.

quit Leave *Mail* and update the file, folder, or system mailbox your were reading. Messages that you have examined are marked as "read" and messages that existed when you started are marked as "old." If you were editing your system mailbox and if you have set the binary option *hold*, all messages which have not been deleted, saved, or mboxed will be retained in your system mailbox. If you were editing your system mailbox and you did *not* have *hold* set, all messages which have not been deleted, saved, or preserved will be moved to the file *mbox* in your home directory.

replyFrame a reply to a single message. The reply will be sent to the person who sent you the message to which you are replying, plus all the people who received the original message, except you. You can add people using the **~t** and **~c** tilde escapes. The subject in your reply is formed by prefacing the subject in the original message with "Re:" unless it already began thus. If the original message included a "reply-to" header field, the reply will go *only* to the recipient named by "reply-to." You type in your message using the same conventions available to you through the **mail** command.

save It is often useful to be able to save messages on related topics in a file. The **save** command gives you ability to do this. The **save** command takes as argument a lit of message numbers, followed by the name of the file on which to save the messages. The messages are appended to the named file, thus allowing one to keep several messages in the file, stored in the order they were put there. The **save** command can be abbreviated to **s**. An example of the **save** command relative to our running example is:

    s 1 2 tuitionmail

Saved messages are not automatically saved in *mbox* at quit time, nor are they selected by the **next** command described above, unless explicitly specified.

set Set an option or give an option a value. Used to customize *Mail*. Section 5.3 contains a list of the options. Options can be *binary*, in which case they are *on* or *off*, or *valued*. To set a binary option *option on*, do

set option

To give the valued option *option* the value *value*, do

set option=value

Several options can be specified in a single **set** command.

shell The **shell** command allows you to escape to the shell. **Shell** invokes an interactive shell and allows you to type commands to it. When you leave the shell, you will return to *Mail*. The shell used is a default assumed by *Mail*; you can override this default by setting the valued option "SHELL," eg:

set SHELL=/bin/csh

source

The **source** command reads *Mail* commands from a file. It is useful when you are trying to fix your ".mailrc" file and you need to re-read it.

top The **top** command takes a message list and prints the first five lines of each addressed message. It may be abbreviated to **to**. If you wish, you can change the number of lines that **top** prints out by setting the valued option "toplines." On a CRT terminal,

set toplines=10

might be preferred.

type Print a list of messages on your terminal. If you have set the option *crt* to a number and the total number of lines in the messages you are printing exceed that specified by *crt*, the messages will be printed by a terminal paging program such as *more*.

undelete

The **undelete** command causes a message that had been deleted previously to regain its initial status. Only messages that have been deleted may be undeleted. This command may be abbreviated to **u**.

unset

Reverse the action of setting a binary or valued option.

visual

It is often useful to be able to invoke one of two editors, based on the type of terminal one is using. To invoke a display oriented editor, you can use the **visual** command. The operation of the **visual** command is otherwise identical to that of the **edit** command.

Both the **edit** and **visual** commands assume some default text editors. These default editors can be overridden by the valued options "EDITOR" and "VISUAL" for the standard and screen editors. You might want to do:

set EDITOR=/usr/ucb/ex VISUAL=/usr/ucb/vi

write

The **save** command always writes the entire message, including the headers, into the file. If you want to write just the message itself, you can use the **write** command. The **write** command has the same syntax as the **save** command, and can be abbreviated to simply **w**. Thus, we could write the second message by doing:

w 2 file.c

As suggested by this example, the **write** command is useful for such tasks as sending and receiving source program text over the message system.

z *Mail* presents message headers in windowfuls as described under the **headers** command. You can move *Mail's* attention forward to the next window by giving the

z+

command. Analogously, you can move to the previous window with:

z–

5.3. Custom options

Throughout this manual, we have seen examples of binary and valued options. This section describes each of the options in alphabetical order, including some that you have not seen yet. To avoid confusion, please note that the options are either all lower case letters or all upper case letters. When I start a sentence such as: "Ask" causes *Mail* to prompt you for a subject header, I am only capitalizing "ask" as a courtesy to English.

EDITOR
: The valued option "EDITOR" defines the pathname of the text editor to be used in the **edit** command and ~e. If not defined, a standard editor is used.

SHELL
: The valued option "SHELL" gives the path name of your shell. This shell is used for the **!** command and ~! escape. In addition, this shell expands file names with shell metacharacters like * and ? in them.

VISUAL
: The valued option "VISUAL" defines the pathname of your screen editor for use in the **visual** command and ~v escape. A standard screen editor is used if you do not define one.

append
: The "append" option is binary and causes messages saved in *mbox* to be appended to the end rather than prepended. Normally, *Mail*will *mbox* in the same order that the system puts messages in your system mailbox. By setting "append," you are requesting that *mbox* be appended to regardless. It is in any event quicker to append.

ask "Ask" is a binary option which causes *Mail* to prompt you for the subject of each message you send. If you respond with simply a newline, no subject field will be sent.

askcc
: "Askcc" is a binary option which causes you to be prompted for additional carbon copy recipients at the end of each message. Responding with a newline shows your satisfaction with the current list.

autoprint
: "Autoprint" is a binary option which causes the **delete** command to behave like **dp** – thus, after deleting a message, the next one will be typed automatically. This is useful to quickly scanning and deleting messages in your mailbox.

debug
: The binary option "debug" causes debugging information to be displayed. Use of this option is the same as useing the

–d command line flag.

dot "Dot" is a binary option which, if set, causes *Mail* to interpret a period alone on a line as the terminator of a message you are sending.

escape
: To allow you to change the escape character used when sending mail, you can set the valued option "escape." Only the first character of the "escape" option is used, and it must be doubled if it is to appear as the first character of a line of your message. If you change your escape character, then ~ loses all its special meaning, and need no longer be doubled at the beginning of a line.

folder
The name of the directory to use for storing folders of messages. If this name begins with a '/' *Mail* considers it to be an absolute pathname; otherwise, the folder directory is found relative to your home directory.

hold The binary option "hold" causes messages that have been read but not manually dealt with to be held in the system mailbox. This prevents such messages from being automatically swept into your mbox.

ignore
The binary option "ignore" causes RUBOUT characters from your terminal to be ignored and echoed as @'s while you are sending mail. RUBOUT characters retain their original meaning in *Mail* command mode. Setting the "ignore" option is equivalent to supplying the **-i** flag on the command line as described in section 6.

ignoreeof
An option related to "dot" is "ignoreeof" which makes *Mail* refuse to accept a control-d as the end of a message. "Ignoreeof" also applies to *Mail* command mode.

keep The "keep" option causes *Mail* to truncate your system mailbox instead of deleting it when it is empty. This is useful if you elect to protect your mailbox, which you would do with the shell command:

chmod 600 /usr/spool/mail/yourname

where *yourname* is your login name. If you do not do this, anyone can probably read your mail, although people usually don't.

keepsave
When you **save** a message, *Mail* usually discards it when you **quit**. To retain all saved messages, set the "keepsave" option.

metoo
When sending mail to an alias, *Mail* makes sure that if you are included in the alias, that mail will not be sent to you. This is useful if a single alias is being used by all members of the group. If however, you wish to receive a copy of all the messages you send to the alias, you can set the binary option "metoo."

noheader
The binary option "noheader" suppresses the printing of the version and headers when *Mail* is first invoked. Setting this option is the same as using **-N** on the command line.

nosave
Normally, when you abort a message with two RUBOUTs, *Mail* copies the partial letter to the file "dead.letter" in your home directory. Setting the binary option "nosave" prevents this.

quiet The binary option "quiet" suppresses the printing of the version when *Mail* is first invoked, as well as printing the for example "Message 4:" from the **type** command.

record
If you love to keep records, then the valued option "record" can be set to the name of a file to save your outgoing mail. Each new message you send is appended to the end of the file.

screen
When *Mail* initially prints the message headers, it determines the number to print by looking at the speed of your terminal. The faster your terminal, the more it prints. The valued option "screen" overrides this calculation and specifies how many message headers you want printed. This number is also used for scrolling with the **z** command.

sendmail
To alternate delivery system, set the "sendmail" option to the full pathname of the program to use. Note: this is not for everyone! Most people should use the default delivery system.

toplines
The valued option "toplines" defines the number of lines that the "top" command will print

out instead of the default five lines.

verbose

The binary option "verbose" causes *Mail* to invoke sendmail with the **–v** flag, which causes it to go into versbose mode and announce expansion of aliases, etc. Setting the "verbose" option is equivalent to invoking *Mail* with the **–v** flag as described in section 6.

6. Command line options

This section describes command line options for *Mail* and what they are used for.

–N Suppress the initial printing of headers.

–d Turn on debugging information. Not of general interest.

–f file Show the messages in *file* instead of your system mailbox. If *file* is omitted, *Mail* reads *mbox* in your home directory.

–i Ignore tty interrupt signals. Useful on noisy phone lines, which generate spurious RUBOUT or DELETE characters. It's usually more effective to change your interrupt character to control–c, for which see the *stty* shell command.

–n Inhibit reading of /usr/lib/Mail.rc. Not generally useful, since /usr/lib/Mail.rc is usually empty.

–s string
Used for sending mail. *String* is used as the subject of the message being composed. If *string* contains blanks, you must surround it with quote marks.

–u name
Read *names's* mail instead of your own. Unwitting others often neglect to protect their mailboxes, but discretion is advised. Essentially, **–u user** is a shorthand way of doing **–f /usr/spool/user**.

–v Use the **–v** flag when invoking sendmail. This feature may also be enabled by setting the the option "verbose".

The following command line flags are also recognized, but are intended for use by programs invoking *Mail* and not for people.

–T file
Arrange to print on *file* the contents of the *article-id* fields of all messages that were either read or deleted. **–T** is for the *readnews* program and should NOT be used for reading your mail.

–h number
Pass on hop count information. *Mail* will take the number, increment it, and pass it with **–h** to the mail delivery system. **–h** only has effect when sending mail and is used for network mail forwarding.

–r name
Used for network mail forwarding: interpret *name* as the sender of the message. The *name* and **–r** are simply sent along to the mail delivery system. Also, *Mail* will wait for the message to be sent and return the exit status. Also restricts formatting of message.

Note that **–h** and **–r**, which are for network mail forwarding, are not used in practice since mail forwarding is now handled separately. They may disappear soon.

7. Format of messages

This section describes the format of messages. Messages begin with a *from* line, which consists of the word "From" followed by a user name, followed by anything, followed by a date in the format returned by the *ctime* library routine described in section 3 of the Unix Programmer's Manual. A possible *ctime* format date is:

Tue Dec 1 10:58:23 1981

The *ctime* date may be optionally followed by a single space and a time zone indication, which should be three capital letters, such as PDT.

Following the *from* line are zero or more *header field* lines. Each header field line is of the form:

name: information

Name can be anything, but only certain header fields are recognized as having any meaning. The recognized header fields are: *article-id, bcc, cc, from, reply-to, sender, subject,* and *to*. Other header fields are also significant to other systems; see, for example, the current Arpanet message standard for much more on this topic. A header field can be continued onto following lines by making the first character on the following line a space or tab character.

If any headers are present, they must be followed by a blank line. The part that follows is called the *body* of the message, and must be ASCII text, not containing null characters. Each line in the message body must be terminated with an ASCII newline character and no line may be longer than 512 characters. If binary data must be passed through the mail system, it is suggested that this data be encoded in a system which encodes six bits into a printable character. For example, one could use the upper and lower case letters, the digits, and the characters comma and period to make up the 64 characters. Then, one can send a 16-bit binary number as three characters. These characters should be packed into lines, preferably lines about 70 characters long as long lines are transmitted more efficiently.

The message delivery system always adds a blank line to the end of each message. This blank line must not be deleted.

The UUCP message delivery system sometimes adds a blank line to the end of a message each time it is forwarded through a machine.

It should be noted that some network transport protocols enforce limits to the lengths of messages.

8. Glossary

This section contains the definitions of a few phrases peculiar to *Mail*.

alias An alternative name for a person or list of people.

flag An option, given on the command line of *Mail*, prefaced with a –. For example, **–f** is a flag.

header field
At the beginning of a message, a line which contains information that is part of the structure of the message. Popular header fields include ***to***, ***cc***, and ***subject***.

mail
A collection of messages. Often used in the phrase, "Have you read your mail?"

mailbox
The place where your mail is stored, typically in the directory /usr/spool/mail.

message
A single letter from someone, initially stored in your ***mailbox***.

message list
A string used in *Mail* command mode to describe a sequence of messages.

option
A piece of special purpose information used to tailor *Mail* to your taste. Options are specified with the **set** command.

9. Summary of commands, options, and escapes

This section gives a quick summary of the *Mail* commands, binary and valued options, and tilde escapes.

The following table describes the commands:

Command	*Description*
!	Single command escape to shell
-	Back up to previous message
Print	Type message with ignored fields
Reply	Reply to author of message only
Type	Type message with ignored fields
alias	Define an alias as a set of user names
alternates	List other names you are known by
chdir	Change working directory, home by default
copy	Copy a message to a file or folder
delete	Delete a list of messages
dt	Delete current message, type next message
endif	End of conditional statement; see **if**
edit	Edit a list of messages
else	Start of else part of conditional; see **if**
exit	Leave mail without changing anything
file	Interrogate/change current mail file
folder	Same as **file**
folders	List the folders in your folder directory
from	List headers of a list of messages
headers	List current window of messages
help	Print brief summary of *Mail* commands
hold	Same as **preserve**
if	Conditional execution of *Mail* commands
ignore	Set/examine list of ignored header fields
list	List valid *Mail* commands
local	List other names for the local host
mail	Send mail to specified names
mbox	Arrange to save a list of messages in *mbox*
next	Go to next message and type it
preserve	Arrange to leave list of messages in system mailbox
quit	Leave *Mail*; update system mailbox, *mbox* as appropriate
reply	Compose a reply to a message
save	Append messages, headers included, on a file
set	Set binary or valued options
shell	Invoke an interactive shell
top	Print first so many (5 by default) lines of list of messages
type	Print messages
undelete	Undelete list of messages
unset	Undo the operation of a **set**
visual	Invoke visual editor on a list of messages
write	Append messages to a file, don't include headers
z	Scroll to next/previous screenful of headers

The following table describes the options. Each option is shown as being either a binary or valued option.

Option	*Type*	*Description*
EDITOR	*valued*	Pathname of editor for ~e and **edit**
SHELL	*valued*	Pathname of shell for **shell**, ~! and **!**
VISUAL	*valued*	Pathname of screen editor for ~v, **visual**
append	*binary*	Always append messages to end of *mbox*
ask	*binary*	Prompt user for Subject: field when sending
askcc	*binary*	Prompt user for additional Cc's at end of message
autoprint	*binary*	Print next message after **delete**
crt	*valued*	Minimum number of lines before using *more*
debug	*binary*	Print out debugging information
dot	*binary*	Accept . alone on line to terminate message input
escape	*valued*	Escape character to be used instead of ~
folder	*valued*	Directory to store folders in
hold	*binary*	Hold messages in system mailbox by default
ignore	*binary*	Ignore RUBOUT while sending mail
ignoreeof	*binary*	Don't terminate letters/command input with **↑D**
keep	*binary*	Don't unlink system mailbox when empty
keepsave	*binary*	Don't delete **saved** messages by default
metoo	*binary*	Include sending user in aliases
noheader	*binary*	Suppress initial printing of version and headers
nosave	*binary*	Don't save partial letter in *dead.letter*
quiet	*binary*	Suppress printing of *Mail* version and message numbers
record	*valued*	File to save all outgoing mail in
screen	*valued*	Size of window of message headers for **z**, etc.
sendmail	*valued*	Choose alternate mail delivery system
toplines	*valued*	Number of lines to print in **top**
verbose	*binary*	Invoke sendmail with the **–v** flag

The following table summarizes the tilde escapes available while sending mail.

Escape	*Arguments*	*Description*
~!	*command*	Execute shell command
~c	*name ...*	Add names to Cc: field
~d		Read *dead.letter* into message
~e		Invoke text editor on partial message
~f	*messages*	Read named messages
~h		Edit the header fields
~m	*messages*	Read named messages, right shift by tab
~p		Print message entered so far
~q		Abort entry of letter; like RUBOUT
~r	*filename*	Read file into message
~s	*string*	Set Subject: field to *string*
~t	*name ...*	Add names to To: field
~v		Invoke screen editor on message
~w	*filename*	Write message on file
~|	*command*	Pipe message through *command*
~~	*string*	Quote a ~ in front of *string*

The following table shows the command line flags that *Mail* accepts:

Flag	Description
–N	Suppress the initial printing of headers
–T *file*	Article-id's of read/deleted messages to *file*
–d	Turn on debugging
–f *file*	Show messages in *file* or ~/*mbox*
–h *number*	Pass on hop count for mail forwarding
–i	Ignore tty interrupt signals
–n	Inhibit reading of /usr/lib/Mail.rc
–r *name*	Pass on *name* for mail forwarding
–s *string*	Use *string* as subject in outgoing mail
–u *name*	Read *name's* mail instead of your own
–v	Invoke sendmail with the –v flag

Notes: **–T**, **–d**, **–h**, and **–r** are not for human use.

10. Conclusion

Mail is an attempt to provide a simple user interface to a variety of underlying message systems. Thanks are due to the many users who contributed ideas and testing to *Mail*.

Typing Documents on the UNIX System: Using the –ms Macros with Troff and Nroff

M. E. Lesk

Text Formatting
Phototypesetting

ABSTRACT

This document describes a set of easy-to-use macros for preparing documents on the UNIX system. Documents may be produced on either the phototypesetter or a on a computer terminal, without changing the input.

The macros provide facilities for paragraphs, sections (optionally with automatic numbering), page titles, footnotes, equations, tables, two-column format, and cover pages for papers.

This memo includes, as an appendix, the text of the "Guide to Preparing Documents with –ms" which contains additional examples of features of –ms.

This manual is a revision of, and replaces, "Typing Documents on UNIX," dated November 22, 1974.

September 16, 1986

Typing Documents on the UNIX System: Using the –ms Macros with Troff and Nroff

M. E. Lesk

Text Formatting
Phototypesetting

Introduction. This memorandum describes a package of commands to produce papers using the ***troff*** and ***nroff*** formatting programs on the UNIX system. As with other ***roff***-derived programs, text is prepared interspersed with formatting commands. However, this package, which itself is written in ***troff*** commands, provides higher-level commands than those provided with the basic ***troff*** program. The commands available in this package are listed in Appendix A.

Text. Type normally, except that instead of indenting for paragraphs, place a line reading ".PP" before each paragraph. This will produce indenting and extra space.

Alternatively, the command .LP that was used here will produce a left-aligned (block) paragraph. The paragraph spacing can be changed: see below under "Registers."

Beginning. For a document with a paper-type cover sheet, the input should start as follows:

[optional overall format .RP – see below]
.TL
Title of document (one or more lines)
.AU
Author(s) (may also be several lines)
.AI
Author's institution(s)
.AB
Abstract; to be placed on the cover sheet of a paper.
Line length is 5/6 of normal; use .ll here to change.
.AE (abstract end)
text ... (begins with .PP, which see)

To omit some of the standard headings (e.g. no abstract, or no author's institution) just omit the corresponding fields and command lines. The word ABSTRACT can be suppressed by writing ".AB no" for ".AB". Several interspersed .AU and .AI lines can be used for multiple authors. The headings are not compulsory: beginning with a .PP command is perfectly OK and will just start printing an ordinary paragraph. ***Warning:*** You can't just begin a document with a line of text. Some –ms command must precede any text input. When in doubt, use .LP to get proper initialization, although any of the commands .PP, .LP, .TL, .SH, .NH is good enough. Figure 1 shows the legal arrangement of commands at the start of a document.

Cover Sheets and First Pages. The first line of a document signals the general format of the first page. In particular, if it is ".RP" a cover sheet with title and abstract is prepared. The default format is useful for scanning drafts.

In general –ms is arranged so that only one form of a document need be stored, containing all information; the first command gives the format, and unnecessary items for that format are ignored.

Warning: don't put extraneous material between the .TL and .AE commands. Processing of the titling items is special, and other data placed in them may not behave as you expect. Don't forget that some –ms command must precede any input text.

Page headings. The –ms macros, by default, will print a page heading containing a page number (if greater than 1). A default page footer is provided only in ***nroff***, where the date is used. The user can make minor adjustments to the page headings/footings by redefining the strings LH, CH, and RH which are the left, center and right portions of the page headings, respectively; and the strings LF, CF, and RF, which are the left, center and right portions of the page footer. For more complex formats, the user can redefine the macros PT and BT, which are invoked respectively at the top and bottom of each page. The margins (taken from registers HM and FM for the top and bottom margin respectively) are normally 1 inch; the page header/footer are in the middle of that space. The user who redefines these macros should be careful not to change parameters such as point size or font without resetting them to default values.

Multi-column formats. If you place the command ".2C" in your document, the document will be printed in double column format beginning at that point. This feature is not too useful in computer terminal output, but is often desirable on the typesetter. The command ".1C" will go back to one-column format and also skip to a new page. The ".2C" command is actually a special case of the command

.MC [column width [gutter width]]

which makes multiple columns with the specified column and gutter width; as many columns as will fit across the page are used. Thus triple, quadruple, ... column pages can be printed. Whenever the number of columns is changed (except going from full width to some larger number of columns) a new page is started.

Headings. To produce a special heading, there are two commands. If you type

```
.NH
type section heading here
may be several lines
```

you will get automatically numbered section headings (1, 2, 3, ...), in boldface. For example,

```
.NH
Care and Feeding of Department Heads
```

produces

1. Care and Feeding of Department Heads

Alternatively,

```
.SH
Care and Feeding of Directors
```

will print the heading with no number added:

Care and Feeding of Directors

Every section heading, of either type, should be followed by a paragraph beginning with .PP or .LP, indicating the end of the heading. Headings may contain more than one line of text.

The .NH command also supports more complex numbering schemes. If a numerical argument is given, it is taken to be a "level" number and an appropriate sub-section number is generated. Larger level numbers indicate deeper sub-sections, as in this example:

```
.NH
Erie-Lackawanna
.NH 2
Morris and Essex Division
.NH 3
Gladstone Branch
.NH 3
Montclair Branch
.NH 2
Boonton Line
```

generates:

2. Erie-Lackawanna

2.1. Morris and Essex Division

2.1.1. Gladstone Branch

2.1.2. Montclair Branch

2.2. Boonton Line

An explicit ".NH 0" will reset the numbering of level 1 to one, as here:

```
.NH 0
Penn Central
```

1. Penn Central

Indented paragraphs. (Paragraphs with hanging numbers, e.g. references.) The sequence

```
.IP [1]
Text for first paragraph, typed
normally for as long as you would
like on as many lines as needed.
.IP [2]
Text for second paragraph, ...
```

produces

[1] Text for first paragraph, typed normally for as long as you would like on as many lines as needed.

[2] Text for second paragraph, ...

A series of indented paragraphs may be followed by an ordinary paragraph beginning with .PP or .LP, depending on whether you wish indenting or not. The command .LP was used here.

More sophisticated uses of .IP are also possible. If the label is omitted, for example, a plain block indent is produced.

```
.IP
This material will
just be turned into a
block indent suitable for quotations or
such matter.
.LP
```

will produce

> This material will just be turned into a block indent suitable for quotations or such matter.

If a non-standard amount of indenting is required, it may be specified after the label (in character positions) and will remain in effect until the next .PP or .LP. Thus, the general form of the .IP command contains two additional fields: the label and the indenting length. For example,

```
.IP first: 9
Notice the longer label, requiring larger
indenting for these paragraphs.
.IP second:
And so forth.
.LP
```

produces this:

first: Notice the longer label, requiring larger indenting for these paragraphs.

second: And so forth.

It is also possible to produce multiple nested indents; the command .RS indicates that the next .IP starts from the current indentation level. Each .RE will eat up one level of indenting so you should balance .RS and .RE commands. The .RS command should be thought of as "move right" and the .RE command as "move left". As an example

```
.IP 1.
Bell Laboratories
.RS
.IP 1.1
Murray Hill
.IP 1.2
Holmdel
.IP 1.3
Whippany
.RS
.IP 1.3.1
Madison
.RE
.IP 1.4
Chester
.RE
.LP
```

will result in

1. Bell Laboratories
    - 1.1 Murray Hill
    - 1.2 Holmdel
    - 1.3 Whippany
        - 1.3.1 Madison
    - 1.4 Chester

All of these variations on .LP leave the right margin untouched. Sometimes, for purposes such as setting off a quotation, a paragraph indented on both right and left is required.

> A single paragraph like this is obtained by preceding it with .QP. More complicated material (several paragraphs) should be bracketed with .QS and .QE.

Emphasis. To get italics (on the typesetter) or underlining (on the terminal) say

```
.I
as much text as you want
can be typed here
.R
```

as was done for *these three words.* The .R command restores the normal (usually

Roman) font. If only one word is to be italicized, it may be just given on the line with the .I command,

.I word

and in this case no .R is needed to restore the previous font. **Boldface** can be produced by

.B
Text to be set in boldface
goes here
.R

and also will be underlined on the terminal or line printer. As with .I, a single word can be placed in boldface by placing it on the same line as the .B command.

A few size changes can be specified similarly with the commands .LG (make larger), .SM (make smaller), and .NL (return to normal size). The size change is two points; the commands may be repeated for increased effect (here one .NL canceled two .SM commands).

If actual underlining as opposed to italicizing is required on the typesetter, the command

.UL word

will underline a word. There is no way to underline multiple words on the typesetter.

Footnotes. Material placed between lines with the commands .FS (footnote) and .FE (footnote end) will be collected, remembered, and finally placed at the bottom of the current page*. By default, footnotes are 11/12th the length of normal text, but this can be changed using the FL register (see below).

Displays and Tables. To prepare displays of lines, such as tables, in which the lines should not be re-arranged, enclose them in the commands .DS and .DE

.DS
table lines, like the
examples here, are placed
between .DS and .DE
.DE

By default, lines between .DS and .DE are indented and left-adjusted. You can also center lines, or retain the left margin. Lines bracketed by .DS C and .DE commands are centered (and not re-arranged); lines bracketed by .DS L and .DE are left-adjusted, not indented, and not re-arranged. A plain .DS is equivalent to .DS I, which indents and left-adjusts. Thus,

these lines were preceded
by .DS C and followed by
a .DE command;

whereas

these lines were preceded
by .DS L and followed by
a .DE command.

Note that .DS C centers each line; there is a variant .DS B that makes the display into a left-adjusted block of text, and then centers that entire block. Normally a display is kept together, on one page. If you wish to have a long display which may be split across page boundaries, use .CD, .LD, or .ID in place of the commands .DS C, .DS L, or .DS I respectively. An extra argument to the .DS I or .DS command is taken as an amount to indent. Note: it is tempting to assume that .DS R will right adjust lines, but it doesn't work.

Boxing words or lines. To draw rectangular boxes around words the command

.BX word

will print $\boxed{\text{word}}$ as shown. The boxes will not be neat on a terminal, and this should not be used as a substitute for italics.

Longer pieces of text may be boxed by enclosing them with .B1 and .B2: .B1 text... .B2 as has been done here.

Keeping blocks together. If you wish to keep a table or other block of lines together on a page, there are "keep - release" commands. If a block of lines preceded by .KS and followed by .KE does not fit on the remainder of the current page, it will begin on a new page. Lines bracketed by .DS and .DE commands are automatically kept together this way. There is also a "keep floating" command: if the block to be kept together is preceded by .KF instead of .KS and does not fit on the current page, it will be moved down through the text until the top of the next

* Like this.

page. Thus, no large blank space will be introduced in the document.

Nroff/Troff commands. Among the useful commands from the basic formatting programs are the following. They all work with both typesetter and computer terminal output:

.bp - begin new page.
.br - "break", stop running text from line to line.
.sp n - insert n blank lines.
.na - don't adjust right margins.

Date. By default, documents produced on computer terminals have the date at the bottom of each page; documents produced on the typesetter don't. To force the date, say ".DA". To force no date, say ".ND". To lie about the date, say ".DA July 4, 1776" which puts the specified date at the bottom of each page. The command

.ND May 8, 1945

in ".RP" format places the specified date on the cover sheet and nowhere else. Place this line before the title.

Signature line. You can obtain a signature line by placing the command .SG in the document. The authors' names will be output in place of the .SG line. An argument to .SG is used as a typing identification line, and placed after the signatures. The .SG command is ignored in released paper format.

Registers. Certain of the registers used by –ms can be altered to change default settings. They should be changed with .nr commands, as with

.nr PS 9

to make the default point size 9 point. If the effect is needed immediately, the normal *troff* command should be used in addition to changing the number register.

Register	Defines	Takes effect	Default
PS	point size	next para.	10
VS	line spacing	next para.	12 pts
LL	line length	next para.	6''
LT	title length	next para.	6''
PD	para. spacing	next para.	0.3 VS
PI	para. indent	next para.	5 ens
FL	footnote length	next FS	11/12 LL
CW	column width	next 2C	7/15 LL
GW	intercolumn gap	next 2C	1/15 LL
PO	page offset	next page	26/27''
HM	top margin	next page	1''
FM	bottom margin	next page	1''

You may also alter the strings LH, CH, and RH which are the left, center, and right headings respectively; and similarly LF, CF, and RF which are strings in the page footer. The page number on *output* is taken from register PN, to permit changing its output style. For more complicated headers and footers the macros PT and BT can be redefined, as explained earlier.

Accents. To simplify typing certain foreign words, strings representing common accent marks are defined. They precede the letter over which the mark is to appear. Here are the strings:

Input	Output	Input	Output
*'e	é	*~a	ã
*`e	è	*Ce	ě
*:u	ü	*,c	ç
*^e	ê		

Use. After your document is prepared and stored on a file, you can print it on a terminal with the command*

nroff –ms file

and you can print it on the typesetter with the command

troff –ms file

(many options are possible). In each case, if your document is stored in several files, just list all the filenames where we have used "file". If equations or tables are used, *eqn* and/or *tbl* must be invoked as preprocessors.

References and further study. If you have to do Greek or mathematics, see *eqn* [1] for equation setting. To aid *eqn* users, *–ms* provides definitions of .EQ and .EN which normally center the equation and set it off slightly. An argument on .EQ is taken to be an equation number and placed in the right margin near the equation. In addition, there are three special arguments to EQ: the letters C, I, and L indicate centered (default), indented, and left adjusted equations, respectively. If there is both a format argument and an equation number, give the format argu-

* If .2C was used, pipe the *nroff* output through *col;* make the first line of the input ".pi /usr/bin/col."

ment first, as in

.EQ L (1.3a)

for a left-adjusted equation numbered (1.3a).

Similarly, the macros .TS and .TE are defined to separate tables (see [2]) from text with a little space. A very long table with a heading may be broken across pages by beginning it with .TS H instead of .TS, and placing the line .TH in the table data after the heading. If the table has no heading repeated from page to page, just use the ordinary .TS and .TE macros.

To learn more about *troff* see [3] for a general introduction, and [4] for the full details (experts only). Information on related UNIX commands is in [5]. For jobs that do not seem well-adapted to –ms, consider other macro packages. It is often far easier to write a specific macro packages for such tasks as imitating particular journals than to try to adapt –ms.

Acknowledgment. Many thanks are due to Brian Kernighan for his help in the design and implementation of this package, and for his assistance in preparing this manual.

References

[1] B. W. Kernighan and L. L. Cherry, *Typesetting Mathematics — Users Guide (2nd edition),* Bell Laboratories Computing Science Report no. 17.

[2] M. E. Lesk, *Tbl — A Program to Format Tables,* Bell Laboratories Computing Science Report no. 45.

[3] B. W. Kernighan, *A Troff Tutorial,* Bell Laboratories, 1976.

[4] J. F. Ossanna, *Nroff/Troff Reference Manual,* Bell Laboratories Computing Science Report no. 51.

[5] K. Thompson and D. M. Ritchie, *UNIX Programmer's Manual,* Bell Laboratories, 1978.

Appendix A
List of Commands

1C	Return to single column format.
2C	Start double column format.
AB	Begin abstract.
AE	End abstract.
AI	Specify author's institution.
AU	Specify author.
B	Begin boldface.
DA	Provide the date on each page.
DE	End display.
DS	Start display (also CD, LD, ID).
EN	End equation.
EQ	Begin equation.
FE	End footnote.
FS	Begin footnote.
I	Begin italics.
IP	Begin indented paragraph.
KE	Release keep.
KF	Begin floating keep.
KS	Start keep.
LG	Increase type size.
LP	Left aligned block paragraph.
ND	Change or cancel date.
NH	Specify numbered heading.
NL	Return to normal type size.
PP	Begin paragraph.
R	Return to regular font (usually Roman).
RE	End one level of relative indenting.
RP	Use released paper format.
RS	Relative indent increased one level.
SG	Insert signature line.
SH	Specify section heading.
SM	Change to smaller type size.
TL	Specify title.
UL	Underline one word.

Register Names

The following register names are used by –ms internally. Independent use of these names in one's own macros may produce incorrect output. Note that no lower case letters are used in any –ms internal name.

Number registers used in –ms

:	DW	GW	HM	IQ	LL	NA	OJ	PO	T.	TV
#T	EF	H1	HT	IR	LT	NC	PD	PQ	TB	VS
1T	FL	H3	IK	KI	MM	NF	PF	PX	TD	YE
AV	FM	H4	IM	L1	MN	NS	PI	RO	TN	YY
CW	FP	H5	IP	LE	MO	OI	PN	ST	TQ	ZN

String registers used in –ms

'	A5	CB	DW	EZ	I	KF	MR	R1	RT	TL
`	AB	CC	DY	FA	I1	KQ	ND	R2	S0	TM
^	AE	CD	E1	FE	I2	KS	NH	R3	S1	TQ
~	AI	CF	E2	FJ	I3	LB	NL	R4	S2	TS
:	AU	CH	E3	FK	I4	LD	NP	R5	SG	TT
,	B	CM	E4	FN	I5	LG	OD	RC	SH	UL
1C	BG	CS	E5	FO	ID	LP	OK	RE	SM	WB
2C	BT	CT	EE	FQ	IE	ME	PP	RF	SN	WH
A1	C	D	EL	FS	IM	MF	PT	RH	SY	WT
A2	C1	DA	EM	FV	IP	MH	PY	RP	TA	XD
A3	C2	DE	EN	FY	IZ	MN	QF	RQ	TE	XF
A4	CA	DS	EQ	HO	KE	MO	R	RS	TH	XK

Order of Commands in Input

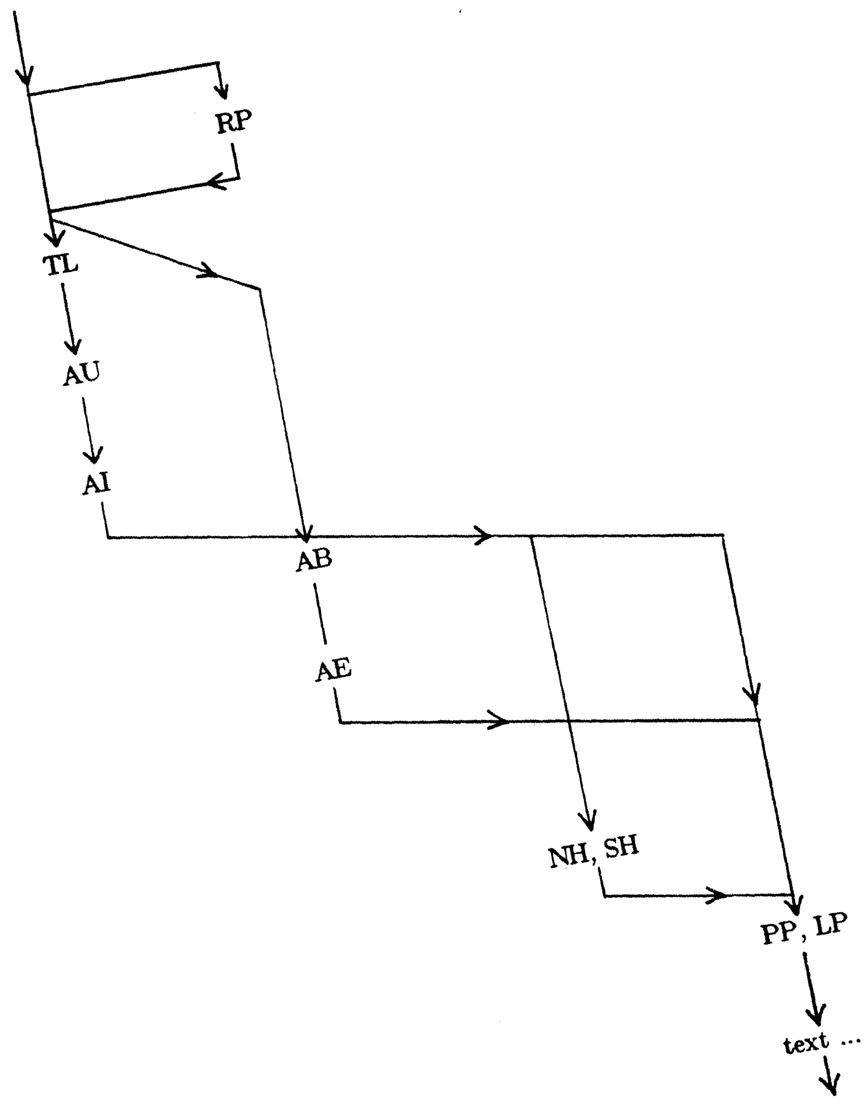


Figure 1

A Guide to Preparing Documents with −ms

M. E. Lesk

Bell Laboratories

August 1978

This guide gives some simple examples of document preparation on Bell Labs computers, emphasizing the use of the *−ms* macro package. It enormously abbreviates information in

1. *Typing Documents on UNIX and GCOS,* by M. E. Lesk;
2. *Typesetting Mathematics – User's Guide,* by B. W. Kernighan and L. L. Cherry; and
3. *Tbl – A Program to Format Tables,* by M. E. Lesk.

These memos are all included in the *UNIX Programmer's Manual, Volume 2.* The new user should also have *A Tutorial Introduction to the UNIX Text Editor,* by B. W. Kernighan.

For more detailed information, read *Advanced Editing on UNIX* and *A Troff Tutorial,* by B. W. Kernighan, and (for experts) *Nroff/Troff Reference Manual* by J. F. Ossanna. Information on related commands is found (for UNIX users) in *UNIX for Beginners* by B. W. Kernighan and the *UNIX Programmer's Manual* by K. Thompson and D. M. Ritchie.

Contents

Throughout the examples, input is shown in
this Helvetica sans serif font
while the resulting output is shown in
this Times Roman font.

UNIX Document no. 1111

Commands for a TM

```
.TM 1978-5b3 99999 99999-11
.ND April 1, 1976
.TL
The Role of the Allen Wrench in Modern
Electronics
.AU "MH 2G-111" 2345
J. Q. Pencilpusher
.AU "MH 1K-222" 5432
X. Y. Hardwired
.AI
.MH
.OK
Tools
Design
.AB
This abstract should be short enough to
fit on a single page cover sheet.
It must attract the reader into sending for
the complete memorandum.
.AE
.CS 10 2 12 5 6 7
.NH
Introduction.
.PP
Now the first paragraph of actual text ...
...
Last line of text.
.SG MH-1234-JQP/XYH-unix
.NH
References ...
```

Commands not needed in a particular format are ignored.

Bell Laboratories Cover Sheet for TM

This information is for employees of Bell Laboratories. (GEI 13.9-3)

Title- The Role of the Allen Wrench in Modern Electronics

Date- April 1, 1976

TM- 1978-5b3

Other Keywords- Tools
Design

Author	Location	Ext.	
J. Q. Pencilpusher	MH 2G-111	2345	Charging Case- 99999
X. Y. Hardwired	MH 1K-222	5432	Filing Case- 99999a

ABSTRACT

This abstract should be short enough to fit on a single page cover sheet. It must attract the reader into sending for the complete memorandum.

Pages Text	10	Other	2	Total	12
No. Figures	5	No. Tables	6	No. Refs.	7

E-1932-U (6-73) SEE REVERSE SIDE FOR DISTRIBUTION LIST

A Released Paper with Mathematics

```
.EQ
delim $$
.EN
.RP

... (as for a TM)

.CS 10 2 12 5 6 7
.NH
Introduction
.PP
The solution to the torque handle equation
.EQ (1)
sum from 0 to inf F ( x sub i ) == G ( x )
.EN
is found with the transformation $ x == rho over
theta $ where $ rho == G prime (x) $ and $theta$
is derived ...
```

The Role of the Allen Wrench
in Modern Electronics

J. Q. Pencilpusher

X. Y. Hardwired

Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

This abstract should be short enough to fit on a single page cover sheet. It must attract the reader into sending for the complete memorandum.

April 1, 1976

The Role of the Allen Wrench
in Modern Electronics

J. Q. Pencilpusher

X. Y. Hardwired

Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction

The solution to the torque handle equation

$$\sum_{0}^{\infty} F(x_i) = G(x) \tag{1}$$

is found with the transformation $x = \frac{\rho}{\theta}$ where $\rho = G'(x)$ and θ is derived from well-known principles.

An Internal Memorandum

```
.IM
.ND January 24, 1956
.TL
The 1956 Consent Decree
.AU
Able, Baker &
Charley, Attys.
.PP
Plaintiff, United States of America, having filed
its complaint herein on January 14, 1949; the
defendants having appeared and filed their
answer to such complaint denying the
substantive allegations thereof; and the parties,
by their attorneys, ...
```

Bell Laboratories

Subject: The 1956 Consent Decree date: January 24, 1956

from: Able, Baker &
Charley, Attys.

Plaintiff, United States of America, having filed its complaint herein on January 14, 1949; the defendants having appeared and filed their answer to such complaint denying the substantive allegations thereof; and the parties, by their attorneys, having severally consented to the entry of this Final Judgment without trial or adjudication of any issues of fact or law herein and without this Final Judgment constituting any evidence or admission by any party in respect of any such issues;

Now, therefore before any testimony has been taken herein, and without trial or adjudication of any issue of fact or law herein, and upon the consent of all parties hereto, it is hereby

Ordered, adjudged and decreed as follows:

I. [Sherman Act]

This Court has jurisdiction of the subject matter herein and of all the parties hereto. The complaint states a claim upon which relief may be granted against each of the defendants under Sections 1, 2 and 3 of the Act of Congress of July 2, 1890, entitled "An act to protect trade and commerce against unlawful restraints and monopolies," commonly known as the Sherman Act, as amended.

II. [Definitions]

For the purposes of this Final Judgment:

(a) "Western" shall mean the defendant Western Electric Company, Incorporated.

Other formats possible (specify before .TL) are: .MR ("memo for record"), .MF ("memo for file"), .EG ("engineer's notes") and .TR (Computing Science Tech. Report).

Headings

```
.NH
Introduction.
.PP
text text text
```

1. Introduction

text text text

```
.SH
Appendix I
.PP
text text text
```

Appendix I

text text text

A Simple List

```
.IP 1.
J. Pencilpusher and X. Hardwired,
.I
A New Kind of Set Screw,
.R
Proc. IEEE
.B 75
(1976), 23-255.
.IP 2.
H. Nails and R. Irons,
.I
Fasteners for Printed Circuit Boards,
.R
Proc. ASME
.B 23
(1974), 23-24.
.LP (terminates list)
```

1. J. Pencilpusher and X. Hardwired, *A New Kind of Set Screw,* Proc. IEEE **75** (1976), 23-255.
2. H. Nails and R. Irons, *Fasteners for Printed Circuit Boards,* Proc. ASME **23** (1974), 23-24.

Displays

```
text text text text text text
.DS
and now
for something
completely different
.DE
text text text text text text
```

hoboken harrison newark roseville avenue grove street east orange brick church orange highland avenue mountain station south orange maplewood millburn short hills summit new providence

    and now
    for something
    completely different

murray hill berkeley heights gillette stirling millington lyons basking ridge bernardsville far hills peapack gladstone

Options: .DS L: left-adjust; .DS C: line-by-line center; .DS B: make block, then center.

Footnotes

```
Among the most important occupants
of the workbench are the long-nosed pliers.
Without these basic tools*
.FS
* As first shown by Tiger & Leopard
(1975).
.FE
few assemblies could be completed. They may
lack the popular appeal of the sledgehammer
```

Among the most important occupants of the workbench are the long-nosed pliers. Without these basic tools* few assemblies could be completed. They may lack the popular appeal of the sledgehammer

* As first shown by Tiger & Leopard (1975).

Multiple Indents

```
This is ordinary text to point out
the margins of the page.
.IP 1.
First level item
.RS
.IP a)
Second level.
.IP b)
Continued here with another second
level item, but somewhat longer.
.RE
.IP 2.
Return to previous value of the
indenting at this point.
.IP 3.
Another
line.
```

This is ordinary text to point out the margins of the page.

1. First level item
   a) Second level.
   b) Continued here with another second level item, but somewhat longer.
2. Return to previous value of the indenting at this point.
3. Another line.

Keeps

Lines bracketed by the following commands are kept together, and will appear entirely on one page:

.KS	not moved	.KF	may float
.KE	through text	.KE	in text

Double Column

```
.TL
The Declaration of Independence
.2C
.PP
When in the course of human events, it becomes
necessary for one people to dissolve the
political bonds which have connected them with
another, and to assume among the powers of the
earth the separate and equal station to which
the laws of Nature and of Nature's God entitle
them, a decent respect to the opinions of
```

The Declaration of Independence

When in the course of human events, it becomes necessary for one people to dissolve the political bonds which have connected them with another, and to assume among the powers of the earth the separate and equal station to which the laws of Nature and of Nature's God entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation.

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their creator with certain unalienable rights, that among these are life, liberty, and the pursuit of happiness. That to secure these rights, governments are instituted among men,

Equations

```
A displayed equation is marked
with an equation number at the right margin
by adding an argument to the EQ line:
.EQ (1.3)
x sup 2 over a sup 2 ~=~ sqrt {p z sup 2 +qz+r}
.EN
```

A displayed equation is marked with an equation number at the right margin by adding an argument to the EQ line:

$$\frac{x^2}{a^2} = \sqrt{pz^2+qz+r} \qquad (1.3)$$

```
.EQ I (2.2a)
bold V bar sub nu~=~left [ pile {a above b above
c } right] + left [ matrix { col { A(11) above .
above . } col { . above . above .} col {. above .
above A(33) }} right ] cdot left [ pile { alpha
above beta above gamma } right]
.EN
```

$$\bar{\mathbf{V}}_\nu = \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} A(11) & . & . \\ . & . & . \\ . & . & A(33) \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \qquad (2.2a)$$

```
.EQ L
F hat ( chi ) ~ mark =~ |del V |sup 2
.EN
.EQ L
lineup =~ {left ( {partial V} over {partial x} right )
} sup 2 + { left ( {partial V} over {partial y} right
) } sup 2 ~~~~~~ lambda -> inf
.EN
```

$$\hat{F}(x) = |\nabla V|^2$$

$$= \left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2 \qquad \lambda\to\infty$$

```
$ a dot $, $ b dotdot$, $ xi tilde times y vec$:
```

$\dot{a}$, $\ddot{b}$, $\tilde{\xi}\times\vec{y}$. (with delim $$ on, see panel 3).

See also the equations in the second table, panel 8.

Some Registers You Can Change

Line length
  .nr LL 7i

Title length
  .nr LT 7i

Point size
  .nr PS 9

Vertical spacing
  .nr VS 11

Column width
  .nr CW 3i

Intercolumn spacing
  .nr GW .5i

Margins – head and foot
  .nr HM .75i
  .nr FM .75i

Paragraph indent
  .nr PI 2n

Paragraph spacing
  .nr PD 0

Page offset
  .nr PO 0.5i

Page heading
  .ds CH Appendix (center)
  .ds RH 7-25-76 (right)
  .ds LH Private (left)

Page footer
  .ds CF Draft
  .ds LF
  .ds RF similar

Page numbers
  .nr % 3

Tables

```
.TS          (Ⓣ indicates a tab)
allbox;
c s s
c c c
n n n.
AT&T Common Stock
YearⓉPriceⓉDividend
1971Ⓣ41-54Ⓣ$2.60
2Ⓣ41-54Ⓣ2.70
3Ⓣ46-55Ⓣ2.87
4Ⓣ40-53Ⓣ3.24
5Ⓣ45-52Ⓣ3.40
6Ⓣ51-59Ⓣ.95*
.TE
* (first quarter only)
```

AT&T Common Stock		
Year	Price	Dividend
1971	41-54	$2.60
2	41-54	2.70
3	46-55	2.87
4	40-53	3.24
5	45-52	3.40
6	51-59	.95*

* (first quarter only)

The meanings of the key-letters describing the alignment of each entry are:

c	center	**n**	numerical
r	right-adjust	**a**	subcolumn
l	left-adjust	**s**	spanned

The global table options are **center, expand, box, doublebox, allbox, tab** (*x*) and **linesize** (*n*).

```
.TS          (with delim $$ on, see panel 3)
doublebox, center;
c c
l l.
NameⓉDefinition
.sp
GammaⓉ$ GAMMA (z) = int sub 0 sup inf \
  t sup {z-1} e sup -t dt$
SineⓉ$ sin (x) = 1 over 2 i ( e sup ix - e sup -ix )$
ErrorⓉ$ roman erf (z) = 2 over sqrt pi \
  int sub 0 sup z e sup {-t sup 2} dt$
BesselⓉ$ J sub 0 (z) = 1 over pi \
  int sub 0 sup pi cos ( z sin theta ) d theta $
ZetaⓉ$ zeta (s) = \
  sum from k=1 to inf k sup -s ~~( Re~s > 1)$
.TE
```

Name	Definition
Gamma	$\Gamma(z)=\int_0^\infty t^{z-1}e^{-t}\,dt$
Sine	$\sin(x)=\frac{1}{2i}(e^{ix}-e^{-ix})$
Error	$\mathrm{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$
Bessel	$J_0(z)=\frac{1}{\pi}\int_0^\pi \cos(z\sin\theta)\,d\theta$
Zeta	$\zeta(s)=\sum_{k=1}^\infty k^{-s}$ (Re $s>1$)

Usage

Documents with just text:
troff -ms files
With equations only:
eqn files | troff -ms
With tables only:
tbl files | troff -ms
With both tables and equations:
tbl files | eqn | troff -ms

The above generates STARE output on GCOS: replace **– st** with **– ph** for typesetter output.

A Revised Version of −ms

Bill Tuthill

Computing Services
University of California
Berkeley, CA 94720

The −ms macros have been slightly revised and rearranged. Because of the rearrangement, the new macros can be read by the computer in about half the time required by the previous version of −ms. This means that output will begin to appear between ten seconds and several minutes more quickly, depending on the system load. On long files, however, the savings in total time are not substantial. The old version of −ms is still available as −mos.

Several bugs in −ms have been fixed, including a bad problem with the .1C macro, minor difficulties with boxed text, a break induced by .EQ before initialization, the failure to set tab stops in displays, and several bothersome errors in the **refer** macros. Macros used only at Bell Laboratories have been removed. There are a few extensions to previous −ms macros, and a number of new macros, but all the documented −ms macros still work exactly as they did before, and have the same names as before. Output produced with −ms should look like output produced with −mos.

One important new feature is automatically numbered footnotes. Footnote numbers are printed by means of a pre-defined string (**), which you invoke separately from .FS and .FE. Each time it is used, this string increases the footnote number by one, whether or not you use .FS and .FE in your text. Footnote numbers will be superscripted on the phototypesetter and on daisy-wheel terminals, but on low-resolution devices (such as the lpr and a crt), they will be bracketed. If you use ** to indicate numbered footnotes, then the .FS macro will automatically include the footnote number at the bottom of the page. This footnote, for example, was produced as follows:[1]

```
This footnote, for example, was produced as follows:\**
.FS
    ...
.FE
```

If you are using ** to number footnotes, but want a particular footnote to be marked with an asterisk or a dagger, then give that mark as the first argument to .FS: †

```
then give that mark as the first argument to .FS: \(dg
.FS  \(dg
    ...
.FE
```

Footnote numbering will be temporarily suspended, because the ** string is not used. Instead of a dagger, you could use an asterisk * or double dagger ‡, represented as \(dd.

Another new feature is a macro for printing theses according to Berkeley standards. This macro is called .TM, which stands for thesis mode. (It is much like the .th macro in −me.) It will put page numbers in the upper right-hand corner; number the first page; suppress the date; and doublespace everything except quotes, displays, and keeps. Use it at the top of each file making up your thesis. Calling .TM defines the

[1] If you never use the "**" string, no footnote numbers will appear anywhere in the text, including down here. The output footnotes will look exactly like footnotes produced with −mos.

† In the footnote, the dagger will appear where the footnote number would otherwise appear, as on the left.

.CT macro for chapter titles, which skips to a new page and moves the pagenumber to the center footer. The .P1 (P one) macro can be used even without thesis mode to print the header on page 1, which is suppressed except in thesis mode. If you want roman numeral page numbering, use an ".af PN i" request.

There is a new macro especially for bibliography entries, called .XP, which stands for exdented paragraph. It will exdent the first line of the paragraph by \n(PI units, usually 5n (the same as the indent for the first line of a .PP). Most bibliographies are printed this way. Here are some examples of exdented paragraphs:

Lumley, Lyle S., *Sex in Crustaceans: Shell Fish Habits,* Harbinger Press, Tampa Bay and San Diego, October 1979. 243 pages. The pioneering work in this field.

Leffadinger, Harry A., "Mollusk Mating Season: 52 Weeks, or All Year?" in *Acta Biologica,* vol. 42, no. 11, November 1980. A provocative thesis, but the conclusions are wrong.

Of course, you will have to take care of italicizing the book title and journal, and quoting the title of the journal article. Indentation or exdentation can be changed by setting the value of number register PI.

If you need to produce endnotes rather than footnotes, put the references in a file of their own. This is similar to what you would do if you were typing the paper on a conventional typewriter. Note that you can use automatic footnote numbering without actually having .FS and .FE pairs in your text. If you place footnotes in a separate file, you can use .IP macros with ** as a hanging tag; this will give you numbers at the left-hand margin. With some styles of endnotes, you would want to use .PP rather then .IP macros, and specify ** before the reference begins.

There are four new macros to help produce a table of contents. Table of contents entries must be enclosed in .XS and .XE pairs, with optional .XA macros for additional entries; arguments to .XS and .XA specify the page number, to be printed at the right. A final .PX macro prints out the table of contents. Here is a sample of typical input and output text:

```
.XS ii
Introduction
.XA 1
Chapter 1: Review of the Literature
.XA 23
Chapter 2: Experimental Evidence
.XE
.PX
```

Table of Contents

Introduction	ii
Chapter 1: Review of the Literature	1
Chapter 2: Experimental Evidence	23

The .XS and .XE pairs may also be used in the text, after a section header for instance, in which case page numbers are supplied automatically. However, most documents that require a table of contents are too long to produce in one run, which is necessary if this method is to work. It is recommended that you do a table of contents after finishing your document. To print out the table of contents, use the .PX macro; if you forget it, nothing will happen.

As an aid in producing text that will format correctly with both **nroff** and **troff**, there are some new string definitions that define quotation marks and dashes for each of these two formatting programs. The *- string will yield two hyphens in **nroff**, but in **troff** it will produce an em dash— like this one. The *Q and *U strings will produce " and " in **troff**, but " in **nroff**. (In typesetting, the double quote is traditionally considered bad form.)

There are now a large number of optional foreign accent marks defined by the -ms macros. All the accent marks available in -mos are present, and they all work just as they always did. However, there are better definitions available by placing .AM at the beginning of your document. Unlike the -mos accent marks, the accent strings should come ***after*** the letter being accented. Here is a list of the diacritical marks, with examples of what they look like.

name of accent	input	output
acute accent	e*´	é
grave accent	e*`	è
circumflex	o*^	ô
cedilla	c*,	ç
tilde	n*~	ñ
question	*?	¿
exclamation	*!	¡
umlaut	u*:	ü
digraph s	*8	ß
haček	c*v	č
macron	a*_	ā
underdot	s*.	ṣ
o-slash	o*/	ø
angstrom	a*o	å
yogh	kni*3t	kniȝt
Thorn	*(Th	Þ
thorn	*(th	þ
Eth	*(D-	Ð
eth	*(d-	ð
hooked o	*q	ǫ
ae ligature	*(ae	æ
AE ligature	*(Ae	Æ
oe ligature	*(oe	œ
OE ligature	*(Oe	Œ

If you want to use these new diacritical marks, don't forget the .AM at the top of your file. Without it, some will not print at all, and others will be placed on the wrong letter.

It is also possible to produce custom headers and footers that are different on even and odd pages. The .OH and .EH macros define odd and even headers, while .OF and .EF define odd and even footers. Arguments to these four macros are specified as with .tl. This document was produced with:

```
.OH  \fIThe -mx Macros´´Page %\fP´
.EH  \fIPage %´´The -mx Macros\fP´
```

Note that it would be a error to have an apostrophe in the header text; if you need one, you will have to use a different delimiter around the left, center, and right portions of the title. You can use any character as a delimiter, provided it doesn't appear elsewhere in the argument to .OH, .EH, .OF, or EF.

The -ms macros work in conjunction with the **tbl**, **eqn**, and **refer** preprocessors. Macros to deal with these items are read in only as needed, as are the thesis macros (.TM), the special accent mark definitions (.AM), table of contents macros (.XS and .XE), and macros to format the optional cover page. The code for the -ms package lives in /usr/lib/tmac/tmac.s, and sourced files reside in the directory /usr/ucb/lib/ms.

WRITING PAPERS WITH NROFF USING –ME

Eric P. Allman

Electronics Research Laboratory
University of California, Berkeley
Berkeley, California 94720

This document describes the text processing facilities available on the UNIX† operating system via NROFF† and the –me macro package. It is assumed that the reader already is generally familiar with the UNIX operating system and a text editor such as **ex**. This is intended to be a casual introduction, and as such not all material is covered. In particular, many variations and additional features of the –me macro package are not explained. For a complete discussion of this and other issues, see *The –me Reference Manual* and *The NROFF/TROFF Reference Manual.*

NROFF, a computer program that runs on the UNIX operating system, reads an input file prepared by the user and outputs a formatted paper suitable for publication or framing. The input consists of *text,* or words to be printed, and *requests,* which give instructions to the NROFF program telling how to format the printed copy.

Section 1 describes the basics of text processing. Section 2 describes the basic requests. Section 3 introduces displays. Annotations, such as footnotes, are handled in section 4. The more complex requests which are not discussed in section 2 are covered in section 5. Finally, section 6 discusses things you will need to know if you want to typeset documents. If you are a novice, you probably won't want to read beyond section 4 until you have tried some of the basic features out.

When you have your raw text ready, call the NROFF formatter by typing as a request to the UNIX shell:

nroff –me –T*type files*

where *type* describes the type of terminal you are outputting to. Common values are **dtc** for a DTC 300s (daisy-wheel type) printer and **lpr** for the line printer. If the **–T** flag is omitted, a "lowest common denominator" terminal is assumed; this is good for previewing output on most terminals. A complete description of options to the NROFF command can be found in *The NROFF/TROFF Reference Manual.*

The word *argument* is used in this manual to mean a word or number which appears on the same line as a request which modifies the meaning of that request. For example, the request

.sp

spaces one line, but

.sp 4

spaces four lines. The number **4** is an *argument* to the **.sp** request which says to space four lines instead of one. Arguments are separated from the request and from each other by spaces.

†UNIX, NROFF, and TROFF are Trademarks of Bell Laboratories

1. Basics of Text Processing

The primary function of NROFF is to ***collect*** words from input lines, ***fill*** output lines with those words, ***justify*** the right hand margin by inserting extra spaces in the line, and output the result. For example, the input:

Now is the time
for all good men
to come to the aid
of their party.
Four score and seven
years ago,...

will be read, packed onto output lines, and justified to produce:

Now is the time for all good men to come to the aid of their party. Four score and seven years ago,...

Sometimes you may want to start a new output line even though the line you are on is not yet full; for example, at the end of a paragraph. To do this you can cause a ***break***, which starts a new output line. Some requests cause a break automatically, as do blank input lines and input lines beginning with a space.

Not all input lines are text to be formatted. Some of the input lines are ***requests*** which describe how to format the text. Requests always have a period or an apostrophe (" ′ ") as the first character of the input line.

The text formatter also does more complex things, such as automatically numbering pages, skipping over page folds, putting footnotes in the correct place, and so forth.

I can offer you a few hints for preparing text for input to NROFF. First, keep the input lines short. Short input lines are easier to edit, and NROFF will pack words onto longer lines for you anyhow. In keeping with this, it is helpful to begin a new line after every period, comma, or phrase, since common corrections are to add or delete sentences or phrases. Second, do not put spaces at the end of lines, since this can sometimes confuse the NROFF processor. Third, do not hyphenate words at the end of lines (except words that should have hyphens in them, such as "mother-in-law"); NROFF is smart enough to hyphenate words for you as needed, but is not smart enough to take hyphens out and join a word back together. Also, words such as "mother-in-law" should not be broken over a line, since then you will get a space where not wanted, such as "mother- in-law".

2. Basic Requests

2.1. Paragraphs

Paragraphs are begun by using the **.pp** request. For example, the input:

.pp
Now is the time for all good men
to come to the aid of their party.
Four score and seven years ago,...

produces a blank line followed by an indented first line. The result is:

Now is the time for all good men to come to the aid of their party. Four score and seven years ago,...

Notice that the sentences of the paragraphs ***must not*** begin with a space, since blank lines and lines begining with spaces cause a break. For example, if I had typed:

```
.pp
Now is the time for all good men
    to come to the aid of their party.
Four score and seven years ago,...
```

The output would be:

```
  Now is the time for all good men
to come to the aid of their party. Four score and seven years ago,...
```

A new line begins after the word "men" because the second line began with a space character.

There are many fancier types of paragraphs, which will be described later.

2.2. Headers and Footers

Arbitrary headers and footers can be put at the top and bottom of every page. Two requests of the form **.he** *title* and **.fo** *title* define the titles to put at the head and the foot of every page, respectively. The titles are called *three-part* titles, that is, there is a left-justified part, a centered part, and a right-justified part. To separate these three parts the first character of *title* (whatever it may be) is used as a delimiter. Any character may be used, but backslash and double quote marks should be avoided. The percent sign is replaced by the current page number whenever found in the title. For example, the input:

```
.he ''%''
.fo 'Jane Jones''My Book'
```

results in the page number centered at the top of each page, "Jane Jones" in the lower left corner, and "My Book" in the lower right corner.

2.3. Double Spacing

NROFF will double space output text automatically if you use the request **.ls 2**, as is

done in this section. You can revert to single spaced mode by typing **.ls 1**.

2.4. Page Layout

A number of requests allow you to change the way the printed copy looks, sometimes called the *layout* of the output page. Most of these requests adjust the placing of "white space" (blank lines or spaces). In these explanations, characters in italics should be replaced with values you wish to use; bold characters represent characters which should actually be typed.

The **.bp** request starts a new page.

The request **.sp** *N* leaves *N* lines of blank space. *N* can be omitted (meaning skip a single line) or can be of the form *N***i** (for *N* inches) or *N***c** (for *N* centimeters). For example, the input:

```
.sp 1.5i
My thoughts on the subject
.sp
```

leaves one and a half inches of space, followed by the line "My thoughts on the subject", followed by a single blank line.

The **.in** *+N* request changes the amount of white space on the left of the page (the *indent*). The argument *N* can be of the form **+***N* (meaning leave *N* spaces more than you are already leaving), **–***N* (meaning leave less than you do now), or just *N* (meaning leave exactly *N* spaces). *N* can be of the form *N***i** or *N***c** also. For example, the input:

```
initial text
.in 5
some text
.in +1i
more text
.in -2c
final text
```

produces "some text" indented exactly five spaces from the left margin, "more text" indented five spaces plus one inch from the left margin (fifteen spaces on a pica typewriter), and "final text" indented five spaces plus one inch minus two centimeters from the margin. That is, the output is:

```
initial text
     some text
               more text
          final text
```

The **.ti** *+N* (temporary indent) request is used like **.in** *+N* when the indent should apply to one line only, after which it should revert to the previous indent. For example, the input:

```
.in 1i
.ti 0
Ware, James R. The Best of Confucius,
Halcyon House, 1950.
An excellent book containing translations of
most of Confucius´ most delightful sayings.
A definite must for anyone interested in the early foundations
of Chinese philosophy.
```

produces:

Ware, James R. The Best of Confucius, Halcyon House, 1950. An excellent book containing translations of most of Confucius' most delightful sayings. A definite must for anyone interested in the early foundations of Chinese philosophy.

Text lines can be centered by using the **.ce** request. The line after the **.ce** is centered (horizontally) on the page. To center more than one line, use **.ce** *N* (where *N* is the number of lines to center), followed by the *N* lines. If you want to center many lines but don't want to count them, type:

```
.ce 1000
lines to center
.ce 0
```

The **.ce 0** request tells NROFF to center zero more lines, in other words, stop centering.

All of these requests cause a break; that is, they always start a new line. If you want to start a new line without performing any other action, use **.br**.

2.5. Underlining

Text can be underlined using the **.ul** request. The **.ul** request causes the next input line to be underlined when output. You can underline multiple lines by stating a count of *input* lines to underline, followed by those lines (as with the **.ce** request). For example, the input:

```
.ul 2
Notice that these two input lines
are underlined.
```

will underline those eight words in NROFF. (In TROFF they will be set in italics.)

3. Displays

Displays are sections of text to be set off from the body of the paper. Major quotes, tables, and figures are types of displays, as are all the examples used in this document. All displays except centered blocks are output single spaced.

3.1. Major Quotes

Major quotes are quotes which are several lines long, and hence are set in from the rest of the text without quote marks around them. These can be generated using the commmands **.(q** and **.)q** to surround the quote. For example, the input:

```
As Weizenbaum points out:
.(q
It is said that to explain is to explain away.
This maxim is nowhere so well fulfilled
as in the areas of computer programming,...
.)q
```

generates as output:

As Weizenbaum points out:

> It is said that to explain is to explain away. This maxim is nowhere so well fulfilled as in the areas of computer programming,...

3.2. Lists

A *list* is an indented, single spaced, unfilled display. Lists should be used when the material to be printed should not be filled and justified like normal text, such as columns of figures or the examples used in this paper. Lists are surrounded by the requests **.(l** and **.)l**. For example, type:

```
Alternatives to avoid deadlock are:
.(l
Lock in a specified order
Detect deadlock and back out one process
Lock all resources needed before proceeding
.)l
```

will produce:

Alternatives to avoid deadlock are:

    Lock in a specified order
    Detect deadlock and back out one process
    Lock all resources needed before proceeding

3.3. Keeps

A *keep* is a display of lines which are kept on a single page if possible. An example of where you would use a keep might be a diagram. Keeps differ from lists in that lists may be broken over a page boundary whereas keeps will not.

Blocks are the basic kind of keep. They begin with the request **.(b** and end with the request **.)b**. If there is not room on the current page for everything in the block, a new page is begun. This has the unpleasant effect of leaving blank space at the bottom of the page. When this is not appropriate, you can use the alternative, called *floating keeps.*

Floating keeps move relative to the text. Hence, they are good for things which will be referred to by name, such as "See figure 3". A floating keep will appear at the bottom of the current page if it will fit; otherwise, it will appear at the top of the next page. Floating keeps begin with the line **.(z** and end with the line **.)z**. For an example of a floating keep, see figure 1. The **.hl** request is used to draw a horizontal line so that the figure stands

```
.(z
.hl
Text of keep to be floated.
.sp
.ce
Figure 1.  Example of a Floating Keep.
.hl
.)z
```

Figure 1. Example of a Floating Keep.

out from the text.

3.4. Fancier Displays

Keeps and lists are normally collected in *nofill* mode, so that they are good for tables and such. If you want a display in fill mode (for text), type **.(l F** (Throughout this section, comments applied to **.(l** also apply to **.(b** and **.(z**). This kind of display will be indented from both margins. For example, the input:

```
.(l F
And now boys and girls,
a newer, bigger, better toy than ever before!
Be the first on your block to have your own computer!
Yes kids, you too can have one of these modern
data processing devices.
You too can produce beautifully formatted papers
without even batting an eye!
.)l
```

will be output as:

> And now boys and girls, a newer, bigger, better toy than ever before! Be the first on your block to have your own computer! Yes kids, you too can have one of these modern data processing devices. You too can produce beautifully formatted papers without even batting an eye!

Lists and blocks are also normally indented (floating keeps are normally left justified). To get a left-justified list, type **.(l L**. To get a list centered line-for-line, type **.(l C**. For example, to get a filled, left justified list, enter:

```
.(l L F
text of block
.)l
```

The input:

```
.(l
first line of unfilled display
more lines
.)l
```

produces the indented text:

first line of unfilled display
more lines

Typing the character **L** after the **.(l** request produces the left justified result:

first line of unfilled display
more lines

Using **C** instead of **L** produces the line-at-a-time centered output:

first line of unfilled display
more lines

Sometimes it may be that you want to center several lines as a group, rather than centering them one line at a time. To do this use centered blocks, which are surrounded by the requests **.(c** and **.)c**. All the lines are centered as a unit, such that the longest line is centered and the rest are lined up around that line. Notice that lines do not move relative to each other using centered blocks, whereas they do using the **C** argument to keeps.

Centered blocks are *not* keeps, and may be used in conjunction with keeps. For example, to center a group of lines as a unit and keep them on one page, use:

```
.(b L
.(c
first line of unfilled display
more lines
.)c
.)b
```

to produce:

first line of unfilled display
more lines

If the block requests (**.(b** and **.)b**) had been omitted the result would have been the same, but with no guarantee that the lines of the centered block would have all been on one page. Note the use of the **L** argument to **.(b**; this causes the centered block to center within the entire line rather than within the line minus the indent. Also, the center requests must be nested *inside* the keep requests.

4. Annotations

There are a number of requests to save text for later printing. *Footnotes* are printed at the bottom of the current page. *Delayed text* is intended to be a variant form of footnote; the text is printed only when explicitly called for, such as at the end of each chapter. *Indexes* are a type of delayed text having a tag (usually the page number) attached to each entry after a row of dots. Indexes are also saved until called for explicitly.

4.1. Footnotes

Footnotes begin with the request **.(f** and end with the request **.)f**. The current footnote number is maintained automatically, and can be used by typing ******, to produce a footnote number[1]. The number is automatically incremented after every footnote. For example, the input:

[1]Like this.

```
.(q
A man who is not upright
and at the same time is presumptuous;
one who is not diligent and at the same time is ignorant;
one who is untruthful and at the same time is incompetent;
such men I do not count among acquaintances.\**
.(f
\**James R. Ware,
.ul
The Best of Confucius,
Halcyon House, 1950.
Page 77.
.)f
.)q
```

generates the result:

> A man who is not upright and at the same time is presumptuous; one who is not diligent and at the same time is ignorant; one who is untruthful and at the same time is incompetent; such men I do not count among acquaintances.[2]

It is important that the footnote appears *inside* the quote, so that you can be sure that the footnote will appear on the same page as the quote.

4.2. Delayed Text

Delayed text is very similar to a footnote except that it is printed when called for explicitly. This allows a list of references to appear (for example) at the end of each chapter, as is the convention in some disciplines. Use ***#** on delayed text instead of ***** as on footnotes.

If you are using delayed text as your standard reference mechanism, you can still use footnotes, except that you may want to reference them with special characters* rather than numbers.

4.3. Indexes

An "index" (actually more like a table of contents, since the entries are not sorted alphabetically) resembles delayed text, in that it is saved until called for. However, each entry has the page number (or some other tag) appended to the last line of the index entry after a row of dots.

Index entries begin with the request **.(x** and end with **.)x**. The **.)x** request may have a argument, which is the value to print as the "page number". It defaults to the current page number. If the page number given is an underscore ("_") no page number or line of dots is printed at all. To get the line of dots without a page number, type **.)x ""**, which specifies an explicitly null page number.

The **.xp** request prints the index.

For example, the input:

[2]James R. Ware, *The Best of Confucius,* Halcyon House, 1950. Page 77.

*Such as an asterisk.

```
.(x
Sealing wax
.)x
.(x
Cabbages and kings
.)x _
.(x
Why the sea is boiling hot
.)x 2.5a
.(x
Whether pigs have wings
.)x ""
.(x
This is a terribly long index entry, such as might be used
for a list of illustrations, tables, or figures; I expect it to
take at least two lines.
.)x
.xp
```

generates:

The **.(x** request may have a single character argument, specifying the "name" of the index; the normal index is **x**. Thus, several "indicies" may be maintained simultaneously (such as a list of tables, table of contents, etc.).

Notice that the index must be printed at the *end* of the paper, rather than at the beginning where it will probably appear (as a table of contents); the pages may have to be physically rearranged after printing.

5. Fancier Features

A large number of fancier requests exist, notably requests to provide other sorts of paragraphs, numbered sections of the form **1.2.3** (such as used in this document), and multicolumn output.

5.1. More Paragraphs

Paragraphs generally start with a blank line and with the first line indented. It is possible to get left-justified block-style paragraphs by using **.lp** instead of **.pp**, as demonstrated by the next paragraph.

Sometimes you want to use paragraphs that have the *body* indented, and the first line exdented (opposite of indented) with a label. This can be done with the **.ip** request. A word specified on the same line as **.ip** is printed in the margin, and the body is lined up at a prespecified position (normally five spaces). For example, the input:

```
.ip one
This is the first paragraph.
Notice how the first line
of the resulting paragraph lines up
with the other lines in the paragraph.
.ip two
And here we are at the second paragraph already.
You may notice that the argument to .ip
appears
in the margin.
.lp
We can continue text...
```

produces as output:

one This is the first paragraph. Notice how the first line of the resulting paragraph lines up with the other lines in the paragraph.

two And here we are at the second paragraph already. You may notice that the argument to **.ip** appears in the margin.

We can continue text without starting a new indented paragraph by using the **.lp** request.

If you have spaces in the label of a **.ip** request, you must use an "unpaddable space" instead of a regular space. This is typed as a backslash character ("\") followed by a space. For example, to print the label "Part 1", enter:

```
.ip "Part\ 1"
```

If a label of an indented paragraph (that is, the argument to **.ip**) is longer than the space allocated for the label, **.ip** will begin a new line after the label. For example, the input:

```
.ip longlabel
This paragraph had a long label.
The first character of text on the first line
will not line up with the text on second and subsequent lines,
although they will line up with each other.
```

will produce:

longlabel
    This paragraph had a long label. The first character of text on the first line will not line up with the text on second and subsequent lines, although they will line up with each other.

It is possible to change the size of the label by using a second argument which is the size of the label. For example, the above example could be done correctly by saying:

```
.ip longlabel 10
```

which will make the paragraph indent 10 spaces for this paragraph only. If you have many paragraphs to indent all the same amount, use the ***number register*** **ii**. For example, to leave one inch of space for the label, type:

```
.nr ii 1i
```

somewhere before the first call to **.ip**. Refer to the reference manual for more information.

If **.ip** is used with no argument at all no hanging tag will be printed. For example, the input:

```
.ip [a]
This is the first paragraph of the example.
We have seen this sort of example before.
.ip
This paragraph is lined up with the previous paragraph,
but it has no tag in the margin.
```

produces as output:

[a] This is the first paragraph of the example. We have seen this sort of example before.

This paragraph is lined up with the previous paragraph, but it has no tag in the margin.

A special case of **.ip** is **.np**, which automatically numbers paragraphs sequentially from 1. The numbering is reset at the next **.pp**, **.lp**, or **.sh** (to be described in the next section) request. For example, the input:

```
.np
This is the first point.
.np
This is the second point.
Points are just regular paragraphs
which are given sequence numbers automatically
by the .np request.
.pp
This paragraph will reset numbering by .np.
.np
For example,
we have reverted to numbering from one now.
```

generates:

(1) This is the first point.

(2) This is the second point. Points are just regular paragraphs which are given sequence numbers automatically by the .np request.

This paragraph will reset numbering by .np.

(1) For example, we have reverted to numbering from one now.

5.2. Section Headings

Section numbers (such as the ones used in this document) can be automatically generated using the **.sh** request. You must tell **.sh** the *depth* of the section number and a section title. The depth specifies how many numbers are to appear (separated by decimal points) in the section number. For example, the section number **4.2.5** has a depth of three.

Section numbers are incremented in a fairly intuitive fashion. If you add a number (increase the depth), the new number starts out at one. If you subtract section numbers (or keep the same number) the final number is incremented. For example, the input:

```
.sh 1 "The Preprocessor"
.sh 2 "Basic Concepts"
.sh 2 "Control Inputs"
.sh 3
.sh 3
.sh 1 "Code Generation"
.sh 3
```

produces as output the result:

1. The Preprocessor
1.1. Basic Concepts
1.2. Control Inputs
1.2.1.
1.2.2.
2. Code Generation
2.1.1.

You can specify the section number to begin by placing the section number after the section title, using spaces instead of dots. For example, the request:

```
.sh 3 "Another section" 7 3 4
```

will begin the section numbered **7.3.4**; all subsequent **.sh** requests will number relative to this number.

There are more complex features which will cause each section to be indented proportionally to the depth of the section. For example, if you enter:

.nr si *N*

each section will be indented by an amount *N*. *N* must have a scaling factor attached, that is, it must be of the form *Nx*, where *x* is a character telling what units *N* is in. Common values for *x* are **i** for inches, **c** for centimeters, and **n** for *ens* (the width of a single character). For example, to indent each section one-half inch, type:

```
.nr si 0.5i
```

After this, sections will be indented by one-half inch per level of depth in the section number. For example, this document was produced using the request

```
.nr si 3n
```

at the beginning of the input file, giving three spaces of indent per section depth.

Section headers without automatically generated numbers can be done using:

```
.uh "Title"
```

which will do a section heading, but will put no number on the section.

5.3. Parts of the Basic Paper

There are some requests which assist in setting up papers. The **.tp** request initializes for a title page. There are no headers or footers on a title page, and unlike other pages you can space down and leave blank space at the top. For example, a typical title page might appear as:

```
.tp
.sp 2i
.(l C
THE GROWTH OF TOENAILS
IN UPPER PRIMATES
.sp
by
.sp
Frank N. Furter
.)l
.bp
```

The request **.th** sets up the environment of the NROFF processor to do a thesis, using the rules established at Berkeley. It defines the correct headers and footers (a page number in the upper right hand corner only), sets the margins correctly, and double spaces.

The **.+c** *T* request can be used to start chapters. Each chapter is automatically numbered from one, and a heading is printed at the top of each chapter with the chapter number and the chapter name *T*. For example, to begin a chapter called "Conclusions", use the request:

.+c "CONCLUSIONS"

which will produce, on a new page, the lines

CHAPTER 5
CONCLUSIONS

with appropriate spacing for a thesis. Also, the header is moved to the foot of the page on the first page of a chapter. Although the **.+c** request was not designed to work only with the **.th** request, it is tuned for the format acceptable for a PhD thesis at Berkeley.

If the title parameter *T* is omitted from the **.+c** request, the result is a chapter with no heading. This can also be used at the beginning of a paper; for example, **.+c** was used to generate page one of this document.

Although papers traditionally have the abstract, table of contents, and so forth at the front of the paper, it is more convenient to format and print them last when using NROFF. This is so that index entries can be collected and then printed for the table of contents (or whatever). At the end of the paper, issue the **.++ P** request, which begins the preliminary part of the paper. After issuing this request, the **.+c** request will begin a preliminary section of the paper. Most notably, this prints the page number restarted from one in lower case Roman numbers. **.+c** may be used repeatedly to begin different parts of the front material for example, the abstract, the table of contents, acknowledgments, list of illustrations, etc. The request **.++ B** may also be used to begin the bibliographic section at the end of the paper. For example, the paper might appear as outlined in figure 2. (In this figure, comments begin with the sequence **\"**.)

5.4. Equations and Tables

Two special UNIX programs exist to format special types of material. **Eqn** and **neqn** set equations for the phototypesetter and NROFF respectively. **Tbl** arranges to print extremely pretty tables in a variety of formats. This document will only describe the embellishments to the standard features; consult the reference manuals for those processors for a description of their use.

The **eqn** and **neqn** programs are described fully in the document *Typesetting Mathematics – Users' Guide* by Brian W. Kernighan and Lorinda L. Cherry. Equations are centered, and are kept on one page. They are introduced by the **.EQ** request and terminated by the **.EN** request.

The **.EQ** request may take an equation number as an optional argument, which is printed vertically centered on the right hand side of the equation. If the equation becomes too long it should be split between two lines. To do this, type:

```
.EQ (eq 34)
text of equation 34
.EN C
.EQ
continuation of equation 34
.EN
```

The **C** on the **.EN** request specifies that the equation will be continued.

The **tbl** program produces tables. It is fully described (including numerous examples) in the document *Tbl – A Program to Format Tables* by M. E. Lesk. Tables begin with the **.TS** request and end with the **.TE** request. Tables are normally kept on a single page. If you have a table which is too big to fit on a single page, so that you know it will extend to several pages, begin the table with the request **.TS H** and put the request **.TH** after the part of the table which you want duplicated at the top of every page that the table is printed on. For example, a table definition for a long table might look like:

```
.th                              \" set for thesis mode
.fo 'DRAFT''                     \" define footer for each page
.tp                              \" begin title page
.(l C                            \" center a large block
THE GROWTH OF TOENAILS
IN UPPER PRIMATES
.sp
by
.sp
Frank Furter
.)l                              \" end centered part
.+c INTRODUCTION                 \" begin chapter named "INTRODUCTION"
.(x t                            \" make an entry into index 't'
Introduction
.)x                              \" end of index entry
text of chapter one
.+c "NEXT CHAPTER"               \" begin another chapter
.(x t                            \" enter into index 't' again
Next Chapter
.)x
text of chapter two
.+c CONCLUSIONS
.(x t
Conclusions
.)x
text of chapter three
.++ B                            \" begin bibliographic information
.+c BIBLIOGRAPHY                 \" begin another 'chapter'
.(x t
Bibliography
.)x
text of bibliography
.++ P                            \" begin preliminary material
.+c "TABLE OF CONTENTS"
.xp t                            \" print index 't' collected above
.+c PREFACE                      \" begin another preliminary section
text of preface
```

Figure 2. Outline of a Sample Paper

```
.TS H
c s s
n n n.
THE TABLE TITLE
.TH
text of the table
.TE
```

5.5. Two Column Output

You can get two column output automatically by using the request **.2c**. This causes everything after it to be output in two-column form. The request **.bc** will start a new column; it differs from **.bp** in that **.bp** may leave a totally blank column when it starts a new page. To revert to single column output, use **.1c**.

5.6. Defining Macros

A *macro* is a collection of requests and text which may be used by stating a simple request. Macros begin with the line **.de** *xx* (where *xx* is the name of the macro to be defined) and end with the line consisting of two dots. After defining the macro, stating the line *.xx* is the same as stating all the other lines. For example, to define a macro that spaces 3 lines and then centers the next input line, enter:

```
.de SS
.sp 3
.ce
..
```

and use it by typing:

```
.SS
Title Line
(beginning of text)
```

Macro names may be one or two characters. In order to avoid conflicts with names in -me, always use upper case letters as names. The only names to avoid are **TS**, **TH**, **TE**, **EQ**, and **EN**.

5.7. Annotations Inside Keeps

Sometimes you may want to put a footnote or index entry inside a keep. For example, if you want to maintain a "list of figures" you will want to do something like:

```
.(z
.(c
text of figure
.)c
.ce
Figure 5.
.(x f
Figure 5
.)x
.)z
```

which you may hope will give you a figure with a label and an entry in the index **f** (presumably a list of figures index). Unfortunately, the index entry is read and interpreted when the keep is read, not when it is printed, so the page number in the index is likely to be wrong. The solution is to use the magic string **\!** at the beginning of all the lines dealing with the index. In other words, you should use:

.(z
.(c
Text of figure
.)c
.ce
Figure 5.
\!.(x f
\!Figure 5
\!.)x
.)z

which will defer the processing of the index until the figure is output. This will guarantee that the page number in the index is correct. The same comments apply to blocks (with **.(b** and **.)b**) as well.

6. TROFF and the Photosetter

With a little care, you can prepare documents that will print nicely on either a regular terminal or when phototypeset using the TROFF formatting program.

6.1. Fonts

A *font* is a style of type. There are three fonts that are available simultaneously, Times Roman, Times Italic, and Times Bold, plus the special math font. The normal font is Roman. Text which would be underlined in NROFF with the **.ul** request is set in italics in TROFF.

There are ways of switching between fonts. The requests **.r**, **.i**, and **.b** switch to Roman, italic, and bold fonts respectively. You can set a single word in some font by typing (for example):

.i word

which will set *word* in italics but does not affect the surrounding text. In NROFF, italic and bold text is underlined.

Notice that if you are setting more than one word in whatever font, you must surround that word with double quote marks (' " ') so that it will appear to the NROFF processor as a single word. The quote marks will not appear in the formatted text. If you do want a quote mark to appear, you should quote the entire string (even if a single word), and use *two* quote marks where you want one to appear. For example, if you want to produce the text:

"*Master Control*"

in italics, you must type:

.i """Master Control\|"""

The \| produces a very narrow space so that the "l" does not overlap the quote sign in TROFF, like this:

"*Master Control*"

There are also several "pseudo-fonts" available. The input:

.(b
.u underlined
.bi "bold italics"
.bx "words in a box"
.)b

generates

underlined
bold italics
words in a box

In NROFF these all just underline the text. Notice that pseudo font requests set only the single parameter in the pseudo font; ordinary font requests will begin setting all text in the special font if you do not provide a parameter. No more than one word should appear with these three font requests in the middle of lines. This is because of the way TROFF justifies text. For example, if you were to issue the requests:

.bi "some bold italics"
and
.bx "words in a box"

in the middle of a line TROFF would produce ***sommebbbddiitaalliiccss*** and words in a box, which I think you will agree does not look good.

The second parameter of all font requests is set in the original font. For example, the font request:

.b bold face

generates "bold" in bold font, but sets "face" in the font of the surrounding text, resulting in:

boldface.

To set the two words **bold** and **face** both in **bold face**, type:

.b "bold face"

You can mix fonts in a word by using the special sequence **\c** at the end of a line to indicate "continue text processing"; this allows input lines to be joined together without a space inbetween them. For example, the input:

.u under \c
.i italics

generates under*italics*, but if we had typed:

.u under
.i italics

the result would have been under *italics* as two words.

6.2. Point Sizes

The phototypesetter supports different sizes of type, measured in points. The default point size is 10 points for most text, 8 points for footnotes. To change the pointsize, type:

.sz $+N$

where N is the size wanted in points. The ***vertical spacing*** (distance between the bottom of most letters (the ***baseline***) between adjacent lines) is set to be proportional to the type size.

Warning: changing point sizes on the phototypesetter is a slow mechanical operation. Size changes should be considered carefully.

6.3. Quotes

It is conventional when using the typesetter to use pairs of grave and acute accents to generate double quotes, rather than the double quote character ('"'). This is because it looks better to use grave and acute accents; for example, compare "quote" to "quote".

In order to make quotes compatible between the typesetter and terminals, you may use the sequences ***(lq** and ***(rq** to stand for the left and right quote respectively. These both appear as " on most terminals, but are typeset as " and " respectively. For example, use:

```
\*(lqSome things aren't true
even if they did happen.\*(rq
```

to generate the result:

> "Some things aren't true even if they did happen."

As a shorthand, the special font request:

```
.q "quoted text"
```

will generate "quoted text". Notice that you must surround the material to be quoted with double quote marks if it is more than one word.

Acknowledgments

I would like to thank Bob Epstein, Bill Joy, and Larry Rowe for having the courage to use the -me macros to produce non-trivial papers during the development stages; Ricki Blau, Pamela Humphrey, and Jim Joyce for their help with the documentation phase; and the plethora of people who have contributed ideas and have given support for the project.

This document was TROFF'ed on September 16, 1986 and applies to version 1.1 of the -me macros.

–ME REFERENCE MANUAL

Release 1.1/25

Eric P. Allman

Electronics Research Laboratory
University of California, Berkeley
Berkeley, California 94720

This document describes in extremely terse form the features of the **–me** macro package for version seven NROFF/TROFF. Some familiarity is assumed with those programs, specifically, the reader should understand breaks, fonts, pointsizes, the use and definition of number registers and strings, how to define macros, and scaling factors for ens, points, **v**'s (vertical line spaces), etc.

For a more casual introduction to text processing using NROFF, refer to the document *Writing Papers with NROFF using –me.*

There are a number of macro parameters that may be adjusted. Fonts may be set to a font number only. In NROFF font 8 is underlined, and is set in bold font in TROFF (although font 3, bold in TROFF, is not underlined in NROFF). Font 0 is no font change; the font of the surrounding text is used instead. Notice that fonts 0 and 8 are "pseudo-fonts"; that is, they are simulated by the macros. This means that although it is legal to set a font register to zero or eight, it is not legal to use the escape character form, such as:

    \f8

All distances are in basic units, so it is nearly always necessary to use a scaling factor. For example, the request to set the paragraph indent to eight one-en spaces is:

    .nr pi 8n

and not

    .nr pi 8

which would set the paragraph indent to eight basic units, or about 0.02 inch. Default parameter values are given in brackets in the remainder of this document.

Registers and strings of the form **$*x*** may be used in expressions but should not be changed. Macros of the form **$*x*** perform some function (as described) and may be redefined to change this function. This may be a sensitive operation; look at the body of the original macro before changing it.

All names in –me follow a rigid naming convention. The user may define number registers, strings, and macros, provided that s/he uses single character upper case names or double character names consisting of letters and digits, with at least one upper case letter. In no case should special characters be used in user-defined names.

On daisy wheel type printers in twelve pitch, the **–rx1** flag can be stated to make lines default to one eighth inch (the normal spacing for a newline in twelve-pitch). This is normally too small for easy readability, so the default is to space one sixth inch.

†NROFF and TROFF are Trademarks of Bell Laboratories.

This documentation was TROFF'ed on September 16, 1986 and applies to version 1.1/25 of the –me macros.

1. Paragraphing

These macros are used to begin paragraphs. The standard paragraph macro is **.pp**; the others are all variants to be used for special purposes.

The first call to one of the paragraphing macros defined in this section or the **.sh** macro (defined in the next session) ***initializes*** the macro processor. After initialization it is not possible to use any of the following requests: **.sc**, **.lo**, **.th**, or **.ac**. Also, the effects of changing parameters which will have a global effect on the format of the page (notably page length and header and footer margins) are not well defined and should be avoided.

.lp — Begin left-justified paragraph. Centering and underlining are turned off if they were on, the font is set to **\n(pf** [1] the type size is set to **\n(pp** [10p], and a **\n(ps** space is inserted before the paragraph [0.35v in TROFF, 1v or 0.5v in NROFF depending on device resolution]. The indent is reset to **\n($i** [0] plus **\n(po** [0] unless the paragraph is inside a display. (see **.ba**). At least the first two lines of the paragraph are kept together on a page.

.pp — Like **.lp**, except that it puts **\n(pi** [5n] units of indent. This is the standard paragraph macro.

.ip *T I* — Indented paragraph with hanging tag. The body of the following paragraph is indented *I* spaces (or **\n(ii** [5n] spaces if *I* is not specified) more than a non-indented paragraph (such as with **.pp**) is. The title *T* is exdented (opposite of indented). The result is a paragraph with an even left edge and *T* printed in the margin. Any spaces in *T* must be unpaddable. If *T* will not fit in the space provided, **.ip** will start a new line.

.np — A variant of .ip which numbers paragraphs. Numbering is reset after a **.lp**, **.pp**, or **.sh**. The current paragraph number is in **\n($p**.

2. Section Headings

Numbered sections are similiar to paragraphs except that a section number is automatically generated for each one. The section numbers are of the form **1.2.3**. The *depth* of the section is the count of numbers (separated by decimal points) in the section number.

Unnumbered section headings are similar, except that no number is attached to the heading.

.sh *+N T a b c d e f* — Begin numbered section of depth *N*. If *N* is missing the current depth (maintained in the number register **\n($0**) is used. The values of the individual parts of the section number are maintained in **\n($1** through **\n($6**. There is a **\n(ss** [1v] space before the section. *T* is printed as a section title in font **\n(sf** [8] and size **\n(sp** [10p]. The "name" of the section may be accessed via ***($n**. If **\n(si** is non-zero, the base indent is set to **\n(si** times the section depth, and the section title is exdented. (See **.ba**.) Also, an additional indent of **\n(so** [0] is added to the section title (but not to the body of the section). The font is then set to the paragraph font, so that more information may occur on the line with the section number and title. **.sh** insures that there is enough room to print the section head plus the beginning of a paragraph (about 3 lines total). If *a* through *f* are specified, the section number is set to that number rather than incremented automatically. If any of *a* through *f* are a hyphen that number is not reset. If *T* is a single underscore ("_") then the section depth and numbering is reset, but the base indent is not reset and nothing is printed out. This is useful to automatically coordinate section numbers with chapter numbers.

.sx *+N* — Go to section depth *N* [**–1**], but do not print the number and title, and do not increment the section number at level *N*. This has the effect of starting a new paragraph at level *N*.

.uh *T* — Unnumbered section heading. The title *T* is printed with the same rules for spacing, font, etc., as for **.sh**.

.$p *T B N* — Print section heading. May be redefined to get fancier headings. *T* is the title passed on the **.sh** or **.uh** line; *B* is the section number for this section, and *N* is the depth of this section. These parameters are not always present; in particular, **.sh** passes all three, **.uh** passes only the first, and **.sx** passes three, but the first two are null strings. Care should be taken if this macro is redefined; it is quite complex and subtle.

.$0 *T B N* — This macro is called automatically after every call to **.$p**. It is normally undefined, but may be used to automatically put every section title into the table of contents or for some similiar function. *T* is the section title for the section title which was just printed, *B* is the section number, and *N* is the section depth.

.$1 – .$6 — Traps called just before printing that depth section. May be defined to (for example) give variable spacing before sections. These macros are called from **.$p**, so if you redefine that macro you may lose this feature.

3. Headers and Footers

Headers and footers are put at the top and bottom of every page automatically. They are set in font **\n(tf** [3] and size **\n(tp** [10p]. Each of the definitions apply as of the *next* page. Three-part titles must be quoted if there are two blanks adjacent anywhere in the title or more than eight blanks total.

The spacing of headers and footers are controlled by three number registers. **\n(hm** [4v] is the distance from the top of the page to the top of the header, **\n(fm** [3v] is the distance from the bottom of the page to the bottom of the footer, **\n(tm** [7v] is the distance from the top of the page to the top of the text, and **\n(bm** [6v] is the distance from the bottom of the page to the bottom of the text (nominal). The macros **.m1**, **.m2**, **.m3**, and **.m4** are also supplied for compatibility with ROFF documents.

.he *'l'm'r'* — Define three-part header, to be printed on the top of every page.

.fo *'l'm'r'* — Define footer, to be printed at the bottom of every page.

.eh *'l'm'r'* — Define header, to be printed at the top of every even-numbered page.

.oh *'l'm'r'* — Define header, to be printed at the top of every odd-numbered page.

.ef *'l'm'r'* — Define footer, to be printed at the bottom of every even-numbered page.

.of *'l'm'r'* — Define footer, to be printed at the bottom of every odd-numbered page.

.hx — Suppress headers and footers on the next page.

.m1 *+N* — Set the space between the top of the page and the header [4v].

.m2 *+N* — Set the space between the header and the first line of text [2v].

.m3 *+N* — Set the space between the bottom of the text and the footer [2v].

.m4 *+N* — Set the space between the footer and the bottom of the page [4v].

.ep — End this page, but do not begin the next page. Useful for forcing out footnotes, but other than that hardly every used. Must be followed by a **.bp** or the end of input.

.$h — Called at every page to print the header. May be redefined to provide fancy (e.g., multi-line) headers, but doing so loses the function of the **.he**, **.fo**, **.eh**, **.oh**, **.ef**, and **.of** requests, as well as the chapter-style title feature of **.+c**.

.$f Print footer; same comments apply as in **.$h**.

.$H A normally undefined macro which is called at the top of each page (after outputing the header, initial saved floating keeps, etc.); in other words, this macro is called immediately before printing text on a page. It can be used for column headings and the like.

4. Displays

All displays except centered blocks and block quotes are preceeded and followed by an extra **\n(bs** [same as **\n(ps**] space. Quote spacing is stored in a separate register; centered blocks have no default initial or trailing space. The vertical spacing of all displays except quotes and centered blocks is stored in register **\n($R** instead of **\n($r**.

.(l *m f* Begin list. Lists are single spaced, unfilled text. If *f* is **F**, the list will be filled. If *m* [I] is **I** the list is indented by **\n(bi** [4n]; if **M** the list is indented to the left margin; if **L** the list is left justified with respect to the text (different from **M** only if the base indent (stored in **\n($i** and set with **.ba**) is not zero); and if **C** the list is centered on a line-by-line basis. The list is set in font **\n(df** [0]. Must be matched by a **.)l**. This macro is almost like **.(b** except that no attempt is made to keep the display on one page.

.)l End list.

.(q Begin major quote. These are single spaced, filled, moved in from the text on both sides by **\n(qi** [4n], preceeded and followed by **\n(qs** [same as **\n(bs**] space, and are set in point size **\n(qp** [one point smaller than surrounding text].

.)q End major quote.

.(b *m f* Begin block. Blocks are a form of *keep*, where the text of a keep is kept together on one page if possible (keeps are useful for tables and figures which should not be broken over a page). If the block will not fit on the current page a new page is begun, *unless* that would leave more than **\n(bt** [0] white space at the bottom of the text. If **\n(bt** is zero, the threshold feature is turned off. Blocks are not filled unless *f* is **F**, when they are filled. The block will be left-justified if *m* is **L**, indented by **\n(bi** [4n] if *m* is **I** or absent, centered (line-for-line) if *m* is **C**, and left justified to the margin (not to the base indent) if *m* is **M**. The block is set in font **\n(df** [0].

.)b End block.

.(z *m f* Begin floating keep. Like **.(b** except that the keep is *floated* to the bottom of the page or the top of the next page. Therefore, its position relative to the text changes. The floating keep is preceeded and followed by **\n(zs** [1v] space. Also, it defaults to mode **M**.

.)z End floating keep.

.(c Begin centered block. The next keep is centered as a block, rather than on a line-by-line basis as with **.(b C**. This call may be nested inside keeps.

.)c End centered block.

5. Annotations

.(d Begin delayed text. Everything in the next keep is saved for output later with **.pd**, in a manner similar to footnotes.

.)d *n* — End delayed text. The delayed text number register **\n($d** and the associated string ***#** are incremented if ***#** has been referenced.

.pd — Print delayed text. Everything diverted via **.(d** is printed and truncated. This might be used at the end of each chapter.

.(f — Begin footnote. The text of the footnote is floated to the bottom of the page and set in font **\n(ff** [1] and size **\n(fp** [8p]. Each entry is preceeded by **\n(fs** [0.2v] space, is indented **\n(fi** [3n] on the first line, and is indented **\n(fu** [0] from the right margin. Footnotes line up underneath two columned output. If the text of the footnote will not all fit on one page it will be carried over to the next page.

.)f *n* — End footnote. The number register **\n($f** and the associated string ****** are incremented if they have been referenced.

.$s — The macro to output the footnote seperator. This macro may be redefined to give other size lines or other types of separators. Currently it draws a 1.5i line.

.(x *x* — Begin index entry. Index entries are saved in the index *x* [**x**] until called up with **.xp.** Each entry is preceeded by a **\n(xs** [0.2v] space. Each entry is "undented" by **\n(xu** [0.5i]; this register tells how far the page number extends into the right margin.

.)x *P A* — End index entry. The index entry is finished with a row of dots with *A* [null] right justified on the last line (such as for an author's name), followed by P [**\n%**]. If *A* is specified, *P* must be specified; **\n%** can be used to print the current page number. If *P* is an underscore, no page number and no row of dots are printed.

.xp *x* — Print index *x* [**x**]. The index is formated in the font, size, and so forth in effect at the time it is printed, rather than at the time it is collected.

6. Columned Output

.2c *+S N* — Enter two-column mode. The column separation is set to *+S* [4n, 0.5i in ACM mode] (saved in **\n($s**). The column width, calculated to fill the single column line length with both columns, is stored in **\n($l**. The current column is in **\n($c**. You can test register **\n($m** [1] to see if you are in single column or double column mode. Actually, the request enters *N* [2] columned output.

.1c — Revert to single-column mode.

.bc — Begin column. This is like **.bp** except that it begins a new column on a new page only if necessary, rather than forcing a whole new page if there is another column left on the current page.

7. Fonts and Sizes

.sz *+P* — The pointsize is set to *P* [10p], and the line spacing is set proportionally. The ratio of line spacing to pointsize is stored in **\n($r**. The ratio used internally by displays and annotations is stored in **\n($R** (although this is not used by **.sz**).

.r *W X* — Set *W* in roman font, appending *X* in the previous font. To append different font requests, use *X* = **\c**. If no parameters, change to roman font.

.i *W X* — Set *W* in italics, appending *X* in the previous font. If no parameters, change to italic font. Underlines in NROFF.

.b *W X*	Set *W* in bold font and append *X* in the previous font. If no parameters, switch to bold font. In NROFF, underlines.
.rb *W X*	Set *W* in bold font and append *X* in the previous font. If no parameters, switch to bold font. **.rb** differs from **.b** in that **.rb** does not underline in NROFF.
.u *W X*	Underline *W* and append *X*. This is a true underlining, as opposed to the **.ul** request, which changes to "underline font" (usually italics in TROFF). It won't work right if *W* is spread or broken (including hyphenated). In other words, it is safe in nofill mode only.
.q *W X*	Quote *W* and append *X*. In NROFF this just surrounds *W* with double quote marks ('" '), but in TROFF uses directed quotes.
.bi *W X*	Set *W* in bold italics and append *X*. Actually, sets *W* in italic and overstrikes once. Underlines in NROFF. It won't work right if *W* is spread or broken (including hyphenated). In other words, it is safe in nofill mode only.
.bx *W X*	Sets *W* in a box, with *X* appended. Underlines in NROFF. It won't work right if *W* is spread or broken (including hyphenated). In other words, it is safe in nofill mode only.

8. Roff Support

.ix *+N*	Indent, no break. Equivalent to **'in** *N*.
.bl *N*	Leave *N* contiguous white space, on the next page if not enough room on this page. Equivalent to a **.sp** *N* inside a block.
.pa *+N*	Equivalent to **.bp**.
.ro	Set page number in roman numerals. Equivalent to **.af % i**.
.ar	Set page number in arabic. Equivalent to **.af % 1**.
.n1	Number lines in margin from one on each page.
.n2 *N*	Number lines from *N*, stop if *N* = 0.
.sk	Leave the next output page blank, except for headers and footers. This is used to leave space for a full-page diagram which is produced externally and pasted in later. To get a partial-page paste-in display, say **.sv** *N*, where *N* is the amount of space to leave; this space will be output immediately if there is room, and will otherwise be output at the top of the next page. However, be warned: if *N* is greater than the amount of available space on an empty page, no space will ever be output.

9. Preprocessor Support

.EQ *m T*	Begin equation. The equation is centered if *m* is **C** or omitted, indented **\n(bi** [4n] if *m* is **I**, and left justified if *m* is **L**. *T* is a title printed on the right margin next to the equation. See *Typesetting Mathematics – User's Guide* by Brian W. Kernighan and Lorinda L. Cherry.
.EN *c*	End equation. If *c* is **C** the equation must be continued by immediately following with another **.EQ**, the text of which can be centered along with this one. Otherwise, the equation is printed, always on one page, with **\n(es** [0.5v in TROFF, 1v in NROFF] space above and below it.
.TS *h*	Table start. Tables are single spaced and kept on one page if possible. If you have a large table which will not fit on one page, use *h* = **H** and follow the header part (to be printed on every page of the table) with a **.TH**. See *Tbl – A Program to Format Tables* by M. E. Lesk.

.TH With **.TS H**, ends the header portion of the table.

.TE Table end. Note that this table does not float, in fact, it is not even guaranteed to stay on one page if you use requests such as **.sp** intermixed with the text of the table. If you want it to float (or if you use requests inside the table), surround the entire table (including the **.TS** and **.TE** requests) with the requests **.(z** and **.)z**.

10. Miscellaneous

.re Reset tabs. Set to every 0.5i in TROFF and every 0.8i in NROFF.

.ba *+N* Set the base indent to *+N* [0] (saved in **\n($i**). All paragraphs, sections, and displays come out indented by this amount. Titles and footnotes are unaffected. The **.sh** request performs a **.ba** request if **\n(si** [0] is not zero, and sets the base indent to **\n(si*\n($0**.

.xl *+N* Set the line length to *N* [6.0i]. This differs from **.ll** because it only affects the current environment.

.ll *+N* Set line length in all environments to *N* [6.0i]. This should not be used after output has begun, and particularly not in two-columned output. The current line length is stored in **\n($l**.

.hl Draws a horizontal line the length of the page. This is useful inside floating keeps to differentiate between the text and the figure.

.lo This macro loads another set of macros (in **/usr/lib/me/local.me**) which is intended to be a set of locally defined macros. These macros should all be of the form **.*** *X*, where *X* is any letter (upper or lower case) or digit.

11. Standard Papers

.tp Begin title page. Spacing at the top of the page can occur, and headers and footers are supressed. Also, the page number is not incremented for this page.

.th Set thesis mode. This defines the modes acceptable for a doctoral dissertation at Berkeley. It double spaces, defines the header to be a single page number, and changes the margins to be 1.5 inch on the left and one inch on the top. **.++** and **.+c** should be used with it. This macro must be stated before initialization, that is, before the first call of a paragraphing macro or **.sh**.

.++ *m H* This request defines the section of the paper which we are entering. The section type is defined by *m*. **C** means that we are entering the chapter portion of the paper, **A** means that we are entering the appendix portion of the paper, **P** means that the material following should be the preliminary portion (abstract, table of contents, etc.) portion of the paper, **AB** means that we are entering the abstract (numbered independently from 1 in Arabic numerals), and **B** means that we are entering the bibliographic portion at the end of the paper. Also, the variants **RC** and **RA** are allowed, which specify renumbering of pages from one at the beginning of each chapter or appendix, respectively. The *H* parameter defines the new header. If there are any spaces in it, the entire header must be quoted. If you want the header to have the chapter number in it, Use the string **\\\\n(ch**. For example, to number appendixes **A.1** etc., type **.++ RA** *''***\\\\n(ch.***%'*. Each section (chapter, appendix, etc.) should be preceeded by the **.+c** request. It should be mentioned that it is easier when using TROFF to put the front material at the end of the paper, so that the table of contents can be collected and output; this material can then be physically moved to the beginning of the paper.

.+c *T* — Begin chapter with title *T*. The chapter number is maintained in **\n(ch**. This register is incremented every time **.+c** is called with a parameter. The title and chapter number are printed by **.$c**. The header is moved to the footer on the first page of each chapter. If *T* is omitted, **.$c** is not called; this is useful for doing your own "title page" at the beginning of papers without a title page proper. **.$c** calls **.$C** as a hook so that chapter titles can be inserted into a table of contents automatically. The footnote numbering is reset to one.

.$c** *T* — Print chapter number (from **\n(ch**) and *T*. This macro can be redefined to your liking. It is defined by default to be acceptable for a PhD thesis at Berkeley. This macro calls **$C, which can be defined to make index entries, or whatever.

.$C *K N T* — This macro is called by **.$c**. It is normally undefined, but can be used to automatically insert index entries, or whatever. *K* is a keyword, either "Chapter" or "Appendix" (depending on the **.++** mode); *N* is the chapter or appendix number, and *T* is the chapter or appendix title.

.ac *A N* — This macro (short for **.acm**) sets up the NROFF environment for photo-ready papers as used by the ACM. This format is 25% larger, and has no headers or footers. The author's name *A* is printed at the bottom of the page (but off the part which will be printed in the conference proceedings), together with the current page number and the total number of pages *N*. Additionally, this macro loads the file **/usr/lib/me/acm.me**, which may later be augmented with other macros useful for printing papers for ACM conferences. It should be noted that this macro will not work correctly in TROFF, since it sets the page length wider than the physical width of the phototypesetter roll.

12. Predefined Strings

****** — Footnote number, actually *[\n($f*]. This macro is incremented after each call to **.)f**.

***#** — Delayed text number. Actually [**\n($d**].

***[** — Superscript. This string gives upward movement and a change to a smaller point size if possible, otherwise it gives the left bracket character ('['). Extra space is left above the line to allow room for the superscript.

***]** — Unsuperscript. Inverse to ***[**. For example, to produce a superscript you might type **x*[2*]**, which will produce $\mathbf{x}^2$.

***<** — Subscript. Defaults to '<' if half-carriage motion not possible. Extra space is left below the line to allow for the subscript.

***>** — Inverse to ***<**.

***(dw** — The day of the week, as a word.

***(mo** — The month, as a word.

***(td** — Today's date, directly printable. The date is of the form September 16, 1986. Other forms of the date can be used by using **\n(dy** (the day of the month; for example, 16), ***(mo** (as noted above) or **\n(mo** (the same, but as an ordinal number; for example, September is 9), and **\n(yr** (the last two digits of the current year).

***(lq** — Left quote marks. Double quote in NROFF.

***(rq** — Right quote.

*-	¾ em dash in TROFF; two hyphens in NROFF.

13. Special Characters and Marks

There are a number of special characters and diacritical marks (such as accents) available through -me. To reference these characters, you must call the macro **.sc** to define the characters before using them.

.sc	Define special characters and diacritical marks, as described in the remainder of this section. This macro must be stated before initialization.

The special characters available are listed below.

Name	Usage	Example	
Acute accent	*'	a*'	á
Grave accent	*`	e*`	è
Umlat	*:	u*:	ü
Tilde	*~	n*~	ñ
Caret	*^	e*^	ê
Cedilla	*,	c*,	ç
Czech	*v	e*v	ě
Circle	*o	A*o	Å
There exists	*(qe		∃
For all	*(qa		∀

Acknowledgments

I would like to thank Bob Epstein, Bill Joy, and Larry Rowe for having the courage to use the -me macros to produce non-trivial papers during the development stages; Ricki Blau, Pamela Humphrey, and Jim Joyce for their help with the documentation phase; and the plethora of people who have contributed ideas and have given support for the project.

Writing Tools - The STYLE and DICTION Programs

L. L. Cherry

W. Vesterman

Livingston College

Rutgers University

readability

ABSTRACT

Text processing systems are now in heavy use in many companies to format documents. With many documents stored on line, it has become possible to use computers to study writing style itself and to help writers produce better written and more readable prose. The system of programs described here is an initial step toward such help. It includes programs and a data base designed to produce a stylistic profile of writing at the word and sentence level. The system measures readability, sentence and word length, sentence type, word usage, and sentence openers. It also locates common examples of wordy phrasing and bad diction. The system is useful for evaluating a document's style, locating sentences that may be difficult to read or excessively wordy, and determining a particular writer's style over several documents.

September 12, 1986

1. Introduction

Computers have become important in the document preparation process, with programs to check for spelling errors and to format documents. As the amount of text stored on line increases, it becomes feasible and attractive to study writing style and to attempt to help the writer in producing readable documents. The system of writing tools described here is a first step toward such help. The system includes programs and a data base to analyze writing style at the word and sentence level. We use the term "style" in this paper to describe the results of a writer's particular choices among individual words and sentence forms. Although many judgements of style are subjective, particularly those of word choice, there are some objective measures that experts agree lead to good style. Three programs have been written to measure some of the objectively definable characteristics of writing style and to identify some commonly misused or unnecessary phrases. Although a document that conforms to the stylistic rules is not guaranteed to be coherent and readable, one that violates all of the rules is likely to be difficult or tedious to read. The program STYLE calculates readability, sentence length variability, sentence type, word usage and sentence openers at a rate of about 400 words per second on a PDP11/70 running the UNIX† Operating System. It assumes that the sentences are well-formed, i. e. that each sentence has a verb and that the subject and verb agree in number. DICTION identifies phrases that are either bad usage or unnecessarily wordy. EXPLAIN acts as a thesaurus for the phrases found by DICTION. Sections 2, 3, and 4 describe the programs; Section 5 gives the results on a cross-section of technical documents; Section 6 discusses accuracy and problems; Section 7 gives implementation details.

2. STYLE

The program STYLE reads a document and prints a summary of readability indices, sentence length and type, word usage, and sentence openers. It may also be used to locate all sentences in a document longer than a given length, of readability index higher than a given number, those containing a passive verb, or those beginning with an expletive. STYLE is based on the system for finding English word classes or parts of speech, PARTS [1]. PARTS is a set of programs that uses a small dictionary (about 350 words) and suffix rules to partially assign word classes to English text. It then uses experimentally derived rules of word order to assign word classes to all words in the text with an accuracy of about 95%. Because PARTS uses only a small dictionary and general rules, it works on text about any subject, from physics to psychology. Style measures have been built into the output phase of the programs that make up PARTS. Some of the measures are simple counters of the word classes found by PARTS; many are more complicated. For example, the verb count is the total number of verb phrases. This includes phrases like:

has been going
was only going
to go

each of which each counts as one verb. Figure 1 shows the output of STYLE run on a paper by Kernighan and Mashey about the UNIX programming environment [2]. As the example shows, STYLE output is in five parts. After a brief discussion of sentences, we will describe the parts in order.

2.1. What is a sentence?

Readers of documents have little trouble deciding where the sentences end. People don't even have to stop and think about uses of the character "." in constructions like 1.25, A. J. Jones, Ph.D., i. e., or etc. . When a computer reads a document, finding the end of sentences is not as easy. First we must throw away the printer's marks and formatting commands that litter the text in computer form. Then STYLE defines a sentence as a string of words ending in one of:

† UNIX is a trademark of Bell Laboratories.

programming environment	
readability grades:	
	(Kincaid) 12.3 (auto) 12.8 (Coleman-Liau) 11.8 (Flesch) 13.5 (46.3)
sentence info:	
	no. sent 335 no. wds 7419
	av sent leng 22.1 av word leng 4.91
	no. questions 0 no. imperatives 0
	no. nonfunc wds 4362 58.8% av leng 6.38
	short sent (<17) 35% (118) long sent (>32) 16% (55)
	longest sent 82 wds at sent 174; shortest sent 1 wds at sent 117
sentence types:	
	simple 34% (114) complex 32% (108)
	compound 12% (41) compound-complex 21% (72)
word usage:	
	verb types as % of total verbs
	tobe 45% (373) aux 16% (133) inf 14% (114)
	passives as % of non-inf verbs 20% (144)
	types as % of total
	prep 10.8% (804) conj 3.5% (262) adv 4.8% (354)
	noun 26.7% (1983) adj 18.7% (1388) pron 5.3% (393)
	nominalizations 2 % (155)
sentence beginnings:	
	subject opener: noun (63) pron (43) pos (0) adj (58) art (62) tot 67%
	prep 12% (39) adv 9% (31)
	verb 0% (1) sub_conj 6% (20) conj 1% (5)
	expletives 4% (13)

Figure 1

. ! ? /.

The end marker "/." may be used to indicate an imperative sentence. Imperative sentences that are not so marked are not identified as imperative. STYLE properly handles numbers with embedded decimal points and commas, strings of letters and numbers with embedded decimal points used for naming computer file names, and the common abbreviations listed in Appendix 1. Numbers that end sentences, like the preceding sentence, cause a sentence break if the next word begins with a capital letter. Initials only cause a sentence break if the next word begins with a capital and is found in the dictionary of function words used by PARTS. So the string

J. D. JONES

does not cause a break, but the string

... system H. The ...

does. With these rules most sentences are broken at the proper place, although occasionally either two sentences are called one or a fragment is called a sentence. More on this later.

2.2. Readability Grades

The first section of STYLE output consists of four readability indices. As Klare points out in [3] readability indices may be used to estimate the reading skills needed by the reader to understand a document. The readability indices reported by STYLE are based on measures of sentence and word lengths. Although the indices may not measure whether the document is coherent and well organized, experience has shown that high indices seem to be indicators of stylistic difficulty. Documents with short sentences and short words have low scores; those with long sentences and many polysyllabic words have high scores. The 4 formulae reported are Kincaid Formula [4], Automated Readability Index [5], Coleman-Liau Formula [6] and a normalized version of Flesch Reading Ease Score [7]. The formulae differ because they were experimentally derived using different texts and subject groups. We will discuss each of the formulae briefly; for a more detailed discussion the reader should see [3].

The Kincaid Formula, given by:

$$Reading\_Grade = 11.8 * syl\_per\_wd + .39 * wds\_per\_sent - 15.59$$

was based on Navy training manuals that ranged in difficulty from 5.5 to 16.3 in reading grade level. The score reported by this formula tends to be in the mid-range of the 4 scores. Because it is based on adult training manuals rather than school book text, this formula is probably the best one to apply to technical documents.

The Automated Readability Index (ARI), based on text from grades 0 to 7, was derived to be easy to automate. The formula is:

$$Reading\_Grade = 4.71 * let\_per\_wd + .5 * wds\_per\_sent - 21.43$$

ARI tends to produce scores that are higher than Kincaid and Coleman-Liau but are usually slightly lower than Flesch.

The Coleman-Liau Formula, based on text ranging in difficulty from .4 to 16.3, is:

$$Reading\_Grade = 5.89 * let\_per\_wd - .3 * sent\_per\_100\_wds - 15.8$$

Of the four formulae this one usually gives the lowest grade when applied to technical documents.

The last formula, the Flesch Reading Ease Score, is based on grade school text covering grades 3 to 12. The formula, given by:

$$Reading\_Score = 206.835 - 84.6 * syl\_per\_wd - 1.015 * wds\_per\_sent$$

is usually reported in the range 0 (very difficult) to 100 (very easy). The score reported by STYLE is scaled to be comparable to the other formulas, except that the maximum grade level reported is set to 17. The Flesch score is usually the highest of the 4 scores on technical documents.

Coke [8] found that the Kincaid Formula is probably the best predictor for technical documents; both ARI and Flesch tend to overestimate the difficulty; Coleman-Liau tend to underestimate. On text in the range of grades 7 to 9 the four formulas tend to be about the same. On easy text the Coleman-Liau formula is probably preferred since it is reasonably accurate at the lower grades and it is safer to present text that is a little too easy than a little too hard.

If a document has particularly difficult technical content, especially if it includes a lot of mathematics, it is probably best to make the text very easy to read, i.e. a lower readability index by shortening the sentences and words. This will allow the reader to concentrate on the technical content and not the long sentences. The user should remember that these indices are estimators; they should not be taken as absolute numbers. STYLE called with "–r number" will print all sentences with an Automated Readability Index equal to or greater than "number".

2.3. Sentence length and structure

The next two sections of STYLE output deal with sentence length and structure. Almost all books on writing style or effective writing emphasize the importance of variety in sentence length and structure for good writing. Ewing's first rule in discussing style in the book *Writing for Results* [9] is:

> "Vary the sentence structure and length of your sentences."

Leggett, Mead and Charvat break this rule into 3 in *Prentice-Hall Handbook for Writers* [10] as follows:

"34a. Avoid the overuse of short simple sentences."
"34b. Avoid the overuse of long compound sentences."
"34c. Use various sentence structures to avoid monotony and increase effectiveness."

Although experts agree that these rules are important, not all writers follow them. Sample technical documents have been found with almost no sentence length or type variability. One document had 90% of its sentences about the same length as the average; another was made up almost entirely of simple sentences (80%).

The output sections labeled "sentence info" and "sentence types" give both length and structure measures. STYLE reports on the number and average length of both sentences and words, and number of questions and imperative sentences (those ending in "/."). The measures of non-function words are an attempt to look at the content words in the document. In English non-function words are nouns, adjectives, adverbs, and non-auxiliary verbs; function words are prepositions, conjunctions, articles, and auxiliary verbs. Since most function words are short, they tend to lower the average word length. The average length of non-function words may be a more useful measure for comparing word choice of different writers than the total average word length. The percentages of short and long sentences measure sentence length variability. Short sentences are those at least 5 words less than the average; long sentences are those at least 10 words longer than the average. Last in the sentence information section is the length and location of the longest and shortest sentences. If the flag "–l number" is used, STYLE will print all sentences longer than "number".

Because of the difficulties in dealing with the many uses of commas and conjunctions in English, sentence type definitions vary slightly from those of standard textbooks, but still measure the same constructional activity.

1. A simple sentence has one verb and no dependent clause.
2. A complex sentence has one independent clause and one dependent clause, each with one verb. Complex sentences are found by identifying sentences that contain either a subordinate conjunction or a clause beginning with words like "that" or "who". The preceding sentence has such a clause.
3. A compound sentence has more than one verb and no dependent clause. Sentences joined by ";" are also counted as compound.
4. A compound-complex sentence has either several dependent clauses or one dependent clause and a compound verb in either the dependent or independent clause.

Even using these broader definitions, simple sentences dominate many of the technical documents that have been tested, but the example in Figure 1 shows variety in both sentence structure and sentence length.

2.4. Word Usage

The word usage measures are an attempt to identify some other constructional features of writing style. There are many different ways in English to say the same thing. The constructions differ from one another in the form of the words used. The following sentences all convey approximately the same meaning but differ in word usage:

The cxio program is used to perform all communication between the systems.
The cxio program performs all communications between the systems.
The cxio program is used to communicate between the systems.
The cxio program communicates between the systems.
All communication between the systems is performed by the cxio program.

The distribution of the parts of speech and verb constructions helps identify overuse of particular constructions. Although the measures used by STYLE are crude, they do point out problem areas. For each category, STYLE reports a percentage and a raw count. In addition to looking at the percentage, the user may find it useful to compare the raw count with the number of sentences. If, for example, the number of infinitives is almost equal to the number of sentences, then many of the sentences in the document are constructed like the first and third in the preceding example. The user may want to transform some of these sentences into another form. Some of the implications of the word usage measures are discussed below.

Verbs are measured in several different ways to try to determine what types of verb constructions are most frequent in the document. Technical writing tends to contain many passive verb constructions and other usage of the verb "to be". The category of verbs labeled "tobe" measures both passives and sentences of the form:

subject tobe predicate

In counting verbs, whole verb phrases are counted as one verb. Verb phrases containing auxiliary verbs are counted in the category "aux". The verb phrases counted here are those whose tense is not simple present or simple past. It might eventually be useful to do more detailed measures of verb tense or mood. Infinitives are listed as "inf". The percentages reported for these three categories are based on the total number of verb phrases found. These categories are not mutually exclusive; they cannot be added, since, for example, "to be going" counts as both "tobe" and "inf". Use of these three types of verb constructions varies significantly among authors.

STYLE reports passive verbs as a percentage of the finite verbs in the document. Most style books warn against the overuse of passive verbs. Coleman [11] has shown that sentences with active verbs are easier to learn than those with passive verbs. Although the inverted object-subject order of the passive voice seems to emphasize the object, Coleman's experiments showed that there is little difference in retention by word position. He also showed that the direct object of an active verb is retained better than the subject of a passive verb. These experiments support the advice of the style books suggesting that writers should try to use active verbs wherever possible. The flag "–p" causes STYLE to print all sentences containing passive verbs.

Pronouns
add cohesiveness and connectivity to a document by providing back-reference. They are often a short-hand notation for something previously mentioned, and therefore connect the sentence containing the pronoun with the word to which the pronoun refers. Although there are other mechanisms for such connections, documents with no pronouns tend to be wordy and to have little connectivity.

Adverbs
can provide transition between sentences and order in time and space. In performing these functions, adverbs, like pronouns, provide connectivity and cohesiveness.

Conjunctions
provide parallelism in a document by connecting two or more equal units. These units may be whole sentences, verb phrases, nouns, adjectives, or prepositional phrases. The compound and compound-complex sentences reported under sentence type are parallel structures. Other uses of parallel structures are indicated by the degree that the number of conjunctions reported under word usage exceeds the compound sentence measures.

Nouns and Adjectives.
A ratio of nouns to adjectives near unity may indicate the over-use of modifiers. Some technical writers qualify every noun with one or more adjectives. Qualifiers in phrases like "simple linear single-link network model" often lend more obscurity than precision to a text.

Nominalizations
are verbs that are changed to nouns by adding one of the suffixes "ment", "ance", "ence", or "ion". Examples are accomplishment, admittance, adherence, and abbreviation. When a writer transforms a nominalized sentence to a non-nominalized sentence, she/he increases the

effectiveness of the sentence in several ways. The noun becomes an active verb and frequently one complicated clause becomes two shorter clauses. For example,

> Their inclusion of this provision is admission of the importance of the system.
> When they included this provision, they admitted the importance of the system.

Coleman found that the transformed sentences were easier to learn, even when the transformation produced sentences that were slightly longer, provided the transformation broke one clause into two. Writers who find their document contains many nominalizations may want to transform some of the sentences to use active verbs.

2.5. Sentence openers

Another agreed upon principle of style is variety in sentence openers. Because STYLE determines the type of sentence opener by looking at the part of speech of the first word in the sentence, the sentences counted under the heading "subject opener" may not all really begin with the subject. However, a large percentage of sentences in this category still indicates lack of variety in sentence openers. Other sentence opener measures help the user determine if there are transitions between sentences and where the subordination occurs. Adverbs and conjunctions at the beginning of sentences are mechanisms for transition between sentences. A pronoun at the beginning shows a link to something previously mentioned and indicates connectivity.

The location of subordination can be determined by comparing the number of sentences that begin with a subordinator with the number of sentences with complex clauses. If few sentences start with subordinate conjunctions then the subordination is embedded or at the end of the complex sentences. For variety the writer may want to transform some sentences to have leading subordination.

The last category of openers, expletives, is commonly overworked in technical writing. Expletives are the words "it" and "there", usually with the verb "to be", in constructions where the subject follows the verb. For example,

> There are three streets used by the traffic.
> There are too many users on this system.

This construction tends to emphasize the object rather than the subject of the sentence. The flag "–e" will cause STYLE to print all sentences that begin with an expletive.

3. DICTION

The program DICTION prints all sentences in a document containing phrases that are either frequently misused or indicate wordiness. The program, an extension of Aho's FGREP [12] string matching program, takes as input a file of phrases or patterns to be matched and a file of text to be searched. A data base of about 450 phrases has been compiled as a default pattern file for DICTION. Before attempting to locate phrases, the program maps upper case letters to lower case and substitutes blanks for punctuation. Sentence boundaries were deemed less critical in DICTION than in STYLE, so abbreviations and other uses of the character "." are not treated specially. DICTION brackets all pattern matches in a sentence with the characters "[" "]" . Although many of the phrases in the default data base are correct in some contexts, in others they indicate wordiness. Some examples of the phrases and suggested alternatives are:

Phrase	Alternative
a large number of	many
arrive at a decision	decide
collect together	collect
for this reason	so
pertaining to	about
through the use of	by or with
utilize	use
with the exception of	except

Appendix 2 contains a complete list of the default file. Some of the entries are short forms of problem phrases. For example, the phrase "the fact" is found in all of the following and is sufficient to point out the wordiness to the user:

Phrase	Alternative
accounted for by the fact that	caused by
an example of this is the fact that	thus
based on the fact that	because
despite the fact that	although
due to the fact that	because
in light of the fact that	because
in view of the fact that	since
notwithstanding the fact that	although

Entries in Appendix 2 preceded by "~" are not matched. See Section 7 for details on the use of "~".

The user may supply her/his own pattern file with the flag "–f patfile". In this case the default file will be loaded first, followed by the user file. This mechanism allows users to suppress patterns contained in the default file or to include their own pet peeves that are not in the default file. The flag "–n" will exclude the default file altogether. In constructing a pattern file, blanks should be used before and after each phrase to avoid matching substrings in words. For example, to find all occurrences of the word "the", the pattern " the " should be used. The blanks cause only the word "the" to be matched and not the string "the" in words like there, other, and therefore. One side effect of surrounding the words with blanks is that when two phrases occur without intervening words, only the first will be matched.

4. EXPLAIN

The last program, EXPLAIN, is an interactive thesaurus for phrases found by DICTION. The user types one of the phrases bracketed by DICTION and EXPLAIN responds with suggested substitutions for the phrase that will improve the diction of the document.

5. Results

5.1. STYLE

To get baseline statistics and check the program's accuracy, we ran STYLE on 20 technical documents. There were a total of 3287 sentences in the sample. The shortest document was 67 sentences long; the longest 339 sentences. The documents covered a wide range of subject matter, including theoretical computing, physics, psychology, engineering, and affirmative action. Table 1 gives the range, median, and standard deviation of the various style measures. As you will note most of the measurements have a fairly wide range of values across the sample documents.

As a comparison, Table 2 gives the median results for two different technical authors, a sample of instructional material, and a sample of the Federalist Papers. The two authors show similar

Table 1
Text Statistics on 20 Technical Documents

	variable	minimum	maximum	mean	standard deviation
Readability	Kincaid	9.5	16.9	13.3	2.2
	automated	9.0	17.4	13.3	2.5
	Cole-Liau	10.0	16.0	12.7	1.8
	Flesch	8.9	17.0	14.4	2.2
sentence info.	av sent length	15.5	30.3	21.6	4.0
	av word length	4.61	5.63	5.08	.29
	av nonfunction length	5.72	7.30	6.52	.45
	short sent	23%	46%	33%	5.9
	long sent	7%	20%	14%	2.9
sentence types	simple	31%	71%	49%	11.4
	complex	19%	50%	33%	8.3
	compound	2%	14%	7%	3.3
	compound-complex	2%	19%	10%	4.8
verb types	tobe	26%	64%	44.7%	10.3
	auxiliary	10%	40%	21%	8.7
	infinitives	8%	24%	15.1%	4.8
	passives	12%	50%	29%	9.3
word usage	prepositions	10.1%	15.0%	12.3%	1.6
	conjunction	1.8%	4.8%	3.4%	.9
	adverbs	1.2%	5.0%	3.4%	1.0
	nouns	23.6%	31.6%	27.8%	1.7
	adjectives	15.4%	27.1%	21.1%	3.4
	pronouns	1.2%	8.4%	2.5%	1.1
	nominalizations	2%	5%	3.3%	.8
sentence openers	prepositions	6%	19%	12%	3.4
	adverbs	0%	20%	9%	4.6
	subject	56%	85%	70%	8.0
	verbs	0%	4%	1%	1.0
	subordinating conj	1%	12%	5%	2.7
	conjunctions	0%	4%	0%	1.5
	expletives	0%	6%	2%	1.7

styles, although author 2 uses somewhat shorter sentences and longer words than author 1. Author 1 uses all types of sentences, while author 2 prefers simple and complex sentences, using few compound or compound-complex sentences. The other major difference in the styles of these authors is the location of subordination. Author 1 seems to prefer embedded or trailing subordination, while author 2 begins many sentences with the subordinate clause. The documents tested for both authors 1 and 2 were technical documents, written for a technical audience. The instructional documents, which are written for craftspeople, vary surprisingly little from the two technical samples. The sentences and words are a little longer, and they contain many passive and auxiliary verbs, few adverbs, and almost no pronouns. The instructional documents contain many imperative sentences, so there are many sentence with verb openers. The sample of Federalist Papers contrasts with the other samples in almost every way.

5.2. DICTION

In the few weeks that DICTION has been available to users about 35,000 sentences have been run with about 5,000 string matches. The authors using the program seem to make the suggested changes about 50-75% of the time. To date, almost 200 of the 450 strings in the default file have been matched. Although most of these phrases are valid and correct in some contexts, the 50-75%

Table 2
Text Statistics on Single Authors

	variable	author 1	author 2	inst.	FED
readability	Kincaid	11.0	10.3	10.8	16.3
	automated	11.0	10.3	11.9	17.8
	Coleman-Liau	9.3	10.1	10.2	12.3
	Flesch	10.3	10.7	10.1	15.0
sentence info	av sent length	22.64	19.61	22.78	31.85
	av word length	4.47	4.66	4.65	4.95
	av nonfunction length	5.64	5.92	6.04	6.87
	short sent	35%	43%	35%	40%
	long sent	18%	15%	16%	21%
sentence types	simple	36%	43%	40%	31%
	complex	34%	41%	37%	34%
	compound	13%	7%	4%	10%
	compound-complex	16%	8%	14%	25%
verb type	tobe	42%	43%	45%	37%
	auxiliary	17%	19%	32%	32%
	infinitives	17%	15%	12%	21%
	passives	20%	19%	36%	20%
word usage	prepositions	10.0%	10.8%	12.3%	15.9%
	conjunctions	3.2%	2.4%	3.9%	3.4%
	adverbs	5.05%	4.6%	3.5%	3.7%
	nouns	27.7%	26.5%	29.1%	24.9%
	adjectives	17.0%	19.0%	15.4%	12.4%
	pronouns	5.3%	4.3%	2.1%	6.5%
	nominalizations	1%	2%	2%	3%
sentence openers	prepositions	11%	14%	6%	5%
	adverbs	9%	9%	6%	4%
	subject	65%	59%	54%	66%
	verb	3%	2%	14%	2%
	subordinating conj	8%	14%	11%	3%
	conjunction	1%	0%	0%	3%
	expletives	3%	3%	0%	3%

change rate seems to show that the phrases are used much more often than concise diction warrants.

6. Accuracy

6.1. Sentence Identification

The correctness of the STYLE output on the 20 document sample was checked in detail. STYLE misidentified 129 sentence fragments as sentences and incorrectly joined two or more sentences 75 times in the 3287 sentence sample. The problems were usually because of nonstandard formatting commands, unknown abbreviations, or lists of non-sentences. An impossibly long sentence found as the longest sentence in the document usually is the result of a long list of non-sentences.

6.2. Sentence Types

Style correctly identified sentence type on 86.5% of the sentences in the sample. The type distribution of the sentences was 52.5% simple, 29.9% complex, 8.5% compound and 9% compound-complex. The program reported 49.5% simple, 31.9% complex, 8% compound and

10.4% compound-complex. Looking at the errors on the individual documents, the number of simple sentences was under-reported by about 4% and the complex and compound-complex were over-reported by 3% and 2%, respectively. The following matrix shows the programs output vs. the actual sentence type.

		Program Results			
		simple	complex	compound	comp-complex
Actual	simple	1566	132	49	17
Sentence	complex	47	892	6	65
Type	compound	40	6	207	23
	comp-complex	0	52	5	249

The system's inability to find imperative sentences seems to have little effect on most of the style statistics. A document with half of its sentences imperative was run, with and without the imperative end marker. The results were identical except for the expected errors of not finding verbs as sentence openers, not counting the imperative sentences, and a slight difference (1%) in the number of nouns and adjectives reported.

6.3. Word Usage

The accuracy of identifying word types reflects that of PARTS, which is about 95% correct. The largest source of confusion is between nouns and adjectives. The verb counts were checked on about 20 sentences from each document and found to be about 98% correct.

7. Technical Details

7.1. Finding Sentences

The formatting commands embedded in the text increase the difficulty of finding sentences. Not all text in a document is in sentence form; there are headings, tables, equations and lists, for example. Headings like "Finding Sentences" above should be discarded, not attached to the next sentence. However, since many of the documents are formatted to be phototypeset, and contain font changes, which usually operate on the most important words in the document, discarding all formatting commands is not correct. To improve the programs' ability to find sentence boundaries, the deformatting program, DEROFF [13], has been given some knowledge of the formatting packages used on the UNIX operating system. DEROFF will now do the following:

1. Suppress all formatting macros that are used for titles, headings, author's name, etc.
2. Suppress the arguments to the macros for titles, headings, author's name, etc.
3. Suppress displays, tables, footnotes and text that is centered or in no-fill mode.
4. Substitute a place holder for equations and check for hidden end markers. The place holder is necessary because many typists and authors use the equation setter to change fonts on important words. For this reason, header files containing the definition of the EQN delimiters must also be included as input to STYLE. End markers are often hidden when an equation ends a sentence and the period is typed inside the EQN delimiters.
5. Add a "." after lists. If the flag –ml is also used, all lists are suppressed. This is a separate flag because of the variety of ways the list macros are used. Often, lists are sentences that should be included in the analysis. The user must determine how lists are used in the document to be analyzed.

Both STYLE and DICTION call DEROFF before they look at the text. The user should supply the –ml flag if the document contains many lists of non-sentences that should be skipped.

7.2. Details of DICTION

The program DICTION is based on the string matching program FGREP. FGREP takes as input a file of patterns to be matched and a file to be searched and outputs each line that contains any of the patterns with no indication of which pattern was matched. The following changes have been added to FGREP:

1. The basic unit that DICTION operates on is a sentence rather than a line. Each sentence that contains one of the patterns is output.
2. Upper case letters are mapped to lower case.
3. Punctuation is replaced by blanks.
4 All pattern matches in the sentence are found and surrounded with "[" "]" .
5. A method for suppressing a string match has been added. Any pattern that begins with "~" will not be matched. Because the matching algorithm finds the longest substring, the suppression of a match allows words in some correct contexts not to be matched while allowing the word in another context to be found. For example, the word "which" is often incorrectly used instead of "that" in restrictive clauses. However, "which" is usually correct when preceded by a preposition or ",". The default pattern file suppresses the match of the common prepositions or a double blank followed by "which" and therefore matches only the suspect uses. The double blank accounts for the replaced comma.

8. Conclusions

A system of writing tools that measure some of the objective characteristics of writing style has been developed. The tools are sufficiently general that they may be applied to documents on any subject with equal accuracy. Although the measurements are only of the surface structure of the text, they do point out problem areas. In addition to helping writers produce better documents, these programs may be useful for studying the writing process and finding other formulae for measuring readability.

References

1. L. L. Cherry, "PARTS - A System for Assigning Word Classes to English Text," submitted *Communications of the ACM.*
2. B. W. Kernighan and J. R. Mashey, "The UNIX Programming Environment," *Software - Practice & Experience* , **9**, 1-15 (1979).
3. G. R. Klare, "Assessing Readability," *Reading Research Quarterly,* 1974-1975, **10** , 62-102.
4. E. A. Smith and P. Kincaid, "Derivation and validation of the automated readability index for use with technical materials," *Human Factors,* 1970, 12, 457-464.
5. J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (Automated Readability Index, Fog count, and Flesch Reading Ease Formula) for Navy enlisted personnel," Navy Training Command Research Branch Report 8-75, Feb., 1975.
6. M. Coleman and T. L. Liau, "A Computer Readability Formula Designed for Machine Scoring," *Journal of Applied Psychology,* 1975, 60, 283-284.
7. R. Flesch, "A New Readability Yardstick," *Journal of Applied Psychology,* 1948, 32, 221-233.
8. E. U. Coke, private communication.
9. D. W. Ewing, *Writing for Results,* John Wiley & Sons, Inc., New York, N. Y. (1974).
10. G. Leggett, C. D. Mead and W. Charvat, *Prentice-Hall Handbook for Writers,* Seventh Edition, Prentice-Hall Inc., Englewood Cliffs, N. J. (1978).
11. E. B. Coleman, "Learning of Prose Written in Four Grammatical Transformations," *Journal of Applied Psychology,* 1965, vol. 49, no. 5, pp. 332-341.
12. A. V. Aho and M. J. Corasick, "Efficient String Matching: an aid to Bibliographic Search," *Communications of the ACM,* **18**, (6), 333-340, June 1975.
13. Bell Laboratories, *"UNIX TIME-SHARING SYSTEM: UNIX PROGRAMMER'S MANUAL,"* Seventh Edition, Vol. 1 (January 1979).

Appendix 1

STYLE Abbreviations

a. d.
A. M.
a. m.
b. c.
Ch.
ch.
ckts.
dB.
Dept.
dept.
Depts.
depts.
Dr.
Drs.
e. g.
Eq.
eq.
et al.
etc.
Fig.
fig.
Figs.
figs.
ft.
i. e.
in.
Inc.
Jr.
jr.
mi.
Mr.
Mrs.
Ms.
No.
no.
Nos.
nos.
P. M.
p. m.
Ph. D.
Ph. d.
Ref.
ref.
Refs.
refs.
St.
vs.
yr.

Appendix 2

Default DICTION Patterns

a great deal of
a large number of
a lot of
a majority of
a need for
a number of
a particular preference for
a preference for
a small number of
a tendency to
abovementioned
absolutely complete
absolutely essential
accomplished
accordingly
activate
actual
added increments
adequate enough
advent
afford an opportunity
aggregate
all of
all throughout
along the line
an indication of
analyzation
and etc
and or
another additional
any and all
arrive at a
as a matter of fact
as a method of
as good or better than
as of now
as per
as regards
as related to
as to
assistance
assistance to
assistance to
assuming that
at a later date
at about
at above
at all times
at an early date
at below
at the present
at the time when
at this point in time
at this time
at which time
at your earliest convenience
authorization
awful
basic fundamentals
basically
be cognizant of
being as
being that
brief in duration
bring to a conclusion
but that
but what
by means of
by the use of
carry out experiments
center about
center around
center portion
check into
check on
check up on
circle around
close proximity
collaborate together
collect together
combine together
come to an end
commence
common accord
compensation
completely eliminated
comprise
concerning
conduct an investigation of
conjecture
connect up
consensus of opinion
consequent result
consolidate together
construct
contemplate
continue on
continue to remain
could of
count up
couple together
debate about
decide on
deleterious effect
demean
demonstrate
depreciate in value
deserving of
desirable benefits
desirous of
different than
discontinue
disutility
divide up
doubt but
due to
duly noted
during the time that
each and every
early beginnings
effectuate
emotional feelings
empty out
enclosed herein
enclosed herewith
end result
end up
endeavor
enter in
enter into
enthused
entirely complete
equally good as
essentially
eventuate
every now and then
exactly identical
experiencing difficulty
fabricate
face up to
facilitate
facts and figures
fast in action
fearful of
fearful that
few in number
file away
final completion
final ending
final outcome
final result
finalize
find it interesting to know
first and foremost
first beginnings
first initiated
firstly
follow after
following after
for the purpose of
for the reason that
for the simple reason that
for this reason
for your information
from the point of view of
full and complete
generally agreed
good and
got to
gratuitous
greatly minimize
head up
help but
helps in the production of
hopeful
if and when
if at all possible
impact
implement
important essentials
importantly
in a large measure
in a position to
in accordance
in advance of
in agreement with
in all cases
in back of
in behalf of
in behind
in between
in case
in close proximity
in conflict with
in conjunction with
in connection with
in fact
in large measure
in many cases
in most cases
in my opinion I think
in order to
in rare cases
in reference to
in regard to
in regards to
in relation with
in short supply
in size
in terms of
in the amount of
in the case of
in the course of
in the event
in the field of
in the form of
in the instance of
in the interim
in the last analysis
in the matter of
in the near future
in the neighborhood of
in the not too distant future
in the proximity of
in the range of
in the same way as described
in the shape of
in the vicinity of
in this case
in view of the
in violation of
inasmuch as
indicate
indicative of
initialize
initiate
injurious to
inquire
inside of
institute a
intents and purposes
intermingle
irregardless
is defined as
is used to control
is when
is where
it is incumbent
it stands to reason
it was noted that if
joint cooperation
joint partnership
just exactly
kind of
know about
last but not least
later on
leaving out of consideration
liable
link up
literally
little doubt that
lose out on
lots of
main essentials
make a
make adjustments to
make an
make application to
make contact with
make mention of
make out a list of
make the acquaintance of
make the adjustment
manner
maximum possible
meaningful
meet up with
melt down
melt up
methodology
might of
minimize as far as possible
minor importance
miss out on
modification

more preferable
most unique
must of
mutual cooperation
necessary requisite
necessitate
need for
nice
not be un
not in a position to
not of a high order of accuracy
not un
notwithstanding
of considerable magnitude
of that
of the opinion that
off of
on a few occasions
on account of
on behalf of
on the grounds that
on the occasion
on the part of
one of the
open up
operates to correct
outside of
over with
overall
past history
perceptive of
perform a measurement
perform the measurement
permits the reduction of
personalize
pertaining to
physical size
plan ahead
plan for the future
plan in advance
plan on
present a conclusion
present a report
presently
prior to
prioritize
proceed to
procure
productive of
prolong the duration
protrude out from
provided that
pursuant to
put to use in
range all the way from
reason is because
reason why
recur again
reduce down
refer back
reference to this
reflective of
regarding
regretful
reinitiate
relative to
repeat again
representative of
resultant effect
resume again
retreat back
return again
return back
revert back
seal off
seems apparent
send a communication
short space of time
should of
single unit
situation
so as to
sort of
spell out
still continue
still remain
subsequent
substantially in agreement
succeed in
suggestive of
superior than
surrounding circumstances
take appropriate
take cognizance of
take into consideration
termed as
terminate
termination
the author
the authors
the case that
the fact
the foregoing
the foreseeable future
the fullest possible extent
the majority of
the nature
the necessity of
the only difference being that
the order of
the point that
the truth is
there are not many
through the medium of
through the use of
throughout the entire
time interval
to summarize the above
total effect of all this
totality
transpire
true facts
try and
ultimate end
under a separate cover
under date of
under separate cover
under the necessity to
underlying purpose
undertake a study
uniformly consistent
unique
until such time as
up to this time
upshot
utilize
very
very complete
very unique
vital
which
with a view to
with reference to
with regard to
with the exception of
with the object of
with the result that
with this in mind, it is clear that
within the realm of possibility
without further delay
worth while
would of
ing behavior
wise
~ which
~ about which
~ after which
~ at which
~ between which
~ by which
~ for which
~ from which
~ in which
~ into which
~ of which
~ on which
~ on which
~ over which
~ through which
~ to which
~ under which
~ upon which
~ with which
~ without which
~clockwise
~likewise
~otherwise

NROFF/TROFF User's Manual

Joseph F. Ossanna

Bell Laboratories
Murray Hill, New Jersey 07974

Introduction

NROFF and TROFF are text processors under the PDP-11 UNIX Time-Sharing System[1] that format text for typewriter-like terminals and for a Graphic Systems phototypesetter, respectively. They accept lines of text interspersed with lines of format control information and format the text into a printable, paginated document having a user-designed style. NROFF and TROFF offer unusual freedom in document styling, including: arbitrary style headers and footers; arbitrary style footnotes; multiple automatic sequence numbering for paragraphs, sections, etc; multiple column output; dynamic font and point-size control; arbitrary horizontal and vertical local motions at any point; and a family of automatic overstriking, bracket construction, and line drawing functions.

NROFF and TROFF are highly compatible with each other and it is almost always possible to prepare input acceptable to both. Conditional input is provided that enables the user to embed input expressly destined for either program. NROFF can prepare output directly for a variety of terminal types and is capable of utilizing the full resolution of each terminal.

Usage

The general form of invoking NROFF (or TROFF) at UNIX command level is

nroff *options files* (or **troff** *options files*)

where *options* represents any of a number of option arguments and *files* represents the list of files containing the document to be formatted. An argument consisting of a single minus (–) is taken to be a file name corresponding to the standard input. If no file names are given input is taken from the standard input. The options, which may appear in any order so long as they appear before the files, are:

Option	*Effect*
–o*list*	Print only pages whose page numbers appear in *list*, which consists of comma-separated numbers and number ranges. A number range has the form *N*-*M* and means pages *N* through *M;* a initial -*N* means from the beginning to page *N;* and a final *N*- means from *N* to the end.
–n*N*	Number first generated page *N*.
–s*N*	Stop every *N* pages. NROFF will halt prior to every *N* pages (default *N*=1) to allow paper loading or changing, and will resume upon receipt of a newline. TROFF will stop the phototypesetter every *N* pages, produce a trailer to allow changing cassettes, and will resume after the phototypesetter START button is pressed.
–m*name*	Prepends the macro file **/usr/lib/tmac.***name* to the input *files*.
–r*aN*	Register *a* (one-character) is set to *N*.
–i	Read standard input after the input files are exhausted.
–q	Invoke the simultaneous input-output mode of the **rd** request.

NROFF Only

−T*name* Specifies the name of the output terminal type. Currently defined names are **37** for the (default) Model 37 Teletype®, **tn300** for the GE TermiNet 300 (or any terminal without half-line capabilities), **300S** for the DASI-300S, **300** for the DASI-300, and **450** for the DASI-450 (Diablo Hyterm).

−e Produce equally-spaced words in adjusted lines, using full terminal resolution.

TROFF Only

−t Direct output to the standard output instead of the phototypesetter.

−f Refrain from feeding out paper and stopping phototypesetter at the end of the run.

−w Wait until phototypesetter is available, if currently busy.

−b TROFF will report whether the phototypesetter is busy or available. No text processing is done.

−a Send a printable (ASCII) approximation of the results to the standard output.

−p*N* Print all characters in point size *N* while retaining all prescribed spacings and motions, to reduce phototypesetter elasped time.

−g Prepare output for the Murray Hill Computation Center phototypesetter and direct it to the standard output.

Each option is invoked as a separate argument; for example,

nroff −o*4,8−10* **−T***300S* **−m***abc file1 file2*

requests formatting of pages 4, 8, 9, and 10 of a document contained in the files named *file1* and *file2*, specifies the output terminal as a DASI-300S, and invokes the macro package *abc*.

Various pre- and post-processors are available for use with NROFF and TROFF. These include the equation preprocessors NEQN and EQN[2] (for NROFF and TROFF respectively), and the table-construction preprocessor TBL[3]. A reverse-line postprocessor COL[4] is available for multiple-column NROFF output on terminals without reverse-line ability; COL expects the Model 37 Teletype escape sequences that NROFF produces by default. TK[4] is a 37 Teletype simulator postprocessor for printing NROFF output on a Tektronix 4014. TCAT[4] is phototypesetter-simulator postprocessor for TROFF that produces an approximation of phototypesetter output on a Tektronix 4014. For example, in

tbl *files* | **eqn** | **troff −t** *options* | **tcat**

the first | indicates the piping of TBL's output to EQN's input; the second the piping of EQN's output to TROFF's input; and the third indicates the piping of TROFF's output to TCAT. GCAT[4] can be used to send TROFF (**−g**) output to the Murray Hill Computation Center.

The remainder of this manual consists of: a Summary and Index; a Reference Manual keyed to the index; and a set of Tutorial Examples. Another tutorial is [5].

Joseph F. Ossanna

References

[1] K. Thompson, D. M. Ritchie, *UNIX Programmer's Manual,* Sixth Edition (May 1975).

[2] B. W. Kernighan, L. L. Cherry, *Typesetting Mathematics — User's Guide (Second Edition),* Bell Laboratories internal memorandum.

[3] M. E. Lesk, *Tbl — A Program to Format Tables,* Bell Laboratories internal memorandum.

[4] Internal on-line documentation, on UNIX.

[5] B. W. Kernighan, *A TROFF Tutorial,* Bell Laboratories internal memorandum.

SUMMARY AND INDEX

Request Form	*Initial Value**	*If No Argument*	*Notes#*	*Explanation*
1. General Explanation				
2. Font and Character Size Control				
.ps ±*N*	10 point	previous	E	Point size; also **\s**±*N*.†
.ss *N*	12/36 em	ignored	E	Space-character size set to *N*/36 em.†
.cs *F N M*	off	-	P	Constant character space (width) mode (font *F*).†
.bd *F N*	off	-	P	Embolden font *F* by *N*–1 units.†
.bd S *F N*	off	-	P	Embolden Special Font when current font is *F*.†
.ft *F*	Roman	previous	E	Change to font *F* = *x*, *xx*, or 1-4. Also **\f***x*, **\f(***xx*, **\f***N*.
.fp *N F*	R,I,B,S	ignored	-	Font named *F* mounted on physical position 1≤*N*≤4.
3. Page Control				
.pl ±*N*	11 in	11 in	**v**	Page length.
.bp ±*N*	*N*=1	-	B‡,**v**	Eject current page; next page number *N*.
.pn ±*N*	*N*=1	ignored	-	Next page number *N*.
.po ±*N*	0; 26/27 in	previous	**v**	Page offset.
.ne *N*	-	*N*=1*V*	D,**v**	Need *N* vertical space (*V* = vertical spacing).
.mk *R*	none	internal	D	Mark current vertical place in register *R*.
.rt ±*N*	none	internal	D,**v**	Return *(upward only)* to marked vertical place.
4. Text Filling, Adjusting, and Centering				
.br	-	-	B	Break.
.fi	fill	-	B,E	Fill output lines.
.nf	fill	-	B,E	No filling or adjusting of output lines.
.ad *c*	adj,both	adjust	E	Adjust output lines with mode *c*.
.na	adjust	-	E	No output line adjusting.
.ce *N*	off	*N*=1	B,E	Center following *N* input text lines.
5. Vertical Spacing				
.vs *N*	1/6in;12pts	previous	E,**p**	Vertical base line spacing (*V*).
.ls *N*	*N*=1	previous	E	Output *N*–1 *V*s after each text output line.
.sp *N*	-	*N*=1*V*	B,**v**	Space vertical distance *N in either direction.*
.sv *N*	-	*N*=1*V*	**v**	Save vertical distance *N*.
.os	-	-	-	Output saved vertical distance.
.ns	space	-	D	Turn no-space mode on.
.rs	-	-	D	Restore spacing; turn no-space mode off.
6. Line Length and Indenting				
.ll ±*N*	6.5 in	previous	E,**m**	Line length.
.in ±*N*	*N*=0	previous	B,E,**m**	Indent.
.ti ±*N*	-	ignored	B,E,**m**	Temporary indent.
7. Macros, Strings, Diversion, and Position Traps				
.de *xx yy*	-	*.yy*=..	-	Define or redefine macro *xx;* end at call of *yy*.
.am *xx yy*	-	*.yy*=..	-	Append to a macro.
.ds *xx string*	-	ignored	-	Define a string *xx* containing *string*.
.as *xx string*	-	ignored	-	Append *string* to string *xx*.

*Values separated by ";" are for NROFF and TROFF respectively.

#Notes are explained at the end of this Summary and Index

†No effect in NROFF.

‡The use of " ´ " as control character (instead of ".") suppresses the break function.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.rm *xx*	-	ignored	-	Remove request, macro, or string.
.rn *xx yy*	-	ignored	-	Rename request, macro, or string *xx* to *yy*.
.di *xx*	-	end	D	Divert output to macro *xx*.
.da *xx*	-	end	D	Divert and append to *xx*.
.wh *N xx*	-	-	v	Set location trap; negative is w.r.t. page bottom.
.ch *xx N*	-	-	v	Change trap location.
.dt *N xx*	-	off	D,v	Set a diversion trap.
.it *N xx*	-	off	E	Set an input-line count trap.
.em *xx*	none	none	-	End macro is *xx*.

8. Number Registers

.nr $R \pm N M$	-	u		Define and set number register *R*; auto-increment by *M*.
.af *R c*	arabic	-	-	Assign format to register *R* (*c*=**1**, **i**, **I**, **a**, **A**).
.rr *R*	-	-	-	Remove register *R*.

9. Tabs, Leaders, and Fields

.ta *Nt* ...	0.8; 0.5in	none	E,m	Tab settings; *left* type, unless *t*=**R**(right), **C**(centered).
.tc *c*	none	none	E	Tab repetition character.
.lc *c*	.	none	E	Leader repetition character.
.fc *a b*	off	off	-	Set field delimiter *a* and pad character *b*.

10. Input and Output Conventions and Character Translations

.ec *c*	\	\	-	Set escape character.
.eo	on	-	-	Turn off escape character mechanism.
.lg *N*	-; on	on	-	Ligature mode on if $N>0$.
.ul *N*	off	N=1	E	Underline (italicize in TROFF) *N* input lines.
.cu *N*	off	N=1	E	Continuous underline in NROFF; like **ul** in TROFF.
.uf *F*	Italic	Italic	-	Underline font set to *F* (to be switched to by **ul**).
.cc *c*	.	.	E	Set control character to *c*.
.c2 *c*	'	'	E	Set nobreak control character to *c*.
.tr *abcd....*	none	-	O	Translate *a* to *b*, etc. on output.

11. Local Horizontal and Vertical Motions, and the Width Function

12. Overstrike, Bracket, Line-drawing, and Zero-width Functions

13. Hyphenation.

.nh	hyphenate	-	E	No hyphenation.
.hy *N*	hyphenate	hyphenate	E	Hyphenate; N = mode.
.hc *c*	\%	\%	E	Hyphenation indicator character *c*.
.hw *word1* ...		ignored	-	Exception words.

14. Three Part Titles.

.tl *'left'center'right'*		-	-	Three part title.
.pc *c*	%	off	-	Page number character.
.lt $\pm N$	6.5 in	previous	E,m	Length of title.

15. Output Line Numbering.

.nm $\pm N M S I$		off	E	Number mode on or off, set parameters.
.nn *N*	-	N=1	E	Do not number next *N* lines.

16. Conditional Acceptance of Input

.if *c anything*			-	- If condition *c* true, accept *anything* as input, for multi-line use \{*anything*\}.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.if !*c anything*	-	-		If condition *c* false, accept *anything*.
.if *N anything*	-	u		If expression $N > 0$, accept *anything*.
.if !*N anything*	-	u		If expression $N \leq 0$, accept *anything*.
.if *'string1 'string2 ' anything*		-		If *string1* identical to *string2*, accept *anything*.
.if ! *'string1 'string2 ' anything*		-		If *string1* not identical to *string2*, accept *anything*.
.ie *c anything*	-	u		If portion of if-else; all above forms (like **if**).
.el *anything*	-	-		Else portion of if-else.

17. Environment Switching.

.ev *N*	*N*=0	previous	-	Environment switched (*push down*).

18. Insertions from the Standard Input

.rd *prompt*	-	*prompt*=BEL-		Read insertion.
.ex	-	-	-	Exit from NROFF/TROFF.

19. Input/Output File Switching

.so *filename*		-	-	Switch source file *(push down)*.
.nx *filename*		end-of-file	-	Next file.
.pi *program*		-	-	Pipe output to *program* (NROFF only).

20. Miscellaneous

.mc *c N*	-	off	E,m	Set margin character *c* and separation *N*.
.tm *string*	-	newline	-	Print *string* on terminal (UNIX standard message output).
.ig *yy*	-	*.yy*=..	-	Ignore till call of *yy*.
.pm *t*	-	all	-	Print macro names and sizes; if *t* present, print only total of sizes.
.fl	-	-	B	Flush output buffer.

21. Output and Error Messages

Notes-

B Request normally causes a break.
D Mode or relevant parameters associated with current diversion level.
E Relevant parameters are a part of the current environment.
O Must stay in effect until logical output.
P Mode must be still or again in effect at the time of physical output.
v,p,m,u Default scale indicator; if not specified, scale indicators are *ignored*.

Alphabetical Request and Section Number Cross Reference

ad	4	cc	10	ds	7	fc	9	ie	16	ll	6	nh	13	pi	19	rn	7	ta	9	vs	5
af	8	ce	4	dt	7	fi	4	if	16	ls	5	nm	15	pl	3	rr	8	tc	9	wh	7
am	7	ch	7	ec	10	fl	20	ig	20	lt	14	nn	15	pm	20	rs	5	ti	6		
as	7	cs	2	el	16	fp	2	in	6	mc	20	nr	8	pn	3	rt	3	tl	14		
bd	2	cu	10	em	7	ft	2	it	7	mk	3	ns	5	po	3	so	19	tm	20		
bp	3	da	7	eo	10	hc	13	lc	9	na	4	nx	19	ps	2	sp	5	tr	10		
br	4	de	7	ev	17	hw	13	lg	10	ne	3	os	5	rd	18	ss	2	uf	10		
c2	10	di	7	ex	18	hy	13	li	10	nf	4	pc	14	rm	7	sv	5	ul	10		

Escape Sequences for Characters, Indicators, and Functions

Section Reference	*Escape Sequence*	*Meaning*
10.1	****	\\ (to prevent or delay the interpretation of \\ )
10.1	**\e**	Printable version of the *current* escape character.
2.1	**\\´**	´ (acute accent); equivalent to **\\(aa**
2.1	**\\`**	` (grave accent); equivalent to **\\(ga**
2.1	**\\–**	– Minus sign in the *current* font
7	**\\.**	Period (dot) (see **de**)
11.1	\\(space)	Unpaddable space-size space character
11.1	**\0**	Digit width space
11.1	**\\\|**	1/6 em narrow space character (zero width in NROFF)
11.1	**\\^**	1/12 em half-narrow space character (zero width in NROFF)
4.1	**\\&**	Non-printing, zero width character
10.6	**\\!**	Transparent line indicator
10.7	**\\"**	Beginning of comment
7.3	**\\$***N*	Interpolate argument $1 \leq N \leq 9$
13	**\\%**	Default optional hyphenation character
2.1	**\\(***xx*	Character named *xx*
7.1	*****x*, ***(***xx*	Interpolate string *x* or *xx*
9.1	**\a**	Non-interpreted leader character
12.3	**\b**´*abc...*´	Bracket building function
4.2	**\c**	Interrupt text processing
11.1	**\d**	Forward (down) 1/2 em vertical motion (1/2 line in NROFF)
2.2	**\f***x*,**\f(***xx*,**\f***N*	Change to font named *x* or *xx*, or position *N*
11.1	**\h**´*N*´	Local horizontal motion; move right *N* *(negative left)*
11.3	**\k***x*	Mark horizontal *input* place in register *x*
12.4	**\l**´*Nc*´	Horizontal line drawing function (optionally with *c*)
12.4	**\L**´*Nc*´	Vertical line drawing function (optionally with *c*)
8	**\n***x*,**\n(***xx*	Interpolate number register *x* or *xx*
12.1	**\o**´*abc...*´	Overstrike characters *a, b, c, ...*
4.1	**\p**	Break and spread output line
11.1	**\r**	Reverse 1 em vertical motion (reverse line in NROFF)
2.3	**\s***N*, **\s±***N*	Point-size change function
9.1	**\t**	Non-interpreted horizontal tab
11.1	**\u**	Reverse (up) 1/2 em vertical motion (1/2 line in NROFF)
11.1	**\v**´*N*´	Local vertical motion; move down *N* *(negative up)*
11.2	**\w**´*string*´	Interpolate width of *string*
5.2	**\x**´*N*´	Extra line-space function *(negative before, positive after)*
12.2	**\z***c*	Print *c* with zero width (without spacing)
16	**\\{**	Begin conditional input
16	**\\}**	End conditional input
10.7	\\(newline)	Concealed (ignored) newline
-	*X*	*X*, any character *not* listed above

The escape sequences ****, **\\.**, **\\"**, **\\$**, *****, **\a**, **\n**, **\t**, and \\(newline) are interpreted in *copy mode* (§7.2).

Predefined General Number Registers

Section Reference	*Register Name*	*Description*
3	**%**	Current page number.
11.2	**ct**	Character type (set by *width* function).
7.4	**dl**	Width (maximum) of last completed diversion.
7.4	**dn**	Height (vertical size) of last completed diversion.
-	**dw**	Current day of the week (1-7).
-	**dy**	Current day of the month (1-31).
11.3	**hp**	Current horizontal place on *input* line.
15	**ln**	Output line number.
-	**mo**	Current month (1-12).
4.1	**nl**	Vertical position of last printed text base-line.
11.2	**sb**	Depth of string below base line (generated by *width* function).
11.2	**st**	Height of string above base line (generated by *width* function).
-	**yr**	Last two digits of current year.

Predefined Read-Only Number Registers

Section Reference	*Register Name*	*Description*
7.3	**.$**	Number of arguments available at the current macro level.
-	**.A**	Set to 1 in TROFF, if **-a** option used; always 1 in NROFF.
11.1	**.H**	Available horizontal resolution in basic units.
-	**.T**	Set to 1 in NROFF, if **-T** option used; always 0 in TROFF.
11.1	**.V**	Available vertical resolution in basic units.
5.2	**.a**	Post-line extra line-space most recently utilized using **\x**'*N*'.
-	**.c**	Number of *lines* read from current input file.
7.4	**.d**	Current vertical place in current diversion; equal to **nl**, if no diversion.
2.2	**.f**	Current font as physical quadrant (1-4).
4	**.h**	Text base-line high-water mark on current page or diversion.
6	**.i**	Current indent.
6	**.l**	Current line length.
4	**.n**	Length of text portion on previous output line.
3	**.o**	Current page offset.
3	**.p**	Current page length.
2.3	**.s**	Current point size.
7.5	**.t**	Distance to the next trap.
4.1	**.u**	Equal to 1 in fill mode and 0 in nofill mode.
5.1	**.v**	Current vertical line spacing.
11.2	**.w**	Width of previous character.
-	**.x**	Reserved version-dependent register.
-	**.y**	Reserved version-dependent register.
7.4	**.z**	Name of current diversion.

REFERENCE MANUAL

1. General Explanation

1.1. Form of input. Input consists of *text lines*, which are destined to be printed, interspersed with *control lines*, which set parameters or otherwise control subsequent processing. Control lines begin with a *control character*—normally **.** (period) or **´** (acute accent)—followed by a one or two character name that specifies a basic *request* or the substitution of a user-defined *macro* in place of the control line. The control character **´** suppresses the *break* function—the forced output of a partially filled line—caused by certain requests. The control character may be separated from the request/macro name by white space (spaces and/or tabs) for esthetic reasons. Names must be followed by either space or newline. Control lines with unrecognized names are ignored.

Various special functions may be introduced anywhere in the input by means of an *escape* character, normally \\. For example, the function **\n***R* causes the interpolation of the contents of the *number register R* in place of the function; here *R* is either a single character name as in **\n***x*, or left-parenthesis-introduced, two-character name as in **\n(***xx*.

1.2. Formatter and device resolution. TROFF internally uses 432 units/inch, corresponding to the Graphic Systems phototypesetter which has a horizontal resolution of 1/432 inch and a vertical resolution of 1/144 inch. NROFF internally uses 240 units/inch, corresponding to the least common multiple of the horizontal and vertical resolutions of various typewriter-like output devices. TROFF rounds horizontal/vertical numerical parameter input to the actual horizontal/vertical resolution of the Graphic Systems typesetter. NROFF similarly rounds numerical input to the actual resolution of the output device indicated by the **−T** option (default Model 37 Teletype).

1.3. Numerical parameter input. Both NROFF and TROFF accept numerical input with the appended scale indicators shown in the following table, where S is the current type size in points, V is the current vertical line spacing in basic units, and C is a *nominal character width* in basic units.

Scale Indicator	Meaning	Number of basic units TROFF	NROFF
i	Inch	432	240
c	Centimeter	432×50/127	240×50/127
P	Pica = 1/6 inch	72	240/6
m	Em = S points	6×S	C
n	En = Em/2	3×S	C, *same as Em*
p	Point = 1/72 inch	6	240/72
u	Basic unit	1	1
v	Vertical line space	V	V
none	Default, see below		

In NROFF, *both* the em and the en are taken to be equal to the C, which is output-device dependent; common values are 1/10 and 1/12 inch. Actual character widths in NROFF need not be all the same and constructed characters such as −> (→) are often extra wide. The default scaling is ems for the horizontally-oriented requests and functions **ll**, **in**, **ti**, **ta**, **lt**, **po**, **mc**, **\h**, and **\l**; Vs for the vertically-oriented requests and functions **pl**, **wh**, **ch**, **dt**, **sp**, **sv**, **ne**, **rt**, **\v**, **\x**, and **\L**; **p** for the **vs** request; and **u** for the requests **nr**, **if**, and **ie**. *All* other requests ignore any scale indicators. When a number register containing an already appropriately scaled number is interpolated to provide numerical input, the unit scale indicator **u** may need to be appended to prevent an additional inappropriate default scaling. The number, N, may be specified in decimal-fraction form but the parameter finally stored is rounded to an integer number of basic units.

The ***absolute position*** indicator | may be prepended to a number *N* to generate the distance to the vertical or horizontal place *N*. For vertically-oriented requests and functions, |*N* becomes the distance in basic units from the current vertical place on the page or in a ***diversion*** (§7.4) to the the vertical place *N*. For *all* other requests and functions, |*N* becomes the distance from the current horizontal place on the *input* line to the horizontal place *N*. For example,

.sp |3.2c

will space ***in the required direction*** to 3.2 centimeters from the top of the page.

1.4. Numerical expressions. Wherever numerical input is expected an expression involving parentheses, the arithmetic operators +, −, /, *, % (mod), and the logical operators <, >, <=, >=, = (or ==), & (and), : (or) may be used. Except where controlled by parentheses, evaluation of expressions is left-to-right; there is no operator precedence. In the case of certain requests, an initial + or − is stripped and interpreted as an increment or decrement indicator respectively. In the presence of default scaling, the desired scale indicator must be attached to *every* number in an expression for which the desired and default scaling differ. For example, if the number register **x** contains 2 and the current point size is 10, then

.ll (4.25i+\nxP+3)/2u

will set the line length to 1/2 the sum of 4.25 inches + 2 picas + 30 points.

1.5. Notation. Numerical parameters are indicated in this manual in two ways. ±*N* means that the argument may take the forms *N*, +*N*, or −*N* and that the corresponding effect is to set the affected parameter to *N*, to increment it by *N*, or to decrement it by *N* respectively. Plain *N* means that an initial algebraic sign is *not* an increment indicator, but merely the sign of *N*. Generally, unreasonable numerical input is either ignored or truncated to a reasonable value. For example, most requests expect to set parameters to non-negative values; exceptions are **sp**, **wh**, **ch**, **nr**, and **if**. The requests **ps**, **ft**, **po**, **vs**, **ls**, **ll**, **in**, and **lt** restore the *previous* parameter value in the *absence* of an argument.

Single character arguments are indicated by single lower case letters and one/two character arguments are indicated by a pair of lower case letters. Character string arguments are indicated by multi-character mnemonics.

2. Font and Character Size Control

2.1. Character set. The TROFF character set consists of the Graphics Systems Commercial II character set plus a Special Mathematical Font character set—each having 102 characters. These character sets are shown in the attached Table I. All ASCII characters are included, with some on the Special Font. With three exceptions, the ASCII characters are input as themselves, and non-ASCII characters are input in the form \(*xx* where *xx* is a two-character name given in the attached Table II. The three ASCII exceptions are mapped as follows:

ASCII Input		Printed by TROFF	
Character	Name	Character	Name
´	acute accent	’	close quote
\`	grave accent	‘	open quote
−	minus	-	hyphen

The characters ´, \`, and − may be input by \´, \\`, and \− respectively or by their names (Table II). The ASCII characters @, #, ", ´, \`, <, >, \, {, }, ~, ^, and _ exist only on the Special Font and are printed as a 1-em space if that Font is not mounted.

NROFF understands the entire TROFF character set, but can in general print only ASCII characters, additional characters as may be available on the output device, such characters as may be able to be constructed by overstriking or other combination, and those that can reasonably be mapped into other printable characters. The exact behavior is determined by a driving table prepared for each device. The characters ´, \`, and _ print as themselves.

2.2. Fonts. The default mounted fonts are Times Roman (**R**), Times Italic (**I**), Times Bold (**B**), and the Special Mathematical Font (**S**) on physical typesetter positions 1, 2, 3, and 4 respectively. These fonts are used in this document. The ***current*** font, initially Roman, may be changed (among the mounted fonts) by use of the **ft** request, or by imbedding at any desired point either **\f***x*, **\f(***xx*, or **\f***N* where *x* and *xx* are the name of a mounted font and *N* is a numerical font position. It is ***not*** necessary to change to the Special font; characters on that font are automatically handled. A request for a named but not-mounted font is ***ignored***. TROFF can be informed that any particular font is mounted by use of the **fp** request. The list of known fonts is installation dependent. In the subsequent discussion of font-related requests, *F* represents either a one/two-character font name or the numerical font position, 1-4. The current font is available (as numerical position) in the read-only number register **.f**.

NROFF understands font control and normally underlines Italic characters (see §10.5).

2.3. Character size. Character point sizes available on the Graphic Systems typesetter are 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, 22, 24, 28, and 36. This is a range of 1/12 inch to 1/2 inch. The **ps** request is used to change or restore the point size. Alternatively the point size may be changed between any two characters by imbedding a **\s***N* at the desired point to set the size to *N*, or a **\s**$\pm N$ ($1 \leq N \leq 9$) to increment/decrement the size by *N*; **\s0** restores the *previous* size. Requested point size values that are between two valid sizes yield the larger of the two. The current size is available in the **.s** register. NROFF ignores type size control.

Request Form	*Initial Value*	*If No Argument*	*Notes**	*Explanation*
.ps $\pm N$	10 point	previous	E	Point size set to $\pm N$. Alternatively imbed **\s***N* or **\s**$\pm N$. Any positive size value may be requested; if invalid, the next larger valid size will result, with a maximum of 36. A paired sequence $+N, -N$ will work because the previous requested value is also remembered. Ignored in NROFF.
.ss *N*	12/36 em	ignored	E	Space-character size is set to *N*/36 ems. This size is the minimum word spacing in adjusted text. Ignored in NROFF.
.cs *F N M*	off	-	P	Constant character space (width) mode is set on for font *F* (if mounted); the width of every character will be taken to be *N*/36 ems. If *M* is absent, the em is that of the character's point size; if *M* is given, the em is *M*-points. All affected characters are centered in this space, including those with an actual width larger than this space. Special Font characters occurring while the current font is *F* are also so treated. If *N* is absent, the mode is turned off. The mode must be still or again in effect when the characters are physically printed. Ignored in NROFF.
.bd *F N*	off	-	P	The characters in font *F* will be artificially emboldened by printing each one twice, separated by *N*−1 basic units. A reasonable value for *N* is 3 when the character size is in the vicinity of 10 points. If *N* is missing the embolden mode is turned off. The column heads above were printed with **.bd I 3**. The mode must be still or again in effect when the characters are physically printed. Ignored in NROFF.
.bd S *F N*	off	-	P	The characters in the Special Font will be emboldened whenever the current font is *F*. This manual was printed with **.bd S B 3**. The mode must be still or again in effect when the characters are physically printed.

*Notes are explained at the end of the Summary and Index above.

.ft *F*	Roman	previous	E	Font changed to *F*. Alternatively, imbed **\f***F*. The font name **P** is reserved to mean the previous font.
.fp *N F*	R,I,B,S	ignored	-	Font position. This is a statement that a font named *F* is mounted on position *N* (1-4). It is a fatal error if *F* is not known. The phototypesetter has four fonts physically mounted. Each font consists of a film strip which can be mounted on a numbered quadrant of a wheel. The default mounting sequence assumed by TROFF is R, I, B, and S on positions 1, 2, 3 and 4.

3. Page control

Top and bottom margins are *not* automatically provided; it is conventional to define two *macros* and to set *traps* for them at vertical positions 0 (top) and $-N$ (N from the bottom). See §7 and Tutorial Examples §T2. A pseudo-page transition onto the *first* page occurs either when the first *break* occurs or when the first *non-diverted* text processing occurs. Arrangements for a trap to occur at the top of the first page must be completed before this transition. In the following, references to the *current diversion* (§7.4) mean that the mechanism being described works during both ordinary and diverted output (the former considered as the top diversion level).

The useable page width on the Graphic Systems phototypesetter is about 7.54 inches, beginning about 1/27 inch from the left edge of the 8 inch wide, continuous roll paper. The physical limitations on NROFF output are output-device dependent.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.pl $\pm N$	11 in	11 in	**v**	Page length set to $\pm N$. The internal limitation is about 75 inches in TROFF and about 136 inches in NROFF. The current page length is available in the **.p** register.
.bp $\pm N$	N=1	-	B*,**v**	Begin page. The current page is ejected and a new page is begun. If $\pm N$ is given, the new page number will be $\pm N$. Also see request **ns**.
.pn $\pm N$	N=1	ignored	-	Page number. The next page (when it occurs) will have the page number $\pm N$. A **pn** must occur before the initial pseudo-page transition to effect the page number of the first page. The current page number is in the **%** register.
.po $\pm N$	0; 26/27 in†	previous	**v**	Page offset. The current *left margin* is set to $\pm N$. The TROFF initial value provides about 1 inch of paper margin including the physical typesetter margin of 1/27 inch. In TROFF the maximum (line-length)+(page-offset) is about 7.54 inches. See §6. The current page offset is available in the **.o** register.
.ne *N*	-	N=1 V	D,**v**	Need N vertical space. If the distance, D, to the next trap position (see §7.5) is less than N, a forward vertical space of size D occurs, which will spring the trap. If there are no remaining traps on the page, D is the distance to the bottom of the page. If $D < V$, another line could still be output and spring the trap. In a diversion, D is the distance to the *diversion trap*, if any, or is very large.
.mk *R*	none	internal	D	Mark the *current* vertical place in an internal register (both associated with the current diversion level), or in register

*The use of " ´ " as control character (instead of ".") suppresses the break function.

†Values separated by ";" are for NROFF and TROFF respectively.

				R, if given. See **rt** request.
.rt $\pm N$	none	internal	D,v	Return ***upward only*** to a marked vertical place in the current diversion. If $\pm N$ (w.r.t. current place) is given, the place is $\pm N$ from the top of the page or diversion or, if N is absent, to a place marked by a previous **mk**. Note that the **sp** request (§5.3) may be used in all cases instead of **rt** by spacing to the absolute place stored in a explicit register; e. g. using the sequence **.mk** *R* ... **.sp** **|\n***R***u**.

4. Text Filling, Adjusting, and Centering

4.1. Filling and adjusting. Normally, words are collected from input text lines and assembled into a output text line until some word doesn't fit. An attempt is then made the hyphenate the word in effort to assemble a part of it into the output line. The spaces between the words on the output line are then increased to spread out the line to the current *line length* minus any current *indent*. A *word* is any string of characters delimited by the *space* character or the beginning/end of the input line. Any adjacent pair of words that must be kept together (neither split across output lines nor spread apart in the adjustment process) can be tied together by separating them with the *unpaddable space* character "\ " (backslash-space). The adjusted word spacings are uniform in TROFF and the minimum interword spacing can be controlled with the **ss** request (§2). In NROFF, they are normally nonuniform because of quantization to character-size spaces; however, the command line option **−e** causes uniform spacing with full output device resolution. Filling, adjustment, and hyphenation (§13) can all be prevented or controlled. The *text length* on the last line output is available in the **.n** register, and text base-line position on the page for this line is in the **nl** register. The text base-line high-water mark (lowest place) on the current page is in the **.h** register.

An input text line ending with **.**, **?**, or **!** is taken to be the end of a *sentence*, and an additional space character is automatically provided during filling. Multiple inter-word space characters found in the input are retained, except for trailing spaces; initial spaces also cause a *break*.

When filling is in effect, a **\p** may be imbedded or attached to a word to cause a *break* at the *end* of the word and have the resulting output line *spread out* to fill the current line length.

A text input line that happens to begin with a control character can be made to not look like a control line by prefacing it with the non-printing, zero-width filler character **\&**. Still another way is to specify output translation of some convenient character into the control character using **tr** (§10.5).

4.2. Interrupted text. The copying of a input line in *nofill* (non-fill) mode can be *interrupted* by terminating the partial line with a **\c**. The *next* encountered input text line will be considered to be a continuation of the same line of input text. Similarly, a word within *filled* text may be interrupted by terminating the word (and line) with **\c**; the next encountered text will be taken as a continuation of the interrupted word. If the intervening control lines cause a break, any partial line will be forced out along with any partial word.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.br	-	-	B	Break. The filling of the line currently being collected is stopped and the line is output without adjustment. Text lines beginning with space characters and empty text lines (blank lines) also cause a break.
.fi	fill on	-	B,E	Fill subsequent output lines. The register **.u** is 1 in fill mode and 0 in nofill mode.
.nf	fill on	-	B,E	Nofill. Subsequent output lines are *neither* filled *nor* adjusted. Input text lines are copied directly to output lines *without regard* for the current line length.
.ad *c*	adj,both	adjust	E	Line adjustment is begun. If fill mode is not on, adjustment will be deferred until fill mode is back on. If the type indicator *c* is present, the adjustment type is changed as

shown in the following table.

Indicator	Adjust Type
l	adjust left margin only
r	adjust right margin only
c	center
b or **n**	adjust both margins
absent	unchanged

.na	adjust	-	E	Noadjust. Adjustment is turned off; the right margin will be ragged. The adjustment type for **ad** is not changed. Output line filling still occurs if fill mode is on.
.ce *N*	off	*N*=1	B,E	Center the next *N* input text lines within the current (line-length minus indent). If *N*=0, any residual count is cleared. A break occurs after each of the *N* input lines. If the input line is too long, it will be left adjusted.

5. Vertical Spacing

5.1. Base-line spacing. The vertical spacing *(V)* between the base-lines of successive output lines can be set using the **vs** request with a resolution of 1/144 inch = 1/2 point in TROFF, and to the output device resolution in NROFF. *V* must be large enough to accommodate the character sizes on the affected output lines. For the common type sizes (9-12 points), usual typesetting practice is to set *V* to 2 points greater than the point size; TROFF default is 10-point type on a 12-point spacing (as in this document). The current *V* is available in the **.v** register. Multiple-*V* line separation (e. g. double spacing) may be requested with **ls**.

5.2. Extra line-space. If a word contains a vertically tall construct requiring the output line containing it to have extra vertical space before and/or after it, the *extra-line-space* function **\x**'*N*' can be imbedded in or attached to that word. In this and other functions having a pair of delimiters around their parameter (here '), the delimiter choice is arbitrary, except that it can't look like the continuation of a number expression for *N*. If *N* is negative, the output line containing the word will be preceded by *N* extra vertical space; if *N* is positive, the output line containing the word will be followed by *N* extra vertical space. If successive requests for extra space apply to the same line, the maximum values are used. The most recently utilized post-line extra line-space is available in the **.a** register.

5.3. Blocks of vertical space. A block of vertical space is ordinarily requested using **sp**, which honors the *no-space* mode and which does not space *past* a trap. A contiguous block of vertical space may be reserved using **sv**.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.vs *N*	1/6in;12pts	previous	E,p	Set vertical base-line spacing size *V*. Transient *extra* vertical space available with **\x**'*N*' (see above).
.ls *N*	*N*=1	previous	E	*Line* spacing set to ±*N*. *N*−1 *V*s *(blank lines)* are appended to each output text line. Appended blank lines are omitted, if the text or previous appended blank line reached a trap position.
.sp *N*	-	*N*=1*V*	B,v	Space vertically in *either* direction. If *N* is negative, the motion is *backward* (upward) and is limited to the distance to the top of the page. Forward (downward) motion is truncated to the distance to the nearest trap. If the no-space mode is on, no spacing occurs (see **ns**, and **rs** below).
.sv *N*	-	*N*=1*V*	v	Save a contiguous vertical block of size *N*. If the distance to the next trap is greater than *N*, *N* vertical space is output. No-space mode has *no* effect. If this distance is less

				than *N*, no vertical space is immediately output, but *N* is remembered for later output (see **os**). Subsequent **sv** requests will overwrite any still remembered *N*.
.os	-	-	-	Output saved vertical space. No-space mode has *no* effect. Used to finally output a block of vertical space requested by an earlier **sv** request.
.ns	space	-	D	No-space mode turned on. When on, the no-space mode inhibits **sp** requests and **bp** requests *without* a next page number. The no-space mode is turned off when a line of output occurs, or with **rs**.
.rs	space	-	D	Restore spacing. The no-space mode is turned off.
Blank text line.		-	B	Causes a break and output of a blank line exactly like **sp 1**.

6. Line Length and Indenting

The maximum line length for fill mode may be set with **ll**. The indent may be set with **in**; an indent applicable to *only* the *next* output line may be set with **ti**. The line length includes indent space but *not* page offset space. The line-length minus the indent is the basis for centering with **ce**. The effect of **ll**, **in**, or **ti** is delayed, if a partially collected line exists, until after that line is output. In fill mode the length of text on an output line is less than or equal to the line length minus the indent. The current line length and indent are available in registers **.l** and **.i** respectively. The length of *three-part titles* produced by **tl** (see §14) is *independently* set by **lt**.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.ll $\pm N$	6.5 in	previous	E,**m**	Line length is set to $\pm N$. In TROFF the maximum (line-length)+(page-offset) is about 7.54 inches.
.in $\pm N$	N=0	previous	B,E,**m**	Indent is set to $\pm N$. The indent is prepended to each output line.
.ti $\pm N$	-	ignored	B,E,**m**	Temporary indent. The *next* output text line will be indented a distance $\pm N$ with respect to the current indent. The resulting total indent may not be negative. The current indent is not changed.

7. Macros, Strings, Diversion, and Position Traps

7.1. Macros and strings. A *macro* is a named set of arbitrary *lines* that may be invoked by name or with a *trap*. A *string* is a named string of *characters*, *not* including a newline character, that may be interpolated by name at any point. Request, macro, and string names share the *same* name list. Macro and string names may be one or two characters long and may usurp previously defined request, macro, or string names. Any of these entities may be renamed with **rn** or removed with **rm**. Macros are created by **de** and **di**, and appended to by **am** and **da**; **di** and **da** cause normal output to be stored in a macro. Strings are created by **ds** and appended to by **as**. A macro is invoked in the same way as a request; a control line beginning *.xx* will interpolate the contents of macro *xx*. The remainder of the line may contain up to nine *arguments*. The strings *x* and *xx* are interpolated at any desired point with *****x* and ***(***xx* respectively. String references and macro invocations may be nested.

7.2. Copy mode input interpretation. During the definition and extension of strings and macros (not by diversion) the input is read in *copy mode*. The input is copied without interpretation *except* that:

- The contents of number registers indicated by **\n** are interpolated.
- Strings indicated by ***** are interpolated.
- Arguments indicated by **\$** are interpolated.
- Concealed newlines indicated by \\(newline) are eliminated.
- Comments indicated by **\\"** are eliminated.

- **\t** and **\a** are interpreted as ASCII horizontal tab and SOH respectively (§9).
- **** is interpreted as ****.
- **\.** is interpreted as "**.**".

These interpretations can be suppressed by prepending a ****. For example, since **** maps into a ****, **\\n** will copy as **\n** which will be interpreted as a number register indicator when the macro or string is reread.

7.3. Arguments. When a macro is invoked by name, the remainder of the line is taken to contain up to nine arguments. The argument separator is the space character, and arguments may be surrounded by double-quotes to permit imbedded space characters. Pairs of double-quotes may be imbedded in double-quoted arguments to represent a single double-quote. If the desired arguments won't fit on a line, a concealed newline may be used to continue on the next line.

When a macro is invoked the ***input level*** is ***pushed down*** and any arguments available at the previous level become unavailable until the macro is completely read and the previous level is restored. A macro's own arguments can be interpolated at ***any*** point within the macro with **\$***N*, which interpolates the *N*th argument ($1 \leq N \leq 9$). If an invoked argument doesn't exist, a null string results. For example, the macro *xx* may be defined by

```
.de xx      \"begin definition
Today is \\$1 the \\$2.
..          \"end definition
```

and called by

```
.xx Monday 14th
```

to produce the text

Today is Monday the 14th.

Note that the **\$** was concealed in the definition with a prepended ****. The number of currently available arguments is in the **.$** register.

No arguments are available at the top (non-macro) level in this implementation. Because string referencing is implemented as a input-level push down, no arguments are available from *within* a string. No arguments are available within a trap-invoked macro.

Arguments are copied in ***copy mode*** onto a stack where they are available for reference. The mechanism does not allow an argument to contain a direct reference to a *long* string (interpolated at copy time) and it is advisable to conceal string references (with an extra ****) to delay interpolation until argument reference time.

7.4. Diversions. Processed output may be diverted into a macro for purposes such as footnote processing (see Tutorial §T5) or determining the horizontal and vertical size of some text for conditional changing of pages or columns. A single diversion trap may be set at a specified vertical position. The number registers **dn** and **dl** respectively contain the vertical and horizontal size of the most recently ended diversion. Processed text that is diverted into a macro retains the vertical size of each of its lines when reread in *nofill* mode regardless of the current *V*. Constant-spaced (**cs**) or emboldened (**bd**) text that is diverted can be reread correctly only if these modes are again or still in effect at reread time. One way to do this is to imbed in the diversion the appropriate **cs** or **bd** requests with the ***transparent*** mechanism described in §10.6.

Diversions may be nested and certain parameters and registers are associated with the current diversion level (the top non-diversion level may be thought of as the 0th diversion level). These are the diversion trap and associated macro, no-space mode, the internally-saved marked place (see **mk** and **rt**), the current vertical place (**.d** register), the current high-water text base-line (**.h** register), and the current diversion name (**.z** register).

7.5. Traps. Three types of trap mechanisms are available—page traps, a diversion trap, and an input-line-count trap. Macro-invocation traps may be planted using **wh** at any page position including the top. This trap position may be changed using **ch**. Trap positions at or below the bottom of the page have no effect unless or until moved to within the page or rendered effective by an increase in page length. Two traps may be planted at the *same* position only by first planting them at different positions and then moving one of the traps; the first planted trap will conceal the second unless and until the first one is moved (see Tutorial Examples §T5). If the first one is moved back, it again conceals the second trap. The macro associated with a page trap is automatically invoked when a line of text is output whose vertical size *reaches* or *sweeps past* the trap position. Reaching the bottom of a page springs the top-of-page trap, if any, provided there is a next page. The distance to the next trap position is available in the **.t** register; if there are no traps between the current position and the bottom of the page, the distance returned is the distance to the page bottom.

A macro-invocation trap effective in the current diversion may be planted using **dt**. The **.t** register works in a diversion; if there is no subsequent trap a *large* distance is returned. For a description of input-line-count traps, see **it** below.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.de *xx yy*	-	*.yy*=..	-	Define or redefine the macro *xx*. The contents of the macro begin on the next input line. Input lines are copied in *copy mode* until the definition is terminated by a line beginning with *.yy*, whereupon the macro *yy* is called. In the absence of *yy*, the definition is terminated by a line beginning with "..". A macro may contain **de** requests provided the terminating macros differ or the contained definition terminator is concealed. ".." can be concealed as **\\..** which will copy as **\..** and be reread as "..".
.am *xx yy*	-	*.yy*=..	-	Append to macro (append version of **de**).
.ds *xx string*	-	ignored	-	Define a string *xx* containing *string*. Any initial double-quote in *string* is stripped off to permit initial blanks.
.as *xx string*	-	ignored	-	Append *string* to string *xx* (append version of **ds**).
.rm *xx*	-	ignored	-	Remove request, macro, or string. The name *xx* is removed from the name list and any related storage space is freed. Subsequent references will have no effect.
.rn *xx yy*	-	ignored	-	Rename request, macro, or string *xx* to *yy*. If *yy* exists, it is first removed.
.di *xx*	-	end	D	Divert output to macro *xx*. Normal text processing occurs during diversion except that page offsetting is not done. The diversion ends when the request **di** or **da** is encountered without an argument; extraneous requests of this type should not appear when nested diversions are being used.
.da *xx*	-	end	D	Divert, appending to *xx* (append version of **di**).
.wh *N xx*	-	-	v	Install a trap to invoke *xx* at page position *N;* a *negative N* will be interpreted with respect to the page *bottom*. Any macro previously planted at *N* is replaced by *xx*. A zero *N* refers to the *top* of a page. In the absence of *xx*, the first found trap at *N*, if any, is removed.
.ch *xx N*	-	-	v	Change the trap position for macro *xx* to be *N*. In the absence of *N*, the trap, if any, is removed.
.dt *N xx*	-	off	D,v	Install a diversion trap at position *N* in the *current* diversion to invoke macro *xx*. Another **dt** will redefine the diversion trap. If no arguments are given, the diversion trap is removed.

.it *N xx*	-	off	E	Set an input-line-count trap to invoke the macro *xx* after *N* lines of *text* input have been read (control or request lines don't count). The text may be in-line text or text interpolated by inline or trap-invoked macros.
.em *xx*	none	none	-	The macro *xx* will be invoked when all input has ended. The effect is the same as if the contents of *xx* had been at the end of the last file processed.

8. Number Registers

A variety of parameters are available to the user as predefined, named ***number registers*** (see Summary and Index, page 7). In addition, the user may define his own named registers. Register names are one or two characters long and ***do not*** conflict with request, macro, or string names. Except for certain predefined read-only registers, a number register can be read, written, automatically incremented or decremented, and interpolated into the input in a variety of formats. One common use of user-defined registers is to automatically number sections, paragraphs, lines, etc. A number register may be used any time numerical input is expected or desired and may be used in numerical ***expressions*** (§1.4).

Number registers are created and modified using **nr**, which specifies the name, numerical value, and the auto-increment size. Registers are also modified, if accessed with an auto-incrementing sequence. If the registers *x* and *xx* both contain *N* and have the auto-increment size *M*, the following access sequences have the effect shown:

Sequence	Effect on Register	Value Interpolated
n*x*	none	*N*
n(*xx*	none	*N*
n+*x*	*x* incremented by *M*	*N+M*
n−*x*	*x* decremented by *M*	*N-M*
n+(*xx*	*xx* incremented by *M*	*N+M*
n−(*xx*	*xx* decremented by *M*	*N-M*

When interpolated, a number register is converted to decimal (default), decimal with leading zeros, lower-case Roman, upper-case Roman, lower-case sequential alphabetic, or upper-case sequential alphabetic according to the format specified by **af**.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.nr *R* ±*N M*	-	u		The number register *R* is assigned the value ±*N* with respect to the previous value, if any. The increment for auto-incrementing is set to *M*.
.af *R c*	arabic	-	-	Assign format *c* to register *R*. The available formats are:

Format	Numbering Sequence
1	0,1,2,3,4,5,...
001	000,001,002,003,004,005,...
i	0,i,ii,iii,iv,v,...
I	0,I,II,III,IV,V,...
a	0,a,b,c,...,z,aa,ab,...,zz,aaa,...
A	0,A,B,C,...,Z,AA,AB,...,ZZ,AAA,...

An arabic format having *N* digits specifies a field width of *N* digits (example 2 above). The read-only registers and the ***width*** function (§11.2) are always arabic.

.rr *R*	-	ignored	-	Remove register *R*. If many registers are being created dynamically, it may become necessary to remove no longer used registers to recapture internal storage space for newer registers.

9. Tabs, Leaders, and Fields

9.1. Tabs and leaders. The ASCII horizontal tab character and the ASCII SOH (hereafter known as the *leader* character) can both be used to generate either horizontal motion or a string of repeated characters. The length of the generated entity is governed by internal *tab stops* specifiable with **ta**. The default difference is that tabs generate motion and leaders generate a string of periods; **tc** and **lc** offer the choice of repeated character or motion. There are three types of internal tab stops—*left* adjusting, *right* adjusting, and *centering*. In the following table: *D* is the distance from the current position on the *input* line (where a tab or leader was found) to the next tab stop; *next-string* consists of the input characters following the tab (or leader) up to the next tab (or leader) or end of line; and *W* is the width of *next-string*.

Tab type	Length of motion or repeated characters	Location of *next-string*
Left	*D*	Following *D*
Right	*D−W*	Right adjusted within *D*
Centered	*D−W*/2	Centered on right end of *D*

The length of generated motion is allowed to be negative, but that of a repeated character string cannot be. Repeated character strings contain an integer number of characters, and any residual distance is prepended as motion. Tabs or leaders found after the last tab stop are ignored, but may be used as *next-string* terminators.

Tabs and leaders are not interpreted in *copy mode*. **\t** and **\a** always generate a non-interpreted tab and leader respectively, and are equivalent to actual tabs and leaders in *copy mode*.

9.2. Fields. A *field* is contained between a *pair* of *field delimiter* characters, and consists of sub-strings separated by *padding* indicator characters. The field length is the distance on the *input* line from the position where the field begins to the next tab stop. The difference between the total length of all the sub-strings and the field length is incorporated as horizontal padding space that is divided among the indicated padding places. The incorporated padding is allowed to be negative. For example, if the field delimiter is # and the padding indicator is ^, #^*xxx*^*right*# specifies a right-adjusted string with the string *xxx* centered in the remaining space.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.ta *Nt* ...	0.8; 0.5in	none	E,m	Set tab stops and types. *t*=**R**, right adjusting; *t*=**C**, centering; *t* absent, left adjusting. TROFF tab stops are preset every 0.5in.; NROFF every 0.8in. The stop values are separated by spaces, and a value preceded by + is treated as an increment to the previous stop value.
.tc *c*	none	none	E	The tab repetition character becomes *c*, or is removed specifying motion.
.lc *c*	.	none	E	The leader repetition character becomes *c*, or is removed specifying motion.
.fc *a b*	off	off	-	The field delimiter is set to *a*; the padding indicator is set to the *space* character or to *b*, if given. In the absence of arguments the field mechanism is turned off.

10. Input and Output Conventions and Character Translations

10.1. Input character translations. Ways of inputting the graphic character set were discussed in §2.1. The ASCII control characters horizontal tab (§9.1), SOH (§9.1), and backspace (§10.3) are discussed elsewhere.

The newline delimits input lines. In addition, STX, ETX, ENQ, ACK, and BEL are accepted, and may be used as delimiters or translated into a graphic with **tr** (§10.5). *All* others are ignored.

The *escape* character \ introduces *escape sequences*—causes the following character to mean another character, or to indicate some function. A complete list of such sequences is given in the Summary and Index on page 6. \ should not be confused with the ASCII control character ESC of the same name. The escape character \ can be input with the sequence \\. The escape character can be changed with **ec**, and all that has been said about the default \ becomes true for the new escape character. **\e** can be used to print whatever the current escape character is. If necessary or convenient, the escape mechanism may be turned off with **eo**, and restored with **ec**.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.ec *c*	\	\	-	Set escape character to \, or to *c*, if given.
.eo	on	-	-	Turn escape mechanism off.

10.2. Ligatures. Five ligatures are available in the current TROFF character set — **fi**, **fl**, **ff**, **ffi**, and **ffl**. They may be input (even in NROFF) by **\(fi**, **\(fl**, **\(ff**, **\(Fi**, and **\(Fl** respectively. The ligature mode is normally on in TROFF, and *automatically* invokes ligatures during input.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.lg *N*	off; on	on	-	Ligature mode is turned on if *N* is absent or non-zero, and turned off if *N*=0. If *N*=2, only the two-character ligatures are automatically invoked. Ligature mode is inhibited for request, macro, string, register, or file names, and in *copy mode*. No effect in NROFF.

10.3. Backspacing, underlining, overstriking, etc. Unless in *copy mode*, the ASCII backspace character is replaced by a backward horizontal motion having the width of the space character. Underlining as a form of line-drawing is discussed in §12.4. A generalized overstriking function is described in §12.1.

NROFF automatically underlines characters in the *underline* font, specifiable with **uf**, normally that on font position 2 (normally Times Italic, see §2.2). In addition to **ft** and **\f***F*, the underline font may be selected by **ul** and **cu**. Underlining is restricted to an output-device-dependent subset of *reasonable* characters.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.ul *N*	off	*N*=1	E	Underline in NROFF (italicize in TROFF) the next *N* input text lines. Actually, switch to *underline* font, saving the current font for later restoration; *other* font changes within the span of a **ul** will take effect, but the restoration will undo the last change. Output generated by **tl** (§14) *is* affected by the font change, but does *not* decrement *N*. If *N*>1, there is the risk that a trap interpolated macro may provide text lines within the span; environment switching can prevent this.
.cu *N*	off	*N*=1	E	A variant of **ul** that causes *every* character to be underlined in NROFF. Identical to **ul** in TROFF.
.uf *F*	Italic	Italic	-	Underline font set to *F*. In NROFF, *F* may *not* be on position 1 (initially Times Roman).

10.4. Control characters. Both the control character **.** and the *no-break* control character **´** may be changed, if desired. Such a change must be compatible with the design of any macros used in the span of the change, and particularly of any trap-invoked macros.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.cc *c*	.	.	E	The basic control character is set to *c*, or reset to ".".
.c2 *c*	'	'	E	The *nobreak* control character is set to *c*, or reset to "'".

10.5. Output translation. One character can be made a stand-in for another character using **tr**. All text processing (e. g. character comparisons) takes place with the input (stand-in) character which appears to have the width of the final character. The graphic translation occurs at the moment of output (including diversion).

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.tr *abcd....*	none	-	O	Translate *a* into *b*, *c* into *d*, etc. If an odd number of characters is given, the last one will be mapped into the space character. To be consistent, a particular translation must stay in effect from *input* to *output* time.

10.6. Transparent throughput. An input line beginning with a **\!** is read in *copy mode* and *transparently* output (without the initial \!); the text processor is otherwise unaware of the line's presence. This mechanism may be used to pass control information to a post-processor or to imbed control lines in a macro created by a diversion.

10.7. Comments and concealed newlines. An uncomfortably long input line that must stay one line (e. g. a string definition, or nofilled text) can be split into many physical lines by ending all but the last one with the escape ****. The sequence \(newline) is *always* ignored—except in a comment. Comments may be imbedded at the *end* of any line by prefacing them with \". The newline at the end of a comment cannot be concealed. A line beginning with \" will appear as a blank line and behave like **.sp 1**; a comment can be on a line by itself by beginning the line with .\".

11. Local Horizontal and Vertical Motions, and the Width Function

11.1. Local Motions. The functions **\v**'*N*' and **\h**'*N*' can be used for *local* vertical and horizontal motion respectively. The distance *N* may be negative; the *positive* directions are *rightward* and *downward*. A *local* motion is one contained *within* a line. To avoid unexpected vertical dislocations, it is necessary that the *net* vertical local motion within a word in filled text and otherwise within a line balance to zero. The above and certain other escape sequences providing local motion are summarized in the following table.

<table>
<tr><th>Vertical Local Motion</th><th>Effect in TROFF</th><th>NROFF</th><th>Horizontal Local Motion</th><th>Effect in TROFF</th><th>NROFF</th></tr>
<tr><td>\v'N'</td><td colspan="2">Move distance N</td><td>\h'N'
\(space)
\0</td><td colspan="2">Move distance N
Unpaddable space-size space
Digit-size space</td></tr>
<tr><td>\u
\d
\r</td><td>½ em up
½ em down
1 em up</td><td>½ line up
½ line down
1 line up</td><td>\|
\^</td><td>1/6 em space
1/12 em space</td><td>ignored
ignored</td></tr>
</table>

As an example, E^2 could be generated by the sequence **E\s–2\v'–0.4m'2\v'0.4m'\s+2**; it should be noted in this example that the 0.4 em vertical motions are at the smaller size.

11.2. Width Function. The *width* function **\w**'*string*' generates the numerical width of *string* (in basic units). Size and font changes may be safely imbedded in *string*, and will not affect the current environment. For example, **.ti –\w'1. 'u** could be used to temporarily indent leftward a distance equal to the size of the string "**1.** ".

The width function also sets three number registers. The registers **st** and **sb** are set respectively to the highest and lowest extent of *string* relative to the baseline; then, for example, the total *height* of the string is **\n(stu–\n(sbu**. In TROFF the number register **ct** is set to a value between 0 and 3: 0 means that all of the characters in *string* were short lower case characters without descenders (like **e**); 1 means that at least one character has a descender (like **y**); 2 means that at least one character is tall (like **H**); and 3 means that both tall characters and characters with descenders are present.

11.3. Mark horizontal place. The escape sequence **\kx** will cause the *current* horizontal position in the *input line* to be stored in register *x*. As an example, the construction **\kx***word***\h´|\nxu+2u´***word* will embolden *word* by backing up to almost its beginning and overprinting it, resulting in ***word***.

12. Overstrike, Bracket, Line-drawing, and Zero-width Functions

12.1. Overstriking. Automatically centered overstriking of up to nine characters is provided by the *overstrike* function **\o´***string***´**. The characters in *string* overprinted with centers aligned; the total width is that of the widest character. *string* should *not* contain local vertical motion. As examples, **\o´e\´´** produces é, and **\o´\(mo\(sl´** produces ∉.

12.2. Zero-width characters. The function **\z***c* will output *c* without spacing over it, and can be used to produce left-aligned overstruck combinations. As examples, **\z\(ci\(pl** will produce ⊕, and **\(br\z\(rn\(ul\(br** will produce the smallest possible constructed box □.

12.3. Large Brackets. The Special Mathematical Font contains a number of bracket construction pieces (⎧ ⎩ ⎫ ⎭ ⎨ ⎬ | ⌊ ⌋ ⌈ ⌉) that can be combined into various bracket styles. The function **\b´***string***´** may be used to pile up vertically the characters in *string* (the first character on top and the last at the bottom); the characters are vertically separated by 1 em and the total pile is centered 1/2 em above the current baseline (½ line in NROFF). For example, **\b´\(lc\(lf´E\ |\b´\(rc\(rf´\x´−0.5m´\x´0.5m´** produces [E].

12.4. Line drawing. The function **\l´***Nc***´** will draw a string of repeated *c*'s towards the right for a distance *N*. (**\l** is \(lower case L). If *c* looks like a continuation of an expression for *N*, it may insulated from *N* with a **\&**. If *c* is not specified, the _ (baseline rule) is used (underline character in NROFF). If *N* is negative, a backward horizontal motion of size *N* is made *before* drawing the string. Any space resulting from *N*/(size of *c*) having a remainder is put at the beginning (left end) of the string. In the case of characters that are designed to be connected such as baseline-rule _, underrule _, and root-en ‾, the remainder space is covered by over-lapping. If *N* is *less* than the width of *c*, a single *c* is centered on a distance *N*. As an example, a macro to underscore a string can be written

```
.de us
\\$1\l´|0\(ul´
..
```

or one to draw a box around a string

```
.de bx
\(br\ |\\$1\ |\(br\l´|0\(rn\l´|0\(ul´
..
```

such that

```
.ul "underlined words"
```

and

```
.bx "words in a box"
```

yield underlined words and |words in a box|.

The function **\L´***Nc***´** will draw a vertical line consisting of the (optional) character *c* stacked vertically apart 1 em (1 line in NROFF), with the first two characters overlapped, if necessary, to form a continuous line. The default character is the *box rule* | (**\(br**); the other suitable character is the *bold vertical* | (**\(bv**). The line is begun without any initial motion relative to the current base line. A positive *N* specifies a line drawn downward and a negative *N* specifies a line drawn upward. After the line is drawn *no* compensating motions are made; the instantaneous baseline is at the *end* of the line.

The horizontal and vertical line drawing functions may be used in combination to produce large boxes. The zero-width *box-rule* and the ½-em wide *underrule* were *designed* to form corners when using 1-em vertical spacings. For example the macro

```
.de eb
.sp –1        \"compensate for next automatic base-line spacing
.nf           \"avoid possibly overflowing word buffer
\h´–.5n\L´|\\nau–1\l\\n(.lu+1n\(ul\L´–|\\nau+1\l´|0u–.5n\(ul´   \"draw box
.fi
..
```

will draw a box around some text whose beginning vertical place was saved in number register *a* (e. g. using **.mk a**) as done for this paragraph.

13. Hyphenation.

The automatic hyphenation may be switched off and on. When switched on with **hy**, several variants may be set. A *hyphenation indicator* character may be imbedded in a word to specify desired hyphenation points, or may be prepended to suppress hyphenation. In addition, the user may specify a small exception word list.

Only words that consist of a central alphabetic string surrounded by (usually null) non-alphabetic strings are considered candidates for automatic hyphenation. Words that were input containing hyphens (minus), em-dashes (\\(**em**), or hyphenation indicator characters—such as mother-in-law—are *always* subject to splitting after those characters, whether or not automatic hyphenation is on or off.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.nh	hyphenate	-	E	Automatic hyphenation is turned off.
.hy*N*	on,N=1	on,N=1	E	Automatic hyphenation is turned on for $N \geq 1$, or off for $N=0$. If $N=2$, *last* lines (ones that will cause a trap) are not hyphenated. For $N=4$ and 8, the last and first two characters respectively of a word are not split off. These values are additive; i. e. $N=14$ will invoke all three restrictions.
.hc *c*	\%	\%	E	Hyphenation indicator character is set to *c* or to the default \%. The indicator does not appear in the output.
.hw *word1* ...		ignored	-	Specify hyphenation points in words with imbedded minus signs. Versions of a word with terminal *s* are implied; i. e. *dig-it* implies *dig-its*. This list is examined initially *and* after each suffix stripping. The space available is small—about 128 characters.

14. Three Part Titles.

The titling function **tl** provides for automatic placement of three fields at the left, center, and right of a line with a title-length specifiable with **lt**. **tl** may be used anywhere, and is independent of the normal text collecting process. A common use is in header and footer macros.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.tl *'left 'center 'right '*		-	-	The strings *left*, *center*, and *right* are respectively left-adjusted, centered, and right-adjusted in the current title-length. Any of the strings may be empty, and overlapping is permitted. If the page-number character (initially %) is found within any of the fields it is replaced by the current page number having the format assigned to register %. Any character may be used as the string delimiter.

.pc *c*	%	off	-	The page number character is set to *c*, or removed. The page-number register remains **%**.
.lt $\pm N$	6.5 in	previous	E,m	Length of title set to $\pm N$. The line-length and the title-length are *independent.* Indents do not apply to titles; page-offsets do.

15. Output Line Numbering.

Automatic sequence numbering of output lines may be requested with **nm**. When in effect, a three-
digit, arabic number plus a digit-space is prepended to output text lines. The text lines are thus offset
3 by four digit-spaces, and otherwise retain their line length; a reduction in line length may be desired to
keep the right margin aligned with an earlier margin. Blank lines, other vertical spaces, and lines gen-
erated by **tl** are *not* numbered. Numbering can be temporarily suspended with **nn**, or with an **.nm**
6 followed by a later **.nm +0**. In addition, a line number indent *I*, and the number-text separation *S*
may be specified in digit-spaces. Further, it can be specified that only those line numbers that are
multiples of some number *M* are to be printed (the others will appear as blank number fields).

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.nm $\pm N$ *M S I*		off	E	Line number mode. If $\pm N$ is given, line numbering is turned on, and the next output line numbered is numbered $\pm N$. Default values are $M=1$, $S=1$, and $I=0$. Parameters corresponding to missing arguments are unaffected; a non-numeric argument is considered missing. In the absence of all arguments, numbering is turned off; the next line number is preserved for possible further use in number register **ln**.
.nn *N*	-	N=1	E	The next *N* text output lines are not numbered.

9 As an example, the paragraph portions of this section are numbered with $M=3$: **.nm 1 3** was placed
at the beginning; **.nm** was placed at the end of the first paragraph; and **.nm +0** was placed in front
of this paragraph; and **.nm** finally placed at the end. Line lengths were also changed (by **\w´0000´u**)
12 to keep the right side aligned. Another example is **.nm +5 5 x 3** which turns on numbering with the
line number of the next line to be 5 greater than the last numbered line, with $M=5$, with spacing *S*
untouched, and with the indent *I* set to 3.

16. Conditional Acceptance of Input

In the following, *c* is a one-character, built-in *condition* name, **!** signifies *not*, *N* is a numerical expression, *string1* and *string2* are strings delimited by any non-blank, non-numeric character *not* in the strings, and *anything* represents what is conditionally accepted.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.if *c anything*			-	- If condition *c* true, accept *anything* as input; in multi-line case use \\{*anything*\\}.
.if !*c anything*		-	-	If condition *c* false, accept *anything*.
.if *N anything*		-	u	If expression $N > 0$, accept *anything*.
.if !*N anything*		-	u	If expression $N \leq 0$, accept *anything*.
.if *´string1´string2´ anything*			-	If *string1* identical to *string2*, accept *anything*.
.if ! *´string1´string2´ anything*			-	If *string1* not identical to *string2*, accept *anything*.
.ie *c anything*		-	u	If portion of if-else; all above forms (like **if**).
.el *anything*		-	-	Else portion of if-else.

The built-in condition names are:

Condition Name	True If
o	Current page number is odd
e	Current page number is even
t	Formatter is TROFF
n	Formatter is NROFF

If the condition *c* is *true*, or if the number *N* is greater than zero, or if the strings compare identically (including motions and character size and font), *anything* is accepted as input. If a **!** precedes the condition, number, or string comparison, the sense of the acceptance is reversed.

Any spaces between the condition and the beginning of *anything* are skipped over. The *anything* can be either a single input line (text, macro, or whatever) or a number of input lines. In the multi-line case, the first line must begin with a left delimiter \{ and the last line must end with a right delimiter \}.

The request **ie** (if-else) is identical to **if** except that the acceptance state is remembered. A subsequent and matching **el** (else) request then uses the reverse sense of that state. **ie - el** pairs may be nested.

Some examples are:

```
.if e .tl 'Even Page %'''
```

which outputs a title if the page number is even; and

```
.ie \n%>1 \{\
'sp 0.5i
.tl 'Page %'''
'sp |1.2i \}
.el .sp |2.5i
```

which treats page 1 differently from other pages.

17. Environment Switching.

A number of the parameters that control the text processing are gathered together into an *environment*, which can be switched by the user. The environment parameters are those associated with requests noting E in their *Notes* column; in addition, partially collected lines and words are in the environment. Everything else is global; examples are page-oriented parameters, diversion-oriented parameters, number registers, and macro and string definitions. All environments are initialized with default parameter values.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.ev *N*	*N*=0	previous	-	Environment switched to environment $0 \leq N \leq 2$. Switching is done in push-down fashion so that restoring a previous environment *must* be done with **.ev** rather than specific reference.

18. Insertions from the Standard Input

The input can be temporarily switched to the system *standard input* with **rd**, which will switch back when *two* newlines in a row are found (the *extra* blank line is not used). This mechanism is intended for insertions in form-letter-like documentation. On UNIX, the *standard input* can be the user's keyboard, a *pipe*, or a *file*.

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.rd *prompt*	-	*prompt*=BEL	-	Read insertion from the standard input until two newlines in a row are found. If the standard input is the user's keyboard, *prompt* (or a BEL) is written onto the user's

				terminal. **rd** behaves like a macro, and arguments may be placed after *prompt*.
.ex	-	-	-	Exit from NROFF/TROFF. Text processing is terminated exactly as if all input had ended.

If insertions are to be taken from the terminal keyboard *while* output is being printed on the terminal, the command line option **−q** will turn off the echoing of keyboard input and prompt only with BEL. The regular input and insertion input *cannot* simultaneously come from the standard input.

As an example, multiple copies of a form letter may be prepared by entering the insertions for all the copies in one file to be used as the standard input, and causing the file containing the letter to reinvoke itself using **nx** (§19); the process would ultimately be ended by an **ex** in the insertion file.

19. Input/Output File Switching

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.so *filename*		-	-	Switch source file. The top input (file reading) level is switched to *filename*. The effect of an **so** encountered in a macro is not felt until the input level returns to the file level. When the new file ends, input is again taken from the original file. **so**'s may be nested.
.nx *filename*		end-of-file	-	Next file is *filename*. The current file is considered ended, and the input is immediately switched to *filename*.
.pi *program*		-	-	Pipe output to *program* (NROFF only). This request must occur *before* any printing occurs. No arguments are transmitted to *program*.

20. Miscellaneous

Request Form	*Initial Value*	*If No Argument*	*Notes*	*Explanation*
.mc *c N*	-	off	E,m	Specifies that a *margin* character *c* appear a distance *N* to the right of the right margin after each non-empty text line (except those produced by **tl**). If the output line is too-long (as can happen in nofill mode) the character will be appended to the line. If *N* is not given, the previous *N* is used; the initial *N* is 0.2 inches in NROFF and 1 em in TROFF. The margin character used with this paragraph was a 12-point box-rule.
.tm *string*	-	newline	-	After skipping initial blanks, *string* (rest of the line) is read in *copy mode* and written on the user's terminal.
.ig *yy*	-	*.yy=..*	-	Ignore input lines. **ig** behaves exactly like **de** (§7) except that the input is discarded. The input is read in *copy mode*, and any auto-incremented registers will be affected.
.pm *t*	-	all	-	Print macros. The names and sizes of all of the defined macros and strings are printed on the user's terminal; if *t* is given, only the total of the sizes is printed. The sizes is given in *blocks* of 128 characters.
.fl	-	-	B	Flush output buffer. Used in interactive debugging to force output.

21. Output and Error Messages.

The output from **tm**, **pm**, and the prompt from **rd**, as well as various *error* messages are written onto UNIX's *standard message* output. The latter is different from the *standard output*, where NROFF formatted output goes. By default, both are written onto the user's terminal, but they can be independently redirected.

Various *error* conditions may occur during the operation of NROFF and TROFF. Certain less serious errors having only local impact do not cause processing to terminate. Two examples are *word overflow*, caused by a word that is too large to fit into the word buffer (in fill mode), and *line overflow*, caused by an output line that grew too large to fit in the line buffer; in both cases, a message is printed, the offending excess is discarded, and the affected word or line is marked at the point of truncation with a * in NROFF and a ☜ in TROFF. The philosophy is to continue processing, if possible, on the grounds that output useful for debugging may be produced. If a serious error occurs, processing terminates, and an appropriate message is printed. Examples are the inability to create, read, or write files, and the exceeding of certain internal limits that make future output unlikely to be useful.

TUTORIAL EXAMPLES

T1. Introduction

Although NROFF and TROFF have by design a syntax reminiscent of earlier text processors* with the intent of easing their use, it is almost always necessary to prepare at least a small set of macro definitions to describe most documents. Such common formatting needs as page margins and footnotes are deliberately not built into NROFF and TROFF. Instead, the macro and string definition, number register, diversion, environment switching, page-position trap, and conditional input mechanisms provide the basis for user-defined implementations.

The examples to be discussed are intended to be useful and somewhat realistic, but won't necessarily cover all relevant contingencies. Explicit numerical parameters are used in the examples to make them easier to read and to illustrate typical values. In many cases, number registers would really be used to reduce the number of places where numerical information is kept, and to concentrate conditional parameter initialization like that which depends on whether TROFF or NROFF is being used.

T2. Page Margins

As discussed in §3, *header* and *footer* macros are usually defined to describe the top and bottom page margin areas respectively. A trap is planted at page position 0 for the header, and at -*N* (*N* from the page bottom) for the footer. The simplest such definitions might be

```
.de hd          \"define header
'sp 1i
..              \"end definition
.de fo          \"define footer
'bp
..              \"end definition
.wh 0 hd
.wh -1i fo
```

which provide blank 1 inch top and bottom margins. The header will occur on the *first* page, only if the definition and trap exist prior to the initial pseudo-page transition (§3). In fill mode, the output line that springs the footer trap was typically forced out because some part or whole word didn't fit on it. If anything in the footer and header that follows causes a *break*, that word or part word will be forced out. In this and other examples, requests like **bp** and **sp** that normally cause breaks are invoked using the *no-break* control character ´ to avoid this. When the header/footer design contains material requiring independent text processing, the environment may be switched, avoiding most interaction with the running text.

A more realistic example would be

```
.de hd          \"header
.if t .tl ´\(rn´\(rn´ \"troff cut mark
.if \\n%>1 \{\
'sp |0.5i-1     \"tl base at 0.5i
.tl ´´- % -´´   \"centered page number
.ps             \"restore size
.ft             \"restore font
.vs \}          \"restore vs
'sp |1.0i       \"space to 1.0i
.ns             \"turn on no-space mode
..
.de fo          \"footer
.ps 10          \"set footer/header size
.ft R           \"set font
.vs 12p         \"set base-line spacing
.if \\n%=1 \{\
'sp |\\n(.pu-0.5i-1 \"tl base 0.5i up
.tl ´´- % -´´\} \"first page number
'bp
..
.wh 0 hd
.wh -1i fo
```

which sets the size, font, and base-line spacing for the header/footer material, and ultimately restores them. The material in this case is a page number at the bottom of the first page and at the top of the remaining pages. If TROFF is used, a *cut mark* is drawn in the form of *root-en*'s at each margin. The **sp**'s refer to absolute positions to avoid dependence on the base-line spacing. Another reason for this in the footer is that the footer is invoked by printing a line whose vertical spacing swept past the trap position by possibly as much as the base-line spacing. The *no-space* mode is

*For example: P. A. Crisman, Ed., *The Compatible Time-Sharing System,* MIT Press, 1965, Section AH9.01 (Description of RUNOFF program on MIT's CTSS system).

turned on at the end of **hd** to render ineffective accidental occurrences of **sp** at the top of the running text.

The above method of restoring size, font, etc. presupposes that such requests (that set *previous* value) are *not* used in the running text. A better scheme is save and restore both the current *and* previous values as shown for size in the following:

```
.de fo
.nr s1 \\n(.s    \"current size
.ps
.nr s2 \\n(.s    \"previous size
. —              \"rest of footer
..
.de hd
. —              \"header stuff
.ps \\n(s2       \"restore previous size
.ps \\n(s1       \"restore current size
..
```

Page numbers may be printed in the bottom margin by a separate macro triggered during the footer's page ejection:

```
.de bn           \"bottom number
.tl ´´– % –´´    \"centered page number
..
.wh –0.5i–1v bn \"tl base 0.5i up
```

T3. Paragraphs and Headings

The housekeeping associated with starting a new paragraph should be collected in a paragraph macro that, for example, does the desired preparagraph spacing, forces the correct font, size, baseline spacing, and indent, checks that enough space remains for *more than one* line, and requests a temporary indent.

```
.de pg           \"paragraph
.br              \"break
.ft R            \"force font,
.ps 10           \"size,
.vs 12p          \"spacing,
.in 0            \"and indent
.sp 0.4          \"prespace
.ne 1+\\n(.Vu \"want more than 1 line
.ti 0.2i         \"temp indent
..
```

The first break in **pg** will force out any previous partial lines, and must occur before the **vs**. The forcing of font, etc. is partly a defense against prior error and partly to permit things like section heading macros to set parameters only once. The prespacing parameter is suitable for TROFF; a larger space, at least as big as the output device vertical resolution, would be more suitable in NROFF. The choice of remaining space to test for in the **ne** is the smallest amount greater than one line (the **.V** is the available vertical resolution).

A macro to automatically number section headings might look like:

```
.de sc           \"section
. —              \"force font, etc.
.sp 0.4          \"prespace
.ne 2.4+\\n(.Vu \"want 2.4+ lines
.fi
\\n+S.
..
.nr S 0 1        \"init S
```

The usage is **.sc**, followed by the section heading text, followed by **.pg**. The **ne** test value includes one line of heading, 0.4 line in the following **pg**, and one line of the paragraph text. A word consisting of the next section number and a period is produced to begin the heading line. The format of the number may be set by **af** (§8).

Another common form is the labeled, indented paragraph, where the label protrudes left into the indent space.

```
.de lp           \"labeled paragraph
.pg
.in 0.5i         \"paragraph indent
.ta 0.2i 0.5i    \"label, paragraph
.ti 0
\t\\$1\t\c       \"flow into paragraph
..
```

The intended usage is ".lp *label*"; *label* will begin at 0.2 inch, and cannot exceed a length of 0.3 inch without intruding into the paragraph. The label could be right adjusted against 0.4 inch by setting the tabs instead with **.ta 0.4iR 0.5i**. The last line of **lp** ends with **\c** so that it will become a part of the first line of the text that follows.

T4. Multiple Column Output

The production of multiple column pages requires the footer macro to decide whether it was invoked by other than the last column, so that it will begin a new column rather than produce the bottom margin. The header can initialize a column register that the footer will increment and test. The following is arranged for two columns, but is easily modified for more.

```
.de hd           \"header
. ---
.nr cl 0 1       \"init column count
.mk              \"mark top of text
```

```
..
.de fo          \"footer
.ie \\n+(cl<2 \{\
.po +3.4i       \"next column; 3.1+0.3
.rt             \"back to mark
.ns \}          \"no-space mode
.el \{\
.po \\nMu       \"restore left margin
. —
'bp \}
..
.ll 3.1i        \"column width
.nr M \\n(.o    \"save left margin
```

Typically a portion of the top of the first page contains full width text; the request for the narrower line length, as well as another **.mk** would be made where the two column output was to begin.

T5. Footnote Processing

The footnote mechanism to be described is used by imbedding the footnotes in the input text at the point of reference, demarcated by an initial **.fn** and a terminal **.ef**:

> **.fn**
> *Footnote text and control lines...*
> **.ef**

In the following, footnotes are processed in a separate environment and diverted for later printing in the space immediately prior to the bottom margin. There is provision for the case where the last collected footnote doesn't completely fit in the available space.

```
.de hd          \"header
. —
.nr x 0 1       \"init footnote count
.nr y 0-\\nb    \"current footer place
.ch fo -\\nbu   \"reset footer trap
.if \\n(dn .fz  \"leftover footnote
..
.de fo          \"footer
.nr dn 0        \"zero last diversion size
.if \\nx \{\
.ev 1           \"expand footnotes in ev1
.nf             \"retain vertical size
.FN             \"footnotes
.rm FN          \"delete it
.if "\\n(.z"fy" .di \"end overflow diversion
.nr x 0         \"disable fx
.ev \}          \"pop environment
. —
'bp
..
.de fx          \"process footnote overflow
.if \\nx .di fy \"divert overflow
..
.de fn          \"start footnote
.da FN          \"divert (append) footnote
.ev 1           \"in environment 1
.if \\n+x=1 .fs \"if first, include separator
.fi             \"fill mode
..
.de ef          \"end footnote
.br             \"finish output
.nr z \\n(.v    \"save spacing
.ev             \"pop ev
.di             \"end diversion
.nr y -\\n(dn   \"new footer position,
.if \\nx=1 .nr y -(\\n(.v-\\nz) \
                \"uncertainty correction
.ch fo \\nyu    \"y is negative
.if (\\n(nl+1v)>(\\n(.p+\\ny) \
.ch fo \\n(nlu+1v \"it didn't fit
..
.de fs          \"separator
\l'1i'          \"1 inch rule
.br
..
.de fz          \"get leftover footnote
.fn
.nf             \"retain vertical size
.fy             \"where fx put it
.ef
..
.nr b 1.0i      \"bottom margin size
.wh 0 hd        \"header trap
.wh 12i fo      \"footer trap, temp position
.wh -\\nbu fx   \"fx at footer position
.ch fo -\\nbu   \"conceal fx with fo
```

The header **hd** initializes a footnote count register **x**, and sets both the current footer trap position register **y** and the footer trap itself to a nominal position specified in register **b**. In addition, if the register **dn** indicates a leftover footnote, **fz** is invoked to reprocess it. The footnote start macro **fn** begins a diversion (append) in environment 1, and increments the count **x**; if the count is one, the footnote separator **fs** is interpolated. The separator is kept in a separate macro to permit user redefinition. The footnote end macro **ef** restores the previous environment and ends the diversion after saving the spacing size in register **z**. **y** is then decremented by the size of the footnote, available in **dn**; then on the first footnote, **y** is further decremented by the difference in vertical base-line spacings of the two environments, to prevent the late triggering the footer trap from causing the last line of the combined footnotes to overflow. The footer trap is then set to the lower (on the page) of **y** or the current page position (**nl**) plus one line, to allow for printing the reference line. If indicated by **x**, the footer **fo** rereads the footnotes from **FN** in nofill mode in environment 1, and deletes **FN**. If the footnotes were too large to fit, the macro **fx** will be trap-invoked to redivert the overflow into **fy**, and the register **dn** will later indicate to the header whether **fy** is empty. Both **fo** and **fx** are planted in the nominal footer trap position in an order that causes **fx** to be concealed unless the **fo** trap is moved. The footer then terminates the overflow diversion, if necessary, and zeros **x** to disable **fx**, because the uncertainty correction together with a not-too-late triggering of the footer can result in the footnote rereading finishing before reaching the **fx** trap.

A good exercise for the student is to combine the multiple-column and footnote mechanisms.

T6. The Last Page

After the last input file has ended, NROFF and TROFF invoke the *end macro* (§7), if any, and when it finishes, eject the remainder of the page. During the eject, any traps encountered are processed normally. At the *end* of this last page, processing terminates *unless* a partial line, word, or partial word remains. If it is desired that another page be started, the end-macro

```
.de en          \"end-macro
\c
'bp
..
.em en
```

will deposit a null partial word, and effect another last page.

Table I

Font Style Examples

The following fonts are printed in 12-point, with a vertical spacing of 14-point, and with non-alphanumeric characters separated by ¼ em space. The Special Mathematical Font was specially prepared for Bell Laboratories by Graphic Systems, Inc. of Hudson, New Hampshire. The Times Roman, Italic, and Bold are among the many standard fonts available from that company.

Times Roman

abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ
1234567890
! $ % & ( ) ‘ ’ * + - . , / : ; = ? [ ] |
• □ — - _ ¼ ½ ¾ ﬁ ﬂ ﬀ ﬃ ﬄ ° † ′ ¢ ® ©

Times Italic

abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ
1234567890
*! $ % & ( ) ‘ ’ * + - . , / : ; = ? [ ] |*
• □ — - _ ¼ ½ ¾ ﬁ ﬂ ﬀ ﬃ ﬄ ° † ′ ¢ ® ©

Times Bold

abcdefghijklmnopqrstuvwxyz
ABCDEFGHIJKLMNOPQRSTUVWXYZ
1234567890
! $ % & ( ) ‘ ’ * + - . , / : ; = ? [ ] |
• ■ — - _ ¼ ½ ¾ ﬁ ﬂ ﬀ ﬃ ﬄ ° † ′ ¢ ® ©

Special Mathematical Font

" ´ \ ^ _ ` ~ / < > { } # @ + − = ∗
α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ ς τ υ φ χ ψ ω
Γ Δ Θ Λ Ξ Π Σ Υ Φ Ψ Ω
√ ‾ ≥ ≤ ≡ ∼ ≃ ≠ → ← ↑ ↓ × ÷ ± ∪ ∩ ⊂ ⊃ ⊆ ⊇ ∞ ∂
§ ∇ ¬ ∫ ∝ ∅ ∈ ‡ ☞ ☜ ∅ | ○ ⎧ ⎩ ⎫ ⎭ ⎨ ⎬ | ⌊ ⌋ ⌈ ⌉ |

Table II

Input Naming Conventions for ´, `,and − and for Non-ASCII Special Characters

Non-ASCII characters and *minus* on the standard fonts.

Char	Input Name	Character Name	Char	Input Name	Character Name
’	´	close quote	ﬁ	\(fi	ﬁ
‘	`	open quote	ﬂ	\(fl	ﬂ
—	\(em	3/4 Em dash	ﬀ	\(ff	ﬀ
-	–	hyphen or	ﬃ	\(Fi	ﬃ
-	\(hy	hyphen	ﬄ	\(Fl	ﬄ
−	\-	current font minus	°	\(de	degree
•	\(bu	bullet	†	\(dg	dagger
□	\(sq	square	′	\(fm	foot mark
_	\(ru	rule	¢	\(ct	cent sign
¼	\(14	1/4	®	\(rg	registered
½	\(12	1/2	©	\(co	copyright
¾	\(34	3/4			

Non-ASCII characters and ´, `, _, +, −, =, and * on the special font.

The ASCII characters @, #, ", ´, `, <, >, \, {, }, ˜, ˆ, and _ exist *only* on the special font and are printed as a 1-em space if that font is not mounted. The following characters exist only on the special font except for the upper case Greek letter names followed by † which are mapped into upper case English letters in whatever font is mounted on font position one (default Times Roman). The special math plus, minus, and equals are provided to insulate the appearance of equations from the choice of standard fonts.

Char	Input Name	Character Name	Char	Input Name	Character Name
+	\(pl	math plus	λ	\(*l	lambda
−	\(mi	math minus	μ	\(*m	mu
=	\(eq	math equals	ν	\(*n	nu
∗	\(**	math star	ξ	\(*c	xi
§	\(sc	section	ο	\(*o	omicron
´	\(aa	acute accent	π	\(*p	pi
`	\(ga	grave accent	ρ	\(*r	rho
_	\(ul	underrule	σ	\(*s	sigma
/	\(sl	slash (matching backslash)	ς	\(ts	terminal sigma
α	\(*a	alpha	τ	\(*t	tau
β	\(*b	beta	υ	\(*u	upsilon
γ	\(*g	gamma	φ	\(*f	phi
δ	\(*d	delta	χ	\(*x	chi
ε	\(*e	epsilon	ψ	\(*q	psi
ζ	\(*z	zeta	ω	\(*w	omega
η	\(*y	eta	Α	\(*A	Alpha†
θ	\(*h	theta	Β	\(*B	Beta†
ι	\(*i	iota	Γ	\(*G	Gamma
κ	\(*k	kappa	Δ	\(*D	Delta

Char	Input Name	Character Name
Ε	\(*E	Epsilon†
Ζ	\(*Z	Zeta†
Η	\(*Y	Eta†
Θ	\(*H	Theta
Ι	\(*I	Iota†
Κ	\(*K	Kappa†
Λ	\(*L	Lambda
Μ	\(*M	Mu†
Ν	\(*N	Nu†
Ξ	\(*C	Xi
Ο	\(*O	Omicron†
Π	\(*P	Pi
Ρ	\(*R	Rho†
Σ	\(*S	Sigma
Τ	\(*T	Tau†
Υ	\(*U	Upsilon
Φ	\(*F	Phi
Χ	\(*X	Chi†
Ψ	\(*Q	Psi
Ω	\(*W	Omega
√	\(sr	square root
‾	\(rn	root en extender
≥	\(>=	>=
≤	\(<=	<=
≡	\(==	identically equal
≃	\(~=	approx =
∼	\(ap	approximates
≠	\(!=	not equal
→	\(->	right arrow
←	\(<-	left arrow
↑	\(ua	up arrow
↓	\(da	down arrow
×	\(mu	multiply
÷	\(di	divide
±	\(+-	plus-minus
∪	\(cu	cup (union)
∩	\(ca	cap (intersection)
⊂	\(sb	subset of
⊃	\(sp	superset of
⊆	\(ib	improper subset
⊇	\(ip	improper superset
∞	\(if	infinity
∂	\(pd	partial derivative
∇	\(gr	gradient
¬	\(no	not
∫	\(is	integral sign
∝	\(pt	proportional to
∅	\(es	empty set
∈	\(mo	member of
│	\(br	box vertical rule
‡	\(dd	double dagger

Char	Input Name	Character Name
☞	\(rh	right hand
☜	\(lh	left hand
∅	\(bs	Bell System logo
∣	\(or	or
○	\(ci	circle
⎧	\(lt	left top of big curly bracket
⎩	\(lb	left bottom
⎫	\(rt	right top
⎭	\(rb	right bot
⎨	\(lk	left center of big curly bracket
⎬	\(rk	right center of big curly bracket
∣	\(bv	bold vertical
⌊	\(lf	left floor (left bottom of big square bracket)
⌋	\(rf	right floor (right bottom)
⌈	\(lc	left ceiling (left top)
⌉	\(rc	right ceiling (right top)

Summary of Changes to N/TROFF Since October 1976 Manual

Options

-h (Nroff only) Output tabs used during horizontal spacing to speed output as well as reduce output byte count. Device tab settings assumed to be every 8 nominal character widths. The default settings of input (logical) tabs is also initialized to every 8 nominal character widths.

-z Efficiently suppresses formatted output. Only message output will occur (from "tm"s and diagnostics).

Old Requests

.ad c The adjustment type indicator "c" may now also be a number previously obtained from the ".j" register (see below).

.so name The contents of file "name" will be interpolated at the point the "so" is encountered. Previously, the interpolation was done upon return to the file-reading input level.

New Request

.ab text Prints "text" on the message output and terminates without further processing. If "text" is missing, "User Abort." is printed. Does not cause a break. The output buffer is flushed.

.fz F N forces font "F" to be in size N. N may have the form N, +N, or -N. For example,

    .fz 3 -2

will cause an implicit \s-2 every time font 3 is entered, and a corresponding \s+2 when it is left. Special font characters occurring during the reign of font F will have the same size modification. If special characters are to be treated differently,

    .fz S F N

may be used to specify the size treatment of special characters during font F. For example,

    .fz 3 -3
    .fz S 3 -0

will cause automatic reduction of font 3 by 3 points while the special characters would not be affected. Any ".fp" request specifying a font on some position must precede ".fz" requests relating to that position.

New Predefined Number Registers.

.k Read-only. Contains the horizontal size of the text portion (without indent) of the current partially collected output line, if any, in the current environment.

.j Read-only. A number representing the current adjustment mode and type. Can be saved and later given to the "ad" request to restore a previous mode.

.P Read-only. 1 if the current page is being printed, and zero otherwise.

.L Read-only. Contains the current line-spacing parameter ("ls").

c. General register access to the input line-number in the current input file. Contains the same value as the read-only ".c" register.

A TROFF Tutorial

Brian W. Kernighan

Typesetting
Text formatting
NROFF

ABSTRACT

troff is a text-formatting program for driving the Graphic Systems phototypesetter on the UNIX† and GCOS operating systems. This device is capable of producing high quality text; this paper is an example of **troff** output.

The phototypesetter itself normally runs with four fonts, containing roman, italic and bold letters (as on this page), a full greek alphabet, and a substantial number of special characters and mathematical symbols. Characters can be printed in a range of sizes, and placed anywhere on the page.

troff allows the user full control over fonts, sizes, and character positions, as well as the usual features of a formatter — right-margin justification, automatic hyphenation, page titling and numbering, and so on. It also provides macros, arithmetic variables and operations, and conditional testing, for complicated formatting tasks.

This document is an introduction to the most basic use of **troff**. It presents just enough information to enable the user to do simple formatting tasks like making viewgraphs, and to make incremental changes to existing packages of **troff** commands. In most respects, the UNIX formatter **nroff** is identical to **troff**, so this document also serves as a tutorial on **nroff**.

September 17, 1986

† UNIX is a trademark of Bell Laboratories.

A TROFF Tutorial

Brian W. Kernighan

Typesetting
Text formatting
NROFF

1. Introduction

troff [1] is a text-formatting program, written by J. F. Ossanna, for producing high-quality printed output from the phototypesetter on the UNIX and GCOS operating systems. This document is an example of **troff** output.

The single most important rule of using **troff** is not to use it directly, but through some intermediary. In many ways, **troff** resembles an assembly language — a remarkably powerful and flexible one — but nonetheless such that many operations must be specified at a level of detail and in a form that is too hard for most people to use effectively.

For two special applications, there are programs that provide an interface to **troff** for the majority of users. **eqn** [2] provides an easy to learn language for typesetting mathematics; the **eqn** user need know no **troff** whatsoever to typeset mathematics. **tbl** [3] provides the same convenience for producing tables of arbitrary complexity.

For producing straight text (which may well contain mathematics or tables), there are a number of 'macro packages' that define formatting rules and operations for specific styles of documents, and reduce the amount of direct contact with **troff**. In particular, the '–ms' [4] and PWB/MM [5] packages for Bell Labs internal memoranda and external papers provide most of the facilities needed for a wide range of document preparation. (This memo was prepared with '–ms'.) There are also packages for viewgraphs, for simulating the older **roff** formatters on UNIX and GCOS, and for other special applications. Typically you will find these packages easier to use than **troff** once you get beyond the most trivial operations; you should always consider them first.

In the few cases where existing packages don't do the whole job, the solution is ***not*** to write an entirely new set of **troff** instructions from scratch, but to make small changes to adapt packages that already exist.

In accordance with this philosophy of letting someone else do the work, the part of **troff** described here is only a small part of the whole, although it tries to concentrate on the more useful parts. In any case, there is no attempt to be complete. Rather, the emphasis is on showing how to do simple things, and how to make incremental changes to what already exists. The contents of the remaining sections are:

The **troff** described here is the C-language version running on UNIX at Murray Hill, as documented in [1].

To use **troff** you have to prepare not only the actual text you want printed, but some information that tells ***how*** you want it printed. (Readers who use **roff** will find the approach familiar.) For **troff** the text and the formatting information are often intertwined quite intimately. Most commands to **troff** are placed on a line separate from the text itself, beginning with a period (one command per line). For example,

```
Some text.
.ps 14
Some more text.
```

will change the 'point size', that is, the size of the letters being printed, to '14 point' (one point is 1/72 inch) like this:

Some text. Some more text.

Occasionally, though, something special occurs in the middle of a line — to produce

Area = πr^2

you have to type

Area = \(*p\fIr\fR\ /\s8\u2\d\s0

(which we will explain shortly). The backslash character \ is used to introduce **troff** commands and special characters within a line of text.

2. Point Sizes; Line Spacing

As mentioned above, the command .ps sets the point size. One point is 1/72 inch, so 6-point characters are at most 1/12 inch high, and 36-point characters are ½ inch. There are 15 point sizes, listed below.

6 point: Pack my box with five dozen liquor jugs.
7 point: Pack my box with five dozen liquor jugs.
8 point: Pack my box with five dozen liquor jugs.
9 point: Pack my box with five dozen liquor jugs.
10 point: Pack my box with five dozen liquor
11 point: Pack my box with five dozen
12 point: Pack my box with five dozen
14 point: Pack my box with five

16 point 18 point 20 point

22 24 28 36

If the number after .ps is not one of these legal sizes, it is rounded up to the next valid value, with a maximum of 36. If no number follows .ps, **troff** reverts to the previous size, whatever it was. **troff** begins with point size 10, which is usually fine. This document is in 9 point.

The point size can also be changed in the middle of a line or even a word with the in-line command \s. To produce

UNIX runs on a PDP-11/45

type

\s8UNIX\s10 runs on a \s8PDP-\s1011/45

As above, \s should be followed by a legal point size, except that \s0 causes the size to revert to its previous value. Notice that \s1011 can be understood correctly as 'size 10, followed by an 11', if the size is legal, but not otherwise. Be cautious with similar constructions.

Relative size changes are also legal and useful:

\s-2UNIX\s+2

temporarily decreases the size, whatever it is, by two points, then restores it. Relative size changes have the advantage that the size difference is independent of the starting size of the document. The amount of the relative change is restricted to a single digit.

The other parameter that determines what the type looks like is the spacing between lines, which is set independently of the point size. Vertical spacing is measured from the bottom of one line to the bottom of the next. The command to control vertical spacing is .vs. For running text, it is usually best to set the vertical spacing about 20% bigger than the character size. For example, so far in this document, we have used "9 on 11", that is,

```
.ps 9
.vs 11p
```

If we changed to

```
.ps 9
.vs 9p
```

the running text would look like this. After a few lines, you will agree it looks a little cramped. The right vertical spacing is partly a matter of taste, depending on how much text you want to squeeze into a given space, and partly a matter of traditional printing style. By default, **troff** uses 10 on 12.

Point size and vertical spacing make a substantial difference in the amount of text per square inch. This is 12 on 14.

Point size and vertical spacing make a substantial difference in the amount of text per square inch. For example, 10 on 12 uses about twice as much space as 7 on 8. This is 6 on 7, which is even smaller. It packs a lot more words per line, but you can go blind trying to read it.

When used without arguments, .ps and .vs revert to the previous size and vertical spacing respectively.

The command .sp is used to get extra vertical space. Unadorned, it gives you one extra blank line (one .vs, whatever that has been set to). Typically, that's more or less than you want, so .sp can be followed by information about how much space you want —

```
.sp 2i
```

means 'two inches of vertical space'.

```
.sp 2p
```

means 'two points of vertical space'; and

.sp 2

means 'two vertical spaces' — two of whatever .vs is set to (this can also be made explicit with .sp 2v); **troff** also understands decimal fractions in most places, so

.sp 1.5i

is a space of 1.5 inches. These same scale factors can be used after .vs to define line spacing, and in fact after most commands that deal with physical dimensions.

It should be noted that all size numbers are converted internally to 'machine units', which are 1/432 inch (1/6 point). For most purposes, this is enough resolution that you don't have to worry about the accuracy of the representation. The situation is not quite so good vertically, where resolution is 1/144 inch (1/2 point).

3. Fonts and Special Characters

troff and the typesetter allow four different fonts at any one time. Normally three fonts (Times roman, italic and bold) and one collection of special characters are permanently mounted.

abcdefghijklmnopqrstuvwxyz 0123456789
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz 0123456789
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz 0123456789
ABCDEFGHIJKLMNOPQRSTUVWXYZ

The greek, mathematical symbols and miscellany of the special font are listed in Appendix A.

troff prints in roman unless told otherwise. To switch into bold, use the .ft command

.ft B

and for italics,

.ft I

To return to roman, use .ft R; to return to the previous font, whatever it was, use either .ft P or just .ft. The 'underline' command

.ul

causes the next input line to print in italics. .ul can be followed by a count to indicate that more than one line is to be italicized.

Fonts can also be changed within a line or word with the in-line command \f:

bold*face* text

is produced by

\fBbold\fIface\fR text

If you want to do this so the previous font, whatever it was, is left undisturbed, insert extra \fP commands, like this:

\fBbold\fP\fIface\fP\fR text\fP

Because only the immediately previous font is remembered, you have to restore the previous font after each change or you can lose it. The same is true of .ps and .vs when used without an argument.

There are other fonts available besides the standard set, although you can still use only four at any given time. The command .fp tells **troff** what fonts are physically mounted on the typesetter:

.fp 3 H

says that the Helvetica font is mounted on position 3. (For a complete list of fonts and what they look like, see the **troff** manual.) Appropriate .fp commands should appear at the beginning of your document if you do not use the standard fonts.

It is possible to make a document relatively independent of the actual fonts used to print it by using font numbers instead of names; for example, \f3 and .ft˜3 mean 'whatever font is mounted at position 3', and thus work for any setting. Normal settings are roman font on 1, italic on 2, bold on 3, and special on 4.

There is also a way to get 'synthetic' bold fonts by overstriking letters with a slight offset. Look at the .bd command in [1].

Special characters have four-character names beginning with \(, and they may be inserted anywhere. For example,

¼ + ½ = ¾

is produced by

\(14 + \(12 = \(34

In particular, greek letters are all of the form \(*–, where – is an upper or lower case roman letter reminiscent of the greek. Thus to get

$\Sigma(\alpha\times\beta) \rightarrow \infty$

in bare **troff** we have to type

\(*S(\(*a\(mu\(*b) \(-> \(if

That line is unscrambled as follows:

\(*S	Σ
(	(
\(*a	α
\(mu	×
\(*b	β
)	)
\(->	→
\(if	∞

A complete list of these special names occurs in Appendix A.

In **eqn** [2] the same effect can be achieved with the input

SIGMA (alpha times beta) -> inf

which is less concise, but clearer to the uninitiated.

Notice that each four-character name is a single character as far as **troff** is concerned — the 'translate' command

.tr \(mi\(em

is perfectly clear, meaning

.tr -—

that is, to translate - into —.

Some characters are automatically translated into others: grave ` and acute ´ accents (apostrophes) become open and close single quotes '‘’'; the combination of "..." is generally preferable to the double quotes "...". Similarly a typed minus sign becomes a hyphen -. To print an explicit − sign, use \-. To get a backslash printed, use \e.

4. Indents and Line Lengths

troff starts with a line length of 6.5 inches, too wide for 8½×11 paper. To reset the line length, use the .ll command, as in

.ll 6i

As with .sp, the actual length can be specified in several ways; inches are probably the most intuitive.

The maximum line length provided by the typesetter is 7.5 inches, by the way. To use the full width, you will have to reset the default physical left margin ("page offset"), which is normally slightly less than one inch from the left edge of the paper. This is done by the .po command.

.po 0

sets the offset as far to the left as it will go.

The indent command .in causes the left margin to be indented by some specified amount from the page offset. If we use .in to move the left margin in, and .ll to move the right margin to the left, we can make offset blocks of text:

.in 0.3i
.ll −0.3i
text to be set into a block
.ll +0.3i
.in −0.3i

will create a block that looks like this:

> Pater noster qui est in caelis sanctificetur nomen tuum; adveniat regnum tuum; fiat voluntas tua, sicut in caelo, et in terra. ... Amen.

Notice the use of '+' and '−' to specify the amount of change. These change the previous setting by the specified amount, rather than just overriding it. The distinction is quite important: .ll +1i makes lines one inch longer; .ll 1i makes them one inch *long*.

With .in, .ll and .po, the previous value is used if no argument is specified.

To indent a single line, use the 'temporary indent' command .ti. For example, all paragraphs in this memo effectively begin with the command

.ti 3

Three of what? The default unit for .ti, as for most horizontally oriented commands (.ll, .in, .po), is ems; an em is roughly the width of the letter 'm' in the current point size. (Precisely, a em in size *p* is *p* points.) Although inches are usually clearer than ems to people who don't set type for a living, ems have a place: they are a measure of size that is proportional to the current point size. If you want to make text that keeps its proportions regardless of point size, you should use ems for all dimensions. Ems can be specified as scale factors directly, as in .ti 2.5m.

Lines can also be indented negatively if the indent is already positive:

.ti −0.3i

causes the next line to be moved back three tenths of an inch. Thus to make a decorative initial capital, we indent the whole paragraph, then move the letter 'P' back with a .ti command:

> Pater noster qui est in caelis sanctificetur nomen tuum; adveniat regnum tuum; fiat voluntas tua, sicut in caelo, et in terra. ... Amen.

Of course, there is also some trickery to make the 'P' bigger (just a '\s36P\s0'), and to move it down from its normal position (see the section on local motions).

5. Tabs

Tabs (the ASCII 'horizontal tab' character) can be used to produce output in columns, or to set the horizontal position of output. Typically tabs are used only in unfilled text. Tab stops are set by default every half inch from the current indent, but can be changed by the .ta command. To set stops every inch, for example,

.ta 1i 2i 3i 4i 5i 6i

Unfortunately the stops are left-justified only (as on a typewriter), so lining up columns of right-justified numbers can be painful. If you have many numbers, or if you need more complicated table layout, *don't* use **troff** directly; use the **tbl** program described in [3].

For a handful of numeric columns, you can do it this way: Precede every number by enough blanks to make it line up when typed.

```
.nf
.ta 1i 2i 3i
  1 tab   2 tab   3
 40 tab  50 tab  60
700 tab 800 tab 900
.fi
```

Then change each leading blank into the string **\0**. This is a character that does not print, but that has the same width as a digit. When printed, this will produce

1	2	3
40	50	60
700	800	900

It is also possible to fill up tabbed-over space with some character other than blanks by setting the 'tab replacement character' with the .tc command:

```
.ta 1.5i 2.5i
.tc \(ru      (\(ru is "_")
Name tab Age tab
```

produces

Name ________________ Age __________

To reset the tab replacement character to a blank, use .tc with no argument. (Lines can also be drawn with the \l command, described in Section 6.)

troff also provides a very general mechanism called 'fields' for setting up complicated columns. (This is used by **tbl**). We will not go into it in this paper.

6. Local Motions: Drawing lines and characters

Remember 'Area = πr^2' and the big 'P' in the Paternoster. How are they done? **troff** provides a host of commands for placing characters of any size at any place. You can use them to draw special characters or to tune your output for a particular appearance. Most of these commands are straightforward, but messy to read and tough to type correctly.

If you won't use **eqn**, subscripts and superscripts are most easily done with the half-line local motions **\u** and **\d**. To go back up the page half a point-size, insert a **\u** at the desired place; to go down, insert a **\d**. (**\u** and **\d** should always be used in pairs, as explained below.) Thus

Area = \(*pr\u2\d

produces

Area = πr^2

To make the '2' smaller, bracket it with \s–2...\s0. Since **\u** and **\d** refer to the current point size, be sure to put them either both inside or both outside the size changes, or you will get an unbalanced vertical motion.

Sometimes the space given by **\u** and **\d** isn't the right amount. The **\v** command can be used to request an arbitrary amount of vertical motion. The in-line command

\v'(amount)'

causes motion up or down the page by the amount specified in '(amount)'. For example, to move the 'P' down, we used

```
.in +0.6i          (move paragraph in)
.ll -0.3i          (shorten lines)
.ti -0.3i          (move P back)
\v'2'\s36P\s0\v'-2'ater noster qui est
in caelis ...
```

A minus sign causes upward motion, while no sign or a plus sign means down the page. Thus **\v'–2'** causes an upward vertical motion of two line spaces.

There are many other ways to specify the amount of motion —

\v'0.1i'
\v'3p'
\v'–0.5m'

and so on are all legal. Notice that the scale specifier i or p or m goes inside the quotes. Any character can be used in place of the quotes; this is also true of all other **troff** commands described in this section.

Since **troff** does not take within-the-line vertical motions into account when figuring out where it is on the page, output lines can have unexpected positions if the left and right ends aren't at the same vertical position. Thus **\v**, like **\u** and **\d**, should always balance upward vertical motion in a line with the same amount in the downward direction.

Arbitrary horizontal motions are also available — **\h** is quite analogous to **\v**, except that the default scale factor is ems instead of line spaces. As an example,

    \h´−0.1i´

causes a backwards motion of a tenth of an inch. As a practical matter, consider printing the mathematical symbol '>>'. The default spacing is too wide, so **eqn** replaces this by

    >\h´−0.3m´>

to produce >>.

Frequently **\h** is used with the 'width function' **\w** to generate motions equal to the width of some character string. The construction

    \w´thing´

is a number equal to the width of 'thing' in machine units (1/432 inch). All **troff** computations are ultimately done in these units. To move horizontally the width of an 'x', we can say

    \h´\w´x´u´

As we mentioned above, the default scale factor for all horizontal dimensions is m, ems, so here we must have the u for machine units, or the motion produced will be far too large. **troff** is quite happy with the nested quotes, by the way, so long as you don't leave any out.

As a live example of this kind of construction, all of the command names in the text, like **.sp**, were done by overstriking with a slight offset. The commands for **.sp** are

    .sp\h´−\w´.sp´u´\h´1u´.sp

That is, put out '.sp', move left by the width of '.sp', move right 1 unit, and print '.sp' again. (Of course there is a way to avoid typing that much input for each command name, which we will discuss in Section 11.)

There are also several special-purpose **troff** commands for local motion. We have already seen **\0**, which is an unpaddable white space of the same width as a digit. 'Unpaddable' means that it will never be widened or split across a line by line justification and filling. There is also **\(blank)**, which is an unpaddable character the width of a space, **\|**, which is half that width, **\^**, which is one quarter of the width of a space, and **\&**, which has zero width. (This last one is useful, for example, in entering a text line which would otherwise begin with a '.'.)

The command **\o**, used like

    \o´set of characters´

causes (up to 9) characters to be overstruck, centered on the widest. This is nice for accents, as in

    syst\o"e\(ga"me t\o"e\(aa"l\o"e\(aa"phonique

which makes

    système téléphonique

The accents are **\(ga** and **\(aa**, or **\`** and **\´**; remember that each is just one character to **troff**.

You can make your own overstrikes with another special convention, **\z**, the zero-motion command. **\zx** suppresses the normal horizontal motion after printing the single character **x**, so another character can be laid on top of it. Although sizes can be changed within **\o**, it centers the characters on the widest, and there can be no horizontal or vertical motions, so **\z** may be the only way to get what you want:

is produced by

    .sp 2
    \s8\z\(sq\s14\z\(sq\s22\z\(sq\s36\(sq

The **.sp** is needed to leave room for the result.

As another example, an extra-heavy semicolon that looks like

    ; instead of ; or ;

can be constructed with a big comma and a big period above it:

    \s+6\z,\v´−0.25m´.\v´0.25m´\s0

'0.25m' is an empirical constant.

A more ornate overstrike is given by the bracketing function **\b**, which piles up characters vertically, centered on the current baseline. Thus we can get big brackets, constructing them with piled-up smaller pieces:

$$\left\{ \left[\text{x} \right] \right\}$$

by typing in only this:

    .sp
    \b´\(lt\(lk\(lb´ \b´\(lc\(lf´ x \b´\(rc\(rf´ \b´\(rt\(rk\(rb´

troff also provides a convenient facility for drawing horizontal and vertical lines of arbitrary length with arbitrary characters. **\l′1i′** draws a line one inch long, like this: ________________. The length can be followed by the character to use if the _ isn't appropriate; **\l′**0.5i.′ draws a half-inch line of dots: .... The construction **\L** is entirely analogous, except that it draws a vertical line instead of horizontal.

7. Strings

Obviously if a paper contains a large number of occurrences of an acute accent over a letter 'e', typing **\o"e\'"** for each é would be a great nuisance.

Fortunately, **troff** provides a way in which you can store an arbitrary collection of text in a 'string', and thereafter use the string name as a shorthand for its contents. Strings are one of several **troff** mechanisms whose judicious use lets you type a document with less effort and organize it so that extensive format changes can be made with few editing changes.

A reference to a string is replaced by whatever text the string was defined as. Strings are defined with the command .ds. The line

```
.ds e \o"e\'"
```

defines the string e to have the value \o"e\'"

String names may be either one or two characters long, and are referred to by ***x** for one character names or *****(xy for two character names. Thus to get téléphone, given the definition of the string e as above, we can say t*el*ephone.

If a string must begin with blanks, define it as

```
.ds xx "      text
```

The double quote signals the beginning of the definition. There is no trailing quote; the end of the line terminates the string.

A string may actually be several lines long; if **troff** encounters a \ at the end of *any* line, it is thrown away and the next line added to the current one. So you can make a long string simply by ending each line but the last with a backslash:

```
.ds xx this \
is a very \
long string
```

Strings may be defined in terms of other strings, or even in terms of themselves; we will discuss some of these possibilities later.

8. Introduction to Macros

Before we can go much further in **troff**, we need to learn a bit about the macro facility. In its simplest form, a macro is just a shorthand notation quite similar to a string. Suppose we want every paragraph to start in exactly the same way — with a space and a temporary indent of two ems:

```
.sp
.ti +2m
```

Then to save typing, we would like to collapse these into one shorthand line, a **troff** 'command' like

```
.PP
```

that would be treated by **troff** exactly as

```
.sp
.ti +2m
```

.PP is called a *macro*. The way we tell **troff** what .PP means is to *define* it with the .de command:

```
.de PP
.sp
.ti +2m
..
```

The first line names the macro (we used '.PP' for 'paragraph', and upper case so it wouldn't conflict with any name that **troff** might already know about). The last line .. marks the end of the definition. In between is the text, which is simply inserted whenever **troff** sees the 'command' or macro call

```
.PP
```

A macro can contain any mixture of text and formatting commands.

The definition of .PP has to precede its first use; undefined macros are simply ignored. Names are restricted to one or two characters.

Using macros for commonly occurring sequences of commands is critically important. Not only does it save typing, but it makes later changes much easier. Suppose we decide that the paragraph indent is too small, the vertical space is much too big, and roman font should be forced. Instead of changing the whole document, we need only change the definition of .PP to something like

```
.de PP        \" paragraph macro
.sp 2p
.ti +3m
.ft R
..
```

and the change takes effect everywhere we used .PP.

\" is a **troff** command that causes the rest of the line to be ignored. We use it here to add comments to the macro definition (a wise idea once definitions get complicated).

As another example of macros, consider these two which start and end a block of offset, unfilled text, like most of the examples in this paper:

```
.de BS     \" start indented block
.sp
.nf
.in +0.3i
..
.de BE     \" end indented block
.sp
.fi
.in −0.3i
..
```

Now we can surround text like

```
Copy to
John Doe
Richard Roberts
Stanley Smith
```

by the commands .BS and .BE, and it will come out as it did above. Notice that we indented by .in +0.3i instead of .in 0.3i. This way we can nest our uses of .BS and BE to get blocks within blocks.

If later on we decide that the indent should be 0.5i, then it is only necessary to change the definitions of .BS and .BE, not the whole paper.

9. Titles, Pages and Numbering

This is an area where things get tougher, because nothing is done for you automatically. Of necessity, some of this section is a cookbook, to be copied literally until you get some experience.

Suppose you want a title at the top of each page, saying just

```
~~~~left top        center top        right top~~~~
```

In **roff**, one can say

```
.he 'left top'center top'right top'
.fo 'left bottom'center bottom'right bottom'
```

to get headers and footers automatically on every page. Alas, this doesn't work in **troff**, a serious hardship for the novice. Instead you have to do a lot of specification.

You have to say what the actual title is (easy); when to print it (easy enough); and what to do at and around the title line (harder). Taking these in reverse order, first we define a macro .NP (for 'new page') to process titles and the like at the end of one page and the beginning of the next:

```
.de NP
'bp
'sp 0.5i
.tl 'left top'center top'right top'
'sp 0.3i
..
```

To make sure we're at the top of a page, we issue a 'begin page' command 'bp, which causes a skip to top-of-page (we'll explain the ' shortly). Then we space down half an inch, print the title (the use of .tl should be self explanatory; later we will discuss parameterizing the titles), space another 0.3 inches, and we're done.

To ask for .NP at the bottom of each page, we have to say something like 'when the text is within an inch of the bottom of the page, start the processing for a new page.' This is done with a 'when' command .wh:

```
.wh  −1i  NP
```

(No '.' is used before NP; this is simply the name of a macro, not a macro call.) The minus sign means 'measure up from the bottom of the page', so '−1i' means 'one inch from the bottom'.

The **.wh** command appears in the input outside the definition of .NP; typically the input would be

```
.de NP
...
..
.wh −1i NP
```

Now what happens? As text is actually being output, **troff** keeps track of its vertical position on the page, and after a line is printed within one inch from the bottom, the .NP macro is activated. (In the jargon, the **.wh** command sets a *trap* at the specified place, which is 'sprung' when that point is passed.) .NP causes a skip to the top of the next page (that's what the 'bp was for), then prints the title with the appropriate margins.

Why 'bp and 'sp instead of .bp and .sp? The answer is that .sp and .bp, like several other commands, cause a *break* to take place. That is, all the input text collected but not yet printed is flushed out as soon as possible, and the next input line is guaranteed to start a new line of output. If we had used .sp or .bp in the .NP macro, this would cause a break in the middle of the current output line when a new page is started. The effect would be to print the left-over part of that line at the top of the page, followed by the next input line on a new output line. This is *not* what we want. Using ' instead of . for a command tells **troff** that no break is to take place — the output line currently being filled should *not* be forced out before the space or new page.

The list of commands that cause a break is short and natural:

```
.bp  .br  .ce  .fi  .nf  .sp  .in  .ti
```

All others cause *no* break, regardless of whether you use a . or a ′. If you really need a break, add a .br command at the appropriate place.

One other thing to beware of — if you're changing fonts or point sizes a lot, you may find that if you cross a page boundary in an unexpected font or size, your titles come out in that size and font instead of what you intended. Furthermore, the length of a title is independent of the current line length, so titles will come out at the default length of 6.5 inches unless you change it, which is done with the .lt command.

There are several ways to fix the problems of point sizes and fonts in titles. For the simplest applications, we can change .NP to set the proper size and font for the title, then restore the previous values, like this:

```
.de NP
'bp
'sp 0.5i
.ft R          \" set title font to roman
.ps 10         \" and size to 10 point
.lt 6i         \" and length to 6 inches
.tl 'left'center'right'
.ps            \" revert to previous size
.ft P          \" and to previous font
'sp 0.3i
..
```

This version of .NP does *not* work if the fields in the .tl command contain size or font changes. To cope with that requires **troff**'s 'environment' mechanism, which we will discuss in Section 13.

To get a footer at the bottom of a page, you can modify .NP so it does some processing before the ′bp command, or split the job into a footer macro invoked at the bottom margin and a header macro invoked at the top of the page. These variations are left as exercises.

Output page numbers are computed automatically as each page is produced (starting at 1), but no numbers are printed unless you ask for them explicitly. To get page numbers printed, include the character % in the .tl line at the position where you want the number to appear. For example

```
.tl ''- % -''
```

centers the page number inside hyphens, as on this page. You can set the page number at any time with either .bp **n**, which immediately starts a new page numbered **n**, or with .pn **n**, which sets the page number for the next page but doesn't cause a skip to the new page. Again, .bp +**n** sets the page number to n more than its current value; .bp means .bp +1.

10. Number Registers and Arithmetic

troff has a facility for doing arithmetic, and for defining and using variables with numeric values, called *number registers.* Number registers, like strings and macros, can be useful in setting up a document so it is easy to change later. And of course they serve for any sort of arithmetic computation.

Like strings, number registers have one or two character names. They are set by the .nr command, and are referenced anywhere by **\nx** (one character name) or **\n(xy** (two character name).

There are quite a few pre-defined number registers maintained by **troff**, among them % for the current page number; nl for the current vertical position on the page; **dy**, **mo** and **yr** for the current day, month and year; and .s and .f for the current size and font. (The font is a number from 1 to 4.) Any of these can be used in computations like any other register, but some, like .s and .f, cannot be changed with .nr.

As an example of the use of number registers, in the –ms macro package [4], most significant parameters are defined in terms of the values of a handful of number registers. These include the point size for text, the vertical spacing, and the line and title lengths. To set the point size and vertical spacing for the following paragraphs, for example, a user may say

```
.nr PS 9
.nr VS 11
```

The paragraph macro .PP is defined (roughly) as follows:

```
.de PP
.ps \\n(PS       \" reset size
.vs \\n(VSp      \" spacing
.ft R            \" font
.sp 0.5v         \" half a line
.ti +3m
..
```

This sets the font to Roman and the point size and line spacing to whatever values are stored in the number registers PS and VS.

Why are there two backslashes? This is the eternal problem of how to quote a quote. When **troff** originally reads the macro definition, it peels off one backslash to see what's coming next. To ensure that another is left in the definition when the macro is *used*, we have to put in two backslashes in the definition. If only one backslash is used, point size and vertical spacing will be frozen at the time the macro is defined, not when it is used.

Protecting by an extra layer of backslashes is only needed for **\n**, *****, **\$** (which we haven't come to yet), and **** itself. Things like **\s**, **\f**, **\h**, **\v**, and so on do not need an extra backslash, since they are converted by **troff** to an internal code immediately upon being seen.

Arithmetic expressions can appear anywhere that a number is expected. As a trivial example,

```
.nr PS \\n(PS-2
```

decrements PS by 2. Expressions can use the arithmetic operators +, −, *, /, % (mod), the relational operators >, >=, <, <=, =, and != (not equal), and parentheses.

Although the arithmetic we have done so far has been straightforward, more complicated things are somewhat tricky. First, number registers hold only integers. **troff** arithmetic uses truncating integer division, just like Fortran. Second, in the absence of parentheses, evaluation is done left-to-right without any operator precedence (including relational operators). Thus

```
7*-4+3/13
```

becomes '−1'. Number registers can occur anywhere in an expression, and so can scale indicators like p, i, m, and so on (but no spaces). Although integer division causes truncation, each number and its scale indicator is converted to machine units (1/432 inch) before any arithmetic is done, so 1i/2u evaluates to 0.5i correctly.

The scale indicator u often has to appear when you wouldn't expect it — in particular, when arithmetic is being done in a context that implies horizontal or vertical dimensions. For example,

```
.ll 7/2i
```

would seem obvious enough — 3½ inches. Sorry. Remember that the default units for horizontal parameters like .ll are ems. That's really '7 ems / 2 inches', and when translated into machine units, it becomes zero. How about

```
.ll 7i/2
```

Sorry, still no good — the '2' is '2 ems', so '7i/2' is small, although not zero. You *must* use

```
.ll 7i/2u
```

So again, a safe rule is to attach a scale indicator to every number, even constants.

For arithmetic done within a .nr command, there is no implication of horizontal or vertical dimension, so the default units are 'units', and 7i/2 and 7i/2u mean the same thing. Thus

```
.nr ll 7i/2
.ll \\n(llu
```

does just what you want, so long as you don't forget the u on the .ll command.

11. Macros with arguments

The next step is to define macros that can change from one use to the next according to parameters supplied as arguments. To make this work, we need two things: first, when we define the macro, we have to indicate that some parts of it will be provided as arguments when the macro is called. Then when the macro is called we have to provide actual arguments to be plugged into the definition.

Let us illustrate by defining a macro .SM that will print its argument two points smaller than the surrounding text. That is, the macro call

```
.SM TROFF
```

will produce TROFF.

The definition of .SM is

```
.de SM
\s-2\\$1\s+2
..
```

Within a macro definition, the symbol **\\$n** refers to the nth argument that the macro was called with. Thus **\\$1** is the string to be placed in a smaller point size when .SM is called.

As a slightly more complicated version, the following definition of .SM permits optional second and third arguments that will be printed in the normal size:

```
.de SM
\\$3\s-2\\$1\s+2\\$2
..
```

Arguments not provided when the macro is called are treated as empty, so

```
.SM TROFF ),
```

produces TROFF), while

```
.SM TROFF ). (
```

produces (TROFF). It is convenient to reverse the order of arguments because trailing punctuation is much more common than leading.

By the way, the number of arguments that a macro was called with is available in number register **.$**.

The following macro .BD is the one used to make the 'bold roman' we have been using for **troff** command names in text. It combines horizontal motions, width computations, and argument rearrangement.

```
.de BD
\&\\$3\f1\\$1\h´-\w´\\$1´u+1u´\\$1\fP\\$2
..
```

The **\h** and **\w** commands need no extra backslash, as we discussed above. The **\&** is there in case the argument begins with a period.

Two backslashes are needed with the **\\$n** commands, though, to protect one of them when the macro is being defined. Perhaps a second example will make this clearer. Consider a macro called .SH which produces section headings rather like those in this paper, with the sections numbered automatically, and the title in bold in a smaller size. The use is

```
.SH "Section title ..."
```

(If the argument to a macro is to contain blanks, then it must be *surrounded* by double quotes, unlike a string, where only one leading quote is permitted.)

Here is the definition of the .SH macro:

```
.nr SH 0        \" initialize section number
.de SH
.sp 0.3i
.ft B
.nr SH \\n(SH+1      \" increment number
.ps \\n(PS−1 \" decrease PS
\\n(SH. \\$1         \" number. title
.ps \\n(PS           \" restore PS
.sp 0.3i
.ft R
..
```

The section number is kept in number register SH, which is incremented each time just before it is used. (A number register may have the same name as a macro without conflict but a string may not.)

We used **\\n(SH** instead of **\n(SH** and **\\n(PS** instead of **\n(PS**. If we had used **\n(SH**, we would get the value of the register at the time the macro was *defined*, not at the time it was *used*. If that's what you want, fine, but not here. Similarly, by using **\\n(PS**, we get the point size at the time the macro is called.

As an example that does not involve numbers, recall our .NP macro which had a

```
.tl ´left´center´right´
```

We could make these into parameters by using instead

```
.tl ´\\*(LT´\\*(CT´\\*(RT´
```

so the title comes from three strings called LT, CT and RT. If these are empty, then the title will be a blank line. Normally CT would be set with something like

```
.ds CT - % -
```

to give just the page number between hyphens (as on the top of this page), but a user could supply private definitions for any of the strings.

12. Conditionals

Suppose we want the .SH macro to leave two extra inches of space just before section 1, but nowhere else. The cleanest way to do that is to test inside the .SH macro whether the section number is 1, and add some space if it is. The .if command provides the conditional test that we can add just before the heading line is output:

```
.if \\n(SH=1 .sp 2i       \" first section only
```

The condition after the .if can be any arithmetic or logical expression. If the condition is logically true, or arithmetically greater than zero, the rest of the line is treated as if it were text — here a command. If the condition is false, or zero or negative, the rest of the line is skipped.

It is possible to do more than one command if a condition is true. Suppose several operations are to be done before section 1. One possibility is to define a macro .S1 and invoke it if we are about to do section 1 (as determined by an .if).

```
.de S1
--- processing for section 1 ---
..
.de SH
...
.if \\n(SH=1 .S1
...
..
```

An alternate way is to use the extended form of the .if, like this:

```
.if \\n(SH=1 \{--- processing
for section 1 ---\}
```

The braces **\{** and **\}** must occur in the positions shown or you will get unexpected extra lines in your output. **troff** also provides an 'if-else' construction, which we will not go into here.

A condition can be negated by preceding it with !; we get the same effect as above (but less clearly) by using

```
.if !\\n(SH>1 .S1
```

There are a handful of other conditions that can be tested with .if. For example, is the current page even or odd?

```
.if e .tl ''even page title''
.if o .tl ''odd page title''
```

gives facing pages different titles when used inside an appropriate new page macro.

Two other conditions are **t** and **n**, which tell you whether the formatter is **troff** or **nroff**.

```
.if t troff stuff ...
.if n nroff stuff ...
```

Finally, string comparisons may be made in an .if:

```
.if 'string1'string2' stuff
```

does 'stuff' if *string1* is the same as *string2*. The character separating the strings can be anything reasonable that is not contained in either string. The strings themselves can reference strings with *****, arguments with **\\$**, and so on.

13. Environments

As we mentioned, there is a potential problem when going across a page boundary: parameters like size and font for a page title may well be different from those in effect in the text when the page boundary occurs. **troff** provides a very general way to deal with this and similar situations. There are three 'environments', each of which has independently settable versions of many of the parameters associated with processing, including size, font, line and title lengths, fill/nofill mode, tab stops, and even partially collected lines. Thus the titling problem may be readily solved by processing the main text in one environment and titles in a separate one with its own suitable parameters.

The command .ev n shifts to environment n; n must be 0, 1 or 2. The command .ev with no argument returns to the previous environment. Environment names are maintained in a stack, so calls for different environments may be nested and unwound consistently.

Suppose we say that the main text is processed in environment 0, which is where **troff** begins by default. Then we can modify the new page macro .NP to process titles in environment 1 like this:

```
.de NP
.ev 1           \" shift to new environment
.lt 6i          \" set parameters here
.ft R
.ps 10
... any other processing ...
.ev             \" return to previous environment
..
```

It is also possible to initialize the parameters for an environment outside the .NP macro, but the version shown keeps all the processing in one place and is thus easier to understand and change.

14. Diversions

There are numerous occasions in page layout when it is necessary to store some text for a period of time without actually printing it. Footnotes are the most obvious example: the text of the footnote usually appears in the input well before the place on the page where it is to be printed is reached. In fact, the place where it is output normally depends on how big it is, which implies that there must be a way to process the footnote at least enough to decide its size without printing it.

troff provides a mechanism called a diversion for doing this processing. Any part of the output may be diverted into a macro instead of being printed, and then at some convenient time the macro may be put back into the input.

The command .di **xy** begins a diversion — all subsequent output is collected into the macro **xy** until the command .di with no arguments is encountered. This terminates the diversion. The processed text is available at any time thereafter, simply by giving the command

```
.xy
```

The vertical size of the last finished diversion is contained in the built-in number register dn.

As a simple example, suppose we want to implement a 'keep-release' operation, so that text between the commands .KS and .KE will not be split across a page boundary (as for a figure or table). Clearly, when a .KS is encountered, we have to begin diverting the output so we can find out how big it is. Then when a .KE is seen, we decide whether the diverted text will fit on the current page, and print it either there if it fits, or at the top of the next page if it doesn't. So:

```
.de KS      \" start keep
.br         \" start fresh line
.ev 1       \" collect in new environment
.fi         \" make it filled text
.di XX      \" collect in XX
..

.de KE      \" end keep
.br         \" get last partial line
.di         \" end diversion
.if \\n(dn>=\\n(.t .bp   \" bp if doesn't fit
.nf         \" bring it back in no-fill
.XX         \" text
.ev         \" return to normal environment
..
```

Recall that number register **nl** is the current position on the output page. Since output was being diverted, this remains at its value when the diversion started. **dn** is the amount of text in the diversion; **.t** (another built-in register) is the distance to the next trap, which we assume is at the bottom margin of the page. If the diversion is large enough to go past the trap, the **.if** is satisfied, and a **.bp** is issued. In either case, the diverted output is then brought back with **.XX**. It is essential to bring it back in no-fill mode so **troff** will do no further processing on it.

This is not the most general keep-release, nor is it robust in the face of all conceivable inputs, but it would require more space than we have here to write it in full generality. This section is not intended to teach everything about diversions, but to sketch out enough that you can read existing macro packages with some comprehension.

Acknowledgements

I am deeply indebted to J. F. Ossanna, the author of **troff**, for his repeated patient explanations of fine points, and for his continuing willingness to adapt **troff** to make other uses easier. I am also grateful to Jim Blinn, Ted Dolotta, Doug McIlroy, Mike Lesk and Joel Sturman for helpful comments on this paper.

References

[1] J. F. Ossanna, *NROFF/TROFF* User's Manual, Bell Laboratories Computing Science Technical Report 54, 1976.

[2] B. W. Kernighan, *A System for Typesetting Mathematics — User's Guide (Second Edition),* Bell Laboratories Computing Science Technical Report 17, 1977.

[3] M. E. Lesk, *TBL — A Program to Format Tables,* Bell Laboratories Computing Science Technical Report 49, 1976.

[4] M. E. Lesk, *Typing Documents on UNIX,* Bell Laboratories, 1978.

[5] J. R. Mashey and D. W. Smith, *PWB/MM — Programmer's Workbench Memorandum Macros,* Bell Laboratories internal memorandum.

Appendix A: Phototypesetter Character Set

These characters exist in roman, italic, and bold. To get the one on the left, type the four-character name on the right.

<table>
<tr><td>ff</td><td>\(ff</td><td>fi</td><td>\(fi</td><td>fl</td><td>\(fl</td><td>ffi</td><td>\(Fi</td><td>ffl</td><td>\(Fl</td></tr>
<tr><td>_</td><td>\(ru</td><td>—</td><td>\(em</td><td>¼</td><td>\(14</td><td>½</td><td>\(12</td><td>¾</td><td>\(34</td></tr>
<tr><td>©</td><td>\(co</td><td>°</td><td>\(de</td><td>†</td><td>\(dg</td><td>′</td><td>\(fm</td><td>¢</td><td>\(ct</td></tr>
<tr><td>®</td><td>\(rg</td><td>•</td><td>\(bu</td><td>□</td><td>\(sq</td><td>-</td><td>\(hy</td><td></td><td></td></tr>
</table>

(In bold, \(sq is ■.)

The following are special-font characters:

<table>
<tr><td>+</td><td>\(pl</td><td>−</td><td>\(mi</td><td>×</td><td>\(mu</td><td>÷</td><td>\(di</td></tr>
<tr><td>=</td><td>\(eq</td><td>≡</td><td>\(==</td><td>≥</td><td>\(>=</td><td>≤</td><td>\(<=</td></tr>
<tr><td>≠</td><td>\(!=</td><td>±</td><td>\(+-</td><td>¬</td><td>\(no</td><td>/</td><td>\(sl</td></tr>
<tr><td>~</td><td>\(ap</td><td>≃</td><td>\(~=</td><td>∝</td><td>\(pt</td><td>∇</td><td>\(gr</td></tr>
<tr><td>→</td><td>\(-></td><td>←</td><td>\(<-</td><td>↑</td><td>\(ua</td><td>↓</td><td>\(da</td></tr>
<tr><td>∫</td><td>\(is</td><td>∂</td><td>\(pd</td><td>∞</td><td>\(if</td><td>√</td><td>\(sr</td></tr>
<tr><td>⊂</td><td>\(sb</td><td>⊃</td><td>\(sp</td><td>∪</td><td>\(cu</td><td>∩</td><td>\(ca</td></tr>
<tr><td>⊆</td><td>\(ib</td><td>⊇</td><td>\(ip</td><td>∈</td><td>\(mo</td><td>∅</td><td>\(es</td></tr>
<tr><td>´</td><td>\(aa</td><td>`</td><td>\(ga</td><td>○</td><td>\(ci</td><td>∅</td><td>\(bs</td></tr>
<tr><td>§</td><td>\(sc</td><td>‡</td><td>\(dd</td><td>☜</td><td>\(lh</td><td>☞</td><td>\(rh</td></tr>
<tr><td>⎧</td><td>\(lt</td><td>⎫</td><td>\(rt</td><td>⌈</td><td>\(lc</td><td>⌉</td><td>\(rc</td></tr>
<tr><td>⎩</td><td>\(lb</td><td>⎭</td><td>\(rb</td><td>⌊</td><td>\(lf</td><td>⌋</td><td>\(rf</td></tr>
<tr><td>⎨</td><td>\(lk</td><td>⎬</td><td>\(rk</td><td>│</td><td>\(bv</td><td>ς</td><td>\(ts</td></tr>
<tr><td>│</td><td>\(br</td><td>|</td><td>\(or</td><td>_</td><td>\(ul</td><td>‾</td><td>\(rn</td></tr>
<tr><td>∗</td><td>\(**</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

These four characters also have two-character names. The ´ is the apostrophe on terminals; the ` is the other quote mark.

<table>
<tr><td>´</td><td>\´</td><td>`</td><td>\`</td><td>−</td><td>\-</td><td>_</td><td>_</td></tr>
</table>

These characters exist only on the special font, but they do not have four-character names:

" { } < > ~ ^ \ # @

For greek, precede the roman letter by \(* to get the corresponding greek; for example, \(*a is α.

a b g d e z y h i k l m n c o p r s t u f x q w
α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ τ υ φ χ ψ ω

A B G D E Z Y H I K L M N C O P R S T U F X Q W
Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω

Refer — A Bibliography System

Bill Tuthill

Computing Services
University of California
Berkeley, CA 94720

ABSTRACT

Refer is a bibliography system that supports data entry, indexing, retrieval, sorting, runoff, convenient citations, and footnote or endnote numbering. This document assumes you know how to use some Unix editor, and that you are familiar with the **nroff/troff** text formatters.

The **refer** program is a preprocessor for **nroff/troff,** like **eqn** and **tbl,** except that it is used for literature citations, rather than for equations and tables. Given incomplete but sufficiently precise citations, **refer** finds references in a bibliographic database. The complete references are formatted as footnotes, numbered, and placed either at the bottom of the page, or at the end of a chapter.

A number of ancillary programs make **refer** easier to use. The **addbib** program is for creating and extending the bibliographic database; **sortbib** sorts the bibliography by author and date, or other selected criteria; and **roffbib** runs off the entire database, formatting it not as footnotes, but as a bibliography or annotated bibliography.

Once a full bibliography has been created, access time can be improved by making an index to the references with **indxbib**. Then, the **lookbib** program can be used to quickly retrieve individual citations or groups of citations. Creating this inverted index will speed up **refer,** and **lookbib** will allow you to verify that a citation is sufficiently precise to deliver just one reference.

September 18, 1986

Table of Contents

Refer — A Bibliography System

Bill Tuthill

Computing Services
University of California
Berkeley, CA 94720

Introduction

Taken together, the **refer** programs constitute a database system for use with variable-length information. To distinguish various types of bibliographic material, the system uses labels composed of upper case letters, preceded by a percent sign and followed by a space. For example, one document might be given this entry:

%A Joel Kies
%T Document Formatting on Unix Using the -ms Macros
%I Computing Services
%C Berkeley
%D 1980

Each line is called a field, and lines grouped together are called a record; records are separated from each other by a blank line. Bibliographic information follows the labels, containing data to be used by the **refer** system. The order of fields is not important, except that authors should be entered in the same order as they are listed on the document. Fields can be as long as necessary, and may even be continued on the following line(s).

The labels are meaningful to **nroff/troff** macros, and, with a few exceptions, the **refer** program itself does not pay attention to them. This implies that you can change the label codes, if you also change the macros used by **nroff/troff**. The macro package takes care of details like proper ordering, underlining the book title or journal name, and quoting the article's title. Here are the labels used by **refer**, with an indication of what they represent:

%H Header commentary, printed before reference
%A Author's name
%Q Corporate or foreign author (unreversed)
%T Title of article or book
%S Series title
%J Journal containing article
%B Book containing article
%R Report, paper, or thesis (for unpublished material)
%V Volume
%N Number within volume
%E Editor of book containing article
%P Page number(s)
%I Issuer (publisher)
%C City where published
%D Date of publication
%O Other commentary, printed at end of reference
%K Keywords used to locate reference
%L Label used by –k option of **refer**
%X Abstract (used by **roffbib**, not by **refer**)

Only relevant fields should be supplied. Except for %A, each field should be given only once; in the case of multiple authors, the senior author should come first. The %Q is for organizational authors, or authors with Japanese or Arabic names, in which cases the order of names should be preserved. Books should be labeled with the %T, not with the %B, which is reserved for books containing articles. The %J and %B fields should never appear together, although if they do, the %J will override the %B. If there is no author, just an editor, it is best to type the editor in the %A field, as in this example:

%A Bertrand Bronson, ed.

The %E field is used for the editor of a book (%B) containing an article, which has its own author. For unpublished material such as theses, use the %R field; the title in the %T field will be quoted, but the contents of the %R field will not be underlined. Unlike other fields, %H, %O, and %X should contain their own punctuation. Here is a modest example:

```
%A Mike E. Lesk
%T Some Applications of Inverted Indexes on the Unix System
%B Unix Programmer's Manual
%I Bell Laboratories
%C Murray Hill, NJ
%D 1978
%V 2a
%K refer mkey inv hunt
%X Difficult to read paper that dwells on indexing strategies,
giving little practical advice about using \fBrefer\fP.
```

Note that the author's name is given in normal order, without inverting the surname; inversion is done automatically, except when %Q is used instead of %A. We use %X rather than %O for the commentary because we do not want the comment printed all the time. The %O and %H fields are printed by both **refer** and **roffbib;** the %X field is printed only by **roffbib,** as a detached annotation paragraph.

Data Entry with Addbib

The **addbib** program is for creating and extending bibliographic databases. You must give it the filename of your bibliography:

% **addbib database**

Every time you enter **addbib**, it asks if you want instructions. To get them, type **y**; to skip them, type RETURN. **Addbib** prompts for various fields, reads from the keyboard, and writes records containing the **refer** codes to the database. After finishing a field entry, you should end it by typing RETURN. If a field is too long to fit on a line, type a backslash (\) at the end of the line, and you will be able to continue on the following line. Note: the backslash works in this capacity only inside **addbib**.

A field will not be written to the database if nothing is entered into it. Typing a minus sign as the first character of any field will cause **addbib** to back up one field at a time. Backing up is the best way to add multiple authors, and it really helps if you forget to add something important. Fields not contained in the prompting skeleton may be entered by typing a backslash as the last character before RETURN. The following line will be sent verbatim to the database and **addbib** will resume with the next field. This is identical to the procedure for dealing with long fields, but with new fields, don't forget the % key-letter.

Finally, you will be asked for an abstract (or annotation), which will be preserved as the %X field. Type in as many lines as you need, and end with a control-D (hold down the CTRL button, then press the "d" key). This prompting for an abstract can be suppressed with the **-a** command line option.

After one bibliographic record has been completed, **addbib** will ask if you want to continue. If you do, type RETURN; to quit, type **q** or **n** (quit or no). It is also possible to use one of the system editors to correct mistakes made while entering data. After the "Continue?" prompt, type any of the following: **edit, ex, vi,** or **ed** — you will be placed inside the corresponding editor, and returned to **addbib** afterwards, from where you can either quit or add more data.

If the prompts normally supplied by **addbib** are not enough, are in the wrong order, or are too numerous, you can redefine the skeleton by constructing a promptfile. Create some file, to be named after the –p command line option. Place the prompts you want on the left side, followed by a single TAB (control-I), then the **refer** code that is to appear in the bibliographic database. **Addbib** will send the left side to the screen, and the right side, along with data entered, to the database.

Printing the Bibliography

Sortbib is for sorting the bibliography by author (%A) and date (%D), or by data in other fields. It is quite useful for producing bibliographies and annotated bibliographies, which are seldom entered in strict alphabetical order. It takes as arguments the names of up to 16 bibliography files, and sends the sorted records to standard output (the terminal screen), which may be redirected through a pipe or into a file.

The –s*KEYS* flag to **sortbib** will sort by fields whose key-letters are in the *KEYS* string, rather than merely by author and date. Key-letters in *KEYS* may be followed by a '+' to indicate that all such fields are to be used. The default is to sort by senior author and date (printing the senior author last name first), but –sA+D will sort by all authors and then date, and –sATD will sort on senior author, then title, and then date.

Roffbib is for running off the (probably sorted) bibliography. It can handle annotated bibliographies — annotations are entered in the %X (abstract) field. **Roffbib** is a shell script that calls **refer –B** and **nroff –mbib** . It uses the macro definitions that reside in /usr/lib/tmac/tmac.bib, which you can redefine if you know **nroff** and **troff**. Note that **refer** will print the %H and %O commentaries, but will ignore abstracts in the %X field; **roffbib** will print both fields, unless annotations are suppressed with the –x option.

The following command sequence will lineprint the entire bibliography, organized alphabetically by author and date:

% **sortbib database | roffbib | lpr**

This is a good way to proofread the bibliography, or to produce a stand-alone bibliography at the end of a paper. Incidentally, **roffbib** accepts all flags used with **nroff**. For example:

% **sortbib database | roffbib –Tdtc –s1**

will make accent marks work on a DTC daisy-wheel printer, and stop at the bottom of every page for changing paper. The –n and –o flags may also be quite useful, to start page numbering at a selected point, or to produce only specific pages.

Roffbib understands four command-line number registers, which are something like the two-letter number registers in –ms. The –rN1 argument will number references beginning at one (1); use another number to start somewhere besides one. The –rV2 flag will double-space the entire bibliography, while –rV1 will double-space the references, but single-space the annotation paragraphs. Finally, specifying –rL6i changes the line length from 6.5 inches to 6 inches, and saying –rO1i sets the page offset to one inch, instead of zero. (That's a capital O after –r, not a zero.)

Citing Papers with Refer

The **refer** program normally copies input to output, except when it encounters an item of the form:

```
.[
partial citation
.]
```

The partial citation may be just an author's name and a date, or perhaps a title and a keyword, or maybe just a document number. **Refer** looks up the citation in the bibliographic database, and transforms it into a full, properly formatted reference. If the partial citation does not correctly identify a single work (either finding nothing, or more than one reference), a diagnostic message is given. If nothing is found, it will say "No such paper." If more than one reference is found, it will say "Too many hits." Other diagnostic messages can be quite cryptic; if you are in doubt, use **checknr** to verify that all your .['s have matching .]'s.

When everything goes well, the reference will be brought in from the database, numbered, and placed at the bottom of the page. This citation, for example, was produced by:

```
This citation,
.[
lesk inverted indexes
.]
for example, was produced by
```

The .[and .] markers, in essence, replace the .FS and .FE of the –ms macros, and also provide a numbering mechanism. Footnote numbers will be bracketed on the the lineprinter, but superscripted on daisy-wheel terminals and in **troff**. In the reference itself, articles will be quoted, and books and journals will be underlined in **nroff,** and italicized in **troff.**

Sometimes you need to cite a specific page number along with more general bibliographic material. You may have, for instance, a single document that you refer to several times, each time giving a different page citation. This is how you could get "p. 10" in the reference:

```
.[
kies document formatting
%P 10
.]
```

The first line, a partial citation, will find the reference in your bibliography. The second line will insert the page number into the final citation. Ranges of pages may be specified as "%P 56-78".

When the time comes to run off a paper, you will need to have two files: the bibliographic database, and the paper to format. Use a command line something like one of these:

```
% refer –p database paper | nroff –ms
% refer –p database paper | tbl | nroff –ms
% refer –p database paper | tbl | neqn | nroff –ms
```

If other preprocessors are used, **refer** should precede **tbl,** which must in turn precede **eqn** or **neqn** . The –p option specifies a "private" database, which most bibliographies are.

Refer's Command-line Options

Many people like to place references at the end of a chapter, rather than at the bottom of the page. The –e option will accumulate references until a macro sequence of the form

```
.[
$LIST$
.]
```

is encountered (or until the end of file). **Refer** will then write out all references collected up to that point, collapsing identical references. Warning: there is a limit (currently 200) on the number of references that can be accumulated at one time.

It is also possible to sort references that appear at the end of text. The –s*KEYS* flag will sort references by fields whose key-letters are in the *KEYS* string, and permute reference numbers in the text accordingly. It is unnecessary to use –e with it, since –s implies –e. Key-letters in *KEYS* may be followed by a '+' to indicate that all such fields are to be used. The default is to sort by senior author and date, but –sA+D will sort on all authors and then date, and –sA+T will sort by authors and then title.

Refer can also make citations in what is known as the Social or Natural Sciences format. Instead of numbering references, the –l (letter ell) flag makes labels from the senior author's last name and the year of publication. For example, a reference to the paper on Inverted Indexes cited above might appear as

[Lesk1978a]. It is possible to control the number of characters in the last name, and the number of digits in the date. For instance, the command line argument -l6,2 might produce a reference such as [Kernig78c].

Some bibliography standards shun both footnote numbers and labels composed of author and date, requiring some keyword to identify the reference. The -k flag indicates that, instead of numbering references, key labels specified on the %L line should be used to mark references.

The -n flag means to not search the default reference file, located in /usr/dict/papers/Rv7man. Using this flag may make **refer** marginally faster. The -a*n* flag will reverse the first *n* author names, printing Jones, J. A. instead of J. A. Jones. Often -a1 is enough; this will reverse the names of only the senior author. In some versions of **refer** there is also the -f flag to set the footnote number to some predetermined value; for example, -f23 would start numbering with footnote 23.

Making an Index

Once your database is large and relatively stable, it is a good idea to make an index to it, so that references can be found quickly and efficiently. The **indxbib** program makes an inverted index to the bibliographic database (this program is called **pubindex** in the Bell Labs manual). An inverted index could be compared to the thumb cuts of a dictionary — instead of going all the way through your bibliography, programs can move to the exact location where a citation is found.

Indxbib itself takes a while to run, and you will need sufficient disk space to store the indexes. But once it has been run, access time will improve dramatically. Furthermore, large databases of several million characters can be indexed with no problem. The program is exceedingly simple to use:

% **indxbib database**

Be aware that changing your database will require that you run **indxbib** over again. If you don't, you may fail to find a reference that really is in the database.

Once you have built an inverted index, you can use **lookbib** to find references in the database. **Lookbib** cannot be used until you have run **indxbib** . When editing a paper, **lookbib** is very useful to make sure that a citation can be found as specified. It takes one argument, the name of the bibliography, and then reads partial citations from the terminal, returning references that match, or nothing if none match. Its prompt is the greater-than sign.

```
% lookbib database
> lesk inverted indexes
%A  Mike E. Lesk
%T  Some Applications of Inverted Indexes on the Unix System
%J  Unix Programmer's Manual
%I  Bell Laboratories
%C  Murray Hill, NJ
%D  1978
%V  2a
%X  Difficult to read paper that dwells on indexing strategies,
giving little practical advice about using \fBrefer\fP.
>
```

If more than one reference comes back, you will have to give a more precise citation for **refer** . Experiment until you find something that works; remember that it is harmless to overspecify. To get out of the **lookbib** program, type a control-D alone on a line; **lookbib** then exits with an "EOT" message.

Lookbib can also be used to extract groups of related citations. For example, to find all the papers by Brian Kernighan found in the system database, and send the output to a file, type:

% **lookbib /usr/dict/papers/Ind** > **kern.refs**
> **kernighan**
> EOT
% **cat kern.refs**

Your file, "kern.refs", will be full of references. A similar procedure can be used to pull out all papers of some date, all papers from a given journal, all papers containing a certain group of keywords, etc.

Refer Bugs and Some Solutions

The **refer** program will mess up if there are blanks at the end of lines, especially the %A author line. **Addbib** carefully removes trailing blanks, but they may creep in again during editing. Use an editor command — g/ *$/s/// — to remove trailing blanks from your bibliography.

Having bibliographic fields passed through as string definitions implies that interpolated strings (such as accent marks) must have two backslashes, so they can pass through copy mode intact. For instance, the word "téléphone" would have to be represented:

te*`le*´phone

in order to come out correctly. In the %X field, by contrast, you will have to use single backslashes instead. This is because the %X field is not passed through as a string, but as the body of a paragraph macro.

Another problem arises from authors with foreign names. When a name like "Valéry Giscard d'Estaing" is turned around by the –a option of **refer,** it will appear as "d'Estaing, Valéry Giscard," rather than as "Giscard d'Estaing, Valéry." To prevent this, enter names as follows:

%A Vale*´ry Giscard\0d'Estaing
%A Alexander Csoma\0de\0Ko*:ro*:s

(The second is the name of a famous Hungarian linguist.) The backslash-zero is an **nroff/troff** request meaning to insert a digit-width space. It will protect against faulty name reversal, and also against missorting.

Footnote numbers are placed at the end of the line before the .[macro. This line should be a line of text, not a macro. As an example, if the line before the .[is a .R macro, then the .R will eat the footnote number. (The .R is an –ms request meaning change to Roman font.) In cases where the font needs changing, it is necessary to do the following:

```
\fIet al.\fR
.[
awk aho kernighan weinberger
.]
```

Now the reference will be to Aho *et al.* The \fI changes to italics, and the \fR changes back to Roman font. Both these requests are **nroff/troff** requests, not part of –ms. If and when a footnote number is added after this sequence, it will indeed appear in the output.

Internal Details of Refer

You have already read everything you need to know in order to use the **refer** bibliography system. The remaining sections are provided only for extra information, and in case you need to change the way **refer** works.

The output of **refer** is a stream of string definitions, one for each field in a reference. To create string names, percent signs are simply changed to an open bracket, and an [F string is added, containing the footnote number. The %X, %Y and %Z fields are ignored; however, the **annobib** program changes the %X to an .AP (annotation paragraph) macro. The citation used above yields this intermediate output:

```
.ds [F 1
.]-
.ds [A Mike E. Lesk
.ds [T Some Applications of Inverted Indexes on the Unix System
.ds [J Unix Programmer's Manual
.ds [I Bell Laboratories
.ds [C Murray Hill, NJ
.ds [D 1978
.ds [V 2a
.nr [T 0
.nr [A 0
.nr [O 0
.][ 1 journal-article
```

These string definitions are sent to **nroff**, which can use the –ms macros defined in /usr/lib/mx/tmac.xref to take care of formatting things properly. The initializing macro .]– precedes the string definitions, and the labeled macro .][follows. These are changed from the input .[and .] so that running a file twice through **refer** is harmless.

The .][macro, used to print the reference, is given a type-number argument, which is a numeric label indicating the type of reference involved. Here is a list of the various kinds of references:

Field	Value	Kind of Reference
%J	1	Journal Article
%B	3	Article in Book
%R %G	4	Report, Government Report
%I	2	Book
%M	5	Bell Labs Memorandum (undefined)
none	0	Other

The order listed above is indicative of the precedence of the various fields. In other words, a reference that has both the %J and %B fields will be classified as a journal article. If none of the fields listed is present, then the reference will be classified as "other."

The footnote number is flagged in the text with the following sequence, where *number* is the footnote number:

*([.*number**(.]

The *([. and *(.] stand for bracketing or superscripting. In **nroff** with low-resolution devices such as the lpr and a crt, footnote numbers will be bracketed. In **troff,** or on daisy-wheel printers, footnote numbers will be superscripted. Punctuation normally comes before the reference number; this can be changed by using the –P (postpunctuation) option of **refer**.

In some cases, it is necessary to override certain fields in a reference. For instance, each time a work is cited, you may want to specify different page numbers, and you may want to change certain fields. This citation will find the Lesk reference, but will add specific page numbers to the output, even though no page numbers appeared in the original reference.

```
.[
lesk inverted indexes
%P 7-13
%I Computing Services
%O UNX 12.2.2.
.]
```

The %I line will also override any previous publisher information, and the %O line will append some commentary. The **refer** program simply adds the new %P, %I, and %O strings to the output, and later strings definitions cancel earlier ones.

It is also possible to insert an entire citation that does not appear in the bibliographic database. This reference, for example, could be added as follows:

```
.[
%A  Brian Kernighan
%T  A Troff Tutorial
%I  Bell Laboratories
%D  1978
.]
```

This will cause **refer** to interpret the fields exactly as given, without searching the bibliographic database. This practice is not recommended, however, because it's better to add new references to the database, so they can be used again later.

If you want to change the way footnote numbers are printed, signals can be given on the .[and .] lines. For example, to say "See reference (2)," the citation should appear as:

```
See reference
.[(
partial citation
.]),
```

Note that blanks are significant on these signal lines. If a permanent change in the footnote format is desired, it's best to redefine the [. and .] strings.

Changing the Refer Macros

This section is provided for those who wish to rewrite or modify the **refer** macros. This is necessary in order to make output correspond to specific journal requirements, or departmental standards. First there is an explanation of how new macros can be substituted for the old ones. Then several alterations are given as examples. Finally, there is an annotated copy of the **refer** macros used by **roffbib** .

The **refer** macros for **nroff/troff** supplied by the –ms macro package reside in /usr/lib/mx/tmac.xref; they are reference macros, for producing footnotes or endnotes. The **refer** macros used by **roffbib,** on the other hand, reside in /usr/lib/tmac/tmac.bib; they are for producing a stand-alone bibliography.

To change the macros used by **roffbib,** you will need to get your own version of this shell script into the directory where you are working. These two commands will get you a copy of **roffbib** and the macros it uses: †

% **cp /usr/lib/tmac/tmac.bib bibmac**

You can proceed to change bibmac as much as you like. Then when you use **roffbib**, you should specify your own version of the macros, which will be substituted for the normal ones

% **roffbib –m bibmac** *filename*

where *filename* is the name of your bibliography file. Make sure there's a space between –m and **bibmac**.

If you want to modify the **refer** macros for use with **nroff** and the –ms macros, you will need to get a copy of "tmac.xref":

% **cp /usr/lib/ms/s.ref refmac**

These macros are much like "bibmac", except they have .FS and .FE requests, to be used in conjunction with the –ms macros, rather than independently defined .XP and .AP requests. Now you can put this line at the top of the paper to be formatted:

```
.so refmac
```

Your new **refer** macros will override the definitions previously read in by the –ms package. This method works only if "refmac" is in the working directory.

Suppose you didn't like the way dates are printed, and wanted them to be parenthesized, with no comma before. There are five identical lines you will have to change. The first line below is the old way, while the second is the new way:

```
.if !"\\*([D"" , \\*([D\c
.if !"\\*([D"" \& (\\*([D)\c
```

In the first line, there is a comma and a space, but no parentheses. The "\c" at the end of each line indicates to **nroff** that it should continue, leaving no extra space in the output. The "\&" in the second line is the do-nothing character; when followed by a space, a space is sent to the output.

If you need to format a reference in the style favored by the Modern Language Association or Chicago University Press, in the form (city: publisher, date), then you will have to change the middle of the book macro [2 as follows:

```
\& (\c
.if !"\\*([C"" \\*([C:
\\*([I\c
.if !"\\*([D"" , \\*([D\c
)\c
```

This would print (Berkeley: Computing Services, 1982) if all three strings were present. The first line prints a space and a parenthesis; the second prints the city (and a colon) if present; the third always prints the publisher (books must have a publisher, or else they're classified as other); the fourth line prints a comma and the date if present; and the fifth line closes the parentheses. You would need to make similar changes to the other macros as well.

Acknowledgements

Mike Lesk of Bell Laboratories wrote the original **refer** software, including the indexing programs. Al Stangenberger of the Forestry Department wrote the first version of **addbib**, then called **bibin**. Greg Shenaut of the Linguistics Department wrote the original versions of **sortbib** and **roffbib**. All these contributions are greatly appreciated.

Some Applications of Inverted Indexes on the UNIX System

M. E. Lesk

ABSTRACT

I. Some Applications of Inverted Indexes – Overview

This memorandum describes a set of programs which make inverted indexes to UNIX* text files, and their application to retrieving and formatting citations for documents prepared using *troff*.

The indexing and searching programs make keyword indexes to volumes of material too large for linear searching. Searches for combinations of single words can be performed quickly. The programs for general searching are divided into two phases. The first makes an index from the original data; the second searches the index and retrieves items. Both of these phases are further divided into two parts to separate the data-dependent and algorithm dependent code.

The major current application of these programs is the *troff* preprocessor *refer*. A list of 4300 references is maintained on line, containing primarily papers written and cited by local authors. Whenever one of these references is required in a paper, a few words from the title or author list will retrieve it, and the user need not bother to re-enter the exact citation. Alternatively, authors can use their own lists of papers.

This memorandum is of interest to those who are interested in facilities for searching large but relatively unchanging text files on the UNIX system, and those who are interested in handling bibliographic citations with UNIX *troff*.

II. Updating Publication Lists

This section is a brief note describing the auxiliary programs for managing the updating processing. It is written to aid clerical users in maintaining lists of references. Primarily, the programs described permit a large amount of individual control over the content of publication lists while retaining the usefulness of the files to other users.

III. Manual Pages

This section contains the pages from the UNIX programmer's manual dealing with these commands. It is useful for reference.

---

1. Introduction.

The UNIX† system has many utilities (e.g. *grep, awk, lex, egrep, fgrep, ...*) to search through files of text, but most of them are based on a linear scan through the entire file, using some deterministic automaton. This memorandum discusses a program which uses inverted indexes[1] and can thus be used on much larger data bases.

As with any indexing system, of course, there are some disadvantages; once an index is made, the files that have been indexed can not be changed without remaking the index. Thus applications are restricted to those making many searches of relatively stable data. Furthermore, these programs depend on hashing, and can only search for exact matches of whole keywords. It is not possible to look for arithmetic or logical expressions (e.g. "date greater than 1970") or for regular expression searching such as that in *lex.*[2]

Currently there are two uses of this software, the *refer* preprocessor to format references, and the *lookall* command to search through all text files on the UNIX system.

The remaining sections of this memorandum discuss the searching programs and their uses. Section 2 explains the operation of the searching algorithm and describes the data collected for use with the *lookall* command. The more important application, *refer* has a user's description in section 3. Section 4 goes into more detail on reference files for the benefit of those who wish to add references to data bases or write new *troff* macros for use with *refer.* The options to make *refer* collect identical citations, or otherwise relocate and adjust references, are described in section 5. The UNIX manual sections for *refer, lookall,* and associated commands are attached as appendices.

2. Searching.

The indexing and searching process is divided into two phases, each made of two parts. These are shown below.

A. Construct the index.

(1) Find keys — turn the input files into a sequence of tags and keys, where each tag identifies a distinct item in the input and the keys for each such item are the strings under which it is to be indexed.

(2) Hash and sort — prepare a set of inverted indexes from which, given a set of keys, the appropriate item tags can be found quickly.

B. Retrieve an item in response to a query.

(3) Search — Given some keys, look through the files prepared by the hashing and sorting facility and derive the appropriate tags.

(4) Deliver — Given the tags, find the original items. This completes the searching process.

The first phase, making the index, is presumably done relatively infrequently. It should, of course, be done whenever the data being indexed change. In contrast, the second phase, retrieving items, is presumably done often, and must be rapid.

An effort is made to separate code which depends on the data being handled from code which depends on the searching procedure. The search algorithm is involved only in programs (2) and (3), while knowledge of the actual data files is needed only by programs (1) and (4). Thus it is easy to adapt to different data files or different search algorithms.

To start with, it is necessary to have some way of selecting or generating keys from input files. For dealing with files that are basically English, we have a key-making program which

---

† UNIX is a trademark of Bell Laboratories.

[1] D. Knuth, *The Art of Computer Programming: Vol. 3, Sorting and Searching,* Addison-Wesley, Reading, Mass., 1977. See section 6.5.

[2] M. E. Lesk, "Lex — A Lexical Analyzer Generator," Comp. Sci. Tech. Rep. No. 39, Bell Laboratories, Murray Hill, New Jersey, October 1975.

automatically selects words and passes them to the hashing and sorting program (step 2). The format used has one line for each input item, arranged as follows:

name:start,length (tab) key1 key2 key3 ...

where *name* is the file name, *start* is the starting byte number, and *length* is the number of bytes in the entry.

These lines are the only input used to make the index. The first field (the file name, byte position, and byte count) is the tag of the item and can be used to retrieve it quickly. Normally, an item is either a whole file or a section of a file delimited by blank lines. After the tab, the second field contains the keys. The keys, if selected by the automatic program, are any alphanumeric strings which are not among the 100 most frequent words in English and which are not entirely numeric (except for four-digit numbers beginning 19, which are accepted as dates). Keys are truncated to six characters and converted to lower case. Some selection is needed if the original items are very large. We normally just take the first *n* keys, with *n* less than 100 or so; this replaces any attempt at intelligent selection. One file in our system is a complete English dictionary; it would presumably be retrieved for all queries.

To generate an inverted index to the list of record tags and keys, the keys are hashed and sorted to produce an index. What is wanted, ideally, is a series of lists showing the tags associated with each key. To condense this, what is actually produced is a list showing the tags associated with each hash code, and thus with some set of keys. To speed up access and further save space, a set of three or possibly four files is produced. These files are:

File	Contents
entry	Pointers to posting file for each hash code
posting	Lists of tag pointers for each hash code
tag	Tags for each item
key	Keys for each item (optional)

The posting file comprises the real data: it contains a sequence of lists of items posted under each hash code. To speed up searching, the entry file is an array of pointers into the posting file, one per potential hash code. Furthermore, the items in the lists in the posting file are not referred to by their complete tag, but just by an address in the tag file, which gives the complete tags. The key file is optional and contains a copy of the keys used in the indexing.

The searching process starts with a query, containing several keys. The goal is to obtain all items which were indexed under these keys. The query keys are hashed, and the pointers in the entry file used to access the lists in the posting file. These lists are addresses in the tag file of documents posted under the hash codes derived from the query. The common items from all lists are determined; this must include the items indexed by every key, but may also contain some items which are false drops, since items referenced by the correct hash codes need not actually have contained the correct keys. Normally, if there are several keys in the query, there are not likely to be many false drops in the final combined list even though each hash code is somewhat ambiguous. The actual tags are then obtained from the tag file, and to guard against the possibility that an item has false-dropped on some hash code in the query, the original items are normally obtained from the delivery program (4) and the query keys checked against them by string comparison.

Usually, therefore, the check for bad drops is made against the original file. However, if the key derivation procedure is complex, it may be preferable to check against the keys fed to program (2). In this case the optional key file which contains the keys associated with each item is generated, and the item tag is supplemented by a string

;start,length

which indicates the starting byte number in the key file and the length of the string of keys for each item. This file is not usually necessary with the present key-selection program, since the keys always appear in the original document.

There is also an option (**-C***n*) for coordination level searching. This retrieves items which match all but *n* of the query keys. The items are retrieved in the order of the number of keys that they match. Of course, *n* must be less than the number of query keys (nothing is retrieved unless it matches at least one key).

As an example, consider one set of 4377 references, comprising 660,000 bytes. This included 51,000 keys, of which 5,900 were distinct keys. The hash table is kept full to save space (at the expense of time); 995 of 997 possible hash codes were used. The total set of index files (no key file) included 171,000 bytes, about 26% of the original file size. It took 8 minutes of processor time to hash, sort, and write the index. To search for a single query with the resulting index took 1.9 seconds of processor time, while to find the same paper with a sequential linear search using *grep* (reading all of the tags and keys) took 12.3 seconds of processor time.

We have also used this software to index all of the English stored on our UNIX system. This is the index searched by the *lookall* command. On a typical day there were 29,000 files in our user file system, containing about 152,000,000 bytes. Of these 5,300 files, containing 32,000,000 bytes (about 21%) were English text. The total number of 'words' (determined mechanically) was 5,100,000. Of these 227,000 were selected as keys; 19,000 were distinct, hashing to 4,900 (of 5,000 possible) different hash codes. The resulting inverted file indexes used 845,000 bytes, or about 2.6% of the size of the original files. The particularly small indexes are caused by the fact that keys are taken from only the first 50 non-common words of some very long input files.

Even this large *lookall* index can be searched quickly. For example, to find this document by looking for the keys "lesk inverted indexes" required 1.7 seconds of processor time and system time. By comparison, just to search the 800,000 byte dictionary (smaller than even the inverted indexes, let alone the 27,000,000 bytes of text files) with *grep* takes 29 seconds of processor time. The *lookall* program is thus useful when looking for a document which you believe is stored on-line, but do not know where. For example, many memos from our center are in the file system, but it is often difficult to guess where a particular memo might be (it might have several authors, each with many directories, and have been worked on by a secretary with yet more directories). Instructions for the use of the *lookall* command are given in the manual section, shown in the appendix to this memorandum.

The only indexes maintained routinely are those of publication lists and all English files. To make other indexes, the programs for making keys, sorting them, searching the indexes, and delivering answers must be used. Since they are usually invoked as parts of higher-level commands, they are not in the default command directory, but are available to any user in the directory */usr/lib/refer*. Three programs are of interest: *mkey*, which isolates keys from input files; *inv*, which makes an index from a set of keys; and *hunt*, which searches the index and delivers the items. Note that the two parts of the retrieval phase are combined into one program, to avoid the excessive system work and delay which would result from running these as separate processes.

These three commands have a large number of options to adapt to different kinds of input. The user not interested in the detailed description that now follows may skip to section 3, which describes the *refer* program, a packaged-up version of these tools specifically oriented towards formatting references.

Make Keys. The program *mkey* is the key-making program corresponding to step (1) in phase A. Normally, it reads its input from the file names given as arguments, and if there are no arguments it reads from the standard input. It assumes that blank lines in the input delimit separate items, for each of which a different line of keys should be generated. The lines of keys are written on the standard output. Keys are any alphanumeric string in the input not among the most frequent words in English and not entirely numeric (except that all-numeric strings are acceptable if they are between 1900 and 1999). In the output, keys are translated to lower case, and truncated to six characters in length; any associated punctuation is removed. The following flag arguments are recognized by *mkey:*

–c *name*	Name of file of common words; default is */usr/lib/eign*.
–f *name*	Read a list of files from *name* and take each as an input argument.
–i *chars*	Ignore all lines which begin with '%' followed by any character in *chars*.
–k*n*	Use at most *n* keys per input item.
–l*n*	Ignore items shorter than *n* letters long.
–n*m*	Ignore as a key any word in the first *m* words of the list of common English words. The default is 100.
–s	Remove the labels *(file:start,length)* from the output; just give the keys. Used when searching rather than indexing.
–w	Each whole file is a separate item; blank lines in files are irrelevant.

The normal arguments for indexing references are the defaults, which are *-c /usr/lib/eign*, *-n100*, and *-l3*. For searching, the *-s* option is also needed. When the big *lookall* index of all English files is run, the options are *-w*, *-k50*, and *-f (filelist)*. When running on textual input, the *mkey* program processes about 1000 English words per processor second. Unless the *-k* option is used (and the input files are long enough for it to take effect) the output of *mkey* is comparable in size to its input.

Hash and invert. The *inv* program computes the hash codes and writes the inverted files. It reads the output of *mkey* and writes the set of files described earlier in this section. It expects one argument, which is used as the base name for the three (or four) files to be written. Assuming an argument of *Index* (the default) the entry file is named *Index.ia*, the posting file *Index.ib*, the tag file *Index.ic*, and the key file (if present) *Index.id*. The *inv* program recognizes the following options:

–a	Append the new keys to a previous set of inverted files, making new files if there is no old set using the same base name.
–d	Write the optional key file. This is needed when you can not check for false drops by looking for the keys in the original inputs, i.e. when the key derivation procedure is complicated and the output keys are not words from the input files.
–h*n*	The hash table size is *n* (default 997); *n* should be prime. Making *n* bigger saves search time and spends disk space.
–i[u] *name*	Take input from file *name*, instead of the standard input; if **u** is present *name* is unlinked when the sort is started. Using this option permits the sort scratch space to overlap the disk space used for input keys.
–n	Make a completely new set of inverted files, ignoring previous files.
–p	Pipe into the sort program, rather than writing a temporary input file. This saves disk space and spends processor time.
–v	Verbose mode; print a summary of the number of keys which finished indexing.

About half the time used in *inv* is in the contained sort. Assuming the sort is roughly linear, however, a guess at the total timing for *inv* is 250 keys per second. The space used is usually of more importance: the entry file uses four bytes per possible hash (note the **–h** option), and the tag file around 15-20 bytes per item indexed. Roughly, the posting file contains one item for each key instance and one item for each possible hash code; the items are two bytes long if the tag file is less than 65336 bytes long, and the items are four bytes wide if the tag file is greater than 65536 bytes long. Note that to minimize storage, the hash tables should be over-full; for most of the files indexed in this way, there is no other real choice, since the *entry* file must fit in memory.

Searching and Retrieving. The ***hunt*** program retrieves items from an index. It combines, as mentioned above, the two parts of phase (B): search and delivery. The reason why it is efficient to combine delivery and search is partly to avoid starting unnecessary processes, and partly because the delivery operation must be a part of the search operation in any case. Because of the hashing, the search part takes place in two stages: first items are retrieved which have the right hash codes associated with them, and then the actual items are inspected to determine false drops, i.e. to determine if anything with the right hash codes doesn't really have the right keys. Since the original item is retrieved to check on false drops, it is efficient to present it immediately, rather than only giving the tag as output and later retrieving the item again. If there were a separate key file, this argument would not apply, but separate key files are not common.

Input to ***hunt*** is taken from the standard input, one query per line. Each query should be in *mkey -s* output format; all lower case, no punctuation. The ***hunt*** program takes one argument which specifies the base name of the index files to be searched. Only one set of index files can be searched at a time, although many text files may be indexed as a group, of course. If one of the text files has been changed since the index, that file is searched with *fgrep;* this may occasionally slow down the searching, and care should be taken to avoid having many out of date files. The following option arguments are recognized by ***hunt:***

–a	Give all output; ignore checking for false drops.
–C*n*	Coordination level *n;* retrieve items with not more than *n* terms of the input missing; default *C0*, implying that each search term must be in the output items.
–F[**yn***d*]	"–Fy" gives the text of all the items found; "–Fn" suppresses them. "–F*d*" where *d* is an integer gives the text of the first *d* items. The default is *-Fy.*
–g	Do not use *fgrep* to search files changed since the index was made; print an error comment instead.
–i *string*	Take *string* as input, instead of reading the standard input.
–l *n*	The maximum length of internal lists of candidate items is *n;* default 1000.
–o *string*	Put text output ("–Fy") in *string;* of use *only* when invoked from another program.
–p	Print hash code frequencies; mostly for use in optimizing hash table sizes.
–T[**yn***d*]	"–Ty" gives the tags of the items found; "–Tn" suppresses them. "–T*d*" where *d* is an integer gives the first *d* tags. The default is *-Tn*.
–t *string*	Put tag output ("–Ty") in *string;* of use *only* when invoked from another program.

The timing of ***hunt*** is complex. Normally the hash table is overfull, so that there will be many false drops on any single term; but a multi-term query will have few false drops on all terms. Thus if a query is underspecified (one search term) many potential items will be examined and discarded as false drops, wasting time. If the query is overspecified (a dozen search terms) many keys will be examined only to verify that the single item under consideration has that key posted. The variation of search time with number of keys is shown in the table below. Queries of varying length were constructed to retrieve a particular document from the file of references. In the sequence to the left, search terms were chosen so as to select the desired paper as quickly as possible. In the sequence on the right, terms were chosen inefficiently, so that the query did not uniquely select the desired document until four keys had been used. The same document was the target in each case, and the final set of eight keys are also identical; the differences at five, six and seven keys are produced by measurement error, not by the slightly different key lists.

Efficient Keys				Inefficient Keys			
No. keys	Total drops (incl. false)	Retrieved Documents	Search time (seconds)	No. keys	Total drops (incl. false)	Retrieved Documents	Search time (seconds)
1	15	3	1.27	1	68	55	5.96
2	1	1	0.11	2	29	29	2.72
3	1	1	0.14	3	8	8	0.95
4	1	1	0.17	4	1	1	0.18
5	1	1	0.19	5	1	1	0.21
6	1	1	0.23	6	1	1	0.22
7	1	1	0.27	7	1	1	0.26
8	1	1	0.29	8	1	1	0.29

As would be expected, the optimal search is achieved when the query just specifies the answer; however, overspecification is quite cheap. Roughly, the time required by *hunt* can be approximated as 30 milliseconds per search key plus 75 milliseconds per dropped document (whether it is a false drop or a real answer). In general, overspecification can be recommended; it protects the user against additions to the data base which turn previously uniquely-answered queries into ambiguous queries.

The careful reader will have noted an enormous discrepancy between these times and the earlier quoted time of around 1.9 seconds for a search. The times here are purely for the search and retrieval: they are measured by running many searches through a single invocation of the *hunt* program alone. The normal retrieval operation involves using the shell to set up a pipeline through *mkey* to *hunt* and starting both processes; this adds a fixed overhead of about 1.7 seconds of processor time to any single search. Furthermore, remember that all these times are processor times: on a typical morning on our PDP 11/70 system, with about one dozen people logged on, to obtain 1 second of processor time for the search program took between 2 and 12 seconds of real time, with a median of 3.9 seconds and a mean of 4.8 seconds. Thus, although the work involved in a single search may be only 200 milliseconds, after you add the 1.7 seconds of startup processor time and then assume a 4:1 elapsed/processor time ratio, it will be 8 seconds before any response is printed.

3. Selecting and Formatting References for TROFF

The major application of the retrieval software is *refer,* which is a *troff* preprocessor like *eqn*.[3] It scans its input looking for items of the form

```
.[
imprecise citation
.]
```

where an imprecise citation is merely a string of words found in the relevant bibliographic citation. This is translated into a properly formatted reference. If the imprecise citation does not correctly identify a single paper (either selecting no papers or too many) a message is given. The data base of citations searched may be tailored to each system, and individual users may specify their own citation files. On our system, the default data base is accumulated from the publication lists of the members of our organization, plus about half a dozen personal bibliographies that were collected. The present total is about 4300 citations, but this increases steadily. Even now, the data base covers a large fraction of local citations.

For example, the reference for the *eqn* paper above was specified as

[3] B. W. Kernighan and L. L. Cherry, "A System for Typesetting Mathematics," *Comm. Assoc. Comp. Mach.*, vol. 18, pp. 151-157, Bell Laboratories, Murray Hill, New Jersey, March 1975.

```
...
preprocessor like
.I eqn.
.[
kernighan cherry acm 1975
.]
It scans its input looking for items
...
```

This paper was itself printed using *refer*. The above input text was processed by *refer* as well as *tbl* and *troff* by the command

refer memo-file | tbl | troff -ms

and the reference was automatically translated into a correct citation to the ACM paper on mathematical typesetting.

The procedure to use to place a reference in a paper using *refer* is as follows. First, use the *lookbib* command to check that the paper is in the data base and to find out what keys are necessary to retrieve it. This is done by typing *lookbib* and then typing some potential queries until a suitable query is found. For example, had one started to find the *eqn* paper shown above by presenting the query

```
$ lookbib
kernighan cherry
(EOT)
```

lookbib would have found several items; experimentation would quickly have shown that the query given above is adequate. Overspecifying the query is of course harmless. A particularly careful reader may have noticed that "acm" does not appear in the printed citation; we have supplemented some of the data base items with common extra keywords, such as common abbreviations for journals or other sources, to aid in searching.

If the reference is in the data base, the query that retrieved it can be inserted in the text, between **.[** and **.]** brackets. If it is not in the data base, it can be typed into a private file of references, using the format discussed in the next section, and then the **-p** option used to search this private file. Such a command might read (if the private references are called *myfile*)

refer -p myfile document | tbl | eqn | troff -ms . . .

where *tbl* and/or *eqn* could be omitted if not needed. The use of the *-ms* macros[4] or some other macro package, however, is essential. *Refer* only generates the data for the references; exact formatting is done by some macro package, and if none is supplied the references will not be printed.

By default, the references are numbered sequentially, and the *-ms* macros format references as footnotes at the bottom of the page. This memorandum is an example of that style. Other possibilities are discussed in section 5 below.

4. Reference Files.

A reference file is a set of bibliographic references usable with *refer*. It can be indexed using the software described in section 2 for fast searching. What *refer* does is to read the input document stream, looking for imprecise citation references. It then searches through reference files to find the full citations, and inserts them into the document. The format of the full citation is arranged to make it convenient for a macro package, such as the *-ms* macros, to format the reference for printing. Since the format of the final reference is determined by the desired style of output, which is determined by the macros used, *refer* avoids forcing any kind of reference appearance. All it does is define a set of string registers which contain the basic information about the

[4] M. E. Lesk, *Typing Documents on UNIX and GCOS: The -ms Macros for Troff*, 1977.

reference; and provide a macro call which is expanded by the macro package to format the reference. It is the responsibility of the final macro package to see that the reference is actually printed; if no macros are used, and the output of *refer* fed untranslated to *troff*, nothing at all will be printed.

The strings defined by *refer* are taken directly from the files of references, which are in the following format. The references should be separated by blank lines. Each reference is a sequence of lines beginning with % and followed by a key-letter. The remainder of that line, and successive lines until the next line beginning with %, contain the information specified by the key-letter. In general, *refer* does not interpret the information, but merely presents it to the macro package for final formatting. A user with a separate macro package, for example, can add new key-letters or use the existing ones for other purposes without bothering *refer*.

The meaning of the key-letters given below, in particular, is that assigned by the *-ms* macros. Not all information, obviously, is used with each citation. For example, if a document is both an internal memorandum and a journal article, the macros ignore the memorandum version and cite only the journal article. Some kinds of information are not used at all in printing the reference; if a user does not like finding references by specifying title or author keywords, and prefers to add specific keywords to the citation, a field is available which is searched but not printed (K).

The key letters currently recognized by *refer* and *-ms*, with the kind of information implied, are:

Key	Information specified	Key	Information specified
A	Author's name	N	Issue number
B	Title of book containing item	O	Other information
C	City of publication	P	Page(s) of article
D	Date	R	Technical report reference
E	Editor of book containing item	T	Title
G	Government (NTIS) ordering number	V	Volume number
I	Issuer (publisher)		
J	Journal name		
K	Keys (for searching)	X	or
L	Label	Y	or
M	Memorandum label	Z	Information not used by *refer*

For example, a sample reference could be typed as:

```
%T Bounds on the Complexity of the Maximal
Common Subsequence Problem
%Z ctr127
%A A. V. Aho
%A D. S. Hirschberg
%A J. D. Ullman
%J J. ACM
%V 23
%N 1
%P 1-12
%M abcd-78
%D Jan. 1976
```

Order is irrelevant, except that authors are shown in the order given. The output of *refer* is a stream of string definitions, one for each of the fields of each reference, as shown below.

```
.]-
.ds [A authors' names ...
.ds [T title ...
.ds [J journal ...
...
.] [ type-number
```

The special macro **.]–** precedes the string definitions and the special macro **.] [** follows. These are changed from the input **.[** and **.]** so that running the same file through *refer* again is harmless. The **.]–** macro can be used by the macro package to initialize. The **.] [** macro, which should be used to print the reference, is given an argument *type-number* to indicate the kind of reference, as follows:

Value	Kind of reference
1	Journal article
2	Book
3	Article within book
4	Technical report
5	Bell Labs technical memorandum
0	Other

The reference is flagged in the text with the sequence

```
\*([.number\*(.]
```

where *number* is the footnote number. The strings **[.** and **.]** should be used by the macro package to format the reference flag in the text. These strings can be replaced for a particular footnote, as described in section 5. The footnote number (or other signal) is available to the reference macro **.] [** as the string register **[F**.

In some cases users wish to suspend the searching, and merely use the reference macro formatting. That is, the user doesn't want to provide a search key between **.[** and **.]** brackets, but merely the reference lines for the appropriate document. Alternatively, the user can wish to add a few fields to those in the reference as in the standard file, or override some fields. Altering or replacing fields, or supplying whole references, is easily done by inserting lines beginning with **%**; any such line is taken as direct input to the reference processor rather than keys to be searched. Thus

```
.[
key1 key2 key3 ...
%Q New format item
%R Override report name
.]
```

makes the indicates changes to the result of searching for the keys. All of the search keys must be given before the first **%** line.

If no search keys are provided, an entire citation can be provided in-line in the text. For example, if the *eqn* paper citation were to be inserted in this way, rather than by searching for it in the data base, the input would read

```
...
preprocessor like
.I eqn.
.[
%A B. W. Kernighan
%A L. L. Cherry
%T A System for Typesetting Mathematics
%J Comm. ACM
%V 18
%N 3
%P 151-157
%D March 1975
.]
It scans its input looking for items
...
```

This would produce a citation of the same appearance as that resulting from the file search.

As shown, fields are normally turned into *troff* strings. Sometimes users would rather have them defined as macros, so that other *troff* commands can be placed into the data. When this is necessary, simply double the control character **%** in the data. Thus the input

```
.[
%V 23
%%M
Bell Laboratories,
Murray Hill, N.J. 07974
.]
```

is processed by *refer* into

```
.ds [V 23
.de [M
Bell Laboratories,
Murray Hill, N.J. 07974
..
```

The information after **%%M** is defined as a macro to be invoked by **.[M** while the information after **%V** is turned into a string to be invoked by ***([V**. At present *-ms* expects all information as strings.

5. Collecting References and other Refer Options

Normally, the combination of *refer* and *-ms* formats output as *troff* footnotes which are consecutively numbered and placed at the bottom of the page. However, options exist to place the references at the end; to arrange references alphabetically by senior author; and to indicate references by strings in the text of the form [Name1975a] rather than by number. Whenever references are not placed at the bottom of a page identical references are coalesced.

For example, the **–e** option to *refer* specifies that references are to be collected; in this case they are output whenever the sequence

```
.[
$LIST$
.]
```

is encountered. Thus, to place references at the end of a paper, the user would run *refer* with the *-e* option and place the above $LIST$ commands after the last line of the text. *Refer* will then move all the references to that point. To aid in formatting the collected references, *refer* writes the references preceded by the line

.]<

and followed by the line

.]>

to invoke special macros before and after the references.

Another possible option to *refer* is the **–s** option to specify sorting of references. The default, of course, is to list references in the order presented. The **–s** option implies the **–e** option, and thus requires a

.[
$LIST$
.]

entry to call out the reference list. The **–s** option may be followed by a string of letters, numbers, and '+' signs indicating how the references are to be sorted. The sort is done using the fields whose key-letters are in the string as sorting keys; the numbers indicate how many of the fields are to be considered, with '+' taken as a large number. Thus the default is **–sAD** meaning "Sort on senior author, then date." To sort on all authors and then title, specify **–sA+T**. And to sort on two authors and then the journal, write **–sA2J**.

Other options to *refer* change the signal or label inserted in the text for each reference. Normally these are just sequential numbers, and their exact placement (within brackets, as superscripts, etc.) is determined by the macro package. The **–l** option replaces reference numbers by strings composed of the senior author's last name, the date, and a disambiguating letter. If a number follows the **l** as in **–l3** only that many letters of the last name are used in the label string. To abbreviate the date as well the form **-l***m,n* shortens the last name to the first *m* letters and the date to the last *n* digits. For example, the option **–l3,2** would refer to the *eqn* paper (reference 3) by the signal *Ker75a* , since it is the first cited reference by Kernighan in 1975.

A user wishing to specify particular labels for a private bibliography may use the **–k** option. Specifying **–k***x* causes the field *x* to be used as a label. The default is **L**. If this field ends in –, that character is replaced by a sequence letter; otherwise the field is used exactly as given.

If none of the *refer*-produced signals are desired, the **–b** option entirely suppresses automatic text signals.

If the user wishes to override the *-ms* treatment of the reference signal (which is normally to enclose the number in brackets in *nroff* and make it a superscript in *troff*) this can be done easily. If the lines .[or .] contain anything following these characters, the remainders of these lines are used to surround the reference signal, instead of the default. Thus, for example, to say "See reference (2)." and avoid "See reference.[2]" the input might appear

See reference
.[(
imprecise citation ...
.]).

Note that blanks are significant in this construction. If a permanent change is desired in the style of reference signals, however, it is probably easier to redefine the strings [. and .] (which are used to bracket each signal) than to change each citation.

Although normally *refer* limits itself to retrieving the data for the reference, and leaves to a macro package the job of arranging that data as required by the local format, there are two special options for rearrangements that can not be done by macro packages. The **–c** option puts fields into all upper case (CAPS-SMALL CAPS in *troff* output). The key-letters indicated what information is to be translated to upper case follow the **c**, so that **–cAJ** means that authors' names and journals are to be in caps. The **–a** option writes the names of authors last name first, that is *A. D. Hall, Jr.* is written as *Hall, A. D. Jr*. The citation form of the *Journal of the ACM*, for example, would require both **–cA** and **–a** options. This produces authors' names in the style

KERNIGHAN, B. W. AND CHERRY, L. L. for the previous example. The **–a** option may be followed by a number to indicate how many author names should be reversed; **–a1** (without any **–c** option) would produce *Kernighan, B. W. and L. L. Cherry,* for example.

Finally, there is also the previously-mentioned **–p** option to let the user specify a private file of references to be searched before the public files. Note that *refer* does not insist on a previously made index for these files. If a file is named which contains reference data but is not indexed, it will be searched (more slowly) by *refer* using *fgrep*. In this way it is easy for users to keep small files of new references, which can later be added to the public data bases.

Updating Publication Lists

M. E. Lesk

1. Introduction.

This note describes several commands to update the publication lists. The data base consisting of these lists is kept in a set of files in the directory */usr/dict/papers* on the Version 7 UNIX† system. The reason for having special commands to update these files is that they are indexed, and the only reasonable way to find the items to be updated is to use the index. However, altering the files destroys the usefulness of the index, and makes further editing difficult. So the recommended procedure is to

(1) Prepare additions, deletions, and changes in separate files.

(2) Update the data base and reindex.

Whenever you make changes, etc. it is necessary to run the "add & index" step before logging off; otherwise the changes do not take effect. The next section shows the format of the files in the data base. After that, the procedures for preparing additions, preparing changes, preparing deletions, and updating the public data base are given.

2. Publication Format.

The format of a data base entry is given completely in "Some Applications of Inverted Indexes on UNIX" by M. E. Lesk, the first part of this report, (also TM 77-1274-17) and is summarized here via a few examples. In each example, first the output format for an item is shown, and then the corresponding data base entry.

Journal article:

> A. V. Aho, D. J. Hirschberg, and J. D. Ullman, "Bounds on the Complexity of the Maximal Common Subsequence Problem," *J. Assoc. Comp. Mach.*, vol. 23, no. 1, pp. 1-12 (Jan. 1976).
>
> %T Bounds on the Complexity of the Maximal Common
> Subsequence Problem
> %A A. V. Aho
> %A D. S. Hirschberg
> %A J. D. Ullman
> %J J. Assoc. Comp. Mach.
> %V 23
> %N 1
> %P 1-12
> %D Jan. 1976
> %M TM 75-1271-7

† UNIX is a trademark of Bell Laboratories.

Conference proceedings:

B. Prabhala and R. Sethi, "Efficient Computation of Expressions with Common Subexpressions," *Proc. 5th ACM Symp. on Principles of Programming Languages*, pp. 222-230, Tucson, Ariz. (January 1978).

%A B. Prabhala
%A R. Sethi
%T Efficient Computation of Expressions with
Common Subexpressions
%J Proc. 5th ACM Symp. on Principles
of Programming Languages
%C Tucson, Ariz.
%D January 1978
%P 222-230

Book:

B. W. Kernighan and P. J. Plauger, *Software Tools,* Addison-Wesley, Reading, Mass. (1976).

%T Software Tools
%A B. W. Kernighan
%A P. J. Plauger
%I Addison-Wesley
%C Reading, Mass.
%D 1976

Article within book:

J. W. de Bakker, "Semantics of Programming Languages," pp. 173-227 in *Advances in Information Systems Science, Vol. 2,* ed. J. T. Tou, Plenum Press, New York, N. Y. (1969).

%A J. W. de Bakker
%T Semantics of programming languages
%E J. T. Tou
%B Advances in Information Systems Science, Vol. 2
%I Plenum Press
%C New York, N. Y.
%D 1969
%P 173-227

Technical Report:

F. E. Allen, "Bibliography on Program Optimization," Report RC-5767, IBM T. J. Watson Research Center, Yorktown Heights, N. Y. (1975).

%A F. E. Allen
%D 1975
%T Bibliography on Program Optimization
%R Report RC-5767
%I IBM T. J. Watson Research Center
%C Yorktown Heights, N. Y.

Technical Memorandum:

A. V. Aho, B. W. Kernighan and P. J. Weinberg, "AWK – Pattern Scanning and Processing Language", TM 77-1271-5, TM 77-1273-12, TM 77-3444-1 (1977).

%T AWK – Pattern Scanning and Processing Language
%A A. V. Aho
%A B. W. Kernighan
%A P. J. Weinberger
%M TM 77-1271-5, TM 77-1273-12, TM 77-3444-1
%D 1977

Other forms of publication can be entered similarly. Note that conference proceedings are entered as if journals, with the conference name on a *%J* line. This is also sometimes appropriate for obscure publications such as series of lecture notes. When something is both a report and an article, or both a memorandum and an article, enter all necessary information for both; see the first article above, for example. Extra information (such as "In preparation" or "Japanese translation") should be placed on a line beginning *%O*. The most common use of %O lines now is for "Also in ..." to give an additional reference to a secondary appearance of the same paper.

Some of the possible fields of a citation are:

Letter	Meaning	Letter	Meaning
A	Author	K	Extra keys
B	Book including item	N	Issue number
C	City of publication	O	Other
D	Date	P	Page numbers
E	Editor of book	R	Report number
I	Publisher (issuer)	T	Title of item
J	Journal name	V	Volume number

Note that *%B* is used to indicate the title of a book containing the article being entered; when an item is an entire book, the title should be entered with a *%T* as usual.

Normally, the order of items does not matter. The only exception is that if there are multiple authors (%A lines) the order of authors should be that on the paper. If a line is too long, it may be continued on to the next line; any line not beginning with % or . (dot) is assumed to be a continuation of the previous line. Again, see the first article above for an example of a long title. Except for authors, do not repeat any items; if two %J lines are given, for example, the first is ignored. Multiple items on the same file should be separated by blank lines.

Note that in formatted printouts of the file, the exact appearance of the items is determined by a set of macros and the formatting programs. Do not try to adjust fonts, punctuation, etc. by editing the data base; it is wasted effort. In case someone has a real need for a differently-formatted output, a new set of macros can easily be generated to provide alternative appearances of the citations.

3. Updating and Re-indexing.

This section describes the commands that are used to manipulate and change the data base. It explains the procedures for (a) finding references in the data base, (b) adding new references, (c) changing existing references, and (d) deleting references. Remember that all changes, additions, and deletions are done by preparing separate files and then running an 'update and reindex' step.

Checking what's there now. Often you will want to know what is currently in the data base. There is a special command ***lookbib*** to look for things and print them out. It searches for articles based on words in the title, or the author's name, or the date. For example, you could find the first paper above with lookbib aho ullman maximal subsequence 1976

or

lookbib aho ullman hirschberg

If you don't give enough words, several items will be found; if you spell some wrong, nothing will be found. There are around 4300 papers in the public file; you should always use this command to check when you are not sure whether a certain paper is there or not.

Additions. To add new papers, just type in, on one or more files, the citations for the new papers. Remember to check first if the papers are already in the data base. For example, if a paper has a previous memo version, this should be treated as a change to an existing entry, rather than a new entry. If several new papers are being typed on the same file, be sure that there is a blank line between each two papers.

Changes. To change an item, it should be extracted onto a file. This is done with the command

pub.chg key1 key2 key3 ...

where the items key1, key2, key3, etc. are a set of keys that will find the paper, as in the *lookbib* command. That is, if

lookbib johnson yacc cstr

will find a item (to, in this case, Computing Science Technical Report No. 32, "YACC: Yet Another Compiler-Compiler," by S. C. Johnson) then

pub.chg johnson yacc cstr

will permit you to edit the item. The *pub.chg* command extracts the item onto a file named "bibxxx" where "xxx" is a 3-digit number, e.g. "bib234". The command will print the file name it has chosen. If the set of keys finds more than one paper (or no papers) an error message is printed and no file is written. Each reference to be changed must be extracted with a separate *pub.chg* command, and each will be placed on a separate file. You should then edit the "bibxxx" file as desired to change the item, using the UNIX editor. Do not delete or change the first line of the file, however, which begins %# and is a special code line to tell the update program which item is being altered. You may delete or change other lines, or add lines, as you wish. The changes are not actually made in the public data base until you run the update command *pub.run* (see below). Thus, if after extracting an item and modifying it, you decide that you'd rather leave things as they were, delete the "bibxxx" file, and your change request will disappear.

Deletions. To delete an entry from the data base, type the command

pub.del key1 key2 key3 ...

where the items key1, key2, etc. are a set of keys that will find the paper, as with the *lookbib* command. That is, if

lookbib Aho hirschberg ullman

will find a paper,

pub.del aho hirschberg ullman

deletes it. Note that upper and lower case are equivalent in keys. The *pub.del* command will print the entry being deleted. It also gives the name of a "bibxxx" file on which the deletion command is stored. The actual deletion is not done until the changes, additions, etc. are processed, as with the *pub.chg* command. If, after seeing the item to be deleted, you change your mind about throwing it away, delete the "bibxxx" file and the delete request disappears. Again, if the list of keys does not uniquely identify one paper, an error message is given.

Remember that the default versions of the commands described here edit a public data base. Do not delete items unless you are sure deletion is proper; usually this means that there are duplicate entries for the same paper. Otherwise, view requests for deletion with skepticism; even if one person has no need for a particular item in the data base, someone else may want it there.

If an item is correct, but should not appear in the "List of Publications" as normally produced, add the line

%K DNL

to the item. This preserves the item intact, but implies "Do Not List" to the to the commands that print publication lists. The DNL line is normally used for some technical reports, minor memoranda, or other low-grade publications.

Update and reindex. When you have completed a session of changes, you should type the command

pub.run file1 file2 ...

where the names "file1", ... are the new files of additions you have prepared. You need not list the "bibxxx" files representing changes and deletions; they are processed automatically. All of the new items are edited into the standard public data base, and then a new index is made. This process takes about 15 minutes; during this time, searches of the data base will be slower.

Normally, you should execute *pub.run* just before you logoff after performing some edit requests. However, if you don't, the various change request files remain in your directory until you finally do execute *pub.run*. When the changes are processed, the "bibxxx" files are deleted. It is not desirable to wait too long before processing changes, however, to avoid conflicts with someone else who wishes to change the same file. If executing *pub.run* produces the message "File bibxxx too old" it means that someone else has been editing the same file between the time you prepared your changes, and the time you typed *pub.run*. You must delete such old change files and re-enter them.

Note that although *pub.run* discards the "bibxxx" files after processing them, your files of additions are left around even after *pub.run* is finished. If they were typed in only for purposes of updating the data base, you may delete them after they have been processed by *pub.run*.

Example. Suppose, for example, that you wish to

(1) Add to the data base the memos "The Dilogarithm Function of a Real Argument" by R. Morris, and "UNIX Software Distribution by Communication Link," by M. E. Lesk and A. S. Cohen;

(2) Delete from the data base the item "Cheap Typesetters", by M. E. Lesk, SIGLASH Newsletter, 1973; and

(3) Change "J. Assoc. Comp. Mach." to "Jour. ACM" in the citation for Aho, Hirschberg, and Ullman shown above.

The procedure would be as follows. First, you would make a file containing the additions, here called "new.1", in the normal way using the UNIX editor. In the script shown below, the computer prompts are in italics.

```
$ ed new.1
?
a
%T The Dilogarithm Function of a Real Argument
%A Robert Morris
%M TM 78-1271-1
%D 1978

%T UNIX Software Distribution by Communication Link
%A M. E. Lesk
%A A. S. Cohen
%M TM 78-1274-1, 78-8234-1
%D 1978
w new.1
199
q
```

Next you would specify the deletion, which would be done with the *pub.del* command:

```
$ pub.del lesk cheap typesetters siglash
```

to which the computer responds:

```
Will delete: (file bib176)

%T Cheap Typesetters
%A M. E. Lesk
%J ACM SIGLASH Newsletter
%V 6
%N 4
%P 14-16
%D October 1973
```

And then you would extract the Aho, Hirschberg and Ullman paper. The dialogue involved is shown below. First run *pub.chg* to extract the paper; it responds by printing the citation and informing you that it was placed on file *bib123*. That file is then edited.

```
$ pub.chg aho hirschberg ullman
Extracting as file bib123
%T Bounds on the Complexity of the Maximal
Common Subsequence Problem
%A A. V. Aho
%A D. S. Hirschberg
%A J. D. Ullman
%J J. Assoc. Comp. Mach.
%V 23
%N 1
%P 1-12
%M TM 75-1271-7
%D Jan. 1976

$ ed bib123
312
/Assoc/s/ J/ Jour/p
%J Jour. Assoc. Comp. Mach.
s/Assoc.*/ACM/p
%J Jour. ACM
1,$p
%# /usr/dict/papers/p76 233 245 change
%T Bounds on the Complexity of the Maximal
Common Subsequence Problem
%A A. V. Aho
%A D. S. Hirschberg
%A J. D. Ullman
%J Jour. ACM
%V 23
%N 1
%P 1-12
%M TM 75-1271-7
%D Jan. 1976

w
292
q
$
```

Finally, execute *pub.run*, making sure to remember that you have prepared a new file "new.1":

```
$ pub.run new.1
```

and about fifteen minutes later the new index would be complete and all the changes would be included.

4. Printing a Publication List

There are two commands for printing a publication list, depending on whether you want to print one person's list, or the list of many people. To print a list for one person, use the *pub.indiv* command:

```
pub.indiv M Lesk
```

This runs off the list for M. Lesk and puts it in file "output". Note that no '.' is given after the initial. In case of ambiguity two initials can be used. Similarly, to get the list for group of people, say pub.org xxx

which prints all the publications of the members of organization *xxx*, taking the names for the list in the file */usr/dict/papers/centlist/xxx*. This command should normally be run in the background; it takes perhaps 15 minutes. Two options are available with these commands:

pub.indiv -p M Lesk

prints only the papers, leaving out unpublished notes, patents, etc. Also

pub.indiv -t M Lesk | gcat

prints a typeset copy, instead of a computer printer copy. In this case it has been directed to an alternate typesetter with the 'gcat' command. These options may be used together, and may be used with the ***pub.org*** command as well. For example, to print only the papers for all of organization zzz and typeset them, you could type

pub.center -t -p zzz | gcat &

These publication lists are printed double column with a citation style taken from a set of publication list macros; the macros, of course, can be changed easily to adjust the format of the lists.

Tbl — A Program to Format Tables

M. E. Lesk

ABSTRACT

Tbl is a document formatting preprocessor for *troff* or *nroff* which makes even fairly complex tables easy to specify and enter. It is available on the UNIX† system and on Honeywell 6000 GCOS. Tables are made up of columns which may be independently centered, right-adjusted, left-adjusted, or aligned by decimal points. Headings may be placed over single columns or groups of columns. A table entry may contain equations, or may consist of several rows of text. Horizontal or vertical lines may be drawn as desired in the table, and any table or element may be enclosed in a box. For example:

1970 Federal Budget Transfers (in billions of dollars)			
State	Taxes collected	Money spent	Net
New York	22.91	21.35	–1.56
New Jersey	8.33	6.96	–1.37
Connecticut	4.12	3.10	–1.02
Maine	0.74	0.67	–0.07
California	22.29	22.42	+0.13
New Mexico	0.70	1.49	+0.79
Georgia	3.30	4.28	+0.98
Mississippi	1.15	2.32	+1.17
Texas	9.33	11.13	+1.80

Introduction.

Tbl turns a simple description of a table into a *troff* or *nroff* [1] program (list of commands) that prints the table. *Tbl* may be used on the UNIX [2] system and on the Honeywell 6000 GCOS system. It attempts to isolate a portion of a job that it can successfully handle and leave the remainder for other programs. Thus *tbl* may be used with the equation formatting program *eqn* [3] or various layout macro packages [4,5,6], but does not duplicate their functions.

This memorandum is divided into two parts. First we give the rules for preparing *tbl* input; then some examples are shown. The description of rules is precise but technical, and the beginning user may prefer to read the examples first, as they show some common table arrangements. A section explaining how to invoke *tbl* precedes the examples. To avoid repetition, henceforth read *troff* as "*troff* or *nroff*."

The input to *tbl* is text for a document, with tables preceded by a ".TS" (table start) command and followed by a ".TE" (table end) command. *Tbl* processes the tables, generating *troff* formatting commands, and leaves the remainder of the text unchanged. The ".TS" and ".TE" lines are copied, too, so that *troff* page layout macros (such as the memo formatting macros [4])

† UNIX is a trademark of Bell Laboratories.

can use these lines to delimit and place tables as they see fit. In particular, any arguments on the ".TS" or ".TE" lines are copied but otherwise ignored, and may be used by document layout macro commands.

The format of the input is as follows:

text
.TS
table
.TE
text
.TS
table
.TE
text
. . .

where the format of each table is as follows:

.TS
options ;
format .
data
.TE

Each table is independent, and must contain formatting information followed by the data to be entered in the table. The formatting information, which describes the individual columns and rows of the table, may be preceded by a few options that affect the entire table. A detailed description of tables is given in the next section.

Input commands.

As indicated above, a table contains, first, global options, then a format section describing the layout of the table entries, and then the data to be printed. The format and data are always required, but not the options. The various parts of the table are entered as follows:

1) OPTIONS. There may be a single line of options affecting the whole table. If present, this line must follow the .TS line immediately and must contain a list of option names separated by spaces, tabs, or commas, and must be terminated by a semicolon. The allowable options are:

center	— center the table (default is left-adjust);
expand	— make the table as wide as the current line length;
box	— enclose the table in a box;
allbox	— enclose each item in the table in a box;
doublebox	— enclose the table in two boxes;
tab (*x*)	— use *x* instead of tab to separate data items.
linesize (*n*)	— set lines or rules (e.g. from **box**) in *n* point type;
delim (*xy*)	— recognize *x* and *y* as the *eqn* delimiters.

The *tbl* program tries to keep boxed tables on one page by issuing appropriate "need" (*.ne*) commands. These requests are calculated from the number of lines in the tables, and if there are spacing commands embedded in the input, these requests may be inaccurate; use normal *troff* procedures, such as keep-release macros, in that case. The user who must have a multi-page boxed table should use macros designed for this purpose, as explained below under 'Usage.'

2) FORMAT. The format section of the table specifies the layout of the columns. Each line in this section corresponds to one line of the table (except that the last line corresponds to all following lines up to the next .T&, if any — see below), and each line contains a key-letter for each column of the table. It is good practice to separate the key letters for each column by spaces or tabs. Each key-letter is one of the following:

L or **l** to indicate a left-adjusted column entry;

R or **r** to indicate a right-adjusted column entry;

C or **c** to indicate a centered column entry;

N or **n** to indicate a numerical column entry, to be aligned with other numerical entries so that the units digits of numbers line up;

A or **a** to indicate an alphabetic subcolumn; all corresponding entries are aligned on the left, and positioned so that the widest is centered within the column (see example on page 12);

S ori **s** to indicate a spanned heading, i.e. to indicate that the entry from the previous column continues across this column (not allowed for the first column, obviously); or

^ to indicate a vertically spanned heading, i.e. to indicate that the entry from the previous row continues down through this row. (Not allowed for the first row of the table, obviously).

When numerical alignment is specified, a location for the decimal point is sought. The rightmost dot (.) adjacent to a digit is used as a decimal point; if there is no dot adjoining a digit, the rightmost digit is used as a units digit; if no alignment is indicated, the item is centered in the column. However, the special non-printing character string \& may be used to override unconditionally dots and digits, or to align alphabetic data; this string lines up where a dot normally would, and then disappears from the final output. In the example below, the items shown at the left will be aligned (in a numerical column) as shown on the right:

```
13              13
4.2              4.2
26.4.12       26.4.12
abc              abc
abc\&           abc
43\&3.22        433.22
749.12         749.12
```

Note: If numerical data are used in the same column with wider **L** or **r** type table entries, the widest *number* is centered relative to the wider **L** or **r** items (**L** is used instead of **l** for readability; they have the same meaning as key-letters). Alignment within the numerical items is preserved. This is similar to the behavior of **a** type data, as explained above. However, alphabetic subcolumns (requested by the **a** key-letter) are always slightly indented relative to **L** items; if necessary, the column width is increased to force this. This is not true for **n** type entries.

Warning: the **n** and **a** items should not be used in the same column.

For readability, the key-letters describing each column should be separated by spaces. The end of the format section is indicated by a period. The layout of the key-letters in the format section resembles the layout of the actual data in the table. Thus a simple format might appear as:

```
c s s
l n n .
```

which specifies a table of three columns. The first line of the table contains a heading centered across all three columns; each remaining line contains a left-adjusted item in the first column followed by two columns of numerical data. A sample table in this format might be:

	Overall title	
Item-a	34.22	9.1
Item-b	12.65	.02
Items: c,d,e	23	5.8
Total	69.87	14.92

There are some additional features of the key-letter system:

Horizontal lines

— A key-letter may be replaced by '_' (underscore) to indicate a horizontal line in place of the corresponding column entry, or by '=' to indicate a double horizontal line. If an adjacent column contains a horizontal line, or if there are vertical lines adjoining this column, this horizontal line is extended to meet the nearby lines. If any data entry is provided for this column, it is ignored and a warning message is printed.

Vertical lines

— A vertical bar may be placed between column key-letters. This will cause a vertical line between the corresponding columns of the table. A vertical bar to the left of the first key-letter or to the right of the last one produces a line at the edge of the table. If two vertical bars appear between key-letters, a double vertical line is drawn.

Space between columns

— A number may follow the key-letter. This indicates the amount of separation between this column and the next column. The number normally specifies the separation in *ens* (one en is about the width of the letter 'n').* If the "expand" option is used, then these numbers are multiplied by a constant such that the table is as wide as the current line length. The default column separation number is 3. If the separation is changed the worst case (largest space requested) governs.

Vertical spanning

— Normally, vertically spanned items extending over several rows of the table are centered in their vertical range. If a key-letter is followed by **t** or **T**, any corresponding vertically spanned item will begin at the top line of its range.

Font changes

— A key-letter may be followed by a string containing a font name or number preceded by the letter **f** or **F**. This indicates that the corresponding column should be in a different font from the default font (usually Roman). All font names are one or two letters; a one-letter font name should be separated from whatever follows by a space or tab. The single letters **B**, **b**, **I**, and **i** are shorter synonyms for **fB** and **fI**. Font change commands given with the table entries override these specifications.

Point size changes

— A key-letter may be followed by the letter **p** or **P** and a number to indicate the point size of the corresponding table entries. The number may be a signed digit, in which case it is taken as an increment or decrement from the current point size. If both a point size and a column separation value are given, one or more blanks must separate them.

Vertical spacing changes

— A key-letter may be followed by the letter **v** or **V** and a number to indicate the vertical line spacing to be used within a multi-line corresponding table entry. The number may be a signed digit, in which case it is taken as an increment or decrement from the current vertical spacing. A column separation value must be separated by blanks or some other specification from a vertical spacing request. This request has no effect unless the corresponding table entry is a text block (see below).

* More precisely, an en is a number of points (1 point = 1/72 inch) equal to half the current type size.

Column width indication

— A key-letter may be followed by the letter **w** or **W** and a width value in parentheses. This width is used as a minimum column width. If the largest element in the column is not as wide as the width value given after the **w**, the largest element is assumed to be that wide. If the largest element in the column is wider than the specified value, its width is used. The width is also used as a default line length for included text blocks. Normal *troff* units can be used to scale the width value; if none are used, the default is ens. If the width specification is a unitless integer the parentheses may be omitted. If the width value is changed in a column, the *last* one given controls.

Equal width columns

— A key-letter may be followed by the letter **e** or **E** to indicate equal width columns. All columns whose key-letters are followed by **e** or **E** are made the same width. This permits the user to get a group of regularly spaced columns.

Note:

The order of the above features is immaterial; they need not be separated by spaces, except as indicated above to avoid ambiguities involving point size and font changes. Thus a numerical column entry in italic font and 12 point type with a minimum width of 2.5 inches and separated by 6 ens from the next column could be specified as

```
np12w(2.5i)fI 6
```

Alternative notation

— Instead of listing the format of successive lines of a table on consecutive lines of the format section, successive line formats may be given on the same line, separated by commas, so that the format for the example above might have been written:

```
c s s, l n n .
```

Default

— Column descriptors missing from the end of a format line are assumed to be **L**. The longest line in the format section, however, defines the number of columns in the table; extra columns in the data are ignored silently.

3) DATA. The data for the table are typed after the format. Normally, each table line is typed as one line of data. Very long input lines can be broken: any line whose last character is \ is combined with the following line (and the \ vanishes). The data for different columns (the table entries) are separated by tabs, or by whatever character has been specified in the option *tabs* option. There are a few special cases:

Troff commands within tables

— An input line beginning with a '.' followed by anything but a number is assumed to be a command to *troff* and is passed through unchanged, retaining its position in the table. So, for example, space within a table may be produced by ".sp" commands in the data.

Full width horizontal lines

— An input *line* containing only the character _ (underscore) or = (equal sign) is taken to be a single or double line, respectively, extending the full width of the *table*.

Single column horizontal lines

— An input table *entry* containing only the character _ or = is taken to be a single or double line extending the full width of the *column*. Such lines are extended to meet horizontal or vertical lines adjoining this column. To obtain these characters explicitly in a column, either precede them by \& or follow them by a space before the usual tab or newline.

Short horizontal lines

— An input table *entry* containing only the string _ is taken to be a single line as wide as the contents of the column. It is not extended to meet adjoining lines.

Vertically spanned items

— An input table entry containing only the character string **\^** indicates that the table entry immediately above spans downward over this row. It is equivalent to a table format key-letter of '^'.

Text blocks

— In order to include a block of text as a table entry, precede it by **T{** and follow it by **T}**. Thus the sequence

```
. . . T{
block of
text
T} . . .
```

is the way to enter, as a single entry in the table, something that cannot conveniently be typed as a simple string between tabs. Note that the **T}** end delimiter must begin a line; additional columns of data may follow after a tab on the same line. See the example on page 11 for an illustration of included text blocks in a table. If more than twenty or thirty text blocks are used in a table, various limits in the *troff* program are likely to be exceeded, producing diagnostics such as 'too many string/macro names' or 'too many number registers.'

Text blocks are pulled out from the table, processed separately by *troff,* and replaced in the table as a solid block. If no line length is specified in the *block of text* itself, or in the table format, the default is to use $L \times C/(N+1)$ where L is the current line length, C is the number of table columns spanned by the text, and N is the total number of columns in the table. The other parameters (point size, font, etc.) used in setting the *block of text* are those in effect at the beginning of the table (including the effect of the ".TS" macro) and any table format specifications of size, spacing and font, using the **p**, **v** and **f** modifiers to the column key-letters. Commands within the text block itself are also recognized, of course. However, *troff* commands within the table data but not within the text block do not affect that block.

Warnings:

— Although any number of lines may be present in a table, only the first 200 lines are used in calculating the widths of the various columns. A multi-page table, of course, may be arranged as several single-page tables if this proves to be a problem. Other difficulties with formatting may arise because, in the calculation of column widths all table entries are assumed to be in the font and size being used when the ".TS" command was encountered, except for font and size changes indicated (a) in the table format section and (b) within the table data (as in the entry \s+3\fIdata\fP\s0). Therefore, although arbitrary *troff* requests may be sprinkled in a table, care must be taken to avoid confusing the width calculations; use requests such as '.ps' with care.

4) ADDITIONAL COMMAND LINES. If the format of a table must be changed after many similar lines, as with sub-headings or summarizations, the ".T&" (table continue) command can be used to change column parameters. The outline of such a table input is:

```
.TS
options ;
format .
data
. . .
.T&
format .
data
.T&
format .
data
.TE
```

as in the examples on pages 10 and 13. Using this procedure, each table line can be close to its corresponding format line.

Warning: it is not possible to change the number of columns, the space between columns, the global options such as *box,* or the selection of columns to be made equal width.

Usage.

On UNIX, *tbl* can be run on a simple table with the command

tbl input-file | troff

but for more complicated use, where there are several input files, and they contain equations and *ms* memorandum layout commands as well as tables, the normal command would be

tbl file-1 file-2 . . . | eqn | troff −ms

and, of course, the usual options may be used on the *troff* and *eqn* commands. The usage for *nroff* is similar to that for *troff,* but only TELETYPE® Model 37 and Diablo-mechanism (DASI or GSI) terminals can print boxed tables directly.

For the convenience of users employing line printers without adequate driving tables or post-filters, there is a special *-TX* command line option to *tbl* which produces output that does not have fractional line motions in it. The only other command line options recognized by *tbl* are *-ms* and *-mm* which are turned into commands to fetch the corresponding macro files; usually it is more convenient to place these arguments on the *troff* part of the command line, but they are accepted by *tbl* as well.

Note that when *eqn* and *tbl* are used together on the same file *tbl* should be used first. If there are no equations within tables, either order works, but it is usually faster to run *tbl* first, since *eqn* normally produces a larger expansion of the input than *tbl.* However, if there are equations within tables (using the *delim* mechanism in *eqn*), *tbl* must be first or the output will be scrambled. Users must also beware of using equations in **n**-style columns; this is nearly always wrong, since *tbl* attempts to split numerical format items into two parts and this is not possible with equations. The user can defend against this by giving the *delim(xx)* table option; this prevents splitting of numerical columns within the delimiters. For example, if the *eqn* delimiters are $$, giving *delim($$)* a numerical column such as "1245 $+− 16$" will be divided after 1245, not after 16.

Tbl limits tables to twenty columns; however, use of more than 16 numerical columns may fail because of limits in *troff,* producing the 'too many number registers' message. *Troff* number registers used by *tbl* must be avoided by the user within tables; these include two-digit names from 31 to 99, and names of the forms #*x*, *x*+, *x*|, ^*x*, and *x*−, where *x* is any lower case letter. The names ##, #−, and #^ are also used in certain circumstances. To conserve number register names, the **n** and **a** formats share a register; hence the restriction above that they may not be used in the same column.

For aid in writing layout macros, *tbl* defines a number register TW which is the table width; it is defined by the time that the ".TE" macro is invoked and may be used in the expansion of that macro. More importantly, to assist in laying out multi-page boxed tables the macro T# is defined to produce the bottom lines and side lines of a boxed table, and then invoked at its end. By use of this macro in the page footer a multi-page table can be boxed. In particular, the *ms* macros can be used to print a multi-page boxed table with a repeated heading by giving the argument H to the ".TS" macro. If the table start macro is written

```
.TS H
```

a line of the form

```
.TH
```

must be given in the table after any table heading (or at the start if none). Material up to the ".TH" is placed at the top of each page of table; the remaining lines in the table are placed on several pages as required. Note that this is *not* a feature of *tbl*, but of the *ms* layout macros.

Examples.

Here are some examples illustrating features of *tbl*. The symbol Ⓣ in the input represents a tab character.

Input:

```
.TS
box;
c c c
l l l.
Language Ⓣ Authors Ⓣ Runs on

Fortran Ⓣ Many Ⓣ Almost anything
PL/1 Ⓣ IBM Ⓣ 360/370
C Ⓣ BTL Ⓣ 11/45,H6000,370
BLISS Ⓣ Carnegie-Mellon Ⓣ PDP-10,11
IDS Ⓣ Honeywell Ⓣ H6000
Pascal Ⓣ Stanford Ⓣ 370
.TE
```

Output:

Language	Authors	Runs on
Fortran	Many	Almost anything
PL/1	IBM	360/370
C	BTL	11/45,H6000,370
BLISS	Carnegie-Mellon	PDP-10,11
IDS	Honeywell	H6000
Pascal	Stanford	370

Input:

```
.TS
allbox;
c s s
c c c
n n n.
AT&T Common Stock
Year Ⓣ Price Ⓣ Dividend
1971 Ⓣ 41-54 Ⓣ $2.60
2 Ⓣ 41-54 Ⓣ 2.70
3 Ⓣ 46-55 Ⓣ 2.87
4 Ⓣ 40-53 Ⓣ 3.24
5 Ⓣ 45-52 Ⓣ 3.40
6 Ⓣ 51-59 Ⓣ .95*
.TE
* (first quarter only)
```

Output:

AT&T Common Stock		
Year	Price	Dividend
1971	41-54	$2.60
2	41-54	2.70
3	46-55	2.87
4	40-53	3.24
5	45-52	3.40
6	51-59	.95*

* (first quarter only)

Input:

```
.TS
box;
c s s
c | c | c
l | l | n.
Major New York Bridges
=
Bridge ⓣ Designer ⓣ Length
_
Brooklyn ⓣ J. A. Roebling ⓣ 1595
Manhattan ⓣ G. Lindenthal ⓣ 1470
Williamsburg ⓣ L. L. Buck ⓣ 1600
_
Queensborough ⓣ Palmer & ⓣ 1182
 ⓣ  Hornbostel
_
 ⓣ ⓣ 1380
Triborough ⓣ O. H. Ammann ⓣ _
 ⓣ ⓣ 383
_
Bronx Whitestone ⓣ O. H. Ammann ⓣ 2300
Throgs Neck ⓣ O. H. Ammann ⓣ 1800
_
George Washington ⓣ O. H. Ammann ⓣ 3500
.TE
```

Output:

Major New York Bridges		
Bridge	Designer	Length
Brooklyn	J. A. Roebling	1595
Manhattan	G. Lindenthal	1470
Williamsburg	L. L. Buck	1600
Queensborough	Palmer & Hornbostel	1182
Triborough	O. H. Ammann	1380
		383
Bronx Whitestone	O. H. Ammann	2300
Throgs Neck	O. H. Ammann	1800
George Washington	O. H. Ammann	3500

Input:

```
.TS
c c
np-2 | n | .
ⓣ Stack
ⓣ _
1 ⓣ 46
ⓣ _
2 ⓣ 23
ⓣ _
3 ⓣ 15
ⓣ _
4 ⓣ 6.5
ⓣ _
5 ⓣ 2.1
ⓣ _
.TE
```

Output:

	Stack
1	46
2	23
3	15
4	6.5
5	2.1

Input:

```
.TS
box;
L L L
L L _
L L |LB
L L _
L L L.
january Ⓣ february Ⓣ march
april Ⓣ may
june Ⓣ july Ⓣ Months
august Ⓣ september
october Ⓣ november Ⓣ december
.TE
```

Output:

january	february	march
april	may	
june	july	**Months**
august	september	
october	november	december

Input:

```
.TS
box;
cfB s s s.
Composition of Foods
_
.T&
c |c s s
c |c s s
c |c |c |c.
Food Ⓣ Percent by Weight
\^ Ⓣ _
\^ Ⓣ Protein Ⓣ Fat Ⓣ Carbo-
\^ Ⓣ \^ Ⓣ \^ Ⓣ hydrate
_
.T&
l |n |n |n.
Apples Ⓣ .4 Ⓣ .5 Ⓣ 13.0
Halibut Ⓣ 18.4 Ⓣ 5.2 Ⓣ . . .
Lima beans Ⓣ 7.5 Ⓣ .8 Ⓣ 22.0
Milk Ⓣ 3.3 Ⓣ 4.0 Ⓣ 5.0
Mushrooms Ⓣ 3.5 Ⓣ .4 Ⓣ 6.0
Rye bread Ⓣ 9.0 Ⓣ .6 Ⓣ 52.7
.TE
```

Output:

Composition of Foods			
Food	Percent by Weight		
	Protein	Fat	Carbo-hydrate
Apples	.4	.5	13.0
Halibut	18.4	5.2	...
Lima beans	7.5	.8	22.0
Milk	3.3	4.0	5.0
Mushrooms	3.5	.4	6.0
Rye bread	9.0	.6	52.7

Input:

```
.TS
allbox;
cfI s  s
c   cw(1i)  cw(1i)
lp9 lp9 lp9.
New York Area Rocks
Era ⓣ Formation ⓣ Age (years)
Precambrian ⓣ Reading Prong ⓣ >1 billion
Paleozoic ⓣ Manhattan Prong ⓣ 400 million
Mesozoic ⓣ T{
.na
Newark Basin, incl.
Stockton, Lockatong, and Brunswick
formations; also Watchungs
and Palisades.
T} ⓣ 200 million
Cenozoic ⓣ Coastal Plain ⓣ T{
On Long Island 30,000 years;
Cretaceous sediments redeposited
by recent glaciation.
.ad
T}
.TE
```

Output:

New York Area Rocks		
Era	Formation	Age (years)
Precambrian	Reading Prong	>1 billion
Paleozoic	Manhattan Prong	400 million
Mesozoic	Newark Basin, incl. Stockton, Lockatong, and Brunswick formations; also Watchungs and Palisades.	200 million
Cenozoic	Coastal Plain	On Long Island 30,000 years; Cretaceous sediments redeposited by recent glaciation.

Input:

```
.EQ
delim $$
.EN

. . .

.TS
doublebox;
c c
l l.
Name ⓣ Definition
.sp
.vs +2p
Gamma ⓣ $GAMMA (z) = int sub 0 sup inf  t sup {z-1} e sup -t dt$
Sine ⓣ $sin (x) = 1 over 2i ( e sup ix - e sup -ix )$
Error ⓣ $ roman erf (z) = 2 over sqrt pi int sub 0 sup z e sup {-t sup 2} dt$
Bessel ⓣ $ J sub 0 (z) = 1 over pi int sub 0 sup pi cos ( z sin theta ) d theta $
Zeta ⓣ $ zeta (s) = sum from k=1 to inf k sup -s ~~( Re~s > 1)$
.vs -2p
.TE
```

Output:

Name	Definition
Gamma	$\Gamma(z)=\int_0^\infty t^{z-1}e^{-t}\,dt$
Sine	$\sin(x)=\frac{1}{2i}(e^{ix}-e^{-ix})$
Error	$\mathrm{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$
Bessel	$J_0(z)=\frac{1}{\pi}\int_0^\pi \cos(z\sin\theta)\,d\theta$
Zeta	$\zeta(s)=\sum_{k=1}^\infty k^{-s}\quad(\mathrm{Re}\ s>1)$

Input:

```
.TS
box, tab(:);
cb s s s s
cp-2 s s s s
c || c | c | c | c
c || c | c | c | c
r2 || n2 | n2 | n2 | n.
Readability of Text
Line Width and Leading for 10-Point Type
=
Line : Set : 1-Point : 2-Point : 4-Point
Width : Solid : Leading : Leading : Leading
_
9 Pica : \-9.3 : \-6.0 : \-5.3 : \-7.1
14 Pica : \-4.5 : \-0.6 : \-0.3 : \-1.7
19 Pica : \-5.0 : \-5.1 : 0.0 : \-2.0
31 Pica : \-3.7 : \-3.8 : \-2.4 : \-3.6
43 Pica : \-9.1 : \-9.0 : \-5.9 : \-8.8
.TE
```

Output:

Readability of Text Line Width and Leading for 10-Point Type				
Line Width	Set Solid	1-Point Leading	2-Point Leading	4-Point Leading
9 Pica	−9.3	−6.0	−5.3	−7.1
14 Pica	−4.5	−0.6	−0.3	−1.7
19 Pica	−5.0	−5.1	0.0	−2.0
31 Pica	−3.7	−3.8	−2.4	−3.6
43 Pica	−9.1	−9.0	−5.9	−8.8

Input:

```
.TS
c s
cip-2 s
l n
a n.
Some London Transport Statistics
(Year 1964)
Railway route miles ⓣ 244
Tube ⓣ 66
Sub-surface ⓣ 22
Surface ⓣ 156
.sp .5
.T&
l r
a r.
Passenger traffic \- railway
Journeys ⓣ 674 million
Average length ⓣ 4.55 miles
Passenger miles ⓣ 3,066 million
.T&
l r
a r.
Passenger traffic \- road
Journeys ⓣ 2,252 million
Average length ⓣ 2.26 miles
Passenger miles ⓣ 5,094 million
.T&
l n
a n.
.sp .5
Vehicles ⓣ 12,521
Railway motor cars ⓣ 2,905
Railway trailer cars ⓣ 1,269
Total railway ⓣ 4,174
Omnibuses ⓣ 8,347
.T&
l n
a n.
.sp .5
Staff ⓣ 73,739
Administrative, etc. ⓣ 5,582
Civil engineering ⓣ 5,134
Electrical eng. ⓣ 1,714
Mech. eng. \- railway ⓣ 4,310
Mech. eng. \- road ⓣ 9,152
Railway operations ⓣ 8,930
Road operations ⓣ 35,946
Other ⓣ 2,971
.TE
```

Output:

Some London Transport Statistics	
(Year 1964)	
Railway route miles	244
Tube	66
Sub-surface	22
Surface	156
Passenger traffic – railway	
Journeys	674 million
Average length	4.55 miles
Passenger miles	3,066 million
Passenger traffic – road	
Journeys	2,252 million
Average length	2.26 miles
Passenger miles	5,094 million
Vehicles	12,521
Railway motor cars	2,905
Railway trailer cars	1,269
Total railway	4,174
Omnibuses	8,347
Staff	73,739
Administrative, etc.	5,582
Civil engineering	5,134
Electrical eng.	1,714
Mech. eng. – railway	4,310
Mech. eng. – road	9,152
Railway operations	8,930
Road operations	35,946
Other	2,971

Input:

```
.ps 8
.vs 10p
.TS
center box;
c s s
ci s s
c c c
lB l n.
New Jersey Representatives
(Democrats)
.sp .5
NameⓉOffice addressⓉPhone
.sp .5
James J. FlorioⓉ23 S. White Horse Pike, Somerdale 08083Ⓣ609-627-8222
William J. HughesⓉ2920 Atlantic Ave., Atlantic City 08401Ⓣ609-345-4844
James J. HowardⓉ801 Bangs Ave., Asbury Park 07712Ⓣ201-774-1600
Frank Thompson, Jr.Ⓣ10 Rutgers Pl., Trenton 08618Ⓣ609-599-1619
Andrew MaguireⓉ115 W. Passaic St., Rochelle Park 07662Ⓣ201-843-0240
Robert A. RoeⓉU.S.P.O., 194 Ward St., Paterson 07510Ⓣ201-523-5152
Henry HelstoskiⓉ666 Paterson Ave., East Rutherford 07073Ⓣ201-939-9090
Peter W. Rodino, Jr.ⓉSuite 1435A, 970 Broad St., Newark 07102Ⓣ201-645-3213
Joseph G. MinishⓉ308 Main St., Orange 07050Ⓣ201-645-6363
Helen S. MeynerⓉ32 Bridge St., Lambertville 08530Ⓣ609-397-1830
Dominick V. DanielsⓉ895 Bergen Ave., Jersey City 07306Ⓣ201-659-7700
Edward J. PattenⓉNatl. Bank Bldg., Perth Amboy 08861Ⓣ201-826-4610
.sp .5
.T&
ci s s
lB l n.
(Republicans)
.sp .5v
Millicent FenwickⓉ41 N. Bridge St., Somerville 08876Ⓣ201-722-8200
Edwin B. ForsytheⓉ301 Mill St., Moorestown 08057Ⓣ609-235-6622
Matthew J. RinaldoⓉ1961 Morris Ave., Union 07083Ⓣ201-687-4235
.TE
.ps 10
.vs 12p
```

Output:

New Jersey Representatives *(Democrats)*		
Name	Office address	Phone
James J. Florio	23 S. White Horse Pike, Somerdale 08083	609-627-8222
William J. Hughes	2920 Atlantic Ave., Atlantic City 08401	609-345-4844
James J. Howard	801 Bangs Ave., Asbury Park 07712	201-774-1600
Frank Thompson, Jr.	10 Rutgers Pl., Trenton 08618	609-599-1619
Andrew Maguire	115 W. Passaic St., Rochelle Park 07662	201-843-0240
Robert A. Roe	U.S.P.O., 194 Ward St., Paterson 07510	201-523-5152
Henry Helstoski	666 Paterson Ave., East Rutherford 07073	201-939-9090
Peter W. Rodino, Jr.	Suite 1435A, 970 Broad St., Newark 07102	201-645-3213
Joseph G. Minish	308 Main St., Orange 07050	201-645-6363
Helen S. Meyner	32 Bridge St., Lambertville 08530	609-397-1830
Dominick V. Daniels	895 Bergen Ave., Jersey City 07306	201-659-7700
Edward J. Patten	Natl. Bank Bldg., Perth Amboy 08861	201-826-4610
	(Republicans)	
Millicent Fenwick	41 N. Bridge St., Somerville 08876	201-722-8200
Edwin B. Forsythe	301 Mill St., Moorestown 08057	609-235-6622
Matthew J. Rinaldo	1961 Morris Ave., Union 07083	201-687-4235

This is a paragraph of normal text placed here only to indicate where the left and right margins are. In this way the reader can judge the appearance of centered tables or expanded tables, and observe how such tables are formatted.

Input:

```
.TS
expand;
c s s s
c c c c
l l n n.
Bell Labs Locations
Name ⓣ Address ⓣ Area Code ⓣ Phone
Holmdel ⓣ Holmdel, N. J. 07733 ⓣ 201 ⓣ 949-3000
Murray Hill ⓣ Murray Hill, N. J. 07974 ⓣ 201 ⓣ 582-6377
Whippany ⓣ Whippany, N. J. 07981 ⓣ 201 ⓣ 386-3000
Indian Hill ⓣ Naperville, Illinois 60540 ⓣ 312 ⓣ 690-2000
.TE
```

Output:

Bell Labs Locations			
Name	Address	Area Code	Phone
Holmdel	Holmdel, N. J. 07733	201	949-3000
Murray Hill	Murray Hill, N. J. 07974	201	582-6377
Whippany	Whippany, N. J. 07981	201	386-3000
Indian Hill	Naperville, Illinois 60540	312	690-2000

Input:

```
.TS
box;
cb   s   s   s
c |c |c   s
ltiw(1i) |ltw(2i) |lp8 |lw(1.5i)p8.
Some Interesting Places
_
Name Ⓣ Description Ⓣ Practical Information
_
T{
American Museum of Natural History
T}Ⓣ T{
The collections fill 11.5 acres (Michelin) or 25 acres (MTA)
of exhibition halls on four floors.  There is a full-sized replica
of a blue whale and the world's largest star sapphire (stolen in 1964).
T}Ⓣ Hours Ⓣ 10-5, ex. Sun 11-5, Wed. to 9
\^ Ⓣ \^ Ⓣ Location Ⓣ T{
Central Park West & 79th St.
T}
\^ Ⓣ \^ Ⓣ Admission Ⓣ Donation: $1.00 asked
\^ Ⓣ \^ Ⓣ Subway Ⓣ AA to 81st St.
\^ Ⓣ \^ Ⓣ Telephone Ⓣ 212-873-4225
_
Bronx Zoo Ⓣ T{
About a mile long and .6 mile wide, this is the largest zoo in America.
A lion eats 18 pounds
of meat a day while a sea lion eats 15 pounds of fish.
T}Ⓣ Hours Ⓣ T{
10-4:30 winter, to 5:00 summer
T}
\^ Ⓣ \^ Ⓣ Location Ⓣ T{
185th St. & Southern Blvd, the Bronx.
T}
\^ Ⓣ \^ Ⓣ Admission Ⓣ $1.00, but Tu,We,Th free
\^ Ⓣ \^ Ⓣ Subway Ⓣ 2, 5 to East Tremont Ave.
\^ Ⓣ \^ Ⓣ Telephone Ⓣ 212-933-1759
_
Brooklyn Museum Ⓣ T{
Five floors of galleries contain American and ancient art.
There are American period rooms and architectural ornaments saved
from wreckers, such as a classical figure from Pennsylvania Station.
T}Ⓣ Hours Ⓣ Wed-Sat, 10-5, Sun 12-5
\^ Ⓣ \^ Ⓣ Location Ⓣ T{
Eastern Parkway & Washington Ave., Brooklyn.
T}
\^ Ⓣ \^ Ⓣ Admission Ⓣ Free
\^ Ⓣ \^ Ⓣ Subway Ⓣ 2,3 to Eastern Parkway.
\^ Ⓣ \^ Ⓣ Telephone Ⓣ 718-638-5000
_
T{
New-York Historical Society
T}Ⓣ T{
All the original paintings for Audubon's
.I
Birds of America
.R
are here, as are exhibits of American decorative arts, New York history,
Hudson River school paintings, carriages, and glass paperweights.
T}Ⓣ Hours Ⓣ T{
Tues-Fri & Sun, 1-5; Sat 10-5
T}
\^ Ⓣ \^ Ⓣ Location Ⓣ T{
Central Park West & 77th St.
T}
\^ Ⓣ \^ Ⓣ Admission Ⓣ Free
\^ Ⓣ \^ Ⓣ Subway Ⓣ AA to 81st St.
\^ Ⓣ \^ Ⓣ Telephone Ⓣ 212-873-3400
.TE
```

Output:

Some Interesting Places			
Name	Description	Practical Information	
American Museum of Natural History	The collections fill 11.5 acres (Michelin) or 25 acres (MTA) of exhibition halls on four floors. There is a full-sized replica of a blue whale and the world's largest star sapphire (stolen in 1964).	Hours Location Admission Subway Telephone	10-5, ex. Sun 11-5, Wed. to 9 Central Park West & 79th St. Donation: $1.00 asked AA to 81st St. 212-873-4225
Bronx Zoo	About a mile long and .6 mile wide, this is the largest zoo in America. A lion eats 18 pounds of meat a day while a sea lion eats 15 pounds of fish.	Hours Location Admission Subway Telephone	10-4:30 winter, to 5:00 summer 185th St. & Southern Blvd, the Bronx. $1.00, but Tu,We,Th free 2, 5 to East Tremont Ave. 212-933-1759
Brooklyn Museum	Five floors of galleries contain American and ancient art. There are American period rooms and architectural ornaments saved from wreckers, such as a classical figure from Pennsylvania Station.	Hours Location Admission Subway Telephone	Wed-Sat, 10-5, Sun 12-5 Eastern Parkway & Washington Ave., Brooklyn. Free 2,3 to Eastern Parkway. 718-638-5000
New-York Historical Society	All the original paintings for Audubon's *Birds of America* are here, as are exhibits of American decorative arts, New York history, Hudson River school paintings, carriages, and glass paperweights.	Hours Location Admission Subway Telephone	Tues-Fri & Sun, 1-5; Sat 10-5 Central Park West & 77th St. Free AA to 81st St. 212-873-3400

Acknowledgments.

Many thanks are due to J. C. Blinn, who has done a large amount of testing and assisted with the design of the program. He has also written many of the more intelligible sentences in this document and helped edit all of it. All phototypesetting programs on UNIX are dependent on the work of J. F. Ossanna, whose assistance with this program in particular has been most helpful. This program is patterned on a table formatter originally written by J. F. Gimpel. The assistance of T. A. Dolotta, B. W. Kernighan, and J. N. Sturman is gratefully acknowledged.

References.

[1] J. F. Ossanna, *NROFF/TROFF User's Manual*, Computing Science Technical Report No. 54, Bell Laboratories, 1976.

[2] K. Thompson and D. M. Ritchie, "The UNIX Time-Sharing System," *Comm. ACM.* **17**, pp. 365–75 (1974).

[3] B. W. Kernighan and L. L. Cherry, "A System for Typesetting Mathematics," *Comm. ACM.* **18**, pp. 151–57 (1975).

[4] M. E. Lesk, *Typing Documents on UNIX*, Bell Laboratories internal memorandum.

[5] M. E. Lesk and B. W. Kernighan, *Computer Typesetting of Technical Journals on UNIX*, Computing Science Technical Report No. 44, Bell Laboratories, July 1976.

[6] J. R. Mashey and D. W. Smith, *PWB/MM — Programmer's Workbench Memorandum Macros,* Bell Laboratories memorandum.

List of Tbl Command Characters and Words

Command	*Meaning*	*Section*
a A	Alphabetic subcolumn	2
allbox	Draw box around all items	1
b B	Boldface item	2
box	Draw box around table	1
c C	Centered column	2
center	Center table in page	1
doublebox	Doubled box around table	1
e E	Equal width columns	2
expand	Make table full line width	1
f F	Font change	2
i I	Italic item	2
l L	Left adjusted column	2
n N	Numerical column	2
nnn	Column separation	2
p P	Point size change	2
r R	Right adjusted column	2
s S	Spanned item	2
t T	Vertical spanning at top	2
tab (*x*)	Change data separator character	1
T{ T}	Text block	3
v V	Vertical spacing change	2
w W	Minimum width value	2
.*xx*	Included *troff* command	3
|	Vertical line	2
||	Double vertical line	2
^	Vertical span	2
\^	Vertical span	3
=	Double horizontal line	2,3
_	Horizontal line	2,3
_	Short horizontal line	3

A System for Typesetting Mathematics

Brian W. Kernighan and Lorinda L. Cherry

ABSTRACT

This paper describes the design and implementation of a system for typesetting mathematics. The language has been designed to be easy to learn and to use by people (for example, secretaries and mathematical typists) who know neither mathematics nor typesetting. Experience indicates that the language can be learned in an hour or so, for it has few rules and fewer exceptions. For typical expressions, the size and font changes, positioning, line drawing, and the like necessary to print according to mathematical conventions are all done automatically. For example, the input

sum from i=0 to infinity x sub i = pi over 2

produces

$$\sum_{i=0}^{\infty} x_i = \frac{\pi}{2}$$

The syntax of the language is specified by a small context-free grammar; a compiler-compiler is used to make a compiler that translates this language into typesetting commands. Output may be produced on either a phototypesetter or on a terminal with forward and reverse half-line motions. The system interfaces directly with text formatting programs, so mixtures of text and mathematics may be handled simply.

This paper is a revision of a paper originally published in CACM, March, 1975.

1. Introduction

"Mathematics is known in the trade as *difficult,* or *penalty, copy* because it is slower, more difficult, and more expensive to set in type than any other kind of copy normally occurring in books and journals." [1]

One difficulty with mathematical text is the multiplicity of characters, sizes, and fonts. An expression such as

$$\lim_{x \to \pi/2} (\tan x)^{\sin 2x} = 1$$

requires an intimate mixture of roman, italic and greek letters, in three sizes, and a special character or two. ("Requires" is perhaps the wrong word, but mathematics has its own typographical conventions which are quite different from those of ordinary text.) Typesetting such an expression by traditional methods is still an essentially manual operation.

A second difficulty is the two dimensional character of mathematics, which the superscript and limits in the preceding example showed in its simplest form. This is carried further by

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{a_2 + \cfrac{b_3}{a_3 + \cdots}}}$$

and still further by

$$\int \frac{dx}{ae^{mx} - be^{-mx}} = \begin{cases} \dfrac{1}{2m\sqrt{ab}} \log \dfrac{\sqrt{a}\,e^{mx} - \sqrt{b}}{\sqrt{a}\,e^{mx} + \sqrt{b}} \\[2ex] \dfrac{1}{m\sqrt{ab}} \tanh^{-1}\left(\dfrac{\sqrt{a}}{\sqrt{b}} e^{mx}\right) \\[2ex] \dfrac{-1}{m\sqrt{ab}} \coth^{-1}\left(\dfrac{\sqrt{a}}{\sqrt{b}} e^{mx}\right) \end{cases}$$

These examples also show line-drawing, built-up characters like braces and radicals, and a spectrum of positioning problems. (Section 6 shows what a user has to type to produce these on our system.)

2. Photocomposition

Photocomposition techniques can be used to solve some of the problems of typesetting mathematics. A phototypesetter is a device which exposes a piece of photographic paper or film, placing characters wherever they are wanted. The Graphic Systems phototypesetter[2] on the UNIX operating system[3] works by shining light through a character stencil. The character is made the right size by lenses, and the light beam directed by fiber optics to the desired place on a piece of photographic paper. The exposed paper is developed and typically used in some form of photo-offset reproduction.

On UNIX, the phototypesetter is driven by a formatting program called TROFF [4]. TROFF was designed for setting running text. It also provides all of the facilities that one needs for doing mathematics, such as arbitrary horizontal and vertical motions, line-drawing, size changing, but the syntax for describing these special operations is difficult to learn, and difficult even for experienced users to type correctly.

For this reason we decided to use TROFF as an "assembly language," by designing a language for describing mathematical expressions, and compiling it into TROFF.

3. Language Design

The fundamental principle upon which we based our language design is that the language should be easy to use by people (for example, secretaries) who know neither mathematics nor typesetting.

This principle implies several things. First, "normal" mathematical conventions about operator precedence, parentheses, and the like cannot be used, for to give special meaning to such characters means that the user has to understand what he or she is typing. Thus the language should not assume, for instance, that parentheses are always balanced, for they are not in the half-open interval $(a,b]$. Nor should it assume that that $\sqrt{a+b}$ can be replaced by $(a+b)^{\frac{1}{2}}$, or that $1/(1-x)$ is better written as $\frac{1}{1-x}$ (or vice versa).

Second, there should be relatively few rules, keywords, special symbols and operators, and the like. This keeps the language easy to learn and remember. Furthermore, there should be few exceptions to the rules that do exist: if something works in one situation, it should work everywhere. If a variable can have a subscript, then a subscript can have a subscript, and so on without limit.

Third, "standard" things should happen automatically. Someone who types "x=y+z+1" should get "$x=y+z+1$". Subscripts and superscripts should automatically be printed in an appropriately smaller size, with no special intervention. Fraction bars have to be made the right length and positioned at the right height. And so on. Indeed a mechanism for overriding default actions has to exist, but its application is the exception, not the rule.

We assume that the typist has a reasonable picture (a two-dimensional representation) of the desired final form, as might be handwritten by the author of a paper. We also assume that the input is typed on a computer terminal much like an ordinary typewriter. This implies an input alphabet of perhaps 100 characters, none of them special.

A secondary, but still important, goal in our design was that the system should be easy to implement, since neither of the authors had any desire to make a long-term project of it. Since our design was not firm, it was also necessary that the program be easy to change at any time.

To make the program easy to build and to change, and to guarantee regularity ("it should work everywhere"), the language is defined by a context-free grammar, described in Section 5. The compiler for the language was built using a compiler-compiler.

A priori, the grammar/compiler-compiler approach seemed the right thing to do. Our subsequent experience leads us to believe that any other course would have been folly. The original language was designed in a few days. Construction of a working system sufficient to try significant examples required perhaps a person-month. Since then, we have spent a modest amount of additional time over several years tuning, adding facilities, and occasionally changing the language as users make criticisms and suggestions.

We also decided quite early that we would let TROFF do our work for us whenever possible. TROFF is quite a powerful program, with a macro facility, text and arithmetic variables, numerical computation and testing, and conditional branching. Thus we have been able to avoid writing a lot of mundane but tricky software. For example, we store no text strings, but simply pass them on to TROFF. Thus we avoid having to write a storage management package. Furthermore, we have been able to isolate ourselves from most details of the particular device and character set currently in use. For example, we let TROFF compute the widths of all strings of characters; we need know nothing about them.

A third design goal is special to our environment. Since our program is only useful for typesetting mathematics, it is necessary that it interface cleanly with the underlying typesetting language for the benefit of users who want to set intermingled mathematics and text (the usual case). The standard mode of operation is that when a document is typed, mathematical expressions are input as part of the text, but marked by user settable delimiters. The program reads this input and treats as comments those things which are not mathematics, simply passing them through untouched. At the same time it converts the mathematical input into the necessary TROFF commands. The resulting ioutput is passed directly to TROFF where the comments and the mathematical parts both become text and/or TROFF commands.

4. The Language

We will not try to describe the language precisely here; interested readers may refer to the appendix for more details. Throughout this section, we will write expressions exactly as they are handed to the typesetting program (hereinafter called "EQN"), except that we won't show the delimiters that the user types to mark the beginning and end of the expression. The interface between EQN and TROFF is described at the end of this section.

As we said, typing x=y+z+1 should produce $x=y+z+1$, and indeed it does. Variables are made italic, operators and digits become roman, and normal spacings between letters and operators are altered slightly to give a more pleasing appearance.

Input is free-form. Spaces and new lines in the input are used by EQN to separate pieces of the input; they are not used to create space in the output. Thus

```
x    =    y
   + z + 1
```

also gives $x=y+z+1$. Free-form input is easier to type initially; subsequent editing is also easier, for an expression may be typed as many short lines.

Extra white space can be forced into the output by several characters of various sizes. A tilde " ~ " gives a space equal to the normal word spacing in text; a circumflex gives half this much, and a tab charcter spaces to the next tab stop.

Spaces (or tildes, etc.) also serve to delimit pieces of the input. For example, to get

$$f(t)=2\pi\int\sin(\omega t)dt$$

we write

```
f(t) = 2 pi int sin ( omega t )dt
```

Here spaces are *necessary* in the input to indicate that *sin, pi, int,* and *omega* are special, and potentially worth special treatment. EQN looks up each such string of characters in a table, and if appropriate gives it a translation. In this case, *pi* and *omega* become their greek equivalents, *int* becomes the integral sign (which must be moved down and enlarged so it looks "right"), and *sin* is made roman, following conventional mathematical practice. Parentheses, digits and operators are automatically made roman wherever found.

Fractions are specified with the keyword *over:*

```
a+b over c+d+e = 1
```

produces

$$\frac{a+b}{c+d+e}=1$$

Similarly, subscripts and superscripts are introduced by the keywords *sub* and *sup:*

$$x^2+y^2=z^2$$

is produced by

```
x sup 2 + y sup 2 = z sup 2
```

The spaces after the 2's are necessary to mark the end of the superscripts; similarly the keyword *sup* has to be marked off by spaces or some equivalent delimiter. The return to the proper baseline is automatic. Multiple levels of subscripts or superscripts are of course allowed: "x sup y sup z" is x^{y^z}. The construct "something *sub* something *sup* something" is recognized as a special case, so "x sub i sup 2" is x_i^2 instead of ${x_i}^2$.

More complicated expressions can now be formed with these primitives:

$$\frac{\partial^2 f}{\partial x^2}=\frac{x^2}{a^2}+\frac{y^2}{b^2}$$

is produced by

```
{partial sup 2 f} over {partial x sup 2} =
x sup 2 over a sup 2 + y sup 2 over b sup 2
```

Braces {} are used to group objects together; in this case they indicate unambiguously what goes over what on the left-hand side of the expression. The language defines the precedence of *sup* to be higher than that of *over*, so no braces are needed to get the correct association on the right side. Braces can always be used when in doubt about precedence.

The braces convention is an example of the power of using a recursive grammar to define the language. It is part of the language that if a construct can appear in some context, then *any expression* in braces can also occur in that context.

There is a *sqrt* operator for making square roots of the appropriate size: "sqrt a+b" produces $\sqrt{a+b}$, and x = {−b +− sqrt{b sup 2 −4ac}} over 2a

is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Since large radicals look poor on our typesetter, *sqrt* is not useful for tall expressions.

Limits on summations, integrals and similar constructions are specified with the keywords *from* and *to*. To get

$$\sum_{i=0}^{\infty} x_i \rightarrow 0$$

we need only type

sum from i=0 to inf x sub i −> 0

Centering and making the Σ big enough and the limits smaller are all automatic. The *from* and *to* parts are both optional, and the central part (e.g., the Σ) can in fact be anything:

lim from {x −> pi /2} (tan~x) = inf

is

$$\lim_{x \rightarrow \pi/2} (\tan x) = \infty$$

Again, the braces indicate just what goes into the *from* part.

There is a facility for making braces, brackets, parentheses, and vertical bars of the right height, using the keywords *left* and *right:*

left [x+y over 2a right]~=~1

makes

$$\left[\frac{x+y}{2a} \right] = 1$$

A *left* need not have a corresponding *right,* as we shall see in the next example. Any characters may follow *left* and *right,* but generally only various parentheses and bars are meaningful.

Big brackets, etc., are often used with another facility, called *piles,* which make vertical piles of objects. For example, to get

$$sign(x) \equiv \left\{ \begin{array}{rll} 1 & \text{if} & x>0 \\ 0 & \text{if} & x=0 \\ -1 & \text{if} & x<0 \end{array} \right.$$

we can type

sign (x) ~==~ left {
 rpile {1 above 0 above −1}
 ~~lpile {if above if above if}
 ~~lpile {x>0 above x=0 above x<0}

The construction "left {" makes a left brace big enough to enclose the "rpile {...}", which is a right-justified pile of "above ... above ...". "lpile" makes a left-justified pile. There are also centered piles. Because of the recursive language definition, a pile can contain any number of elements; any element of a pile can of course contain piles.

Although EQN makes a valiant attempt to use the right sizes and fonts, there are times when the default assumptions are simply not what is wanted. For instance the italic *sign* in the previous example would conventionally be in roman. Slides and transparencies often require larger characters than normal text. Thus we also provide size and font changing commands: "size 12 bold {A~x~=~y}" will produce **A x = y**. *Size* is followed by a number representing a character size in points. (One point is 1/72 inch; this paper is set in 9 point type.)

If necessary, an input string can be quoted in "...", which turns off grammatical significance, and any font or spacing changes that might otherwise be done on it. Thus we can say

lim~ roman "sup" ~x sub n = 0

to ensure that the supremum doesn't become a superscript:

$$\lim \mathrm{sup}\ x_n = 0$$

Diacritical marks, long a problem in traditional typesetting, are straightforward:

$$\underset{\cdot}{x} + \hat{x} + \tilde{y} + \hat{X} + \ddot{Y} = \overline{z+Z}$$

is made by typing

x dot under + x hat + y tilde
+ X hat + Y dotdot = z+Z bar

There are also facilities for globally changing default sizes and fonts, for example for making viewgraphs or for setting chemical equations. The language allows for matrices, and for lining up equations at the same horizontal position.

Finally, there is a definition facility, so a user can say

define name "..."

at any time in the document; henceforth, any occurrence of the token "name" in an expression will be expanded into whatever was inside the double quotes in its definition. This lets users tailor the language to their own specifications, for it is quite possible to redefine keywords like *sup* or *over.* Section 6 shows an example of definitions.

The EQN preprocessor reads intermixed text and equations, and passes its output to TROFF. Since TROFF uses lines beginning with a period as control words (e.g., ".ce" means "center the next output line"), EQN uses the sequence ".EQ" to mark the beginning of an equation and ".EN" to mark the end. The ".EQ" and ".EN" are passed through to TROFF untouched, so they can also be used by a knowledgeable user to center equations, number them automatically, etc. By default, however, ".EQ" and ".EN" are simply ignored by TROFF, so by default equations are printed in-line.

".EQ" and ".EN" can be supplemented by TROFF commands as desired; for example, a centered display equation can be produced with the input:

```
.ce
.EQ
x sub i = y sub i ...
.EN
```

Since it is tedious to type ".EQ" and ".EN" around very short expressions (single letters, for instance), the user can also define two characters to serve as the left and right delimiters of expressions. These characters are recognized anywhere in subsequent text. For example if the left and right delimiters have both been set to "#", the input:

Let #x sub i#, #y# and #alpha# be positive

produces:

Let x_i, y and α be positive

Running a preprocessor is strikingly easy on UNIX. To typeset text stored in file "f ", one issues the command:

```
eqn f | troff
```

The vertical bar connects the output of one process (EQN) to the input of another (TROFF).

5. Language Theory

The basic structure of the language is not a particularly original one. Equations are pictured as a set of "boxes," pieced together in various ways. For example, something with a subscript is just a box followed by another box moved downward and shrunk by an appropriate amount. A fraction is just a box centered above another box, at the right altitude, with a line of correct length drawn between them.

The grammar for the language is shown below. For purposes of exposition, we have collapsed some productions. In the original grammar, there are about 70 productions, but many of these are simple ones used only to guarantee that some keyword is recognized early enough in the parsing process. Symbols in capital letters are terminal symbols; lower case symbols are non-terminals, i.e., syntactic categories. The vertical bar | indicates an alternative; the brackets [ ] indicate optional material. A TEXT is a string of non-blank characters or any string inside double quotes; the other terminal symbols represent literal occurrences of the corresponding keyword.

```
eqn  : box | eqn box
box  : text
     | { eqn }
     | box OVER box
     | SQRT box
     | box SUB box | box SUP box
     | [ L | C | R ]PILE { list }
     | LEFT text eqn [ RIGHT text ]
     | box [ FROM box ] [ TO box ]
     | SIZE text box
     | [ROMAN | BOLD | ITALIC] box
     | box [HAT | BAR | DOT | DOTDOT | TILDE]
     | DEFINE text text
list : eqn | list ABOVE eqn
text : TEXT
```

The grammar makes it obvious why there are few exceptions. For example, the observation that something can be replaced by a more complicated something in braces is implicit in the productions:

```
eqn : box | eqn box
box : text | { eqn }
```

Anywhere a single character could be used, *any* legal construction can be used.

Clearly, our grammar is highly ambiguous. What, for instance, do we do with the input

a over b over c ?

Is it

{a over b} over c

or is it

a over {b over c} ?

To answer questions like this, the grammar is supplemented with a small set of rules that describe the precedence and associativity of operators. In particular, we specify (more or less arbitrarily) that *over* associates to the left, so the first alternative above is the one chosen. On the other hand, *sub* and *sup* bind to the right, because this is closer to standard mathematical practice. That is, we assume x^{a^b} is $x^{(a^b)}$, not $(x^a)^b$.

The precedence rules resolve the ambiguity in a construction like

a sup 2 over b

We define *sup* to have a higher precedence than *over*, so this construction is parsed as $\frac{a^2}{b}$ instead of $a^{\frac{2}{b}}$.

Naturally, a user can always force a particular parsing by placing braces around expressions.

The ambiguous grammar approach seems to be quite useful. The grammar we use is small enough to be easily understood, for it contains none of the productions that would be normally used for resolving ambiguity. Instead the supplemental information about precedence and associativity (also small enough to be understood) provides the compiler-compiler with the information it needs to make a fast, deterministic parser for the specific language we want. When the language is supplemented by the disambiguating rules, it is in fact LR(1) and thus easy to parse[5].

The output code is generated as the input is scanned. Any time a production of the grammar is recognized, (potentially) some TROFF commands are output. For example, when the lexical analyzer reports that it has found a TEXT (i.e., a string of contiguous characters), we have recognized the production:

```
text    : TEXT
```

The translation of this is simple. We generate a local name for the string, then hand the name and the string to TROFF, and let TROFF perform the storage management. All we save is the name of the string, its height, and its baseline.

As another example, the translation associated with the production

```
box    : box OVER box
```

is:

```
Width of output box =
  slightly more than largest input width
Height of output box =
  slightly more than sum of input heights
Base of output box =
  slightly more than height of bottom input box
String describing output box =
  move down;
  move right enough to center bottom box;
  draw bottom box (i.e., copy string for bottom box);
  move up; move left enough to center top box;
  draw top box (i.e., copy string for top box);
  move down and left; draw line full width;
  return to proper base line.
```

Most of the other productions have equally simple semantic actions. Picturing the output as a set of properly placed boxes makes the right sequence of positioning commands quite obvious. The main difficulty is in finding the right numbers to use for esthetically pleasing positioning.

With a grammar, it is usually clear how to extend the language. For instance, one of our users suggested a TENSOR operator, to make constructions like

$${}^{l}_{m}\overset{k}{\underset{n}{\mathrm{T}}}{}^{j}_{i}$$

Grammatically, this is easy: it is sufficient to add a production like

```
box    : TENSOR { list }
```

Semantically, we need only juggle the boxes to the right places.

6. Experience

There are really three aspects of interest—how well EQN sets mathematics, how well it satisfies its goal of being "easy to use," and how easy it was to build.

The first question is easily addressed. This entire paper has been set by the program. Readers can judge for themselves whether it is good enough for their purposes. One of our users commented that although the output is not as good as the best hand-set material, it is still better than average, and much better than the worst. In any case, who cares? Printed books cannot compete with the birds and flowers of illuminated manuscripts on esthetic grounds, either, but they have some clear economic advantages.

Some of the deficiencies in the output could be cleaned up with more work on our part. For example, we sometimes leave too much space between a roman letter and an italic one. If we were willing to keep track of the fonts involved, we could do this better more of the time.

Some other weaknesses are inherent in our output device. It is hard, for instance, to draw a line of an arbitrary length without getting a perceptible overstrike at one end.

As to ease of use, at the time of writing, the system has been used by two distinct groups. One user population consists of mathematicians, chemists, physicists, and computer scientists. Their typical reaction has been something like:

(1) It's easy to write, although I make the following mistakes...

(2) How do I do...?

(3) It botches the following things.... Why don't you fix them?

(4) You really need the following features...

The learning time is short. A few minutes gives the general flavor, and typing a page or two of a paper generally uncovers most of the misconceptions about how it works.

The second user group is much larger, the secretaries and mathematical typists who were the original target of the system. They tend to be enthusiastic converts. They find the language easy to learn (most are largely self-taught), and have little trouble producing the output they want. They are of course less critical of the esthetics of their output than users trained in mathematics. After a transition period, most find using a computer more interesting than a regular typewriter.

The main difficulty that users have seems to be remembering that a blank is a delimiter; even experienced users use blanks where they shouldn't and omit them when they are needed. A common instance is typing

```
f(x sub i)
```

which produces

$$f(x_{i)}$$

instead of

$$f(x_i)$$

Since the EQN language knows no mathematics, it cannot deduce that the right parenthesis is not part of the subscript.

The language is somewhat prolix, but this doesn't seem excessive considering how much is being done, and it is certainly more compact than the corresponding TROFF commands. For example, here is the source for the continued fraction expression in Section 1 of this paper:

```
a sub 0 + b sub 1 over
  {a sub 1 + b sub 2 over
    {a sub 2 + b sub 3 over
      {a sub 3 + ... }}}
```

This is the input for the large integral of Section 1; notice the use of definitions:

```
define emx "{e sup mx}"
define mab "{m sqrt ab}"
define sa "{sqrt a}"
define sb "{sqrt b}"
int dx over {a emx − be sup −mx} ˜=˜
left { lpile {
    1 over {2 mab} ˜log˜
        {sa emx − sb} over {sa emx + sb}
  above
    1 over mab ˜ tanh sup −1 ( sa over sb emx )
  above
    −1 over mab ˜ coth sup −1 ( sa over sb emx )
}
```

As to ease of construction, we have already mentioned that there are really only a few person-months invested. Much of this time has gone into two things—fine-tuning (what is the most esthetically pleasing space to use between the numerator and denominator of a fraction?), and changing things found deficient by our users (shouldn't a tilde be a delimiter?).

The program consists of a number of small, essentially unconnected modules for code generation, a simple lexical analyzer, a canned parser which we did not have to write, and some miscellany associated with input files and the macro facility. The program is now about 1600 lines of C [6], a high-level language reminiscent of BCPL. About 20 percent of these lines are "print" statements, generating the output code.

The semantic routines that generate the actual TROFF commands can be changed to accommodate other formatting languages and devices. For example, in less than 24 hours, one of us changed the entire semantic package to drive NROFF, a variant of TROFF, for typesetting mathematics on teletypewriter devices capable of reverse line motions. Since many potential users do not have access to a typesetter, but still have to type mathematics, this provides a way to get a typed version of the final output which is close enough for debugging purposes, and sometimes even for ultimate use.

7. Conclusions

We think we have shown that it is possible to do acceptably good typesetting of mathematics on a phototypesetter, with an input language that is easy to learn and use and that satisfies many users' demands. Such a package can be implemented in short order, given a compiler-compiler and a decent typesetting program underneath.

Defining a language, and building a compiler for it with a compiler-compiler seems like the only sensible way to do business. Our experience with the use of a grammar and a compiler-compiler has been uniformly favorable. If we had written everything into code directly, we would have been locked into our original design. Furthermore, we would have never been sure where the exceptions and special cases were. But because we have a grammar, we can change our minds readily and still be reasonably sure that if a construction works in one place it will work everywhere.

Acknowledgements

We are deeply indebted to J. F. Ossanna, the author of TROFF, for his willingness to modify TROFF to make our task easier and for his continuous assistance during the development of our program. We are also grateful to A. V. Aho for help with language theory, to S. C. Johnson for aid with the compiler-compiler, and to our early users A. V. Aho, S. I. Feldman, S. C. Johnson, R. W. Hamming, and M. D. McIlroy for their constructive criticisms.

References

[1] *A Manual of Style*, 12th Edition. University of Chicago Press, 1969. p 295.

[2] *Model C/A/T Phototypesetter.* Graphic Systems, Inc., Hudson, N. H.

[3] Ritchie, D. M., and Thompson, K. L., "The UNIX time-sharing system." *Comm. ACM 17,* 7 (July 1974), 365-375.

[4] Ossanna, J. F., TROFF User's Manual. Bell Laboratories Computing Science Technical Report 54, 1977.

[5] Aho, A. V., and Johnson, S. C., "LR Parsing." *Comp. Surv. 6,* 2 (June 1974), 99-124.

[6] B. W. Kernighan and D. M. Ritchie, *The C Programming Language.* Prentice-Hall, Inc., 1978.

Typesetting Mathematics — User's Guide
(Second Edition)

Brian W. Kernighan and Lorinda L. Cherry

ABSTRACT

This is the user's guide for a system for typesetting mathematics, using the phototypesetters on the UNIX† and GCOS operating systems.

Mathematical expressions are described in a language designed to be easy to use by people who know neither mathematics nor typesetting. Enough of the language to set in-line expressions like $\lim_{x \to \pi/2} (\tan x)^{\sin 2x} = 1$ or display equations like

$$G(z) = e^{\ln G(z)} = \exp \left(\sum_{k \geq 1} \frac{S_k z^k}{k} \right) = \prod_{k \geq 1} e^{S_k z^k / k}$$

$$= \left(1 + S_1 z + \frac{S_1^2 z^2}{2!} + \cdots \right) \left(1 + \frac{S_2 z^2}{2} + \frac{S_2^2 z^4}{2^2 \cdot 2!} + \cdots \right) \cdots$$

$$= \sum_{m \geq 0} \left(\sum_{\substack{k_1, k_2, \ldots, k_m \geq 0 \\ k_1 + 2k_2 + \cdots + mk_m = m}} \frac{S_1^{k_1}}{1^{k_1} k_1 !} \frac{S_2^{k_2}}{2^{k_2} k_2 !} \cdots \frac{S_m^{k_m}}{m^{k_m} k_m !} \right) z^m$$

can be learned in an hour or so.

The language interfaces directly with the phototypesetting language TROFF, so mathematical expressions can be embedded in the running text of a manuscript, and the entire document produced in one process. This user's guide is an example of its output.

The same language may be used with the UNIX formatter NROFF to set mathematical expressions on DASI and GSI terminals and Model 37 teletypes.

September 15, 1986

† UNIX is a trademark of Bell Laboratories.

Typesetting Mathematics — User's Guide (Second Edition)

Brian W. Kernighan and Lorinda L. Cherry

1. Introduction

EQN is a program for typesetting mathematics on the Graphics Systems phototypesetters on UNIX and GCOS. The EQN language was designed to be easy to use by people who know neither mathematics nor typesetting. Thus EQN knows relatively little about mathematics. In particular, mathematical symbols like +, −, ×, parentheses, and so on have no special meanings. EQN is quite happy to set garbage (but it will look good).

EQN works as a preprocessor for the typesetter formatter, TROFF[1], so the normal mode of operation is to prepare a document with both mathematics and ordinary text interspersed, and let EQN set the mathematics while TROFF does the body of the text.

On UNIX, EQN will also produce mathematics on DASI and GSI terminals and on Model 37 teletypes. The input is identical, but you have to use the programs NEQN and NROFF instead of EQN and TROFF. Of course, some things won't look as good because terminals don't provide the variety of characters, sizes and fonts that a typesetter does, but the output is usually adequate for proofreading.

To use EQN on UNIX,

```
eqn files | troff
```

GCOS use is discussed in section 26.

2. Displayed Equations

To tell EQN where a mathematical expression begins and ends, we mark it with lines beginning .EQ and .EN. Thus if you type the lines

```
.EQ
x=y+z
.EN
```

your output will look like

$$x=y+z$$

The .EQ and .EN are copied through untouched; they are not otherwise processed by EQN. This means that you have to take care of things like centering, numbering, and so on yourself. The most common way is to use the TROFF and NROFF macro package package '−ms' developed by M. E. Lesk[3], which allows you to center, indent, left-justify and number equations.

With the '−ms' package, equations are centered by default. To left-justify an equation, use .EQ L instead of .EQ. To indent it, use .EQ I. Any of these can be followed by an arbitrary 'equation number' which will be placed at the right margin. For example, the input

```
.EQ I (3.1a)
x = f(y/2) + y/2
.EN
```

produces the output

$$x=f\,(y\,/2)+y\,/2 \qquad (3.1a)$$

There is also a shorthand notation so in-line expressions like π_i^2 can be entered without .EQ and .EN. We will talk about it in section 19.

3. Input spaces

Spaces and newlines within an expression are thrown away by EQN. (Normal text is left absolutely alone.) Thus between .EQ and .EN,

```
x=y+z
```

and

```
x = y + z
```

and

```
x  =  y
   + z
```

and so on all produce the same output

$$x=y+z$$

You should use spaces and newlines freely to make your input equations readable and easy to edit. In particular, very long lines are a bad idea, since they are often hard to fix if you make a mistake.

4. Output spaces

To force extra spaces into the *output*, use a tilde "~" for each space you want:

x~=~y~+~z

gives

$$x\ =\ y\ +\ z$$

You can also use a circumflex "^", which gives a space half the width of a tilde. It is mainly useful for fine-tuning. Tabs may also be used to position pieces of an expression, but the tab stops must be set by TROFF commands.

5. Symbols, Special Names, Greek

EQN knows some mathematical symbols, some mathematical names, and the Greek alphabet. For example,

x=2 pi int sin (omega t)dt

produces

$$x=2\pi\int\sin(\omega t)dt$$

Here the spaces in the input are **necessary** to tell EQN that *int, pi, sin* and *omega* are separate entities that should get special treatment. The *sin*, digit 2, and parentheses are set in roman type instead of italic; *pi* and *omega* are made Greek; and *int* becomes the integral sign.

When in doubt, leave spaces around separate parts of the input. A *very* common error is to type *f(pi)* without leaving spaces on both sides of the *pi*. As a result, EQN does not recognize *pi* as a special word, and it appears as $f\,(pi)$ instead of $f\,(\pi)$.

A complete list of EQN names appears in section 23. Knowledgeable users can also use TROFF four-character names for anything EQN doesn't know about, like *\(bs* for the Bell System sign ∅.

6. Spaces, Again

The only way EQN can deduce that some sequence of letters might be special is if that sequence is separated from the letters on either side of it. This can be done by surrounding a special word by ordinary spaces (or tabs or newlines), as we did in the previous section.

You can also make special words stand out by surrounding them with tildes or circumflexes:

x~=~2~pi~int~sin~(~omega~t~)~dt

is much the same as the last example, except that the tildes not only separate the magic words like *sin, omega,* and so on, but also add extra spaces, one space per tilde:

$$x\ =\ 2\ \pi\ \int\ \sin\ (\ \omega\ t\ )\ dt$$

Special words can also be separated by braces { } and double quotes "...", which have special meanings that we will see soon.

7. Subscripts and Superscripts

Subscripts and superscripts are obtained with the words *sub* and *sup*.

x sup 2 + y sub k

gives

$$x^2+y_k$$

EQN takes care of all the size changes and vertical motions needed to make the output look right. The words *sub* and *sup* must be surrounded by spaces; *x sub2* will give you *xsub*2 instead of x_2. Furthermore, don't forget to leave a space (or a tilde, etc.) to mark the end of a subscript or superscript. A common error is to say something like

y = (x sup 2)+1

which causes

$$y=(x^{2)+1}$$

instead of the intended

$$y=(x^2)+1$$

Subscripted subscripts and superscripted superscripts also work:

x sub i sub 1

is

$$x_{i_1}$$

A subscript and superscript on the same thing are printed one above the other if the subscript comes *first:*

x sub i sup 2

is

$$x_i^2$$

Other than this special case, ***sub*** and ***sup*** group to the right, so ***x sup y sub z*** means x^{y_z}, not ${x^y}_z$.

8. Braces for Grouping

Normally, the end of a subscript or superscript is marked simply by a blank (or tab or tilde, etc.) What if the subscript or superscript is something that has to be typed with blanks in it? In that case, you can use the braces { and } to mark the beginning and end of the subscript or superscript:

e sup {i omega t}

is

$$e^{i\omega t}$$

Rule: Braces can ***always*** be used to force EQN to treat something as a unit, or just to make your intent perfectly clear. Thus:

x sub {i sub 1} sup 2

is

$$x_{i_1}^2$$

with braces, but

x sub i sub 1 sup 2

is

$$x_{i_1^2}$$

which is rather different.

Braces can occur within braces if necessary:

e sup {i pi sup {rho +1}}

is

$$e^{i\pi^{\rho+1}}$$

The general rule is that anywhere you could use some single thing like x, you can use an arbitrarily complicated thing if you enclose it in braces. EQN will look after all the details of positioning it and making it the right size.

In all cases, make sure you have the right number of braces. Leaving one out or adding an extra will cause EQN to complain bitterly.

Occasionally you will have to print braces. To do this, enclose them in double quotes, like "{". Quoting is discussed in more detail in section 14.

9. Fractions

To make a fraction, use the word *over:*

a+b over 2c =1

gives

$$\frac{a+b}{2c}=1$$

The line is made the right length and positioned automatically. Braces can be used to make clear what goes over what:

{alpha + beta} over {sin (x)}

is

$$\frac{\alpha+\beta}{\sin(x)}$$

What happens when there is both an *over* and a ***sup*** in the same expression? In such an apparently ambiguous case, EQN does the ***sup*** before the *over*, so

−b sup 2 over pi

is $\frac{-b^2}{\pi}$ instead of $-b^{\frac{2}{\pi}}$ The rules which decide which operation is done first in cases like this are summarized in section 23. When in doubt, however, ***use braces*** to make clear what goes with what.

10. Square Roots

To draw a square root, use *sqrt:*

sqrt a+b + 1 over sqrt {ax sup 2 +bx+c}

is

$$\sqrt{a+b}+\frac{1}{\sqrt{ax^2+bx+c}}$$

Warning — square roots of tall quantities look lousy, because a root-sign big enough to cover the quantity is too dark and heavy:

sqrt {a sup 2 over b sub 2}

is

$$\sqrt{\frac{a^2}{b_2}}$$

Big square roots are generally better written as something to the power ½:

$$(a^2/b_2)^{½}$$

which is

(a sup 2 /b sub 2) sup half

11. Summation, Integral, Etc.

Summations, integrals, and similar constructions are easy:

sum from i=0 to {i= inf} x sup i

produces

$$\sum_{i=0}^{i=\infty} x^i$$

Notice that we used braces to indicate where the upper part $i=\infty$ begins and ends. No braces were necessary for the lower part $i=0$, because it contained no blanks. The braces will never hurt, and if the *from* and *to* parts contain any blanks, you must use braces around them.

The *from* and *to* parts are both optional, but if both are used, they have to occur in that order.

Other useful characters can replace the *sum* in our example:

int prod union inter

become, respectively,

$$\int \quad \prod \quad \bigcup \quad \bigcap$$

Since the thing before the *from* can be anything, even something in braces, *from-to* can often be used in unexpected ways:

lim from {n −> inf} x sub n =0

is

$$\lim_{n \to \infty} x_n = 0$$

12. Size and Font Changes

By default, equations are set in 10-point type (the same size as this guide), with standard mathematical conventions to determine what characters are in roman and what in italic. Although EQN makes a valiant attempt to use esthetically pleasing sizes and fonts, it is not perfect. To change sizes and fonts, use ***size n*** and ***roman, italic, bold*** and ***fat***. Like ***sub*** and ***sup***, size and font changes affect only the thing that follows them, and revert to the normal situation at the end of it. Thus

bold x y

is

$$\mathbf{x}y$$

and

size 14 bold x = y +
size 14 {alpha + beta}

gives

$$\mathbf{x}=y+\alpha+\beta$$

As always, you can use braces if you want to affect something more complicated than a single letter. For example, you can change the size of an entire equation by

size 12 { ... }

Legal sizes which may follow ***size*** are 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, 22, 24, 28, 36. You can also change the size *by* a given amount; for example, you can say *size +2* to make the size two points bigger, or *size −3* to make it three points smaller. This has the advantage that you don't have to know what the current size is.

If you are using fonts other than roman, italic and bold, you can say *font X* where *X* is a one character TROFF name or number for the font. Since EQN is tuned for roman, italic and bold, other fonts may not give quite as good an appearance.

The *fat* operation takes the current font and widens it by overstriking: *fat grad* is ∇ and *fat {x sub i}* is x_i.

If an entire document is to be in a nonstandard size or font, it is a severe nuisance to have to write out a size and font change for each equation. Accordingly, you can set a "global" size or font which thereafter affects all equations. At the beginning of any equation, you might say, for instance,

```
.EQ
gsize 16
gfont R
...
.EN
```

to set the size to 16 and the font to roman thereafter. In place of R, you can use any of the TROFF font names. The size after *gsize* can be a relative change with + or −.

Generally, *gsize* and *gfont* will appear at the beginning of a document but they can also appear thoughout a document: the global font and size can be changed as often as needed. For example, in a footnote‡ you will typically want the size of equations to match the size of the footnote text, which is two points smaller than the main text. Don't forget to reset the global size at the end of the footnote.

13. Diacritical Marks

To get funny marks on top of letters, there are several words:

x dot	$\dot{x}$
x dotdot	$\ddot{x}$
x hat	$\hat{x}$
x tilde	$\tilde{x}$
x vec	$\vec{x}$
x dyad	$\overleftrightarrow{x}$
x bar	$\bar{x}$
x under	$\underline{x}$

The diacritical mark is placed at the right height. The *bar* and *under* are made the right length for the entire construct, as in $\overline{x+y+z}$; other marks are centered.

14. Quoted Text

Any input entirely within quotes ("...") is not subject to any of the font changes and spacing adjustments normally done by the equation setter. This provides a way to do your own spacing and adjusting if needed:

italic "sin(x)" + sin (x)

is

$$sin(x) + \sin(x)$$

‡Like this one, in which we have a few random expressions like x_i and π^2. The sizes for these were set by the command *gsize −2*.

Quotes are also used to get braces and other EQN keywords printed:

"{ size alpha }"

is

{ *size alpha* }

and

roman "{ size alpha }"

is

{ size alpha }

The construction "" is often used as a place-holder when grammatically EQN needs something, but you don't actually want anything in your output. For example, to make ^{2}He, you can't just type *sup 2 roman He* because a *sup* has to be a superscript *on* something. Thus you must say

"" sup 2 roman He

To get a literal quote use "\"". TROFF characters like \(*bs* can appear unquoted, but more complicated things like horizontal and vertical motions with *h* and *v* should always be quoted. (If you've never heard of *h* and *v*, ignore this section.)

15. Lining Up Equations

Sometimes it's necessary to line up a series of equations at some horizontal position, often at an equals sign. This is done with two operations called *mark* and *lineup*.

The word *mark* may appear once at any place in an equation. It remembers the horizontal position where it appeared. Successive equations can contain one occurrence of the word *lineup*. The place where *lineup* appears is made to line up with the place marked by the previous *mark* if at all possible. Thus, for example, you can say

```
.EQ I
x+y mark = z
.EN
.EQ I
x lineup = 1
.EN
```

to produce

$$x+y=z$$
$$\quad\;\; x=1$$

For reasons too complicated to talk about, when you use EQN and '−ms', use either .EQ I or .EQ L. mark and *lineup* don't work with centered equations. Also bear in mind that *mark* doesn't look ahead;

```
x mark =1
...
x+y lineup =z
```

isn't going to work, because there isn't room for the $x+y$ part after the *mark* remembers where the x is.

16. Big Brackets, Etc.

To get big brackets [ ], braces { }, parentheses ( ), and bars | | around things, use the *left* and *right* commands:

```
left { a over b + 1 right }
 ~=~ left ( c over d right )
 + left [ e right ]
```

is

$$\left\{ \frac{a}{b}+1 \right\} = \left(\frac{c}{d} \right) + \left[e \right]$$

The resulting brackets are made big enough to cover whatever they enclose. Other characters can be used besides these, but the are not likely to look very good. One exception is the *floor* and *ceiling* characters:

```
left floor x over y right floor
<= left ceiling a over b right ceiling
```

produces

$$\left\lfloor \frac{x}{y} \right\rfloor \le \left\lceil \frac{a}{b} \right\rceil$$

Several warnings about brackets are in order. First, braces are typically bigger than brackets and parentheses, because they are made up of three, five, seven, etc., pieces, while brackets can be made up of two, three, etc. Second, big left and right parentheses often look poor, because the character set is poorly designed.

The *right* part may be omitted: a "left something" need not have a corresponding "right something". If the *right* part is omitted, put braces around the thing you want the left bracket to encompass. Otherwise, the resulting brackets may be too large.

If you want to omit the *left* part, things are more complicated, because technically you can't have a *right* without a corresponding *left*. Instead you have to say

```
left "" ..... right )
```

for example. The *left* "" means a "left nothing". This satisfies the rules without hurting your output.

17. Piles

There is a general facility for making vertical piles of things; it comes in several flavors. For example:

```
A ~=~ left [
  pile { a above b above c }
  ~~ pile { x above y above z }
right ]
```

will make

$$A = \left[\begin{matrix} a & x \\ b & y \\ c & z \end{matrix} \right]$$

The elements of the pile (there can be as many as you want) are centered one above another, at the right height for most purposes. The keyword *above* is used to separate the pieces; braces are used around the entire list. The elements of a pile can be as complicated as needed, even containing more piles.

Three other forms of pile exist: *lpile* makes a pile with the elements left-justified; *rpile* makes a right-justified pile; and *cpile* makes a centered pile, just like *pile*. The vertical spacing between the pieces is somewhat larger for *l*-, *r*- and *cpiles* than it is for ordinary piles.

```
roman sign (x)~=~
left {
   lpile {1 above 0 above -1}
   ~~ lpile
   {if~x>0 above if~x=0 above if~x<0}
```

makes

$$\mathrm{sign}(x) = \left\{ \begin{array}{ll} 1 & \text{if } x>0 \\ 0 & \text{if } x=0 \\ -1 & \text{if } x<0 \end{array} \right.$$

Notice the left brace without a matching right one.

18. Matrices

It is also possible to make matrices. For example, to make a neat array like

$$\begin{matrix} x_i & x^2 \\ y_i & y^2 \end{matrix}$$

you have to type

```
matrix {
  ccol { x sub i above y sub i }
  ccol { x sup 2 above y sup 2 }
}
```

This produces a matrix with two centered columns. The elements of the columns are then listed just as for a pile, each element separated by the word *above*. You can also use *lcol* or *rcol* to left or right adjust columns. Each column can be separately adjusted, and there can be as many columns as you like.

The reason for using a matrix instead of two adjacent piles, by the way, is that if the elements of the piles don't all have the same height, they won't line up properly. A matrix forces them to line up, because it looks at the entire structure before deciding what spacing to use.

A word of warning about matrices — *each column must have the same number of elements in it.* The world will end if you get this wrong.

19. Shorthand for In-line Equations

In a mathematical document, it is necessary to follow mathematical conventions not just in display equations, but also in the body of the text, for example by making variable names like x italic. Although this could be done by surrounding the appropriate parts with .EQ and .EN, the continual repetition of .EQ and .EN is a nuisance. Furthermore, with '−ms', .EQ and .EN imply a displayed equation.

EQN provides a shorthand for short in-line expressions. You can define two characters to mark the left and right ends of an in-line equation, and then type expressions right in the middle of text lines. To set both the left and right characters to dollar signs, for example, add to the beginning of your document the three lines

```
.EQ
delim $$
.EN
```

Having done this, you can then say things like

```
Let $alpha sub i$ be the primary
variable, and let $beta$ be zero.
Then we can show that $x sub 1$ is
$>=0$.
```

This works as you might expect — spaces, newlines, and so on are significant in the text, but not in the equation part itself. Multiple equations can occur in a single input line.

Enough room is left before and after a line that contains in-line expressions that something like $\sum_{i=1}^{n} x_i$ does not interfere with the lines surrounding it.

To turn off the delimiters,

```
.EQ
delim off
.EN
```

Warning: don't use braces, tildes, circumflexes, or double quotes as delimiters — chaos will result.

20. Definitions

EQN provides a facility so you can give a frequently-used string of characters a name, and thereafter just type the name instead of the whole string. For example, if the sequence

```
x sub i sub 1 + y sub i sub 1
```

appears repeatedly throughout a paper, you can save re-typing it each time by defining it like this:

```
define xy 'x sub i sub 1 + y sub i sub 1'
```

This makes xy a shorthand for whatever characters occur between the single quotes in the definition. You can use any character instead of quote to mark the ends of the definition, so long as it doesn't appear inside the definition.

Now you can use xy like this:

```
.EQ
f(x) = xy ...
.EN
```

and so on. Each occurrence of xy will expand into what it was defined as. Be careful to leave spaces or their equivalent around the name when you actually use it, so EQN will be able to identify it as special.

There are several things to watch out for. First, although definitions can use previous definitions, as in

```
.EQ
define xi ' x sub i '
define xi1 ' xi sub 1 '
.EN
```

don't define something in terms of itself' A favorite error is to say

```
define X ' roman X '
```

This is a guaranteed disaster, since X *is* now defined in terms of itself. If you say

```
define X ' roman "X" '
```

however, the quotes protect the second X, and everything works fine.

EQN keywords can be redefined. You can make / mean *over* by saying

```
define / ' over '
```

or redefine *over* as / with

```
define over ' / '
```

If you need different things to print on a terminal and on the typesetter, it is sometimes worth defining a symbol differently in NEQN and EQN. This can be done with *ndefine* and *tdefine*. A definition made with *ndefine* only takes effect if you are running NEQN; if you use *tdefine*, the definition only applies for EQN. Names defined with plain *define* apply to both EQN and NEQN.

21. Local Motions

Although EQN tries to get most things at the right place on the paper, it isn't perfect, and occasionally you will need to tune the output to make it just right. Small extra horizontal spaces can be obtained with tilde and circumflex. You can also say *back n* and *fwd n* to move small amounts horizontally. *n* is how far to move in 1/100's of an em (an em is about the width of the letter 'm'.) Thus *back 50* moves back about half the width of an m. Similarly you can move things up or down with *up n* and *down n*. As with *sub* or *sup*, the local motions affect the next thing in the input, and this can be something arbitrarily complicated if it is enclosed in braces.

22. A Large Example

Here is the complete source for the three display equations in the abstract of this guide.

```
.EQ I
G(z)~mark =~ e sup { ln ~ G(z) }
~=~ exp left (
sum from k>=1 {S sub k z sup k} over k right )
~=~  prod from k>=1 e sup {S sub k z sup k /k}
.EN
.EQ I
lineup = left ( 1 + S sub 1 z +
{ S sub 1 sup 2 z sup 2 } over 2! + ... right )
left ( 1+ { S sub 2 z sup 2 } over 2
+ { S sub 2 sup 2 z sup 4 } over { 2 sup 2 cdot 2! }
+ ... right ) ...
.EN
.EQ I
lineup =  sum from m>=0 left (
sum from
pile { k sub 1 ,k sub 2 ,..., k sub m  >=0
above
k sub 1 +2k sub 2 + ... +mk sub m =m}
{ S sub 1 sup {k sub 1} } over {1 sup k sub 1 k sub 1 ! } ~
{ S sub 2 sup {k sub 2} } over {2 sup k sub 2 k sub 2 ! } ~
...
{ S sub m sup {k sub m} } over {m sup k sub m k sub m ! }
right ) z sup m
.EN
```

23. Keywords, Precedences, Etc.

If you don't use braces, EQN will do operations in the order shown in this list.

dyad vec under bar tilde hat dot dotdot
fwd back down up
fat roman italic bold size
sub sup sqrt over
from to

These operations group to the left:

over sqrt left right

All others group to the right.

Digits, parentheses, brackets, punctuation marks, and these mathematical words are converted to Roman font when encountered:

sin cos tan sinh cosh tanh arc
max min lim log ln exp
Re Im and if for det

These character sequences are recognized and translated as shown.

>=	$\geq$
<=	$\leq$
==	$\equiv$
!=	$\neq$
+-	$\pm$
->	$\rightarrow$

<-	$\leftarrow$
<<	$\ll$
>>	$\gg$
inf	∞
partial	∂
half	½
prime	$\prime$
approx	$\approx$
nothing	
cdot	$\cdot$
times	$\times$
del	∇
grad	∇
...	$\cdots$
,...,	$,\ldots,$
sum	$\sum$
int	$\int$
prod	$\prod$
union	$\bigcup$
inter	$\bigcap$

To obtain Greek letters, simply spell them out in whatever case you want:

DELTA	Δ	iota	ι
GAMMA	Γ	kappa	κ
LAMBDA	Λ	lambda	λ
OMEGA	Ω	mu	μ
PHI	Φ	nu	ν
PI	Π	omega	ω
PSI	Ψ	omicron	o
SIGMA	Σ	phi	ϕ
THETA	Θ	pi	π
UPSILON	Υ	psi	ψ
XI	Ξ	rho	ρ
alpha	α	sigma	σ
beta	β	tau	τ
chi	χ	theta	θ
delta	δ	upsilon	υ
epsilon	ϵ	xi	ξ
eta	η	zeta	ζ
gamma	γ		

These are all the words known to EQN (except for characters with names), together with the section where they are discussed.

above	17, 18	lpile	17
back	21	mark	15
bar	13	matrix	18
bold	12	ndefine	20
ccol	18	over	9
col	18	pile	17

cpile	17	rcol	18
define	20	right	16
delim	19	roman	12
dot	13	rpile	17
dotdot	13	size	12
down	21	sqrt	10
dyad	13	sub	7
fat	12	sup	7
font	12	tdefine	20
from	11	tilde	13
fwd	21	to	11
gfont	12	under	13
gsize	12	up	21
hat	13	vec	13
italic	12	˜, ˆ	4, 6
lcol	18	{ }	8
left	16	"..."	8, 14
lineup	15		

24. Troubleshooting

If you make a mistake in an equation, like leaving out a brace (very common) or having one too many (very common) or having a *sup* with nothing before it (common), EQN will tell you with the message

syntax error between lines x and y, file z

where x and y are approximately the lines between which the trouble occurred, and z is the name of the file in question. The line numbers are approximate — look nearby as well. There are also self-explanatory messages that arise if you leave out a quote or try to run EQN on a non-existent file.

If you want to check a document before actually printing it (on UNIX only),

```
eqn files >/dev/null
```

will throw away the output but print the messages.

If you use something like dollar signs as delimiters, it is easy to leave one out. This causes very strange troubles. The program *checkeq* (on GCOS, use *./checkeq* instead) checks for misplaced or missing dollar signs and similar troubles.

In-line equations can only be so big because of an internal buffer in TROFF. If you get a message "word overflow", you have exceeded this limit. If you print the equation as a displayed equation this message will usually go away. The message "line overflow" indicates you have exceeded an even bigger buffer. The only cure for this is to break the equation into two separate ones.

On a related topic, EQN does not break equations by itself — you must split long equations up across multiple lines by yourself, marking each by a separate .EQ ... .EN sequence. EQN does warn about equations that are too long to fit on one line.

25. Use on UNIX

To print a document that contains mathematics on the UNIX typesetter,

eqn files | troff

If there are any TROFF options, they go after the TROFF part of the command. For example,

eqn files | troff −ms

To run the same document on the GCOS typesetter, use

eqn files | troff −g (other options) | gcat

A compatible version of EQN can be used on devices like teletypes and DASI and GSI terminals which have half-line forward and reverse capabilities. To print equations on a Model 37 teletype, for example, use

neqn files | nroff

The language for equations recognized by NEQN is identical to that of EQN, although of course the output is more restricted.

To use a GSI or DASI terminal as the output device,

neqn files | nroff −T*x*

where *x* is the terminal type you are using, such as *300* or *300S*.

EQN and NEQN can be used with the TBL program[2] for setting tables that contain mathematics. Use TBL before [N]EQN, like this:

tbl files | eqn | troff
tbl files | neqn | nroff

26. Acknowledgments

We are deeply indebted to J. F. Ossanna, the author of TROFF, for his willingness to extend TROFF to make our task easier, and for his continuous assistance during the development and evolution of EQN. We are also grateful to A. V. Aho for advice on language design, to S. C. Johnson for assistance with the YACC compiler-compiler, and to all the EQN users who have made helpful suggestions and criticisms.

References

[1] J. F. Ossanna, "NROFF/TROFF User's Manual", Bell Laboratories Computing Science Technical Report #54, 1976.

[2] M. E. Lesk, "Typing Documents on UNIX", Bell Laboratories, 1976.

[3] M. E. Lesk, "TBL — A Program for Setting Tables", Bell Laboratories Computing Science Technical Report #49, 1976.

UNIX Programming — Second Edition

Brian W. Kernighan

Dennis M. Ritchie

ABSTRACT

This paper is an introduction to programming on the UNIX† system. The emphasis is on how to write programs that interface to the operating system, either directly or through the standard I/O library. The topics discussed include

- handling command arguments
- rudimentary I/O; the standard input and output
- the standard I/O library; file system access
- low-level I/O: open, read, write, close, seek
- processes: exec, fork, pipes
- signals — interrupts, etc.

There is also an appendix which describes the standard I/O library in detail.

November 18, 1986

† UNIX is a trademark of Bell Laboratories.

UNIX Programming — Second Edition

Brian W. Kernighan

Dennis M. Ritchie

1. INTRODUCTION

This paper describes how to write programs that interface with the UNIX operating system in a non-trivial way. This includes programs that use files by name, that use pipes, that invoke other commands as they run, or that attempt to catch interrupts and other signals during execution.

The document collects material which is scattered throughout several sections of *The UNIX Programmer's Manual* [1] for Version 7 UNIX. There is no attempt to be complete; only generally useful material is dealt with. It is assumed that you will be programming in C, so you must be able to read the language roughly up to the level of *The C Programming Language* [2]. Some of the material in sections 2 through 4 is based on topics covered more carefully there. You should also be familiar with UNIX itself at least to the level of *UNIX for Beginners* [3].

2. BASICS

2.1. Program Arguments

When a C program is run as a command, the arguments on the command line are made available to the function **main** as an argument count **argc** and an array **argv** of pointers to character strings that contain the arguments. By convention, **argv**[0] is the command name itself, so **argc** is always greater than 0.

The following program illustrates the mechanism: it simply echoes its arguments back to the terminal. (This is essentially the **echo** command.)

```
main(argc, argv)        /* echo arguments */
int argc;
char *argv[];
{
        int i;

        for (i = 1; i < argc; i++)
                printf("%s%c", argv[i], (i<argc-1) ? ' ' : '0');
}
```

argv is a pointer to an array whose individual elements are pointers to arrays of characters; each is terminated by **\0**, so they can be treated as strings. The program starts by printing **argv**[1] and loops until it has printed them all.

The argument count and the arguments are parameters to **main**. If you want to keep them around so other routines can get at them, you must copy them to external variables.

2.2. The "Standard Input" and "Standard Output"

The simplest input mechanism is to read the "standard input," which is generally the user's terminal. The function **getchar** returns the next input character each time it is called. A file may be substituted for the terminal by using the < convention: if **prog** uses **getchar**, then the command line

prog <file

causes **prog** to read **file** instead of the terminal. **prog** itself need know nothing about where its input is coming from. This is also true if the input comes from another program via the **pipe** mechanism:

otherprog |prog

provides the standard input for **prog** from the standard output of **otherprog.**

getchar returns the value **EOF** when it encounters the end of file (or an error) on whatever you are reading. The value of **EOF** is normally defined to be **-1**, but it is unwise to take any advantage of that knowledge. As will become clear shortly, this value is automatically defined for you when you compile a program, and need not be of any concern.

Similarly, **putchar(c)** puts the character **c** on the "standard output," which is also by default the terminal. The output can be captured on a file by using >: if **prog** uses **putchar**,

prog >outfile

writes the standard output on **outfile** instead of the terminal. **outfile** is created if it doesn't exist; if it already exists, its previous contents are overwritten. And a pipe can be used:

prog |otherprog

puts the standard output of **prog** into the standard input of **otherprog.**

The function **printf**, which formats output in various ways, uses the same mechanism as **putchar** does, so calls to **printf** and **putchar** may be intermixed in any order; the output will appear in the order of the calls.

Similarly, the function **scanf** provides for formatted input conversion; it will read the standard input and break it up into strings, numbers, etc., as desired. **scanf** uses the same mechanism as **getchar**, so calls to them may also be intermixed.

Many programs read only one input and write one output; for such programs I/O with **getchar**, **putchar**, **scanf**, and **printf** may be entirely adequate, and it is almost always enough to get started. This is particularly true if the UNIX pipe facility is used to connect the output of one program to the input of the next. For example, the following program strips out all ASCII control characters from its input (except for newline and tab).

```
#include <stdio.h>

main() /* ccstrip: strip non-graphic characters */
{
        int c;
        while ((c = getchar()) != EOF)
                if ((c >= ' ' && c < 0177) || c == '' || c == '0)
                        putchar(c);
        exit(0);
}
```

The line

#include <stdio.h>

should appear at the beginning of each source file. It causes the C compiler to read a file (*/usr/include/stdio.h*) of standard routines and symbols that includes the definition of **EOF**.

If it is necessary to treat multiple files, you can use **cat** to collect the files for you:

cat file1 file2 ... |ccstrip >output

and thus avoid learning how to access files from a program. By the way, the call to **exit** at the end is not necessary to make the program work properly, but it assures that any caller of the program will see a normal termination status (conventionally 0) from the program when it completes. Section 6 discusses status returns in more detail.

3. THE STANDARD I/O LIBRARY

The "Standard I/O Library" is a collection of routines intended to provide efficient and portable I/O services for most C programs. The standard I/O library is available on each system that supports C, so programs that confine their system interactions to its facilities can be transported from one system to another essentially without change.

In this section, we will discuss the basics of the standard I/O library. The appendix contains a more complete description of its capabilities.

3.1. File Access

The programs written so far have all read the standard input and written the standard output, which we have assumed are magically pre-defined. The next step is to write a program that accesses a file that is *not* already connected to the program. One simple example is *wc*, which counts the lines, words and characters in a set of files. For instance, the command

```
wc x.c y.c
```

prints the number of lines, words and characters in **x.c** and **y.c** and the totals.

The question is how to arrange for the named files to be read — that is, how to connect the file system names to the I/O statements which actually read the data.

The rules are simple. Before it can be read or written a file has to be *opened* by the standard library function **fopen**. **fopen** takes an external name (like **x.c** or **y.c**), does some housekeeping and negotiation with the operating system, and returns an internal name which must be used in subsequent reads or writes of the file.

This internal name is actually a pointer, called a *file pointer*, to a structure which contains information about the file, such as the location of a buffer, the current character position in the buffer, whether the file is being read or written, and the like. Users don't need to know the details, because part of the standard I/O definitions obtained by including **stdio.h** is a structure definition called **FILE**. The only declaration needed for a file pointer is exemplified by

```
FILE  *fp, *fopen();
```

This says that **fp** is a pointer to a **FILE**, and **fopen** returns a pointer to a **FILE**. **FILE** is a type name, like **int**, not a structure tag.

The actual call to **fopen** in a program is

```
fp = fopen(name, mode);
```

The first argument of **fopen** is the name of the file, as a character string. The second argument is the mode, also as a character string, which indicates how you intend to use the file. The only allowable modes are read "**r**"), write "**w**"), or append "**a**").

If a file that you open for writing or appending does not exist, it is created (if possible). Opening an existing file for writing causes the old contents to be discarded. Trying to read a file that does not exist is an error, and there may be other causes of error as well (like trying to read a file when you don't have permission). If there is any error, **fopen** will return the null pointer value **NULL** (which is defined as zero in **stdio.h**).

The next thing needed is a way to read or write the file once it is open. There are several possibilities, of which **getc** and **putc** are the simplest. **getc** returns the next character from a file; it needs the file pointer to tell it what file. Thus

```
c = getc(fp)
```

places in **c** the next character from the file referred to by **fp**; it returns **EOF** when it reaches end of file. **putc** is the inverse of **getc**:

```
putc(c, fp)
```

puts the character **c** on the file **fp** and returns **c**. **getc** and **putc** return **EOF** on error.

When a program is started, three files are opened automatically, and file pointers are provided for them. These files are the standard input, the standard output, and the standard error output; the corresponding file pointers are called **stdin**, **stdout**, and **stderr**. Normally these are all connected to the terminal, but may be redirected to files or pipes as described in Section 2.2. **stdin**, **stdout** and **stderr** are pre-defined in the I/O library as the standard input, output and error files; they may be used anywhere an object of type **FILE *** can be. They are constants, however, *not* variables, so don't try to assign to them.

With some of the preliminaries out of the way, we can now write *wc*. The basic design is one that has been found convenient for many programs: if there are command-line arguments, they are processed in order. If there are no arguments, the standard input is processed. This way the program can be used stand-alone or as part of a larger process.

```
#include <stdio.h>

main(argc, argv)        /* wc: count lines, words, chars */
int argc;
char *argv[];
{
        int c, i, inword;
        FILE *fp, *fopen();
        long linect, wordct, charct;
        long tlinect = 0, twordct = 0, tcharct = 0;

        i = 1;
        fp = stdin;
        do {
                if (argc > 1 && (fp=fopen(argv[i], "r")) == NULL) {
                        fprintf(stderr, "wc: can't open %s0, argv[i]);
                        continue;
                }
                linect = wordct = charct = inword = 0;
                while ((c = getc(fp)) != EOF) {
                        charct++;
                        if (c == '0)
                                linect++;
                        if (c == ' ' || c == '' || c == '0)
                                inword = 0;
                        else if (inword == 0) {
                                inword = 1;
                                wordct++;
                        }
                }
                printf("%7ld %7ld %7ld", linect, wordct, charct);
                printf(argc > 1 ? " %s0 : "0, argv[i]);
                fclose(fp);
                tlinect += linect;
                twordct += wordct;
                tcharct += charct;
        } while (++i < argc);
        if (argc > 2)
                printf("%7ld %7ld %7ld total0, tlinect, twordct, tcharct);
        exit(0);
}
```

The function **fprintf** is identical to **printf**, save that the first argument is a file pointer that specifies the file to be written.

The function **fclose** is the inverse of **fopen**; it breaks the connection between the file pointer and the external name that was established by **fopen**, freeing the file pointer for another file. Since there is a limit on the number of files that a program may have open simultaneously, it's a good idea to free things when they are no longer needed. There is also another reason to call **fclose** on an output file — it flushes the buffer in which **putc** is collecting output. **fclose** is called automatically for each open file when a program terminates normally.)

3.2. Error Handling — Stderr and Exit

stderr is assigned to a program in the same way that **stdin** and **stdout** are. Output written on **stderr** appears on the user's terminal even if the standard output is redirected. *wc* writes its diagnostics on **stderr** instead of **stdout** so that if one of the files can't be accessed for some reason, the message finds its way to the user's terminal instead of disappearing down a pipeline or into an output file.

The program actually signals errors in another way, using the function **exit** to terminate program execution. The argument of **exit** is available to whatever process called it (see Section 6), so the success or failure of the program can be tested by another program that uses this one as a sub-process. By convention, a return value of 0 signals that all is well; non-zero values signal abnormal situations.

exit itself calls **fclose** for each open output file, to flush out any buffered output, then calls a routine named **_exit**. The function **_exit** causes immediate termination without any buffer flushing; it may be called directly if desired.

3.3. Miscellaneous I/O Functions

The standard I/O library provides several other I/O functions besides those we have illustrated above.

Normally output with **putc**, etc., is buffered (except to **stderr**); to force it out immediately, use **fflush(fp)**.

fscanf is identical to **scanf**, except that its first argument is a file pointer (as with **fprintf**) that specifies the file from which the input comes; it returns **EOF** at end of file.

The functions **sscanf** and **sprintf** are identical to **fscanf** and **fprintf**, except that the first argument names a character string instead of a file pointer. The conversion is done from the string for **sscanf** and into it for **sprintf**.

fgets(buf, size, fp) copies the next line from **fp**, up to and including a newline, into **buf**; at most **size-1** characters are copied; it returns **NULL** at end of file. **fputs(buf, fp)** writes the string in **buf** onto file **fp**.

The function **ungetc(c, fp)** "pushes back" the character **c** onto the input stream **fp**; a subsequent call to **getc**, **fscanf**, etc., will encounter **c**. Only one character of pushback per file is permitted.

4. LOW-LEVEL I/O

This section describes the bottom level of I/O on the system. The lowest level of I/O in provides no buffering or any other services; it is in fact a direct entry into the operating system. You are entirely on your own, but on the other hand, you have the most control over what happens. And since the calls and usage are quite simple, this isn't as bad as it sounds.

4.1. File Descriptors

In the UNIX operating system, all input and output is done by reading or writing files, because all peripheral devices, even the user's terminal, are files in the file system. This means that a single, homogeneous interface handles all communication between a program and peripheral devices.

In the most general case, before reading or writing a file, it is necessary to inform the system of your intent to do so, a process called "opening" the file. If you are going to write on a file, it may also be necessary to create it. The system checks your right to do so (Does the file exist? Do you have permission to access it?), and if all is well, returns a small positive integer called a *file descriptor*. Whenever I/O is to be done on the file, the file descriptor is used instead of the name to identify the file. (This is roughly analogous to the use of READ(5,...) and WRITE(6,...) in Fortran. All information about an open file is maintained by the system; the user program refers to the file only by the file descriptor.

The file pointers discussed in section 3 are similar in spirit to file descriptors, but file descriptors are more fundamental. A file pointer is a pointer to a structure that contains, among other things, the file descriptor for the file in question.

Since input and output involving the user's terminal are so common, special arrangements exist to make this convenient. When the command interpreter (the "shell") runs a program, it opens three files, with file descriptors 0, 1, and 2, called the standard input, the standard output, and the standard error output. All of these are normally connected to the terminal, so if a program reads file descriptor 0 and writes file descriptors 1 and 2, it can do terminal I/O without worrying about opening the files.

If I/O is redirected to and from files with < and >, as in

prog <infile >outfile

the shell changes the default assignments for file descriptors 0 and 1 from the terminal to the named files. Similar observations hold if the input or output is associated with a pipe. Normally file descriptor 2 remains attached to the terminal, so error messages can go there. In all cases, the file assignments are changed by the shell, not by the program. The program does not need to know where its input comes from nor where its output goes, so long as it uses file 0 for input and 1 and 2 for output.

4.2. Read and Write

All input and output is done by two functions called **read** and **write**. For both, the first argument is a file descriptor. The second argument is a buffer in your program where the data is to come from or go to. The third argument is the number of bytes to be transferred. The calls are

```
n_read = read(fd, buf, n);

n_written = write(fd, buf, n);
```

Each call returns a byte count which is the number of bytes actually transferred. On reading, the number of bytes returned may be less than the number asked for, because fewer than **n** bytes remained to be read. (When the file is a terminal, **read** normally reads only up to the next newline, which is generally less than what was requested.) A return value of zero bytes implies end of file, and **-1** indicates an error of some sort. For writing, the returned value is the number of bytes actually written; it is generally an error if this isn't equal to the number supposed to be written.

The number of bytes to be read or written is quite arbitrary. The two most common values are 1, which means one character at a time ("unbuffered"), and 512, which corresponds to a physical blocksize on many peripheral devices. This latter size will be most efficient, but even character at a time I/O is not inordinately expensive.

Putting these facts together, we can write a simple program to copy its input to its output. This program will copy anything to anything, since the input and output can be redirected to any file or device.

```
#define      BUFSIZE      512/* best size for PDP-11 UNIX */

main() /* copy input to output */
{
        char    buf[BUFSIZE];
```

```
        int     n;

        while ((n = read(0, buf, BUFSIZE)) > 0)
                write(1, buf, n);
        exit(0);
}
```

If the file size is not a multiple of **BUFSIZE**, some **read** will return a smaller number of bytes to be written by **write**; the next call to **read** after that will return zero.

It is instructive to see how **read** and **write** can be used to construct higher level routines like **getchar**, **putchar**, etc. For example, here is a version of **getchar** which does unbuffered input.

```
#define     CMASK     0377/* for making char's > 0 */

getchar()     /* unbuffered single character input */
{
        char c;

        return((read(0, &c, 1) > 0) ? c & CMASK : EOF);
}
```

c *must* be declared **char**, because **read** accepts a character pointer. The character being returned must be masked with **0377** to ensure that it is positive; otherwise sign extension may make it negative. (The constant **0377** is appropriate for the PDP -11 but not necessarily for other machines.)

The second version of **getchar** does input in big chunks, and hands out the characters one at a time.

```
#define     CMASK     0377/* for making char's > 0 */
#define     BUFSIZE   512

getchar()     /* buffered version */
{
        static char     buf[BUFSIZE];
        static char     *bufp = buf;
        static int      n = 0;

        if (n == 0) {  /* buffer is empty */
                n = read(0, buf, BUFSIZE);
                bufp = buf;
        }
        return((--n >= 0) ? *bufp++ & CMASK : EOF);
}
```

4.3. Open, Creat, Close, Unlink

Other than the default standard input, output and error files, you must explicitly open files in order to read or write them. There are two system entry points for this, **open** and **creat** [sic].

open is rather like the **fopen** discussed in the previous section, except that instead of returning a file pointer, it returns a file descriptor, which is just an **int**.

```
int fd;

fd = open(name, rwmode);
```

As with **fopen**, the **name** argument is a character string corresponding to the external file name. The access mode argument is different, however: **rwmode** is 0 for read, 1 for write, and 2 for read and write access. **open** returns **-1** if any error occurs; otherwise it returns a valid file descriptor.

It is an error to try to **open** a file that does not exist. The entry point **creat** is provided to create new files, or to re-write old ones.

fd = creat(name, pmode);

returns a file descriptor if it was able to create the file called **name**, and **-1** if not. If the file already exists, **creat** will truncate it to zero length; it is not an error to **creat** a file that already exists.

If the file is brand new, **creat** creates it with the *protection mode* specified by the **pmode** argument. In the UNIX file system, there are nine bits of protection information associated with a file, controlling read, write and execute permission for the owner of the file, for the owner's group, and for all others. Thus a three-digit octal number is most convenient for specifying the permissions. For example, 0755 specifies read, write and execute permission for the owner, and read and execute permission for the group and everyone else.

To illustrate, here is a simplified version of the UNIX utility *cp*, a program which copies one file to another. (The main simplification is that our version copies only one file, and does not permit the second argument to be a directory.)

```
#define NULL 0
#define BUFSIZE 512
#define PMODE 0644 /* RW for owner, R for group, others */

main(argc, argv)        /* cp: copy f1 to f2 */
int argc;
char *argv[];
{
        int     f1, f2, n;
        char    buf[BUFSIZE];

        if (argc != 3)
                error("Usage: cp from to", NULL);
        if ((f1 = open(argv[1], 0)) == -1)
                error("cp: can't open %s", argv[1]);
        if ((f2 = creat(argv[2], PMODE)) == -1)
                error("cp: can't create %s", argv[2]);

        while ((n = read(f1, buf, BUFSIZE)) > 0)
                if (write(f2, buf, n) != n)
                        error("cp: write error", NULL);
        exit(0);
}

error(s1, s2)   /* print error message and die */
char *s1, *s2;
{
        printf(s1, s2);
        printf("0);
        exit(1);
}
```

As we said earlier, there is a limit (typically 15-25) on the number of files which a program may have open simultaneously. Accordingly, any program which intends to process many files must be prepared to re-use file descriptors. The routine **close** breaks the connection between a file descriptor and an open file, and frees the file descriptor for use with some other file. Termination of a program via **exit** or return from the main program closes all open files.

The function **unlink(filename)** removes the file **filename** from the file system.

4.4. Random Access — Seek and Lseek

File I/O is normally sequential: each **read** or **write** takes place at a position in the file right after the previous one. When necessary, however, a file can be read or written in any arbitrary order. The system call **lseek** provides a way to move around in a file without actually reading or writing:

```
lseek(fd, offset, origin);
```

forces the current position in the file whose descriptor is **fd** to move to position **offset**, which is taken relative to the location specified by **origin**. Subsequent reading or writing will begin at that position. **offset** is a **long**; **fd** and **origin** are **int**'s. **origin** can be 0, 1, or 2 to specify that **offset** is to be measured from the beginning, from the current position, or from the end of the file respectively. For example, to append to a file, seek to the end before writing:

```
lseek(fd, 0L, 2);
```

To get back to the beginning ("rewind"),

```
lseek(fd, 0L, 0);
```

Notice the **0L** argument; it could also be written as **(long) 0**.

With **lseek**, it is possible to treat files more or less like large arrays, at the price of slower access. For example, the following simple function reads any number of bytes from any arbitrary place in a file.

```
get(fd, pos, buf, n) /* read n bytes from position pos */
int fd, n;
long pos;
char *buf;
{
	lseek(fd, pos, 0);	/* get to pos */
	return(read(fd, buf, n));
}
```

In pre-version 7 UNIX, the basic entry point to the I/O system is called **seek**. **seek** is identical to **lseek**, except that its **offset** argument is an **int** rather than a **long**. Accordingly, since PDP -11 integers have only 16 bits, the **offset** specified for **seek** is limited to 65,535; for this reason, **origin** values of 3, 4, 5 cause **seek** to multiply the given offset by 512 (the number of bytes in one physical block) and then interpret **origin** as if it were 0, 1, or 2 respectively. Thus to get to an arbitrary place in a large file requires two seeks, first one which selects the block, then one which has **origin** equal to 1 and moves to the desired byte within the block.

4.5. Error Processing

The routines discussed in this section, and in fact all the routines which are direct entries into the system can incur errors. Usually they indicate an error by returning a value of –1. Sometimes it is nice to know what sort of error occurred; for this purpose all these routines, when appropriate, leave an error number in the external cell **errno**. The meanings of the various error numbers are listed in the introduction to Section II of the *UNIX Programmer's Manual,* so your program can, for example, determine if an attempt to open a file failed because it did not exist or because the user lacked permission to read it. Perhaps more commonly, you may want to print out the reason for failure. The routine **perror** will print a message associated with the value of **errno**; more generally, **sys_errno** is an array of character strings which can be indexed by **errno** and printed by your program.

5. PROCESSES

It is often easier to use a program written by someone else than to invent one's own. This section describes how to execute a program from within another.

5.1. The "System" Function

The easiest way to execute a program from another is to use the standard library routine **system**. **system** takes one argument, a command string exactly as typed at the terminal (except for the newline at the end) and executes it. For instance, to time-stamp the output of a program,

```
main()
{
        system("date");
        /* rest of processing */
}
```

If the command string has to be built from pieces, the in-memory formatting capabilities of **sprintf** may be useful.

Remember that **getc** and **putc** normally buffer their input; terminal I/O will not be properly synchronized unless this buffering is defeated. For output, use **fflush**; for input, see **setbuf** in the appendix.

5.2. Low-Level Process Creation — Execl and Execv

If you're not using the standard library, or if you need finer control over what happens, you will have to construct calls to other programs using the more primitive routines that the standard library's **system** routine is based on.

The most basic operation is to execute another program without *returning*, by using the routine **execl**. To print the date as the last action of a running program, use

```
execl("/bin/date", "date", NULL);
```

The first argument to **execl** is the *file name* of the command; you have to know where it is found in the file system. The second argument is conventionally the program name (that is, the last component of the file name), but this is seldom used except as a place-holder. If the command takes arguments, they are strung out after this; the end of the list is marked by a **NULL** argument.

The **execl** call overlays the existing program with the new one, runs that, then exits. There is *no* return to the original program.

More realistically, a program might fall into two or more phases that communicate only through temporary files. Here it is natural to make the second pass simply an **execl** call from the first.

The one exception to the rule that the original program never gets control back occurs when there is an error, for example if the file can't be found or is not executable. If you don't know where **date** is located, say

```
execl("/bin/date", "date", NULL);
execl("/usr/bin/date", "date", NULL);
fprintf(stderr, "Someone stole 'date'0);
```

A variant of **execl** called **execv** is useful when you don't know in advance how many arguments there are going to be. The call is

```
execv(filename, argp);
```

where **argp** is an array of pointers to the arguments; the last pointer in the array must be **NULL** so **execv** can tell where the list ends. As with **execl**, **filename** is the file in which the program is found, and **argp[0]** is the name of the program. (This arrangement is identical to the **argv** array for program arguments.)

Neither of these routines provides the niceties of normal command execution. There is no automatic search of multiple directories — you have to know precisely where the command is located. Nor do you get the expansion of metacharacters like <, >, *, ?, and [] in the argument list. If you want these, use **execl** to invoke the shell **sh**, which then does all the work. Construct a string **commandline** that contains the complete command as it would have been typed at the terminal, then say

```
execl("/bin/sh", "sh", "-c", commandline, NULL);
```

The shell is assumed to be at a fixed place, **/bin/sh**. Its argument **-c** says to treat the next argument as a whole command line, so it does just what you want. The only problem is in constructing the right information in **commandline**.

5.3. Control of Processes — Fork and Wait

So far what we've talked about isn't really all that useful by itself. Now we will show how to regain control after running a program with **execl** or **execv**. Since these routines simply overlay the new program on the old one, to save the old one requires that it first be split into two copies; one of these can be overlaid, while the other waits for the new, overlaying program to finish. The splitting is done by a routine called **fork**:

```
proc_id = fork();
```

splits the program into two copies, both of which continue to run. The only difference between the two is the value of **proc_id**, the "process id." In one of these processes (the "child"), **proc_id** is zero. In the other (the "parent"), **proc_id** is non-zero; it is the process number of the child. Thus the basic way to call, and return from, another program is

```
if (fork() == 0)
        execl("/bin/sh", "sh", "-c", cmd, NULL);/* in child */
```

And in fact, except for handling errors, this is sufficient. The **fork** makes two copies of the program. In the child, the value returned by **fork** is zero, so it calls **execl** which does the **command** and then dies. In the parent, **fork** returns non-zero so it skips the **execl.** (If there is any error, **fork** returns **-1**).

More often, the parent wants to wait for the child to terminate before continuing itself. This can be done with the function **wait**:

```
int status;

if (fork() == 0)
        execl(...);
wait(&status);
```

This still doesn't handle any abnormal conditions, such as a failure of the **execl** or **fork**, or the possibility that there might be more than one child running simultaneously. (The **wait** returns the process id of the terminated child, if you want to check it against the value returned by **fork**.) Finally, this fragment doesn't deal with any funny behavior on the part of the child (which is reported in **status**). Still, these three lines are the heart of the standard library's **system** routine, which we'll show in a moment.

The **status** returned by **wait** encodes in its low-order eight bits the system's idea of the child's termination status; it is 0 for normal termination and non-zero to indicate various kinds of problems. The next higher eight bits are taken from the argument of the call to **exit** which caused a normal termination of the child process. It is good coding practice for all programs to return meaningful status.

When a program is called by the shell, the three file descriptors 0, 1, and 2 are set up pointing at the right files, and all other possible file descriptors are available for use. When this program calls another one, correct etiquette suggests making sure the same conditions hold. Neither **fork** nor the **exec** calls affects open files in any way. If the parent is buffering output that must come out before output from the child, the parent must flush its buffers before the **execl**. Conversely, if a caller buffers an input stream, the called program will lose any information that has been read by the caller.

5.4. Pipes

A *pipe* is an I/O channel intended for use between two cooperating processes: one process writes into the pipe, while the other reads. The system looks after buffering the data and synchronizing the two processes. Most pipes are created by the shell, as in

ls | pr

which connects the standard output of **ls** to the standard input of **pr**. Sometimes, however, it is most convenient for a process to set up its own plumbing; in this section, we will illustrate how the pipe connection is established and used.

The system call **pipe** creates a pipe. Since a pipe is used for both reading and writing, two file descriptors are returned; the actual usage is like this:

```
int     fd[2];

stat = pipe(fd);
if (stat == -1)
        /* there was an error ... */
```

fd is an array of two file descriptors, where **fd**[0] is the read side of the pipe and **fd**[1] is for writing. These may be used in **read**, **write** and **close** calls just like any other file descriptors.

If a process reads a pipe which is empty, it will wait until data arrives; if a process writes into a pipe which is too full, it will wait until the pipe empties somewhat. If the write side of the pipe is closed, a subsequent **read** will encounter end of file.

To illustrate the use of pipes in a realistic setting, let us write a function called **popen(cmd, mode)**, which creates a process **cmd** (just as **system** does), and returns a file descriptor that will either read or write that process, according to **mode**. That is, the call

```
fout = popen("pr", WRITE);
```

creates a process that executes the **pr** command; subsequent **write** calls using the file descriptor **fout** will send their data to that process through the pipe.

popen first creates the the pipe with a **pipe** system call; it then **fork**s to create two copies of itself. The child decides whether it is supposed to read or write, closes the other side of the pipe, then calls the shell (via **execl**) to run the desired process. The parent likewise closes the end of the pipe it does not use. These closes are necessary to make end-of-file tests work properly. For example, if a child that intends to read fails to close the write end of the pipe, it will never see the end of the pipe file, just because there is one writer potentially active.

```
#include <stdio.h>

#define     READ 0
#define     WRITE       1
#define     tst(a, b)   (mode == READ ? (b) : (a))
static int  popen_pid;

popen(cmd, mode)
char    *cmd;
int     mode;
{
        int p[2];

        if (pipe(p) < 0)
                return(NULL);
        if ((popen_pid = fork()) == 0) {
                close(tst(p[WRITE], p[READ]));
                close(tst(0, 1));
                dup(tst(p[READ], p[WRITE]));
```

```
                close(tst(p[READ], p[WRITE]));
                execl("/bin/sh", "sh", "-c", cmd, 0);
                _exit(1);       /* disaster has occurred if we get here */
        }
        if (popen_pid == -1)
                return(NULL);
        close(tst(p[READ], p[WRITE]));
        return(tst(p[WRITE], p[READ]));
}
```

The sequence of **closes** in the child is a bit tricky. Suppose that the task is to create a child process that will read data from the parent. Then the first **close** closes the write side of the pipe, leaving the read side open. The lines

```
close(tst(0, 1));
dup(tst(p[READ], p[WRITE]));
```

are the conventional way to associate the pipe descriptor with the standard input of the child. The **close** closes file descriptor 0, that is, the standard input. **dup** is a system call that returns a duplicate of an already open file descriptor. File descriptors are assigned in increasing order and the first available one is returned, so the effect of the **dup** is to copy the file descriptor for the pipe (read side) to file descriptor 0; thus the read side of the pipe becomes the standard input. (Yes, this is a bit tricky, but it's a standard idiom.) Finally, the old read side of the pipe is closed.

A similar sequence of operations takes place when the child process is supposed to write from the parent instead of reading. You may find it a useful exercise to step through that case.

The job is not quite done, for we still need a function **pclose** to close the pipe created by **popen**. The main reason for using a separate function rather than **close** is that it is desirable to wait for the termination of the child process. First, the return value from **pclose** indicates whether the process succeeded. Equally important when a process creates several children is that only a bounded number of unwaited-for children can exist, even if some of them have terminated; performing the **wait** lays the child to rest. Thus:

```
#include <signal.h>

pclose(fd)      /* close pipe fd */
int fd;
{
        register r, (*hstat)(), (*istat)(), (*qstat)();
        int     status;
        extern int popen_pid;

        close(fd);
        istat = signal(SIGINT, SIG_IGN);
        qstat = signal(SIGQUIT, SIG_IGN);
        hstat = signal(SIGHUP, SIG_IGN);
        while ((r = wait(&status)) != popen_pid && r != -1);
        if (r == -1)
                status = -1;
        signal(SIGINT, istat);
        signal(SIGQUIT, qstat);
        signal(SIGHUP, hstat);
        return(status);
}
```

The calls to **signal** make sure that no interrupts, etc., interfere with the waiting process; this is the topic of the next section.

The routine as written has the limitation that only one pipe may be open at once, because of the single shared variable **popen_pid**; it really should be an array indexed by file descriptor. A **popen** function, with slightly different arguments and return value is available as part of the standard I/O library discussed below. As currently written, it shares the same limitation.

6. SIGNALS — INTERRUPTS AND ALL THAT

This section is concerned with how to deal gracefully with signals from the outside world (like interrupts), and with program faults. Since there's nothing very useful that can be done from within C about program faults, which arise mainly from illegal memory references or from execution of peculiar instructions, we'll discuss only the outside-world signals: *interrupt*, which is sent when the DEL character is typed; *quit*, generated by the FS character; *hangup*, caused by hanging up the phone; and *terminate*, generated by the *kill* command. When one of these events occurs, the signal is sent to *all* processes which were started from the corresponding terminal; unless other arrangements have been made, the signal terminates the process. In the *quit* case, a core image file is written for debugging purposes.

The routine which alters the default action is called **signal**. It has two arguments: the first specifies the signal, and the second specifies how to treat it. The first argument is just a number code, but the second is the address is either a function, or a somewhat strange code that requests that the signal either be ignored, or that it be given the default action. The include file **signal.h** gives names for the various arguments, and should always be included when signals are used. Thus

```
#include <signal.h>
 ...
signal(SIGINT, SIG_IGN);
```

causes interrupts to be ignored, while

```
signal(SIGINT, SIG_DFL);
```

restores the default action of process termination. In all cases, **signal** returns the previous value of the signal. The second argument to **signal** may instead be the name of a function (which has to be declared explicitly if the compiler hasn't seen it already). In this case, the named routine will be called when the signal occurs. Most commonly this facility is used to allow the program to clean up unfinished business before terminating, for example to delete a temporary file:

```
#include <signal.h>

main()
{
        int onintr();

        if (signal(SIGINT, SIG_IGN) != SIG_IGN)
                signal(SIGINT, onintr);

        /* Process ... */

        exit(0);
}

onintr()
{
        unlink(tempfile);
        exit(1);
}
```

Why the test and the double call to **signal**? Recall that signals like interrupt are sent to *all* processes started from a particular terminal. Accordingly, when a program is to be run noninteractively (started by **&**), the shell turns off interrupts for it so it won't be stopped by interrupts intended for foreground processes. If this program began by announcing that all interrupts were to be sent to the **onintr** routine regardless, that would undo the shell's effort to protect it when run in the background.

The solution, shown above, is to test the state of interrupt handling, and to continue to ignore interrupts if they are already being ignored. The code as written depends on the fact that **signal** returns the previous state of a particular signal. If signals were already being ignored, the process should continue to ignore them; otherwise, they should be caught.

A more sophisticated program may wish to intercept an interrupt and interpret it as a request to stop what it is doing and return to its own command-processing loop. Think of a text editor: interrupting a long printout should not cause it to terminate and lose the work already done. The outline of the code for this case is probably best written like this:

```
#include <signal.h>
#include <setjmp.h>
jmp_buf     sjbuf;

main()
{
        int (*istat)(), onintr();

        istat = signal(SIGINT, SIG_IGN);/* save original status */
        setjmp(sjbuf); /* save current stack position */
        if (istat != SIG_IGN)
                signal(SIGINT, onintr);

        /* main processing loop */
}
onintr()
{
        printf("0nterrupt0);
        longjmp(sjbuf);         /* return to saved state */
}
```

The include file **setjmp.h** declares the type **jmp_buf** an object in which the state can be saved. **sjbuf** is such an object; it is an array of some sort. The **setjmp** routine then saves the state of things. When an interrupt occurs, a call is forced to the **onintr** routine, which can print a message, set flags, or whatever. **longjmp** takes as argument an object stored into by **setjmp**, and restores control to the location after the call to **setjmp**, so control (and the stack level) will pop back to the place in the main routine where the signal is set up and the main loop entered. Notice, by the way, that the signal gets set again after an interrupt occurs. This is necessary; most signals are automatically reset to their default action when they occur.

Some programs that want to detect signals simply can't be stopped at an arbitrary point, for example in the middle of updating a linked list. If the routine called on occurrence of a signal sets a flag and then returns instead of calling **exit** or **longjmp**, execution will continue at the exact point it was interrupted. The interrupt flag can then be tested later.

There is one difficulty associated with this approach. Suppose the program is reading the terminal when the interrupt is sent. The specified routine is duly called; it sets its flag and returns. If it were really true, as we said above, that "execution resumes at the exact point it was interrupted," the program would continue reading the terminal until the user typed another line. This behavior might well be confusing, since the user might not know that the program is reading; he presumably would prefer to have the signal take effect instantly. The method chosen to resolve this difficulty is to terminate the terminal read when execution resumes after the signal, returning an error code which indicates what happened.

Thus programs which catch and resume execution after signals should be prepared for "errors" which are caused by interrupted system calls. (The ones to watch out for are reads from a terminal, **wait**, and **pause**.) A program whose **onintr** program just sets **intflag**, resets the interrupt signal, and returns, should usually include code like the following when it reads the standard input:

```
if (getchar() == EOF)
        if (intflag)
                /* EOF caused by interrupt */
        else
                /* true end-of-file */
```

A final subtlety to keep in mind becomes important when signal-catching is combined with execution of other programs. Suppose a program catches interrupts, and also includes a method (like "!" in the editor) whereby other programs can be executed. Then the code should look something like this:

```
if (fork() == 0)
        execl(...);
signal(SIGINT, SIG_IGN);    /* ignore interrupts */
wait(&status);/* until the child is done */
signal(SIGINT, onintr);        /* restore interrupts */
```

Why is this? Again, it's not obvious but not really difficult. Suppose the program you call catches its own interrupts. If you interrupt the subprogram, it will get the signal and return to its main loop, and probably read your terminal. But the calling program will also pop out of its wait for the subprogram and read your terminal. Having two processes reading your terminal is very unfortunate, since the system figuratively flips a coin to decide who should get each line of input. A simple way out is to have the parent program ignore interrupts until the child is done. This reasoning is reflected in the standard I/O library function **system**:

```
#include <signal.h>

system(s)       /* run command string s */
char *s;
{
        int status, pid, w;
        register int (*istat)(), (*qstat)();

        if ((pid = fork()) == 0) {
                execl("/bin/sh", "sh", "-c", s, 0);
                _exit(127);
        }
        istat = signal(SIGINT, SIG_IGN);
        qstat = signal(SIGQUIT, SIG_IGN);
        while ((w = wait(&status)) != pid && w != -1)
                ;
        if (w == -1)
                status = -1;
        signal(SIGINT, istat);
        signal(SIGQUIT, qstat);
        return(status);
}
```

As an aside on declarations, the function **signal** obviously has a rather strange second argument. It is in fact a pointer to a function delivering an integer, and this is also the type of the signal routine itself. The two values **SIG_IGN** and **SIG_DFL** have the right type, but are chosen so they coincide with no possible actual functions. For the enthusiast, here is how they are defined for the PDP-11; the definitions should be sufficiently ugly and nonportable to encourage use of the include file.

```
#define    SIG_DFL    (int (*)())0
#define    SIG_IGN    (int (*)())1
```

References

[1] K. L. Thompson and D. M. Ritchie, *The UNIX Programmer's Manual,* Bell Laboratories, 1978.

[2] B. W. Kernighan and D. M. Ritchie, *The C Programming Language,* Prentice-Hall, Inc., 1978.

[3] B. W. Kernighan, "UNIX for Beginners — Second Edition." Bell Laboratories, 1978.

Appendix — The Standard I/O Library

D. M. Ritchie

The standard I/O library was designed with the following goals in mind.

1. It must be as efficient as possible, both in time and in space, so that there will be no hesitation in using it no matter how critical the application.
2. It must be simple to use, and also free of the magic numbers and mysterious calls whose use mars the understandability and portability of many programs using older packages.
3. The interface provided should be applicable on all machines, whether or not the programs which implement it are directly portable to other systems, or to machines other than the PDP-11 running a version of UNIX .

1. General Usage

Each program using the library must have the line

#include <stdio.h>

which defines certain macros and variables. The routines are in the normal C library, so no special library argument is needed for loading. All names in the include file intended only for internal use begin with an underscore _ to reduce the possibility of collision with a user name. The names intended to be visible outside the package are

stdin The name of the standard input file

stdout The name of the standard output file

stderr The name of the standard error file

EOF is actually −1, and is the value returned by the read routines on end-of-file or error.

NULL is a notation for the null pointer, returned by pointer-valued functions to indicate an error

FILE expands to **struct _iob** and is a useful shorthand when declaring pointers to streams.

BUFSIZ is a number (viz. 512) of the size suitable for an I/O buffer supplied by the user. See **setbuf**, below.

getc, getchar, putc, putchar, feof, ferror, fileno
are defined as macros. Their actions are described below; they are mentioned here to point out that it is not possible to redeclare them and that they are not actually functions; thus, for example, they may not have breakpoints set on them.

The routines in this package offer the convenience of automatic buffer allocation and output flushing where appropriate. The names **stdin**, **stdout**, and **stderr** are in effect constants and may not be assigned to.

2. Calls

FILE *fopen(filename, type) char *filename, *type;
opens the file and, if needed, allocates a buffer for it. **filename** is a character string specifying the name. **type** is a character string (not a single character). It may be "**r**", "**w**", or "**a**" to indicate intent to read, write, or append. The value returned is a file pointer. If it is **NULL** the attempt to open failed.

FILE *freopen(filename, type, ioptr) char *filename, *type; FILE *ioptr;
The stream named by **ioptr** is closed, if necessary, and then reopened as if by **fopen**. If the attempt to open fails, **NULL** is returned, otherwise **ioptr**, which will now refer to the new file. Often the reopened stream is **stdin** or **stdout**.

int getc(ioptr) FILE *ioptr;

returns the next character from the stream named by **ioptr**, which is a pointer to a file such as returned by **fopen**, or the name **stdin**. The integer **EOF** is returned on end-of-file or when an error occurs. The null character **\0** is a legal character.

int fgetc(ioptr) FILE *ioptr;
acts like **getc** but is a genuine function, not a macro, so it can be pointed to, passed as an argument, etc.

putc(c, ioptr) FILE *ioptr;
putc writes the character **c** on the output stream named by **ioptr**, which is a value returned from **fopen** or perhaps **stdout** or **stderr**. The character is returned as value, but **EOF** is returned on error.

fputc(c, ioptr) FILE *ioptr;
acts like **putc** but is a genuine function, not a macro.

fclose(ioptr) FILE *ioptr;
The file corresponding to **ioptr** is closed after any buffers are emptied. A buffer allocated by the I/O system is freed. **fclose** is automatic on normal termination of the program.

fflush(ioptr) FILE *ioptr;
Any buffered information on the (output) stream named by **ioptr** is written out. Output files are normally buffered if and only if they are not directed to the terminal; however, **stderr** always starts off unbuffered and remains so unless **setbuf** is used, or unless it is reopened.

exit(errcode);
terminates the process and returns its argument as status to the parent. This is a special version of the routine which calls **fflush** for each output file. To terminate without flushing, use **_exit**.

feof(ioptr) FILE *ioptr;
returns non-zero when end-of-file has occurred on the specified input stream.

ferror(ioptr) FILE *ioptr;
returns non-zero when an error has occurred while reading or writing the named stream. The error indication lasts until the file has been closed.

getchar();
is identical to **getc(stdin)**.

putchar(c);
is identical to **putc(c, stdout)**.

char *fgets(s, n, ioptr) char *s; FILE *ioptr;
reads up to **n-1** characters from the stream **ioptr** into the character pointer **s**. The read terminates with a newline character. The newline character is placed in the buffer followed by a null character. **fgets** returns the first argument, or **NULL** if error or end-of-file occurred.

fputs(s, ioptr) char *s; FILE *ioptr;
writes the null-terminated string (character array) **s** on the stream **ioptr**. No newline is appended. No value is returned.

ungetc(c, ioptr) FILE *ioptr;
The argument character **c** is pushed back on the input stream named by **ioptr**. Only one character may be pushed back.

printf(format, a1, ...) char *format;
fprintf(ioptr, format, a1, ...) FILE *ioptr; char *format;
sprintf(s, format, a1, ...)char *s, *format;
printf writes on the standard output. **fprintf** writes on the named output stream. **sprintf** puts characters in the character array (string) named by **s**. The specifications are as described in section **printf**(3) of the *UNIX Programmer's Manual.*

scanf(format, a1, ...) char *format;
fscanf(ioptr, format, a1, ...) FILE *ioptr; char *format;
sscanf(s, format, a1, ...) char *s, *format;

scanf reads from the standard input. **fscanf** reads from the named input stream. **sscanf** reads from the character string supplied as **s**. **scanf** reads characters, interprets them according to a format, and stores the results in its arguments. Each routine expects as arguments a control string **format**, and a set of arguments, *each of which must be a pointer,* indicating where the converted input should be stored.

scanf returns as its value the number of successfully matched and assigned input items. This can be used to decide how many input items were found. On end of file, **EOF** is returned; note that this is different from 0, which means that the next input character does not match what was called for in the control string.

fread(ptr, sizeof(*ptr), nitems, ioptr) FILE *ioptr;

reads **nitems** of data beginning at **ptr** from file **ioptr**. No advance notification that binary I/O is being done is required; when, for portability reasons, it becomes required, it will be done by adding an additional character to the mode-string on the **fopen** call.

fwrite(ptr, sizeof(*ptr), nitems, ioptr) FILE *ioptr;

Like **fread**, but in the other direction.

rewind(ioptr) FILE *ioptr;

rewinds the stream named by **ioptr**. It is not very useful except on input, since a rewound output file is still open only for output.

system(string) char *string;

The **string** is executed by the shell as if typed at the terminal.

getw(ioptr) FILE *ioptr;

returns the next word from the input stream named by **ioptr**. **EOF** is returned on end-of-file or error, but since this a perfectly good integer **feof** and **ferror** should be used. A "word" is 16 bits on the

putw(w, ioptr) FILE *ioptr;

writes the integer **w** on the named output stream.

setbuf(ioptr, buf) FILE *ioptr; char *buf;

setbuf may be used after a stream has been opened but before I/O has started. If **buf** is NULL, the stream will be unbuffered. Otherwise the buffer supplied will be used. It must be a character array of sufficient size:

char buf[BUFSIZ];

fileno(ioptr) FILE *ioptr;

returns the integer file descriptor associated with the file.

fseek(ioptr, offset, ptrname) FILE *ioptr; long offset;

The location of the next byte in the stream named by **ioptr** is adjusted. **offset** is a long integer. If **ptrname** is 0, the offset is measured from the beginning of the file; if **ptrname** is 1, the offset is measured from the current read or write pointer; if **ptrname** is 2, the offset is measured from the end of the file. The routine accounts properly for any buffering. (When this routine is used on systems, the offset must be a value returned from **ftell** and the ptrname must be 0).

long ftell(ioptr) FILE *ioptr;

The byte offset, measured from the beginning of the file, associated with the named stream is returned. Any buffering is properly accounted for. (On systems the value of this call is useful only for handing to **fseek**, so as to position the file to the same place it was when **ftell** was called.)

getpw(uid, buf) char *buf;

The password file is searched for the given integer user ID. If an appropriate line is found, it is copied into the character array **buf**, and 0 is returned. If no line is found corresponding to the user ID then 1 is returned.

char *malloc(num);
allocates **num** bytes. The pointer returned is sufficiently well aligned to be usable for any purpose. **NULL** is returned if no space is available.

char *calloc(num, size);
allocates space for **num** items each of size **size**.. The space is guaranteed to be set to 0 and the pointer is sufficiently well aligned to be usable for any purpose. **NULL** is returned if no space is available .

cfree(ptr) char *ptr;
Space is returned to the pool used by **calloc**. Disorder can be expected if the pointer was not obtained from **calloc**.

The following are macros whose definitions may be obtained by including <**ctype.h**>.

isalpha(c) returns non-zero if the argument is alphabetic.

isupper(c) returns non-zero if the argument is upper-case alphabetic.

islower(c) returns non-zero if the argument is lower-case alphabetic.

isdigit(c) returns non-zero if the argument is a digit.

isspace(c) returns non-zero if the argument is a spacing character: tab, newline, carriage return, vertical tab, form feed, space.

ispunct(c) returns non-zero if the argument is any punctuation character, i.e., not a space, letter, digit or control character.

isalnum(c) returns non-zero if the argument is a letter or a digit.

isprint(c) returns non-zero if the argument is printable — a letter, digit, or punctuation character.

iscntrl(c) returns non-zero if the argument is a control character.

isascii(c) returns non-zero if the argument is an ascii character, i.e., less than octal 0200.

toupper(c) returns the upper-case character corresponding to the lower-case letter **c.**

tolower(c) returns the lower-case character corresponding to the upper-case letter **c**.

Make — A Program for Maintaining Computer Programs

S. I. Feldman

ABSTRACT

In a programming project, it is easy to lose track of which files need to be reprocessed or recompiled after a change is made in some part of the source. *Make* provides a simple mechanism for maintaining up-to-date versions of programs that result from many operations on a number of files. It is possible to tell *Make* the sequence of commands that create certain files, and the list of files that require other files to be current before the operations can be done. Whenever a change is made in any part of the program, the *Make* command will create the proper files simply, correctly, and with a minimum amount of effort.

The basic operation of *Make* is to find the name of a needed target in the description, ensure that all of the files on which it depends exist and are up to date, and then create the target if it has not been modified since its generators were. The description file really defines the graph of dependencies; *Make* does a depth-first search of this graph to determine what work is really necessary.

Make also provides a simple macro substitution facility and the ability to encapsulate commands in a single file for convenient administration.

August 15, 1978

Make — A Program for Maintaining Computer Programs

S. I. Feldman

Introduction

It is common practice to divide large programs into smaller, more manageable pieces. The pieces may require quite different treatments: some may need to be run through a macro processor, some may need to be processed by a sophisticated program generator (e.g., Yacc[1] or Lex[2]). The outputs of these generators may then have to be compiled with special options and with certain definitions and declarations. The code resulting from these transformations may then need to be loaded together with certain libraries under the control of special options. Related maintenance activities involve running complicated test scripts and installing validated modules. Unfortunately, it is very easy for a programmer to forget which files depend on which others, which files have been modified recently, and the exact sequence of operations needed to make or exercise a new version of the program. After a long editing session, one may easily lose track of which files have been changed and which object modules are still valid, since a change to a declaration can obsolete a dozen other files. Forgetting to compile a routine that has been changed or that uses changed declarations will result in a program that will not work, and a bug that can be very hard to track down. On the other hand, recompiling everything in sight just to be safe is very wasteful.

The program described in this report mechanizes many of the activities of program development and maintenance. If the information on inter-file dependences and command sequences is stored in a file, the simple command

```
make
```

is frequently sufficient to update the interesting files, regardless of the number that have been edited since the last "make". In most cases, the description file is easy to write and changes infrequently. It is usually easier to type the *make* command than to issue even one of the needed operations, so the typical cycle of program development operations becomes

think — edit — *make* — test . . .

Make is most useful for medium-sized programming projects; it does not solve the problems of maintaining multiple source versions or of describing huge programs. *Make* was designed for use on Unix, but a version runs on GCOS.

Basic Features

The basic operation of *make* is to update a target file by ensuring that all of the files on which it depends exist and are up to date, then creating the target if it has not been modified since its dependents were. *Make* does a depth-first search of the graph of dependences. The operation of the command depends on the ability to find the date and time that a file was last modified.

To illustrate, let us consider a simple example: A program named *prog* is made by compiling and loading three C-language files *x.c*, *y.c*, and *z.c* with the *lS* library. By convention, the output of the C compilations will be found in files named *x.o*, *y.o*, and *z.o*. Assume that the files *x.c* and *y.c* share some declarations in a file named *defs*, but that *z.c* does not. That is, *x.c* and *y.c* have the line

```
#include "defs"
```

The following text describes the relationships and operations:

```
prog : x.o y.o z.o
	cc x.o y.o z.o -lS -o prog

x.o y.o : defs
```

If this information were stored in a file named *makefile*, the command

```
make
```

would perform the operations needed to recreate *prog* after any changes had been made to any of the four source files *x.c*, *y.c*, *z.c*, or *defs*.

Make operates using three sources of information: a user-supplied description file (as above), file names and "last-modified" times from the file system, and built-in rules to bridge some of the gaps. In our example, the first line says that *prog* depends on three ".*o*" files. Once these object files are current, the second line describes how to load them to create *prog*. The third line says that *x.o* and *y.o* depend on the file *defs*. From the file system, *make* discovers that there are three ".*c*" files corresponding to the needed ".*o*" files, and uses built-in information on how to generate an object from a source file (*i.e.*, issue a "cc -c" command).

The following long-winded description file is equivalent to the one above, but takes no advantage of *make*'s innate knowledge:

```
prog : x.o y.o z.o
	cc x.o y.o z.o -lS -o prog
x.o : x.c defs
	cc -c x.c
y.o : y.c defs
	cc -c y.c
z.o : z.c
	cc -c z.c
```

If none of the source or object files had changed since the last time *prog* was made, all of the files would be current, and the command

```
make
```

would just announce this fact and stop. If, however, the *defs* file had been edited, *x.c* and *y.c* (but not *z.c*) would be recompiled, and then *prog* would be created from the new ".*o*" files. If only the file *y.c* had changed, only it would be recompiled, but it would still be necessary to reload *prog*.

If no target name is given on the *make* command line, the first target mentioned in the description is created; otherwise the specified targets are made. The command

```
make x.o
```

would recompile *x.o* if *x.c* or *defs* had changed.

If the file exists after the commands are executed, its time of last modification is used in further decisions; otherwise the current time is used. It is often quite useful to include rules with mnemonic names and commands that do not actually produce a file with that name. These entries can take advantage of *make*'s ability to generate files and substitute macros. Thus, an entry "save" might be included to copy a certain set of files, or an entry "cleanup" might be used to throw away unneeded intermediate files. In other cases one may maintain a zero-length file purely to keep track of the time at which certain actions were performed. This technique is useful for maintaining remote archives and listings.

Make has a simple macro mechanism for substituting in dependency lines and command strings. Macros are defined by command arguments or description file lines with embedded equal signs. A macro is invoked by preceding the name by a dollar sign; macro names longer than one character must be parenthesized. The name of the macro is either the single character after the dollar sign or a name inside parentheses. The following are valid macro invocations:

```
$(CFLAGS)
$2
$(xy)
$Z
$(Z)
```

The last two invocations are identical. $$ is a dollar sign. All of these macros are assigned values during input, as shown below. Four special macros change values during the execution of the command: $*, $@, $?, and $<. They will be discussed later. The following fragment shows the use:

```
OBJECTS = x.o y.o z.o
LIBES = -lS
prog: $(OBJECTS)
	cc $(OBJECTS)  $(LIBES)  -o prog
  . . .
```

The command

```
make
```

loads the three object files with the *lS* library. The command

```
make "LIBES= -ll -lS"
```

loads them with both the Lex ("-ll") and the Standard ("-lS") libraries, since macro definitions on the command line override definitions in the description. (It is necessary to quote arguments with embedded blanks in UNIX† commands.)

The following sections detail the form of description files and the command line, and discuss options and built-in rules in more detail.

Description Files and Substitutions

A description file contains three types of information: macro definitions, dependency information, and executable commands. There is also a comment convention: all characters after a sharp (#) are ignored, as is the sharp itself. Blank lines and lines beginning with a sharp are totally ignored. If a non-comment line is too long, it can be continued using a backslash. If the last character of a line is a backslash, the backslash, newline, and following blanks and tabs are replaced by a single blank.

A macro definition is a line containing an equal sign not preceded by a colon or a tab. The name (string of letters and digits) to the left of the equal sign (trailing blanks and tabs are stripped) is assigned the string of characters following the equal sign (leading blanks and tabs are stripped.) The following are valid macro definitions:

```
2 = xyz
abc = -ll -ly -lS
LIBES =
```

The last definition assigns LIBES the null string. A macro that is never explicitly defined has the null string as value. Macro definitions may also appear on the *make* command line (see below).

Other lines give information about target files. The general form of an entry is:

```
target1 [target2 . . .] :[:] [dependent1 . . .] [; commands] [# . . .]
[(tab) commands] [# . . .]
  . . .
```

Items inside brackets may be omitted. Targets and dependents are strings of letters, digits, periods, and slashes. (Shell metacharacters "*" and "?" are expanded.) A command is any string

† UNIX is a trademark of Bell Laboratories.

of characters not including a sharp (except in quotes) or newline. Commands may appear either after a semicolon on a dependency line or on lines beginning with a tab immediately following a dependency line.

A dependency line may have either a single or a double colon. A target name may appear on more than one dependency line, but all of those lines must be of the same (single or double colon) type.

1. For the usual single-colon case, at most one of these dependency lines may have a command sequence associated with it. If the target is out of date with any of the dependents on any of the lines, and a command sequence is specified (even a null one following a semicolon or tab), it is executed; otherwise a default creation rule may be invoked.

2. In the double-colon case, a command sequence may be associated with each dependency line; if the target is out of date with any of the files on a particular line, the associated commands are executed. A built-in rule may also be executed. This detailed form is of particular value in updating archive-type files.

If a target must be created, the sequence of commands is executed. Normally, each command line is printed and then passed to a separate invocation of the Shell after substituting for macros. (The printing is suppressed in silent mode or if the command line begins with an @ sign). *Make* normally stops if any command signals an error by returning a non-zero error code. (Errors are ignored if the "–i" flags has been specified on the *make* command line, if the fake target name ".IGNORE" appears in the description file, or if the command string in the description file begins with a hyphen. Some UNIX commands return meaningless status). Because each command line is passed to a separate invocation of the Shell, care must be taken with certain commands (e.g., *cd* and Shell control commands) that have meaning only within a single Shell process; the results are forgotten before the next line is executed.

Before issuing any command, certain macros are set. **$@** is set to the name of the file to be "made". **$?** is set to the string of names that were found to be younger than the target. If the command was generated by an implicit rule (see below), **$<** is the name of the related file that caused the action, and **$*** is the prefix shared by the current and the dependent file names.

If a file must be made but there are no explicit commands or relevant built-in rules, the commands associated with the name ".DEFAULT" are used. If there is no such name, *make* prints a message and stops.

Command Usage

The *make* command takes four kinds of arguments: macro definitions, flags, description file names, and target file names.

    make [flags] [macro definitions] [targets]

The following summary of the operation of the command explains how these arguments are interpreted.

First, all macro definition arguments (arguments with embedded equal signs) are analyzed and the assignments made. Command-line macros override corresponding definitions found in the description files.

Next, the flag arguments are examined. The permissible flags are

–i Ignore error codes returned by invoked commands. This mode is entered if the fake target name ".IGNORE" appears in the description file.

–s Silent mode. Do not print command lines before executing. This mode is also entered if the fake target name ".SILENT" appears in the description file.

–r Do not use the built-in rules.

–n No execute mode. Print commands, but do not execute them. Even lines beginning with an "@" sign are printed.

–t Touch the target files (causing them to be up to date) rather than issue the usual commands.

–q Question. The *make* command returns a zero or non-zero status code depending on whether the target file is or is not up to date.

–p Print out the complete set of macro definitions and target descriptions

–d Debug mode. Print out detailed information on files and times examined.

–f Description file name. The next argument is assumed to be the name of a description file. A file name of "–" denotes the standard input. If there are no "–f" arguments, the file named *makefile* or *Makefile* in the current directory is read. The contents of the description files override the built-in rules if they are present).

Finally, the remaining arguments are assumed to be the names of targets to be made; they are done in left to right order. If there are no such arguments, the first name in the description files that does not begin with a period is "made".

Implicit Rules

The *make* program uses a table of interesting suffixes and a set of transformation rules to supply default dependency information and implied commands. (The Appendix describes these tables and means of overriding them.) The default suffix list is:

.o	Object file
.c	C source file
.e	Efl source file
.r	Ratfor source file
.f	Fortran source file
.s	Assembler source file
.y	Yacc-C source grammar
.yr	Yacc-Ratfor source grammar
.ye	Yacc-Efl source grammar
.l	Lex source grammar

The following diagram summarizes the default transformation paths. If there are two paths connecting a pair of suffixes, the longer one is used only if the intermediate file exists or is named in the description.

.o

.c *.r* *.e* *.f* *.s* *.y* *.yr* *.ye* *.l* *.d*

.y *.l* *.yr* *.ye*

If the file *x.o* were needed and there were an *x.c* in the description or directory, it would be compiled. If there were also an *x.l*, that grammar would be run through Lex before compiling the result. However, if there were no *x.c* but there were an *x.l*, *make* would discard the intermediate C-language file and use the direct link in the graph above.

It is possible to change the names of some of the compilers used in the default, or the flag arguments with which they are invoked by knowing the macro names used. The compiler names are the macros AS, CC, RC, EC, YACC, YACCR, YACCE, and LEX. The command

make CC=newcc

will cause the "newcc" command to be used instead of the usual C compiler. The macros CFLAGS, RFLAGS, EFLAGS, YFLAGS, and LFLAGS may be set to cause these commands to be issued with optional flags. Thus,

```
make "CFLAGS= -O"
```

causes the optimizing C compiler to be used.

Example

As an example of the use of *make,* we will present the description file used to maintain the *make* command itself. The code for *make* is spread over a number of C source files and a Yacc grammar. The description file contains:

```
# Description file for the Make command
P = und -3 | opr -r2      # send to GCOS to be printed
FILES = Makefile version.c defs main.c doname.c misc.c files.c dosys.c     gram.y lex.c gcos.c
OBJECTS = version.o main.o doname.o misc.o files.o dosys.o gram.o
LIBES= -lS
LINT = lint -p
CFLAGS = -O

make:  $(OBJECTS)
	cc $(CFLAGS) $(OBJECTS) $(LIBES) -o make
	size make

$(OBJECTS):  defs
gram.o: lex.c

cleanup:
	-rm *.o gram.c
	-du

install:
	@size make /usr/bin/make
	cp make /usr/bin/make ; rm make

print:  $(FILES)  # print recently changed files
	pr $? | $P
	touch print

test:
	make -dp | grep -v TIME >1zap
	/usr/bin/make -dp | grep -v TIME >2zap
	diff 1zap 2zap
	rm 1zap 2zap

lint :  dosys.c doname.c files.c main.c misc.c version.c gram.c
	$(LINT) dosys.c doname.c files.c main.c misc.c version.c gram.c
	rm gram.c

arch:
	ar uv /sys/source/s2/make.a $(FILES)
```

Make usually prints out each command before issuing it. The following output results from typing the simple command

```
make
```

in a directory containing only the source and description file:

```
cc -c version.c
cc -c main.c
cc -c doname.c
cc -c misc.c
cc -c files.c
cc -c dosys.c
yacc gram.y
mv y.tab.c gram.c
cc -c gram.c
cc version.o main.o doname.o misc.o files.o dosys.o gram.o -lS -o make
13188+3348+3044 = 19580b = 046174b
```

Although none of the source files or grammars were mentioned by name in the description file, *make* found them using its suffix rules and issued the needed commands. The string of digits results from the "size make" command; the printing of the command line itself was suppressed by an @ sign. The @ sign on the *size* command in the description file suppressed the printing of the command, so only the sizes are written.

The last few entries in the description file are useful maintenance sequences. The "print" entry prints only the files that have been changed since the last "make print" command. A zero-length file *print* is maintained to keep track of the time of the printing; the $? macro in the command line then picks up only the names of the files changed since *print* was touched. The printed output can be sent to a different printer or to a file by changing the definition of the *P* macro:

```
make print "P = opr -sp"
      or
make print "P=  cat >zap"
```

Suggestions and Warnings

The most common difficulties arise from *make*'s specific meaning of dependency. If file *x.c* has a "#include "defs"" line, then the object file *x.o* depends on *defs*; the source file *x.c* does not. (If *defs* is changed, it is not necessary to do anything to the file *x.c*, while it is necessary to recreate *x.o*.)

To discover what *make* would do, the "-n" option is very useful. The command

```
make -n
```

orders *make* to print out the commands it would issue without actually taking the time to execute them. If a change to a file is absolutely certain to be benign (e.g., adding a new definition to an include file), the "-t" (touch) option can save a lot of time: instead of issuing a large number of superfluous recompilations, *make* updates the modification times on the affected file. Thus, the command

```
make -ts
```

("touch silently") causes the relevant files to appear up to date. Obvious care is necessary, since this mode of operation subverts the intention of *make* and destroys all memory of the previous relationships.

The debugging flag ("-d") causes *make* to print out a very detailed description of what it is doing, including the file times. The output is verbose, and recommended only as a last resort.

Acknowledgments

I would like to thank S. C. Johnson for suggesting this approach to program maintenance control. I would like to thank S. C. Johnson and H. Gajewska for being the prime guinea pigs during development of *make*.

References

1. S. C. Johnson, "Yacc — Yet Another Compiler-Compiler", Bell Laboratories Computing Science Technical Report #32, July 1978.

2. M. E. Lesk, "Lex — A Lexical Analyzer Generator", Computing Science Technical Report #39, October 1975.

Appendix. Suffixes and Transformation Rules

The *make* program itself does not know what file name suffixes are interesting or how to transform a file with one suffix into a file with another suffix. This information is stored in an internal table that has the form of a description file. If the "–r" flag is used, this table is not used.

The list of suffixes is actually the dependency list for the name ".SUFFIXES"; *make* looks for a file with any of the suffixes on the list. If such a file exists, and if there is a transformation rule for that combination, *make* acts as described earlier. The transformation rule names are the concatenation of the two suffixes. The name of the rule to transform a "*.r*" file to a "*.o*" file is thus "*.r.o*". If the rule is present and no explicit command sequence has been given in the user's description files, the command sequence for the rule ".r.o" is used. If a command is generated by using one of these suffixing rules, the macro **$*** is given the value of the stem (everything but the suffix) of the name of the file to be made, and the macro **$<** is the name of the dependent that caused the action.

The order of the suffix list is significant, since it is scanned from left to right, and the first name that is formed that has both a file and a rule associated with it is used. If new names are to be appended, the user can just add an entry for ".SUFFIXES" in his own description file; the dependents will be added to the usual list. A ".SUFFIXES" line without any dependents deletes the current list. (It is necessary to clear the current list if the order of names is to be changed).

The following is an excerpt from the default rules file:

```
.SUFFIXES : .o .c .e .r .f .y .yr .ye .l .s
YACC=yacc
YACCR=yacc -r
YACCE=yacc -e
YFLAGS=
LEX=lex
LFLAGS=
CC=cc
AS=as -
CFLAGS=
RC=ec
RFLAGS=
EC=ec
EFLAGS=
FFLAGS=
.c.o :
        $(CC) $(CFLAGS) -c $<
.e.o .r.o .f.o :
        $(EC) $(RFLAGS) $(EFLAGS) $(FFLAGS) -c $<
.s.o :
        $(AS) -o $@ $<
.y.o :
        $(YACC) $(YFLAGS) $<
        $(CC) $(CFLAGS) -c y.tab.c
        rm y.tab.c
        mv y.tab.o $@
.y.c :
        $(YACC) $(YFLAGS) $<
        mv y.tab.c $@
```

SYSTEM V/68
ASSEMBLER USER'S GUIDE

Third Edition

First Edition June 1984
Second Edition November 1984

Trademarks

1. INTRODUCTION

This is a reference manual for the MPS/UX resident assembler, *as*. Programmers familiar with the M68000 family of processors should be able to program in *as* by referring to this manual, but this is not a manual for the processor itself. Details about the effects of instructions, meanings of status register bits, handling of interrupts, and many other issues are not dealt with here. This manual, therefore, should be used in conjunction with the following reference manuals:

- M68000 16/32-Bit Microprocessor Programmer's Reference Manual, Fourth Edition; Englewood Cliffs, NJ: PRENTICE-HALL, 1984. This manual is also available from the Motorola Literature Distribution Center, P.O. Box 20912, Phoenix, AZ 85036, part number M68000UM.
- MC68020 32-Bit Microprocessor User's Manual; Englewood Cliffs, NJ: PRENTICE-HALL, 1984. This manual is also available from the Motorola Literature Distribution Center, part number MC68020UM.
- M68000 Family Resident Structured Assembler Reference Manual, part number M68KMASM.
- SYSTEM V/68 User's Manual, part number M68KUNUM.
- SYSTEM V/68 VM04 System Manual, part number M68KVM4SYS. This document includes user manual pages to support the MC68881 floating point co-processor provided in SYSTEM V/68 Release 2, Version 2 from Motorola Corp.

This guide also contains information for users of the SGS M68020 Cross Compilation System. For these users, references to *as*(1) and *cc*(1) should be read as *as20*(1) and *cc20*(1). Information about these commands is provided in the SGS M68020 Cross Compilation System Reference Manual, part number M68KUNASX.

2. WARNINGS

A few important warnings to the *as* user should be emphasized at the outset. Though for the most part there is a direct correspondence between *as* notation and the notation used in the documents listed in the preceding section, several exceptions exist that could lead the unsuspecting user to write incorrect code. In addition to the exceptions described in the following paragraphs, refer also to sections 10 and 11 for information about address mode syntax and machine instructions.

2.1. Comparison Instructions

First, the order of the operands in *compare* instructions follows one convention in the M68000 Programmer's Reference Manual and the opposite convention in *as*. Using the convention of the M68000 Programmer's Reference Manual, one might write

```
CMP.W    D5,D3       Is D3 less than D5?
BLE      IS_LESS     Branch if less.
```

Using the *as* convention, one would write

```
cmp.w    %d3,%d5     # Is d3 less than d5?
ble      is_less     # Branch if less.
```

As follows the convention used by other assemblers supported in the UNIX® operating system (both the 3B20S and the VAX follow this convention). This convention makes for straightforward reading of *compare*-and-*branch* instruction sequences, but does nonetheless lead to the peculiarity that if a *compare* instruction is replaced by a *subtract* instruction, the effect on the condition codes will be entirely different. This may be confusing to programmers who are used to thinking of a comparison as a subtraction whose result is not stored. Users of *as* who become accustomed to the convention will find that both the *compare* and *subtract* notations make sense in their respective contexts.

2.2. Overloading of Opcodes

Another issue that users must be aware of arises from the M68000 processors' use of several different instructions to do more or less the same thing. For example, the M68000 Programmer's Reference Manual lists the instructions **SUB, SUBA, SUBI,** and **SUBQ**, which all have the effect of subtracting their source operand from their destination operand. *As* provides the convenience of allowing all these operations to be specified by a single assembly instruction **sub**. On the basis of the operands given to the **sub** instruction, the *as* assembler selects the appropriate M68000 operation code. The danger created by this convenience is that it could leave the misleading impression that all forms of the **SUB** operation are semantically identical. In fact, they are not. The careful reader of the M68000 Programmer's Reference Manual will notice that whereas **SUB, SUBI,** and **SUBQ** all affect the condition codes in a consistent way, **SUBA** does not affect the condition codes at all. Consequently, the *as* user must be aware that when the destination of a **sub** instruction is an address register (which causes the **sub** to be mapped into the operation code for **SUBA**), the condition codes will not be affected.

3. USE OF THE ASSEMBLER

The SYSTEM V/68 command *as* invokes the assembler and has the following syntax:

as [**-o** *output*] *file*

When *as* is invoked with the **-o** *output* flag, the output of the assembly is put in the file *output*. If the -o flag is not specified, the output is left in a file whose name is formed by removing the **.s** suffix, if there is one, from the input filename and appending a **.o** suffix.

The M68020 cross assembler, *as20*(1), is invoked with the same syntax as *as*(1). For information about additional options for these commands, refer to the *SYSTEM V/68 User's Manual* for *as*(1) and the *SGS M68020 Cross Compilation System Reference Manual* for *as20*(1).

UNIX is a registered trademark of AT&T.

4. GENERAL SYNTAX RULES

4.1. Format of Assembly Language Line

Typical lines of *as* assembly code look like these:

```
# Clear a block of memory at location %a3

        text    2
        move.w  &const,%d1
loop:   clr.l   (%a3)+
        dbf     %d1,loop        # go back for const
                                # repetitions

init2:
        clr.l count;  clr.l credit;  clr.l debit;
```

These general points about the example should be noted:

— An identifier occurring at the beginning of a line and followed by a colon (:) is a *label*. One or more *labels* may precede any assembly language instruction or pseudo-operation. Refer to Section 5.2, "Location Counters and Labels."

— A line of assembly code need not include an instruction. It may consist of a comment alone (introduced by #), a label alone (terminated by :), or it may be entirely blank.

— It is good practice to use tabs to align assembly language operations and their operands into columns, but this is not a requirement of the assembler. An opcode may appear at the beginning of the line, if desired, and spaces may precede a label. A single blank or tab suffices to separate an opcode from its operands. Additional blanks and tabs are ignored by the assembler.

— It is permissible to write several instructions on one line separating them by semicolons. The semicolon is syntactically equivalent to a newline character; however, a semicolon inside a comment is ignored.

4.2. Comments

Comments are introduced by the character # and continue to the end of the line. Comments may appear anywhere and are completely disregarded by the assembler.

4.3. Identifiers

An identifier is a string of characters taken from the set **a-z, A-Z, _, ~, %,** and **0-9**. The first character of an identifier must be a letter (uppercase or lowercase) or an underscore. Uppercase and lowercase letters are distinguished; for example, **con35** and **CON35** are two distinct identifiers.

There is no limit on the length of an identifier.

The value of an identifier is established by the **set** pseudo-operation (refer to Section 8.2, "Symbol Definition Operations") or by using it as a label. Refer to Section 5.2, "Location Counters and Labels".

The tilde character (~) has special significance to the assembler. A ~ used alone, as an identifier, means "the current location". A ~ used as the first character in an identifier becomes a period (.) in the symbol table, allowing symbols such as **.eos** and **.Ofake** to be entered into the symbol table, as required by the Common Object File format (COFF). Information about file formats is provided in the *SYSTEM V/68 User's Manual,* Section 4.

4.4. Register Identifiers

A register identifier is an identifier preceded by the character **%**, and represents one of the MC68000 processor's registers. The predefined register identifiers are;

%d0	%d4	%a0	%a4	%acc	%usp
%d1	%d5	%a1	%a5	%pc	%fp
%d2	%d6	%a2	%a6	%sp	
%d3	%d7	%a3	%a7	%sr	

Note: The identifiers **%a7** and **%sp** represent the same machine register. Likewise, **%a6** and **%fp** are equivalent. Use of both **%a7** and **%sp**, or **%a6** and **%fp**, in the same program may result in confusion.

The current version of the assembler will correctly assemble instructions intended for the M68010. There will be a warning message issued. The following additions will be flagged with warnings:

REGISTERS ADDED FOR THE MC68010	
NAME	DESCRIPTION
%sfc	Source Function Code Register
%dfc	Destination Function Code Register
%vbr	Vector Base Register

The entire register set of the MC68000 and MC68010 is included in the MC68020 register set. The following are new control registers for the MC68020:

MC68020 REGISTERS	
NAME	DESCRIPTION
%caar	Cache Address Register
%cacr	Cache Control Register
%isp	Interrupt Stack Pointer
%msp	Master Stack Pointer

The following are suppressed registers (zero registers) used in various MC68020 addressing modes:

MC68020 ZERO REGISTERS		
SUPPRESSED ADDRESS REGISTERS	SUPPRESSED DATA REGISTERS	SUPPRESSED PROGRAM COUNTER
%za0	%zd0	%zpc
%za1	%zd1	
%za2	%zd2	
%za3	%zd3	
%za4	%zd4	
%za5	%zd5	
%za6	%zd6	
%za7	%zd7	

4.5. Constants

As deals only with integer constants. They may be entered in decimal, octal, or hexadecimal, or they may be entered as character constants. Internally, *as* treats all constants as 32-bit binary two's complement quantities.

4.5.1. Numerical Constants. A decimal constant is a string of digits beginning with a non-zero digit. An octal constant is a string of digits beginning with zero. A hexadecimal constant consists of the characters **0x** or **0X** followed by a string of characters from the set **0-9, a-f,** and **A-F**. In hexadeximal constants, uppercase and lowercase letters are not distinguished.

Examples:

```
set      const,35      # Decimal 35
mov.w    &035,%d1      # Octal 35 (decimal 29)
set      const, 0x35   # Hex 35 (decimal 53)
mov.w    &0xff, %d1    # Hex ff (decimal 255)
```

4.5.2. Character Constants. An ordinary character constant consists of a single-quote character (') followed by an arbitrary ASCII character other than the backslash (\). The value of the constant is equal to the ASCII code for the character. Special meanings of characters are overridden when used in character constants; for example, if '# is used, the # is not treated as introducing a comment.

A special character constant consists of '\ followed by another character. All the special character constants and examples of ordinary character constants are listed in the following table.

CONSTANT	VALUE	MEANING
'\ b	0x08	Backspace
'\ t	0z09	Horizontal Tab
'\ n	0x0a	Newline (Line Feed)
'\ v	0x0b	Vertical Tab
'\ f	0x0c	Form Feed
'\ r	0x0d	Carriage Return
'\\	0x5c	Backslash
' '	0x27	Single Quote
'0	0x30	Zero
'A	0x41	Uppercase A
'a	0x61	Lowercase a

4.6. Other Syntactic Details

A discussion of expression syntax appears in Section 7 of this guide. Information about the syntax of specific components of *as* instructions and pseudo-operations is given in Sections 8, 9, and 10.

5. SEGMENTS, LOCATION COUNTERS, AND LABELS

5.1. Segments

A program in *as* assembly language may be broken into segments known as *text, data* and *bss* segments. The convention regarding the use of these segments is to place instructions in *text* segments, initialized data in *data* segments, and uninitialized data in *bss* segments. However, the assembler does not enforce this convention; for example, it permits intermixing of instructions and data in a *text* segment.

Primarily to simplify compiler code generation, the assembler permits up to four separate *text* segments and four separate *data* segments named **0, 1, 2,** and **3**. The assembly language program may switch freely between them by using assembler pseudo-operations (refer to Section 8.3, "Location Counter Control Operations"). When generating the object file, the assembler concatenates the *text* segments to generate a single *text* segment, and the *data* segments to generate a single *data* segment. Thus, the object file contains only one *text* segment and only one *data* segment. There is always only one *bss* segment and it maps directly into the object file.

Because the assembler keeps together everything from a given segment when generating the object file, the order in which information appears in the object file may not be the same as in the assembly language file. For example, if the data for a program consisted of

```
data    1              # segment 1
short   0x1111
data    0              # segment 0
long    0xff ff ff ff
data    1              # segment 1
byte    0xff
```

then equivalent object code would be generated by

data	0
long	0xffffffff
short	0x1111
byte	0xff

5.2. Location Counters and Labels

The assembler maintains separate *location counters* for the *bss* segment and for each of the *text* and *data* segments. The location counter for a given segment is incremented by one for each byte generated in that segment.

The location counters allow values to be assigned to labels. When an identifier is used as a label in the assembly language input, the current value of the current location counter is assigned to the identifier. The assembler also keeps track of which segment the label appeared in. Thus, the identifier represents a memory location relative to the beginning of a particular segment. Any label relative to the location counter should be within the text segment.

6. TYPES

Identifiers and expressions may have values of different types.

— In the simplest case, an expression (or identifier) may have an *absolute* value, such as 29, -5000, or 262143.

— An expression (or identifier) may have a value relative to the start of a particular segment. Such a value is known as a *relocatable* value. The memory location represented by such an expression cannot be known at assembly time, but the relative values of two such expressions (i.e., the difference between them) can be known if they refer to the same segment.

  Identifiers which appear as labels have *relocatable* values.

— If an identifier is never assigned a value, it is assumed to be an *undefined external*. Such identifiers may be used with the expectation that their values will be defined in another program, and therefore known at load time; but the relative values of *undefined externals* cannot be known.

7. EXPRESSIONS

For conciseness, the following abbreviations are useful:

abs	absolute expression
rel	relocatable expression
ext	undefined external

All constants are absolute expressions. An identifier may be thought of as an expression having the identifier's type. Expressions may be built up from lesser expressions using the operators +, -, *, and /, according to the following type rules:

abs + abs = abs
abs + rel = rel + abs = rel
abs + ext = ext + abs = ext

abs — abs = abs
rel — abs = rel
ext — abs = ext
rel — rel = abs (provided that
the two relocatable expressions are relative to the same segment)

abs * abs = abs

abs / abs = abs

— abs = abs

Note: rel — rel expressions are permitted only within the context of a switch statement (refer to Section 8.5, "Switch Table Operation".) Use of a **rel — rel** expression is dangerous, particularly when dealing with identifiers from *text* segments. The problem is that the assembler will determine the value of the expression before it has resolved all questions concerning span-dependent optimizations.

The unary minus operator takes the highest precedence; the next highest precedence is given to * and /, and lowest precedence is given to + and -. Parentheses may be used to coerce the order of evaluation.

If the result of a division is a positive non-integer, it will be truncated toward zero. If the result is a negative non-integer, the direction of truncation cannot be guaranteed.

8. PSEUDO-OPERATIONS

8.1. Data Initialization Operations

byte *abs,abs,...*

One or more arguments, separated by commas, may be given. The values of the arguments are computed to produce successive bytes in the assembly output.

short *abs,abs,...*

One or more arguments, separated by commas, may be given. The values of the arguments are computed to produce successive 16-bit words in the assembly output.

long *expr,expr,...*

One or more arguments, separated by commas, may be given. Each expression may be *absolute*, *relocatable*, or *undefined external.* A 32-bit quantity is generated for each such argument (in the case of *relocatable* or *undefined external* expressions, the actual value may not be filled in until load time).

Alternatively, the arguments may be bit-field expressions. A bit-field expression has the form

n : *value*

where both *n* and *value* denote *absolute* expressions. The quantity *n* represents a field width; the low-order *n* bits of *value* become the contents of the bit-field.

Successive bit-fields fill up 32-bit long quantities starting with the high-order part. If the sum of the lengths of the bit-fields is less than 32 bits, the assembler creates a 32-bit long with zeroes filling out the low-order bits. For example,

long **4: -1, 16: 0x7f, 12:0, 5000**

and

long **4: -1, 16: 0x7f, 5000**

are equivalent to

long **0xf007f000, 5000**

Bit-fields may not span pairs of 32-bit longs. Thus,

long **24: 0xa, 24: 0xb, 24:0xc**

yields the same thing as

long **0x00000a00, 0x00000b00, 0x00000c00**

space *abs*

The value of *abs* is computed, and the resultant number of bytes of zero data is generated. For example,

space **6**

is equivalent to

byte **0,0,0,0,0,0**

8.2. Symbol Definition Operations

set *identifier,expr*

The value of *identifier* is set equal to *expr*, which may be absolute or relocatable.

comm *identifier,abs*

The named identifier is to be assigned to a common area of size *abs* bytes. If *identifier* is not defined by another program, the loader will allocate space for it.

lcomm *identifier,abs*

The named *identifier* is assigned to a *local common* of size *abs* bytes. This results in allocation of space in the *bss* segment.

The type of *identifier* becomes *relocatable*.

global *identifier*

This causes *identifier* to be externally visible. If *identifier* is defined in the current program, then declaring it global allows the loader to resolve references to *identifier* in other programs.

If *identifier* is not defined in the current program, the assembler expects an external resolution; in this case, therefore, *identifier* is global by default.

8.3. Location Counter Control Operations

data *abs*

The argument, if present, must evaluate to 0, 1, 2, or 3; this indicates the number of the data segment into which assembly is to be directed. If no argument is present, assembly is directed into *data* segment 0.

text *abs*

The argument, if present, must evaluate to 0, 1, 2, or 3; this indicates the number of the *text* segment into which assembly is to be directed. If no argument is present, assembly is directed into *text* segment 0.

Before the first **text** or **data** operation is encountered, assembly is by default directed into *text* segment 0.

org *expr*

The current location counter is set to *expr*. *Expr* must represent a value in the current segment, and must not be less than the current location counter.

even

The current location counter is rounded up to the next even value.

8.4. Symbolic Debugging Operations

The assembler allows for symbolic debugging information to be placed into the object code file with special pseudo-operations. The information typically includes line numbers and information about C language symbols, such as their type and storage class. The C compiler (*cc*(1)) generates symbolic debugging information when the **-g** option is used. Assembler programmers may also include such information in source files.

8.4.1. file and ln. The **file** pseudo-operation passes the name of the source file into the object file symbol table. It has the form

file *filename*

where *filename* consists of one to 14 characters enclosed in quotation marks.

The **ln** pseudo-operation makes a line number table entry in the object file. That is, it associates a line number with a memory location. Usually the memory location is the current location in text. The format is

ln *line*[,*value*]

where *line* is the line number. The optional value is the address in *text, data,* or *bss* to associate with the line number. The default when *value* is omitted (which is usually the case) is the current location in *text*.

8.4.2. Symbol Attribute Operations. The basic symbolic testing pseudo-operations are **def** and **endef**. These operations enclose other pseudo-operations that assign attributes to a symbol and must be paired.

```
def     name
.               # Attribute
.               # Assigning
.               # Operations
endef
```

NOTES

- **def** does not define the symbol, although it does create a symbol table entry. Because an undefined symbol is treated as external, a symbol which appears in a **def**, but which never acquires a value, will ultimately result in an error at link edit time.
- to allow the assembler to calculate the sizes of functions for other tools, each **def/endef** pair that defines a function name must be matched by a **def/endef** pair after the function in which a storage class of **—1** is assigned.

The paragraphs below describe the attribute-assigning operations. Keep in mind that all of these operations apply to symbol *name* which appeared in the opening **def** pseudo-operation.

val *expr*

Assigns the value *expr* to *name*. the type of the expression *expr* determines with which section *name* is associated. If value is ˜, the current location in the *text* section is used.

scl *expr*

Declares the C language type of *name*. The expression *expr* must yield an ABSOLUTE value that corresponds to the C compiler's internal representation of a storage class. The special value *—1* designates the physical end of a function.

type *expr*

Declares the C language type of *name*. The expression *expr* must yield an ABSOLUTE value that corresponds to the C compiler's internal representation of a basic or derived type.

tag *str*

Associates *name* with the structure, enumeration, or union named *str* which must have already been declared with a **def/endef** pair.

line *expr*

Provides the line number of *name*, where *name* is a block symbol. The expression *expr* should yield an ABSOLUTE value that represents a line number.

size *expr*

Gives a size for *name*. The expression *expr* must yield an ABSOLUTE value. When *name* is a structure or an array with a predetermined extent, *expr* gives the size in bytes. For bit fields, the size is in bits.

dim *expr1,expr2,...*

Indicates that *name* is an array. Each of the expressions must yield an ABSOLUTE value that provides the corresponding array dimension.

8.5. Switch Table Operation

The C compiler generates a compact set of instructions for the C language *switch* construct. An example is shown below.

```
        sub.l   &1,%d0
        cmp.l   %d0,&4
        bhi     L%21
        add.w   %d0,%d0
        mov.w   10(%pc,%d0.w),%d0
        jmp     6(%pc,%d0.w)
        swbeg   &5
L%22:
        short   L%15-L%22
        short   L%21-L%22
        short   L%16-L%22
        short   L%21-L%22
        short   L%17-L%22
```

The special **swbeg** pseudo-operation communicates to the assembler that the lines following it contain **rel-rel** subtractions. Remember that ordinarily such subtractions are risky because of span-dependent optimization. In this case, however, the assembler makes special allowances for the subtraction because the compiler guarantees that both symbols will be defined in the current assembler file, and that one of the symbols is a fixed distance away from the current location.

The **swbeg** pseudo-operation takes an argument that looks like an immediate operand. The argument is the number of lines that follow **swbeg** and that contain switch table entries. **Swbeg** inserts two words into text. The first is the **ILLEGAL** instruction code. The second is the number of table entries that follow. The disassembler *dis*(1) needs the **ILLEGAL** instruction as a hint that what follows is a switch table. Otherwise, it would get confused when it tried to decode the table entries, differences between two symbols, as instructions.

9. SPAN-DEPENDENT OPTIMIZATION

The assembler makes certain choices about the object code it generates based on the distance between an instruction and its operand(s). Span-dependent optimization occurs most obviously in the choice of object code for branches and jumps. It also occurs when an operand may be represented by the program counter relative address mode instead of as an absolute 2-word (**long**) address. The span-dependent optimization capability is normally enabled; the **-n** command line flag disables it. When this capability is disabled, the assembler makes worst-case assumptions about the types of object code that must be generated. Span-dependent optimizations are performed only within **text** segment 0. Any reference outside **text** segment 0 is assumed to be worst-case.

The C compiler (*cc*(1)) generates branch instructions without a specific offset size. When the optimizer is used, it identifies branches which could be represented by the short form, and it changes the operation accordingly. The assembler chooses only between long and very-long representations for branches.

For the MC68000 and MC68010 processors, branch instructions, e.g., **bra, bsr,** or **bgt**, can have either a byte or a word pc-relative address operand. A byte or word size specification should be used only when the user is sure that the address intended can be represented in the byte or word allowed. The assembler will take one of these instructions with a size specification and generate the byte or word form of the instruction without asking questions.

Although the largest offset specification allowed for the M68000 and M68010 is a word,* large programs could conceivably have need for a branch to location not reachable by a word displacement. Therefore, equivalent long forms of these instructions might be needed. When the Assembler encounters a branch instructions without a size specification, it tries to choose between the long and very-long forms of the instruction. If the operand can be represented in a word, then the word form of the instruction will be generated. Otherwise, the very-long form will be generated. For unconditional branches, e.g., **br, bra,** and **bsr,** the very-long form is just the equivalent jump (**jmp** and **jsr**) with an absolute address operand (instead of pc-relative). For conditional branches, the equivalent very-long form is a conditional branch around a jump, where the conditional test has been reversed.

The following table summarizes span-dependent optimizations. The assembler chooses only between the long form and the very-long form, while the optimizer chooses between the short and long forms for branches (but not **bsr**).

ASSEMBLER SPAN-DEPENDENT OPTIMIZATIONS			
Instruction	Short Form	Long Form	Very-Long Form
br, bra, bsr	**byte offset**	word offset (See footnote for information about M68020.)	**jmp** or **jsr** with absolute long address
conditional branch	byte offset	word offset (See footnote for information about M68020.)	short conditional branch with reversed condition around **jmp** with absolute long adres
jmp, jsr		pc-relative address	absolute long address
lea.l, pea.l		pc-relative address	absolute long address

* The M68020 allows long word offset, as shown by the syntax for the branch instructions.

For the MC68020 processor, branch instructions can have either a byte, word, or long pc-relative address operand. The assembler still chooses between word and long representations for branches if no byte size specification is given; however, the long form is replaced by a branch long with pc-relative address instead of a jump with absolute long address.

10. ADDRESS MODE SYNTAX

The following table summarizes the *as* syntax for MC68000, MC68010, and MC68020 addressing modes. New addressing modes for the MB68020 are shown with "MC68020 Only" in parentheses beneath the MC6800 notation; modes not specified in this way are for all three processors.

In the table, the following abbreviations are used:

an Address register, where *n* is any digit from 0 through 7.

dn Data register, where *n* is any digit from 0 through 7.

ri Index register *i* may be any address or data register with an optional size designation (i.e., **ri.w** for 16 bits or **ri.l** for 32 bits); default size is **.w**.

scl Optional scale factor that may be multipled time index register in some modes. Values for *scl* are 1, 2, 4, or 8; default is 1.

bd Two's complement base displacement that is added before indirection takes place; size can be 16 or 32 bits.

od Outer displacement that is added as a part of effective address calculation after memory indirection; size can be 16 or 32 bits.

d Two's complement or sign-extended displacement that is added as part of effective address calculation; size may be 8 or 16 bits; when omitted, assembler uses value of zero.

pc Program counter

[ ] Grouping characters used to enclose an indirect expression; required characters. Addressing arguments can occur in any order within the brackets.

( ) Grouping characters used to enclose an entire effective address; required characters. Addressing arguments can occur in any order within the parentheses.

{ } Indicate that a scale factor is optional; not required characters.

It is important to note that expressions used for the **absolute** addressing modes need not be *absolute expressions* in the sense described in Section 6. Although the addresses used in those addressing modes must ultimately be filled in with constants, that can be done later by the loader. There is no need for the assembler to be able to compute them. Indeed, the **Absolute Long** addressing mode is commonly used for accessing *undefined external* addresses.

EFFECTIVE ADDRESS MODES		
M68000 Family Notation	*as* Notation	Effective Address Mode
Dn	%dn	Data Register Direct
An	%an	Address Register Direct
(An)	(%an)	Address Register Indirect
(An)+	(%an)+	Address Register Indirect With Postincrement
—(An)	—(%an)	Address Register Indirect With Predecrement
d(An)	d(%an)	Address Register Indirect With Displacement (*d* signifies a signed 16-bit absolute displacement)
d(An,Ri)	d(%an,%ri.w) d(%an,%ri.1)	Address Register Indirect With Index Plus Displacement (*d* signifies a signed 8-bit absolute displacement)
(bd,An,Ri{*scl}) (MC68020 Only)	(bd,%an,%ri{*ri})	Address Register Direct With Index Plus Base Displacement
([bd,An,Ri{*scl}],od) (MC68020 Only)	(bd,%an,%ri{*scl}],od)	Memory Indirect With Preindexing Plus Base and Outer Displacement
([bd,An],Ri{*scl},od) (MC68020 Only)	([bd,%an],%ri{*scl}],od)	Memory Indirect With Postindexing Plus Base and Outer Displacement
d(PC)	d(%pc)	Program Counter Indirect With Displacement (d signifies 16-bit displacement)
d(PC,Ri)	d(%pc,%rn.1) d(%pc,%rn.w)	Program Counter Direct With Index and Displacement (d signifies 8-bit displacement)
(bd,PC,Ri{*scl}) (MC68020 Only)	(bd,%pc,%ri{*scl})	Program Counter Direct With Index and Base Displacement
([bd,PC],Ri{*scl},od) (MC68020 Only)	([bd,%pc],%ri{*scl},od)	Program counter Memory Indirect With Postindexing Plus Base and Outer Displacement
([bd,PC,Ri{*scl}],od) (MC68020 Only)	([bd,%pc,%ri{*scl}],od)	Program Counter Memory Indirect With Preindexing Plus Base and Outer Displacement

EFFECTIVE ADDRESS MODES		
M68000 Family Notation	*as* Notation	Effective Address Mode
d,PC,Ri*scl],od) (MCC68020 Only)	d,pc,ri*scl],od)	Program Counter Memory Indirect With Preindexing Plus Base and Outer Displacement
xxx.W	xxx	Absolute Short Address (*xxx* signifies an expression yielding a 16-bit memory address)
xxx.L	xxx	Absolute Long Address (*xxx* signifies an expression yielding a 32-bit memory address)
#xxx	&xxx	Immediate Data (*xxx* signifies an absolute constant expression)

In the table above, the index register notation should be understood as **ri.size*scale**, where both size and scale are optional. Refer to Chapter 2 of the *M68000 Family Resident Structured Assembler Reference Manual* for additional information about effective address modes. Section 2 of the *MC68020 32-Bit Microprocessor User's Manual* also provides information about generating effective addresses and assembler syntax.

Note that suppressed address register **%zan** can be used in place of **%an**, suppressed PC register **%zpc** can be used in place of **%pc**, and suppressed data register **%zdn** can be used in place of **%dn**, if suppression is desired.

The new address modes for the MB68020 use two different formats of extension. The brief format provides fast indexed addressing, while the full format provides a number of options in size of displacement and indirection. The assembler will generate the brief format if the effective address expression is not memory indirect, value of displacement is within a byte, and no base or index suppression is specified; otherwise, the assembler will generate the full format.

Some source code variations of the new modes may be redundant with the MC68000 address register indirect, address register indirect with displacement, and program counter with displacement modes. The assembler will select the more efficient mode when redundancy occurs. For example, when the assembler sees the form (**An**), it will generate address register indirect mode (mode 2). The assembler will generate address register indirect with displacement (mode 5) when seeing any of the following forms (as long as bd fits in 16 bits or less):

bd(An)
(bd,An)
(An,bd)

11. MACHINE INSTRUCTIONS

11.1. Instructions For The MC68000/MC68010/MC68020

The following table shows how MC68000/MC68010/MC68020 instructions should be written in order to be understood correctly by the *as* assembler. The entire instruction set can be used for the MC68020. Instructions that are MC68010/MC68020-only or MC68020-only are noted as such in the "OPERATION" column.

Several abbreviations are used in the table:

S The letter **S**, as in **add.S**, stands for one of the operation size attribute letters **b**, **w**, or **l**, representing a byte, word, or long operation.

A The letter **A**, as in **add.A**, stands for one of the address operation size attribute letters **w** or **l**, representing a word or long operation.

CC In the contexts **bCC**, **dbCC**, and **sCC**, the letters CC represent any of the following condition code designations (except that **f** and **t** may not be used in the **bCC** instruction):

cc	carry clear	ls	low or same
cs	carry set	lt	less than
eq	equal	mi	minus
f	false	ne	not equal
ge	greater or equal	pl	plus
gt	greater than	t	true
hi	high	vc	overflow clear
hs	high or same (=cc)	vs	overflow set
le	less or equal		
lo	low (=cs)		

EA This represents an arbitrary effective address.

I An absolute expression, used as an immediate operand.

Q An absolute expression evaluating to a number from 1 to 8.

L A label reference, or any expression representing a memory address in the current segment.

d Two's complement or sign-extended displacement that is added as part of effective address calculation; size may be 8 by 16 bits; when omitted, assembler uses value of zero.

%dx, %dy, %dn Represent data registers.

%ax, %ay, %an Represent address registers.

%rx, %ry, %rn Represent either data or address registers.

%rc Represents control register (**%sfc, %dfc, %cacr, %usr, %vbr, %caar, %msp, %isp**).

offset Either an immediate operand or a data register.

width Either an immediate operand or a data register.

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
ABCD	abcd.d	%dy, %dx —(%ay),—(%ax)	Add Decimal with Extend
ADD	add.S	EA,%dn %dn,EA	Add Binary
ADDA	add.A	EA,%an	Add Address
ADDI	add.S	&I,EA	Add Immediate
ADDQ	add.S	&Q,EA	Add Quick
ADDX	addx.S	%dy,%dx —(%ay),—(%ax)	Add Extended
AND	and.S	EA,%dn %dn,EA	AND Logical
ANDI	and.S	&I,EA	AND Immediate
ANDI to CCR	and.b	&I,%cc	AND Immediate to Condition Codes
ANDI to SR	and.w	&I,%sr	AND Immediate to the Status Register
ASL	asl.S	%dx,%dy &Q,%dy	Arithmetic Shift (Left)
	als.w als.w	&1,EA EA	
ASR	asr.S	%dx,%dy &Q,%dy	Arithmetic Shift (Right)
	asr.w asr.w	&1,EA EA	
Bcc	bCC	L	Branch Conditionally (16-bit Displacement)
	bCC.b	L	Branch Conditionally (Short) (8-bit Displacement)
	bCC.l	L	Branch Conditionally (Long) (32-bit Displacement) (MC68020 Only)

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
BCHG	bchg	%dn,EA &I,EA	Test a Bit and Change NOTE: **bchg** should be written with no suffix. If the second operand is a data register, **.l** is assumed; otherwise, **.b** is.
BCLR	bclr	%dn,EA &I,EA	Test a Bit and Clear NOTE: **bclr** should be written with no suffix. If the second operand is a data register, **.l** is assumed; otherwise, **.b** is.
BFCHG	bfchg	EA{offset:width}	Complement Bit Field (MC68020 Only)
BFCLR	bfclr	EA{offset:width}	Clear Bit Field (MC68020 Only)
BFEXTS	bfexts	EA{offset:width},%dn	Extract Bit Field (Signed) (MC68020 Only)
BFEXTU	bfextu	EA{offset:width},%dn	Extract Bit Field (Unsigned) (MC68020 Only)
BFFFO	bfffo	EA{offset:width},%dn	Find First One in Bit Field (MC68020 Only)
BFINS	bfins	%dn,EA{offset:width}	Insert Bit Field (MC68020 Only)
BFSET	bfset	EA{offset:width}	Set Bit Field (MC68020 Only)
BFTST	bftst	EA{offset:width}	Test Bit Field (MC68020 Only)
BKPT	bkpt	&I	Breakpoint (MC68020 Only)

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
BRA	bra	L	Branch Always (16-bit Displacement)
	bra.b	L	Branch Always (Short) (8-bit Displacement)
	br.l	L	Branch Always (Long) (32-bit Displacement) (MC68020 Only)
	br	L	Same as **bra**
	br.b	L	Same as **bra.b**
BSET	bset	%dn,EA &I,EA	Test a Bit and Set NOTE: **bset** should be written with no suffix. If the second operand is a data register, **.l** is assumed; otherwise **.b** is.
BSR	bsr	L	Branch to Subroutine (16-bit Displacement)
	bsr.b	L	Branch to Subroutine (Short) (8-bit Displacement)
	bsr.l	L	Branch to Subroutine (Long) (32-bit Displacement) (MC68020 Only)
BTST	btst	%dn,EA &I,EA	Test a Bit and Set NOTE: **btst** should be written with no suffix. If the second operand is a data register, **.l** is assumed; otherwise **.b** is.
CALLM	callm	&I,EA	Call Module (MC68020 Only)
CAS	cas	%ds,%dy,EA	Compare and Swap Operands (MC68020 Only)
CAS2	cas2	%dx:%dy,%dx:%dy,%rx:%ry	Compare and Swap Dual Operands (MC68020 Only)

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
CHK	chk.w	EA,%dn	Check Register Against Bounds
	chk.l	EA,%dn	Check Register Against Bounds (Long) (MC68020 Only)
CHK2	chk2.S	EA,%rn	Check Register Against Bounds (MC68020 Only)
CLR	clr.S	EA	Clear an Operand
CMP	cmp.S	%dn,EA	Compare
CMPA	cmp.A	%an,EA	Compare Address
CMPI	cmp.S	EA,&I	Compare Immediate
CMPM	cmp.S	(%ax)+,(%ay)+	Compare Memory
CMP2	cmp.S	%rn,EA	Compare Register Against Bounds (MC68020 Only)*
DBcc	dbCC	%dn,L	Test Condition, Decrement, and Branch
	dbra	%dn,L	Decrement and Branch Always
	dbr	%dn,L	Same as **dbra**
DIVS	divs.w	EA,%dx	Signed Divide 32/16 → 32
	tdivs.l divs.l	EA,%dx EA,%dx	Signed Divide (Long) 32/32 → 32 (MC68020 Only)
	tdivs.l	EA,%dx:%dy	Signed Divide (Long) 32/32 → 32r:32q (MC68020 Only)
	divs.l	EA,%dx:%dy	Signed Divide (Long) 64/32 → 32r:32q (MC68020 Only)

* Note: The order of operands in *as* is the reverse of that in the M68000 Programmer's Reference Manual.

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
DIVU	divu.w	EA,%dn	Unsigned Divide 32/16 → 32
	tdivu.l divu.l	EA,%dx EA,%dx	Unsigned Divide (Long) 32/32 → 32 (MC68020 Only)
	tdivu.l	EA,%dx:%dy	Unsigned Divide (Long) 32/32 → 32r:32q (MC68020 Only)
	divu.l	EA,%dx:%dy	Unsigned Divide (Long) 64/32 → 32r:32q (MC68020 Only)
EOR	eor.S	%dn,EA	Exclusive OR Logical
EORI	eor.S	&I,EA	Exclusive OR Immediate
EORI to CCR	eor.b	&I,%cc	Exclusive OR Immediate to Condition Code Register
EORI to SR	eor.w	&I,%sr	Exclusive OR Immediate to the Status Register
EXG	exg	%rx,%ry	Exchange Registers
EXT	ext.w	%dn	Sign-Extend Low-Order Byte of Data to Word
	ext.l	%dn	Sign-Extend Low-Order Word of Data to Long
	extb.l	%dn	Sign-Extend Low-Order Byte of Data to Long (MC68020 Only)
	extw.l	%dn	Same as **ext.l** (MC68020 Only)
JMP	jmp	EA	Jump
JSR	jsr	EA	Jump to Subroutine
LEA	lea.l	EA,%an	Load Effective Address
LINK	link	%an,&I	Link and Allocate

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
LSL	lsl.S	%dx,%dy &Q,%dy	Logical Shift (Left)
	lsl.w lsl.w	&1,EA EA	
LSR	lsr.S	%dx,%dy &Q,&dy	Logical Shift (Right)
	lsr.w lsr.w	&1,EA EA	
MOVE	mov.S	EA,EA	Move Data from Source to Destination NOTE: If the destination is an address register, the instruction generated is **MOVEA.**
MOVE to CCR	mov.w	EA,%cc	Move to Condition Codes
MOVE from CCR	mov.w	%cc,EA (MC68010/MC68020 Only)	Move from Condition Codes
MOVE to SR	mov.w	EA,%sr	Move to the Status Register
MOVE from SR	mov.w	%sr,EA	Move from the Status Register
MOVE USP	mov.l	%usp,%an %an,%usp	Move User Stack Pointer
MOVEA	mov.A	EA,%an	Move Address
MOVEC to CCR	mov.l	%rn,%rc	Move to Control Register (MC68010/MC68020 Only)
MOVEC from CCR	mov.l	%rc,%rn (MC68010/MC68020 Only)	Move from Control Register

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
MOVEM	movm.A	&I,EA EA,&I	Move Multiple Registers* (See footnote)
MOVEP	movp.A	%dx,d(%ay) d(%ay),%dx	Move Peripheral Data
MOVEQ	mov.l	&I,%dn	Move Quick
MOVES	movs.S movs.S	%rn,EA EA,%rn	Move to/form Address Space (MC68010/MC68020 Only)
MULS	muls.w	EA,%dx	Signed Multiply 16*16 → 32
	tmuls.l muls.l	EA,%dx EA,%dx	Signed Multiply (Long) 32*32 → 32 (MC68020 Only)
	muls.l	EA,%dx:%dy	Signed Multiply (Long) 32*32 → 64 (MC68020 Only)
MULU	mulu.w	EA,%dx	Unsigned Multiply 16*16 → 32
	tmulu.l mulu.l	EA,%dx EA,%dx	Unsigned Multiply (Long) 32*32 → 32 (MC68020 Only)
	mulu.l	EA,%dx:%dy	Unsigned Multiply (Long) 32*32 → 64 (MC68020 Only)
NBCD	nbcd.b	EA	Negate Decimal with Extend
NEG	neg.S	EA	Negate
NEGX	negx.S	EA	Negate with Extend
NOP	nop		No Operation
NOT	not.S	EA	Logical Complement

* The immediate operand is a mask designating which registers are to be moved to memory or which registers are to receive memory data. not all addressing modes are permitted, and the correspondence between mask bits and register numbers depends on the addressing mode used. Refer to the MC68000 Programmer's Reference Manual for details.

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
OR	or.S	EA,%dn %dn,EA	Inclusive OR Logical
ORI	or.S	&I,EA	Inclusive OR Immediate
ORI to CCR	or.b	&I,%cc	Inclusive OR Immediate to Condition Codes
ORI to SR	or.w	&I,%sr	Inclusive OR Immediate to the Status Register
PACK	pack pack	—(%ax),—(%ay),&I %dx,%dy,&I	Pack BCD (MC68020 Only)
PEA	pea.l	EA	Push Effective Address
RESET	reset		Reset External Devices
ROL	rol.S	%dx,%dy &Q,%dy	Rotate (without Extend) (Left)
	rol.w rol.w	&1,EA EA	
ROR	ror.S	%dx,%dy &Q,%dy	Rotate (without Extend) (Right)
	ror.w ror.w	&1,EA EA	
ROXL	roxl.S	%dx,%dy &Q,%dy	Rotate with Extend (Left)
	roxl.w roxl.w	&1,EA EA	
ROXR	roxr.S	%dx,%dy &Q,%dy	Rotate with Extend (Right)
	roxr.w roxr.w	&1,EA EA	
RTD	rtd	&I	Return and Deallocate Parameters (MC68010/MC68020 Only)
RTE	rte		Return from Exception
RTM	rtm	%rn	Return from Module (MC68020 Only)

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
RTR RTS	rtr rts		Return and Restore Condition Codes Return from Subroutine
SBCD	sbcd.b	%dy,%dx —(%ay),—(%ax)	Subtract Decimal with Extend
Scc	sCC.b	EA	Set According to Condition
STOP	stop	&I	Load Status Register and Stop
SUB SUBA SUBI SUBQ SUBX	sub.S sub.A sub.S sub.S subx.S	EA,%dn %dn,EA EA,%an &I,EA &Q,EA %dy,%dx —(%ay),—(%ax)	Subtract Binary Subtract Address Subtract Immediate Subtract Quick Subtract with Extend
SWAP	swap.w	%dn	Swap Register Halves
TAS	tas.b	EA	Test and Set an Operand
TRAP TRAPV TRAPcc	trap trapv tCC tpCC.A	&I &I	Trap Trap on Overflow Trap on Condition (MC68020 Only)
TST	tst.S	EA	Test an Operand
UNLK	unlk	%an	Unlink
UNPK	unpk	—(%ax),—(%ay),&I %dx,%dy,&I	Unpack BCD (MC68020 Only)

11.2. Instructions For the MC68881

The following table shows how the floating point co-processor (MC68881) instructions should be written to be understood by the *as* assembler.

In the table, *fpcc* represents any of the following floating point condition code designations:

TRAP ON UNORDERED	
fpcc	MEANING
ge	greater than or equal
gl	greater or less than
gle	greater or less than or equal
gt	greater than
le	less than or equal
lt	less than
ngt	not greater than
nge	not greater than or equal
nlt	not less than
ngl	not greater or less than
nle	not greater or less than or equal
ngle	not greater or less than or equal
sneq	not equal
sf	never
seq	equal
st	always

NO TRAP ON UNORDERED	
fpcc	MEANING
eq	equal
oge	greater than or equal
ogl	greater or less than
ogt	greater than
ole	less than or equal
olt	less than
or	ordered
t	always
ule	unordered or less or equal
ult	unordered less than
uge	unordered greater than or equal
ueq	unordered equal
ugt	unordered greater than
un	unordered
neq	unordered ore greater or less
f	never

The designation *ccc* represents a group of constants in MC68881 constant ROM which have the following values:

ccc	VALUE
00	pi
0B	log10(2)
0C	e
0D	log2(e)
0D	log10(e)
0F	0.0
10	logn(2)
11	logn(10)
12	10**0
13	10**1
14	10**2
15	10**4
16	10**8
17	10**16
18	10**32
19	10**64
1A	10**128
1B	10**256
1C	10**512
1D	10**1024
1E	10**2048
1F	10**4096

Additional abbreviations used in the table are:

EA	represents and effective address
L	a label reference or any expression representing a memory address in the current segment
I	represents an absolute expression, used as an immediate operand
%dn	represents data register
%fpm,%fpn,%fpq	represents floating point data registers
%control	represents floating point control register
%status	represents floating point status register
%iaddr	represents floating point instruction address register
SF	represents source format letters: b byte integer w word integer l long word integer s single precision d double precision x extended precision p packed binary code decimal
A	represents source format letters w or l
B	represents source format letters b, w, l, s, or p

NOTE: The source format must be specified if more than one source format is permitted or a default source format x is assumed. Source format need not be specified if only one format is permitted by the operation.

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
FABS	fabs.SF fabs.x fabs.x	EA,%fpn %fpm,%fpn %fpn	absolute value function
FACOS	facos.SF facos.x facos.x	EA,%fpn %fpm,%fpn %fpn	arccosine function
FADD	fadd.SF fadd.x	EA,%fpn %fpm,%fpn	floating point add
FASIN	fasin.SF fasin.x fasin.x	EA,%fpn %fpm,%fpn %fpn	arcsine function
FATAN	fatan.SF fatan.x fatan.x	EA,%fpn %fpm,%fpn %fpn	arctangent function
FATANH	fatanh.SF fatanh.x fatanh.x	EA,%fpn %fpm,%fpn %fpn	hyperbolic arctangent function
FBfpcc	fbfpcc.A	L	co-processor branch conditionally
FCMP	fcmp.SF fcmp.x	%fpn,EA %fpn,%fpm	floating point compare
FCOS	fcos.SF fcos.x fcos.x	EA,%fpn %fpm,%fpn %fpn	cosine function
FCOSH	fcosh.SF fcosh.x fcosh.x	EA,%fpn %fpm,%fpn %fpn	hyperbolic cosine function
FDBfpcc	fdbfpcc.w	%dn,L	decrement and branch on condition
FDIV	fdiv.SF fdiv.x	EA,%fpn %fpm,%fpn	floating point divided

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
FETOX	fetox.SF fetox.x fatan.x	EA,%fpn %fpm,%fpn %fpn	e**x function
FETOXM1	fetoxm1.SF fetoxm1.x fetoxm1.x	EA,%fpn %fpm,%fpn %fpn	e**x(x-1) function
FGETEXP	fgetexp.SF fgetexp.x fgetexp.x	EA,%fpn %fpm,%fpn %fpn	get the exponent function
FGETMAN	fgetman.SF fgetman.x fgetman.x	EA,%fpn %fpm,%fpn %fpn	get the mantissa function
FINT	fint.SF fint.x fint.x	EA,%fpn %fpm,%fpn %fpn	integer part function
FLOG2	flog2.SF flog2.x flog2.x	EA,%fpn %fpm,%fpn %fpn	binary log function
FLOG10	flog10.SF flog10.x flog10.x	EA,%fpn %fpm,%fpn %fpn	common log function
FLOGN	flogn.SF flogn.x flogn.x	EA,%fpn %fpm,%fpn %fpn	natural log function
FLOGNP1	flognp1.SF flognp1.x flognp1.x	EA,%fpn %fpm,%fpn %fpn	natural log (x+1) function
FMOD	fmod.SF fmod.x	EA,%fpn %fpm,%fpn	floating point module

MC68000 INSTRUCTION FORMATS		
MNEMONIC	ASSEMBLER SYNTAX	OPERATION
FMOVE	fmov.SF EA,%fpn fmov.x %fpm,%fpn	move to floating point register
	fmov.SF %fpn,EA fmov.p %fpn,EA{&I} fmov.p %fpn,EA{%dn}	move from floating point register to memory
	fmov.l EA,%control fmov.l EA,%status fmov.l EA,%iaddr	move from memory to special register
	fmov.l %control,EA fmov.l %statsu,EA fmov.l %iaddr,EA	move to memory from special register
FMOVECR	fmovcr.x &ccc,%fpn	move a ROM-stored to a floating point register
FMOVEM	fmovm.x EA,&I	move to multiple floating point register
	fmovm.x &I,EA	move from multiple registers to memory
	fmovm.x EA,%dn	move to a data register
	fmovm.x %dn,EA	move a data register to memory
	fmovm.l EA,%control/%sta-tus/%iaddr	move to special registers
	fmovm.l %control/%status/%iaddr,EA	move from special registers

NOTE: The immediate operand is a mask designating which registers are to be moved to memory or which registers are to receive memory data. Not all addressing modes are permitted and the correspondence between mask bvits and register numbers depends on the addressing mode used.

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
FMUL	fmul.SF fmul.x	EA,%fpn %fpm,%fpn	floating point multiply
FNEG	fneg.SF fneg.x fneg.x	EA,%fpn %fpm,%fpn %fpn	negate function
FNOP	fnop		floating point no-op
FREM	frem.SF frem.x	EA,%fpn %fpm,%fpn	floating point remainder
FRESTORE	frestore	EA	restore internal state of co-processor
FSAVE	fsave	EA	co-processor save
FSCALE	fscale.SF fscale.x	EA,%fpn %fpm,%fpn	floating point scale exponent
FSfpcc	fsfpcc.b	EA	set on condition
FSGLDIV	fsgldiv.B fsgldiv.x	EA,%fpn %fpm,%fpn	floating point single precision divide
FSGLMUL	fsglmul.B fsglmul.s	EA,%fpn %fpm,%fpn	floating point single precision multiply
FSIN	fsin.SF fsin.x fsin.x	EA,%fpn %fpm,%fpn %fpn	sine function
FSINCOS	fsincos.SF fsincos.x	EA,%fpn %fpm,%fpn:%fpq	sine/cosine function
FSINH	fsinh.SF fsinh.x fsinh.x	EA,%fpn %fpm,%fpn %fpn	hyperbolic sine function
FSQRT	fsqrt.SF fsqrt.x fsqrt.x	EA,%fpn %fpm,%fpn %fpn	square root function

MC68000 INSTRUCTION FORMATS			
MNEMONIC	ASSEMBLER SYNTAX		OPERATION
FSUB	fsub.SF fsub.x	EA,%fpn %fpm,%fpn	square root function
FTAN	ftan.SF ftan.x ftan.x	EA,%fpn %fpm,%fpn %fpn	tangent function
FTANH	ftanh.SF ftanh.x ftanh.x	EA,%fpn %fpm,%fpn %fpn	hyperbolic tangent function
FTENTOX	ftentox.SF ftentox.x ftentox.x	EA,%fpn %fpm,%fpn %fpn	10**x function
FTfpcc	ftfpcc		trap on condition without a parameter
FTPfpcc	ftpfpcc.A	&I	trap on condition with a parameter
FTST	ftest.SF ftest.x	EA %fpm	floating point test an operand
FTWOTOX	ftwotox.SF ftwotox.x ftwotox.x	EA,%fpn %fpm,%fpn %fpn	2**x function
FYTOX	fytox.SF fytox.x	EA,%fpn %fpm,%fpn	floating point y**x

Screen Updating and Cursor Movement Optimization: A Library Package

Kenneth C. R. C. Arnold

Computer Science Division
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, California 94720

ABSTRACT

This document describes a package of C library functions which allow the user to:

- update a screen with reasonable optimization,
- get input from the terminal in a screen-oriented fashion, and
- independent from the above, move the cursor optimally from one point to another.

These routines all use the **/etc/termcap** database to describe the capabilities of the terminal.

Acknowledgements

This package would not exist without the work of Bill Joy, who, in writing his editor, created the capability to generally describe terminals, wrote the routines which read this database, and, most importantly, those which implement optimal cursor movement, which routines I have simply lifted nearly intact. Doug Merritt and Kurt Shoens also were extremely important, as were both willing to waste time listening to me rant and rave. The help and/or support of Ken Abrams, Alan Char, Mark Horton, and Joe Kalash, was, and is, also greatly appreciated.

Contents

Appendixes

1. Overview

In making available the generalized terminal descriptions in **/etc/termcap**, much information was made available to the programmer, but little work was taken out of one's hands. The purpose of this package is to allow the C programmer to do the most common type of terminal dependent functions, those of movement optimization and optimal screen updating, without doing any of the dirty work, and (hopefully) with nearly as much ease as is necessary to simply print or read things.

The package is split into three parts: (1) Screen updating; (2) Screen updating with user input; and (3) Cursor motion optimization.

It is possible to use the motion optimization without using either of the other two, and screen updating and input can be done without any programmer knowledge of the motion optimization, or indeed the database itself.

1.1. Terminology (or, Words You Can Say to Sound Brilliant)

In this document, the following terminology is kept to with reasonable consistency:

window: An internal representation containing an image of what a section of the terminal screen may look like at some point in time. This subsection can either encompass the entire terminal screen, or any smaller portion down to a single character within that screen.

terminal: Sometimes called ***terminal screen***. The package's idea of what the terminal's screen currently looks like, i.e., what the user sees now. This is a special *screen*:

screen: This is a subset of windows which are as large as the terminal screen, i.e., they start at the upper left hand corner and encompass the lower right hand corner. One of these, *stdscr*, is automatically provided for the programmer.

1.2. Compiling Things

In order to use the library, it is necessary to have certain types and variables defined. Therefore, the programmer must have a line:

#include <curses.h>

at the top of the program source. The header file **<curses.h>** needs to include **<sgtty.h>**, so the one should not do so oneself[1]. Also, compilations should have the following form:

cc [*flags*] file ... **–lcurses –ltermlib**

1.3. Screen Updating

In order to update the screen optimally, it is necessary for the routines to know what the screen currently looks like and what the programmer wants it to look like next. For this purpose, a data type (structure) named *WINDOW* is defined which describes a window image to the routines, including its starting position on the screen (the (y, x) co-ordinates of the upper left hand corner) and its size. One of these (called *curscr* for *current screen*) is a screen image of what the terminal currently looks like. Another screen (called *stdscr*, for *standard screen*) is provided by default to make changes on.

A window is a purely internal representation. It is used to build and store a potential image of a portion of the terminal. It doesn't bear any necessary relation to what is really on the terminal screen. It is more like an array of characters on which to make changes.

When one has a window which describes what some part the terminal should look like, the routine *refresh()* (or *wrefresh()* if the window is not *stdscr*) is called. *refresh()* makes the terminal,

[1] The screen package also uses the Standard I/O library, so **<curses.h>** includes **<stdio.h>**. It is redundant (but harmless) for the programmer to do it, too.

in the area covered by the window, look like that window. Note, therefore, that changing something on a window *does not change the terminal.* Actual updates to the terminal screen are made only by calling *refresh()* or *wrefresh()*. This allows the programmer to maintain several different ideas of what a portion of the terminal screen should look like. Also, changes can be made to windows in any order, without regard to motion efficiency. Then, at will, the programmer can effectively say "make it look like this," and let the package worry about the best way to do this.

1.4. Naming Conventions

As hinted above, the routines can use several windows, but two are automatically given: *curscr*, which knows what the terminal looks like, and *stdscr*, which is what the programmer wants the terminal to look like next. The user should never really access *curscr* directly. Changes should be made to the appropriate screen, and then the routine *refresh()* (or *wrefresh()*) should be called.

Many functions are set up to deal with *stdscr* as a default screen. For example, to add a character to *stdscr*, one calls *addch()* with the desired character. If a different window is to be used, the routine *waddch()* (for window-specific *addch()*) is provided[2]. This convention of prepending function names with a "**w**" when they are to be applied to specific windows is consistent. The only routines which do *not* do this are those to which a window must always be specified.

In order to move the current (y, x) co-ordinates from one point to another, the routines *move()* and *wmove()* are provided. However, it is often desirable to first move and then perform some I/O operation. In order to avoid clumsyness, most I/O routines can be preceded by the prefix "**mv**" and the desired (y, x) co-ordinates then can be added to the arguments to the function. For example, the calls

```
move(y, x);
addch(ch);
```

can be replaced by

```
mvaddch(y, x, ch);
```

and

```
wmove(win, y, x);
waddch(win, ch);
```

can be replaced by

```
mvwaddch(win, y, x, ch);
```

Note that the window description pointer (*win*) comes before the added (y, x) co-ordinates. If such pointers are need, they are always the first parameters passed.

2. Variables

Many variables which are used to describe the terminal environment are available to the programmer. They are:

type	name	description
WINDOW *	curscr	current version of the screen (terminal screen).
WINDOW *	stdscr	standard screen. Most updates are usually done here.
char *	Def_term	default terminal type if type cannot be determined
bool	My_term	use the terminal specification in *Def_term* as terminal, irrelevant of real terminal type
char *	ttytype	full name of the current terminal.
int	LINES	number of lines on the terminal

[2] Actually, *addch()* is really a "#define" macro with arguments, as are most of the "functions" which deal with *stdscr* as a default.

int	COLS	number of columns on the terminal
int	ERR	error flag returned by routines on a fail.
int	OK	error flag returned by routines when things go right.

There are also several "#define" constants and types which are of general usefulness:

reg	storage class "register" (e.g., *reg int i;*)
bool	boolean type, actually a "char" (e.g., *bool doneit;*)
TRUE	boolean "true" flag (1).
FALSE	boolean "false" flag (0).

3. Usage

This is a description of how to actually use the screen package. In it, we assume all updating, reading, etc. is applied to *stdscr*. All instructions will work on any window, with changing the function name and parameters as mentioned above.

3.1. Starting up

In order to use the screen package, the routines must know about terminal characteristics, and the space for *curscr* and *stdscr* must be allocated. These functions are performed by *initscr()*. Since it must allocate space for the windows, it can overflow core when attempting to do so. On this rather rare occasion, *initscr()* returns ERR. *initscr()* must ***always*** be called before any of the routines which affect windows are used. If it is not, the program will core dump as soon as either *curscr* or *stdscr* are referenced. However, it is usually best to wait to call it until after you are sure you will need it, like after checking for startup errors. Terminal status changing routines like *nl()* and *crmode()* should be called after *initscr()*.

Now that the screen windows have been allocated, you can set them up for the run. If you want to, say, allow the window to scroll, use *scrollok()*. If you want the cursor to be left after the last change, use *leaveok()*. If this isn't done, *refresh()* will move the cursor to the window's current (y, x) co-ordinates after updating it. New windows of your own can be created, too, by using the functions *newwin()* and *subwin()*. *delwin()* will allow you to get rid of old windows. If you wish to change the official size of the terminal by hand, just set the variables *LINES* and *COLS* to be what you want, and then call *initscr()*. This is best done before, but can be done either before or after, the first call to *initscr()*, as it will always delete any existing *stdscr* and/or *curscr* before creating new ones.

3.2. The Nitty-Gritty

3.2.1. Output

Now that we have set things up, we will want to actually update the terminal. The basic functions used to change what will go on a window are *addch()* and *move()*. *addch()* adds a character at the current (y, x) co-ordinates, returning ERR if it would cause the window to illegally scroll, i.e., printing a character in the lower right-hand corner of a terminal which automatically scrolls if scrolling is not allowed. *move()* changes the current (y, x) co-ordinates to whatever you want them to be. It returns ERR if you try to move off the window when scrolling is not allowed. As mentioned above, you can combine the two into *mvaddch()* to do both things in one fell swoop.

The other output functions, such as *addstr()* and *printw()*, all call *addch()* to add characters to the window.

After you have put on the window what you want there, when you want the portion of the terminal covered by the window to be made to look like it, you must call *refresh()*. In order to optimize finding changes, *refresh()* assumes that any part of the window not changed since the last *refresh()* of that window has not been changed on the terminal, i.e., that you have not refreshed a portion of the terminal with an overlapping window. If this is not the case, the routine *touchwin()* is provided to make it look like the entire window has been changed, thus making *refresh()* check the whole subsection of the terminal for changes.

If you call *wrefresh()* with *curscr*, it will make the screen look like *curscr* thinks it looks like. This is useful for implementing a command which would redraw the screen in case it get messed up.

3.2.2. Input

Input is essentially a mirror image of output. The complementary function to *addch()* is *getch()* which, if echo is set, will call *addch()* to echo the character. Since the screen package needs to know what is on the terminal at all times, if characters are to be echoed, the tty must be in raw or cbreak mode. If it is not, *getch()* sets it to be cbreak, and then reads in the character.

3.2.3. Miscellaneous

All sorts of fun functions exists for maintaining and changing information about the windows. For the most part, the descriptions in section 5.4. should suffice.

3.3. Finishing up

In order to do certain optimizations, and, on some terminals, to work at all, some things must be done before the screen routines start up. These functions are performed in *getttmode()* and *setterm()*, which are called by *initscr()*. In order to clean up after the routines, the routine *endwin()* is provided. It restores tty modes to what they were when *initscr()* was first called. Thus, anytime after the call to initscr, *endwin()* should be called before exiting.

4. Cursor Motion Optimization: Standing Alone

It is possible to use the cursor optimization functions of this screen package without the overhead and additional size of the screen updating functions. The screen updating functions are designed for uses where parts of the screen are changed, but the overall image remains the same. This includes such programs as **eye** and **vi**[3]. Certain other programs will find it difficult to use these functions in this manner without considerable unnecessary program overhead. For such applications, such as some "*crt hacks*"[4] and optimizing **cat**(1)-type programs, all that is needed is the motion optimizations. This, therefore, is a description of what some of what goes on at the lower levels of this screen package. The descriptions assume a certain amount of familiarity with programming problems and some finer points of C. None of it is terribly difficult, but you should be forewarned.

4.1. Terminal Information

In order to use a terminal's features to the best of a program's abilities, it must first know what they are[5]. The **/etc/termcap** database describes these, but a certain amount of decoding is necessary, and there are, of course, both efficient and inefficient ways of reading them in. The algorithm that the uses is taken from **vi** and is hideously efficient. It reads them in a tight loop into a set of variables whose names are two uppercase letters with some mnemonic value. For example, *HO* is a string which moves the cursor to the "home" position[6]. As there are two types of variables involving ttys, there are two routines. The first, *gettmode()*, sets some variables based upon the tty modes accessed by **gtty**(2) and **stty**(2) The second, *setterm()*, a larger task by reading in the descriptions from the **/etc/termcap** database. This is the way these routines are used by *initscr()*:

---

[3] Eye actually uses these functions, vi does not.

[4] Graphics programs designed to run on character-oriented terminals. I could name many, but they come and go, so the list would be quickly out of date. Recently, there have been programs such as **rocket** and **gun**.

[5] If this comes as any surprise to you, there's this tower in Paris they're thinking of junking that I can let you have for a song.

[6] These names are identical to those variables used in the **/etc/termcap** database to describe each capability. See Appendix A for a complete list of those read, and **termcap**(5) for a full description.

```
if (isatty(0)) {
      gettmode();
      if (sp=getenv("TERM"))
            setterm(sp);
}
else
      setterm(Def_term);
_puts(TI);
_puts(VS);
```

isatty() checks to see if file descriptor 0 is a terminal[7]. If it is, *gettmode()* sets the terminal description modes from a **gtty**(2) *getenv()* is then called to get the name of the terminal, and that value (if there is one) is passed to *setterm()*, which reads in the variables from **/etc/termcap** associated with that terminal. (*getenv()* returns a pointer to a string containing the name of the terminal, which we save in the character pointer *sp*.) If *isatty()* returns false, the default terminal *Def_term* is used. The *TI* and *VS* sequences initialize the terminal (*_puts()* is a macro which uses *tputs()* (see **termcap**(3)) to put out a string). It is these things which *endwin()* undoes.

4.2. Movement Optimizations, or, Getting Over Yonder

Now that we have all this useful information, it would be nice to do something with it[8]. The most difficult thing to do properly is motion optimization. When you consider how many different features various terminals have (tabs, backtabs, non-destructive space, home sequences, absolute tabs, ...) you can see that deciding how to get from here to there can be a decidedly non-trivial task. The editor **vi** uses many of these features, and the routines it uses to do this take up many pages of code. Fortunately, I was able to liberate them with the author's permission, and use them here.

After using *gettmode()* and *setterm()* to get the terminal descriptions, the function *mvcur()* deals with this task. It usage is simple: you simply tell it where you are now and where you want to go. For example

```
mvcur(0, 0, LINES/2, COLS/2)
```

would move the cursor from the home position (0, 0) to the middle of the screen. If you wish to force absolute addressing, you can use the function *tgoto()* from the **termlib**(7) routines, or you can tell *mvcur()* that you are impossibly far away, like Cleveland. For example, to absolutely address the lower left hand corner of the screen from anywhere just claim that you are in the upper right hand corner:

```
mvcur(0, COLS-1, LINES-1, 0)
```

5. The Functions

In the following definitions, "†" means that the "function" is really a "#define" macro with arguments. This means that it will not show up in stack traces in the debugger, or, in the case of such functions as *addch()*, it will show up as it's "**w**" counterpart. The arguments are given to show the order and type of each. Their names are not mandatory, just suggestive.

5.1. Output Functions

[7] *isatty()* is defined in the default C library function routines. It does a **gtty**(2) on the descriptor and checks the return value.

[8] Actually, it *can* be emotionally fulfilling just to get the information. This is usually only true, however, if you have the social life of a kumquat.

addch(ch) †
char *ch;*

waddch(win, ch)
WINDOW **win;*
char *ch;*

Add the character *ch* on the window at the current (y, x) co-ordinates. If the character is a newline ('\n') the line will be cleared to the end, and the current (y, x) co-ordinates will be changed to the beginning off the next line if newline mapping is on, or to the next line at the same x co-ordinate if it is off. A return ('\r') will move to the beginning of the line on the window. Tabs ('\t') will be expanded into spaces in the normal tabstop positions of every eight characters. This returns ERR if it would cause the screen to scroll illegally.

addstr(str) †
char **str;*

waddstr(win, str)
WINDOW **win;*
char **str;*

Add the string pointed to by *str* on the window at the current (y, x) co-ordinates. This returns ERR if it would cause the screen to scroll illegally. In this case, it will put on as much as it can.

box(win, vert, hor)
WINDOW **win;*
char *vert, hor;*

Draws a box around the window using *vert* as the character for drawing the vertical sides, and *hor* for drawing the horizontal lines. If scrolling is not allowed, and the window encompasses the lower right-hand corner of the terminal, the corners are left blank to avoid a scroll.

clear() †

wclear(win)
WINDOW **win;*

Resets the entire window to blanks. If *win* is a screen, this sets the clear flag, which will cause a clear-screen sequence to be sent on the next *refresh()* call. This also moves the current (y, x) co-ordinates to (0, 0).

clearok(scr, boolf) †
WINDOW **scr;*
bool *boolf;*

Sets the clear flag for the screen *scr*. If *boolf* is TRUE, this will force a clear-screen to be printed on the next *refresh()*, or stop it from doing so if *boolf* is FALSE. This only works on screens, and, unlike *clear()*, does not alter the contents of the screen. If *scr* is *curscr*, the next *refresh()* call will cause a clear-screen, even if the window passed to *refresh()* is not a screen.

clrtobot() †

wclrtobot(win)
*WINDOW *win;*

Wipes the window clear from the current (y, x) co-ordinates to the bottom. This does not force a clear-screen sequence on the next refresh under any circumstances. This has no associated "**mv**" command.

clrtoeol() †

wclrtoeol(win)
*WINDOW *win;*

Wipes the window clear from the current (y, x) co-ordinates to the end of the line. This has no associated "**mv**" command.

delch()

wdelch(win)
*WINDOW *win;*

Delete the character at the current (y, x) co-ordinates. Each character after it on the line shifts to the left, and the last character becomes blank.

deleteln()

wdeleteln(win)
*WINDOW *win;*

Delete the current line. Every line below the current one will move up, and the bottom line will become blank. The current (y, x) co-ordinates will remain unchanged.

erase() †

werase(win)
*WINDOW *win;*

Erases the window to blanks without setting the clear flag. This is analagous to *clear()*, except that it never causes a clear-screen sequence to be generated on a *refresh()*. This has no associated "**mv**" command.

insch(c)
char c;

winsch(win, c)
*WINDOW *win;*
char c;

Insert *c* at the current (y, x) co-ordinates Each character after it shifts to the right, and the last character disappears. This returns ERR if it would cause the screen to scroll illegally.

insertln()

winsertln(win)
*WINDOW *win;*

Insert a line above the current one. Every line below the current line will be shifted down, and the bottom line will disappear. The current line will become blank, and the current (y, x) co-ordinates will remain unchanged. This returns ERR if it would cause the screen to scroll illegally.

move(y, x) †
int y, x;

wmove(win, y, x)
*WINDOW *win;*
int y, x;

Change the current (y, x) co-ordinates of the window to (*y, x*). This returns ERR if it would cause the screen to scroll illegally.

overlay(win1, win2)
*WINDOW *win1, *win2;*

Overlay *win1* on *win2*. The contents of *win1*, insofar as they fit, are placed on *win2* at their starting (y, x) co-ordinates. This is done non-destructively, i.e., blanks on *win1* leave the contents of the space on *win2* untouched.

overwrite(win1, win2)
*WINDOW *win1, *win2;*

Overwrite *win1* on *win2*. The contents of *win1*, insofar as they fit, are placed on *win2* at their starting (y, x) co-ordinates. This is done destructively, i.e., blanks on *win1* become blank on *win2*.

printw(fmt, arg1, arg2, ...)
*char *fmt;*

wprintw(win, fmt, arg1, arg2, ...)
*WINDOW *win;*
*char *fmt;*

Performs a *printf()* on the window starting at the current (y, x) co-ordinates. It uses *addstr()* to add the string on the window. It is often advisable to use the field width options of *printf()* to avoid leaving things on the window from earlier calls. This returns ERR if it would cause the screen to scroll illegally.

refresh() †

wrefresh(win)
*WINDOW *win;*

Synchronize the terminal screen with the desired window. If the window is not a screen, only that part covered by it is updated. This returns ERR if it would cause the screen to scroll illegally. In this case, it will update whatever it can without causing the scroll.

standout() †

wstandout(win)
*WINDOW *win;*

standend() †

wstandend(win)
*WINDOW *win;*

Start and stop putting characters onto *win* in standout mode. *standout()* causes any characters added to the window to be put in standout mode on the terminal (if it has that capability). *standend()* stops this. The sequences *SO* and *SE* (or *US* and *UE* if they are not defined) are used (see Appendix A).

5.2. Input Functions

crmode() †

nocrmode() †

Set or unset the terminal to/from cbreak mode.

echo() †

noecho() †

Sets the terminal to echo or not echo characters.

getch() †

wgetch(win)
*WINDOW *win;*

Gets a character from the terminal and (if necessary) echos it on the window. This returns ERR if it would cause the screen to scroll illegally. Otherwise, the character gotten is returned. If *noecho* has been set, then the window is left unaltered. In order to retain control of the terminal, it is necessary to have one of *noecho*, *cbreak*, or *rawmode* set. If you do not set one, whatever routine you call to read characters will set *cbreak* for you, and then reset to the original mode when finished.

getstr(str) †
char **str;*

wgetstr(win, str)
WINDOW **win;*
char **str;*

Get a string through the window and put it in the location pointed to by *str*, which is assumed to be large enough to handle it. It sets tty modes if necessary, and then calls *getch()* (or *wgetch(win)*) to get the characters needed to fill in the string until a newline or EOF is encountered. The newline stripped off the string. This returns ERR if it would cause the screen to scroll illegally.

raw() †

noraw() †

Set or unset the terminal to/from raw mode. On version 7 UNIX[9] this also turns of newline mapping (see *nl()*).

scanw(fmt, arg1, arg2, ...)
char **fmt;*

wscanw(win, fmt, arg1, arg2, ...)
WINDOW **win;*
char **fmt;*

Perform a *scanf()* through the window using *fmt*. It does this using consecutive *getch()*'s (or *wgetch(win)*'s). This returns ERR if it would cause the screen to scroll illegally.

5.3. Miscellaneous Functions

delwin(win)
WINDOW **win;*

Deletes the window from existence. All resources are freed for future use by **calloc**(3). If a window has a *subwin()* allocated window inside of it, deleting the outer window the subwindow is not affected, even though this does invalidate it. Therefore, subwindows should be deleted before their outer windows are.

endwin()

Finish up window routines before exit. This restores the terminal to the state it was before *initscr()* (or *gettmode()* and *setterm()*) was called. It should always be called before exiting. It does not exit. This is especially useful for resetting tty stats when trapping rubouts via **signal**(2).

[9] UNIX is a trademark of Bell Laboratories.

getyx(win, y, x) †
*WINDOW *win;*
int y, x;

Puts the current (y, x) co-ordinates of *win* in the variables *y* and *x*. Since it is a macro, not a function, you do not pass the address of *y* and *x*.

inch() †

winch(win) †
*WINDOW *win;*

Returns the character at the current (y, x) co-ordinates on the given window. This does not make any changes to the window. This has no associated "**mv**" command.

initscr()

Initialize the screen routines. This must be called before any of the screen routines are used. It initializes the terminal-type data and such, and without it, none of the routines can operate. If standard input is not a tty, it sets the specifications to the terminal whose name is pointed to by *Def_term* (initialy "dumb"). If the boolean *My_term* is true, *Def_term* is always used.

leaveok(win, boolf) †
*WINDOW *win;*
bool boolf;

Sets the boolean flag for leaving the cursor after the last change. If *boolf* is TRUE, the cursor will be left after the last update on the terminal, and the current (y, x) co-ordinates for *win* will be changed accordingly. If it is FALSE, it will be moved to the current (y, x) co-ordinates. This flag (initialy FALSE) retains its value until changed by the user.

longname(termbuf, name)
*char *termbuf, *name;*

Fills in *name* with the long (full) name of the terminal described by the termcap entry in *termbuf*. It is generally of little use, but is nice for telling the user in a readable format what terminal we think he has. This is available in the global variable *ttytype*. *Termbuf* is usually set via the termlib routine *tgetent()*.

mvwin(win, y, x)
*WINDOW *win;*
int y, x;

Move the home position of the window *win* from its current starting coordinates to (*y*, *x*). If that would put part or all of the window off the edge of the terminal screen, *mvwin()* returns ERR and does not change anything.

*WINDOW **
newwin(lines, cols, begin_y, begin_x)
int lines, cols, begin_y, begin_x;

Create a new window with *lines* lines and *cols* columns starting at position (*begin_y, begin_x*). If either *lines* or *cols* is 0 (zero), that dimension will be set to (*LINES* - *begin_y*) or (*COLS* - *begin_x*) respectively. Thus, to get a new window of dimensions *LINES* × *COLS*, use *newwin(0, 0, 0, 0)*.

nl() †

nonl() †

Set or unset the terminal to/from nl mode, i.e., start/stop the system from mapping **<RETURN>** to **<LINE-FEED>**. If the mapping is not done, *refresh()* can do more optimization, so it is recommended, but not required, to turn it off.

scrollok(win, boolf) †
*WINDOW *win;*
bool boolf;

Set the scroll flag for the given window. If *boolf* is FALSE, scrolling is not allowed. This is its default setting.

touchwin(win)
*WINDOW *win;*

Make it appear that the every location on the window has been changed. This is usually only needed for refreshes with overlapping windows.

*WINDOW **
subwin(win, lines, cols, begin_y, begin_x)
*WINDOW *win;*
int lines, cols, begin_y, begin_x;

Create a new window with *lines* lines and *cols* columns starting at position (*begin_y, begin_x*) in the middle of the window *win*. This means that any change made to either window in the area covered by the subwindow will be made on both windows. *begin_y, begin_x* are specified relative to the overall screen, not the relative (0, 0) of *win*. If either *lines* or *cols* is 0 (zero), that dimension will be set to (*LINES* - *begin_y*) or (*COLS* - *begin_x*) respectively.

unctrl(ch) †
char ch;

This is actually a debug function for the library, but it is of general usefulness. It returns a string which is a representation of *ch*. Control characters become their upper-case equivalents preceded by a "^". Other letters stay just as they are. To use *unctrl()*, you must have **#include <unctrl.h>** in your file.

5.4. Details

gettmode()

Get the tty stats. This is normally called by *initscr()*.

mvcur(lasty, lastx, newy, newx)
int *lasty, lastx, newy, newx;*

Moves the terminal's cursor from (*lasty, lastx*) to (*newy, newx*) in an approximation of optimal fashion. This routine uses the functions borrowed from *ex* version 2.6. It is possible to use this optimization without the benefit of the screen routines. With the screen routines, this should not be called by the user. *move()* and *refresh()* should be used to move the cursor position, so that the routines know what's going on.

scroll(win)
WINDOW **win;*

Scroll the window upward one line. This is normally not used by the user.

savetty() †

resetty() †

savetty() saves the current tty characteristic flags. *resetty()* restores them to what *savetty()* stored. These functions are performed automatically by *initscr()* and *endwin()*.

setterm(name)
char **name;*

Set the terminal characteristics to be those of the terminal named *name*. This is normally called by *initscr()*.

tstp()

If the new **tty**(4) driver is in use, this function will save the current tty state and then put the process to sleep. When the process gets restarted, it restores the tty state and then calls *wrefresh(curscr)* to redraw the screen. *initscr()* sets the signal SIGTSTP to trap to this routine.

1. Capabilities from termcap

1.1. Disclaimer

The description of terminals is a difficult business, and we only attempt to summarize the capabilities here: for a full description see the paper describing termcap.

1.2. Overview

Capabilities from termcap are of three kinds: string valued options, numeric valued options, and boolean options. The string valued options are the most complicated, since they may include padding information, which we describe now.

Intelligent terminals often require padding on intelligent operations at high (and sometimes even low) speed. This is specified by a number before the string in the capability, and has meaning for the capabilities which have a **P** at the front of their comment. This normally is a number of milliseconds to pad the operation. In the current system which has no true programmable delays, we do this by sending a sequence of pad characters (normally nulls, but can be changed (specified by *PC*)). In some cases, the pad is better computed as some number of milliseconds times the number of affected lines (to the bottom of the screen usually, except when terminals have insert modes which will shift several lines.) This is specified as, e.g., **12***. before the capability, to say 12 milliseconds per affected whatever (currently always line). Capabilities where this makes sense say **P***.

1.3. Variables Set By setterm()

variables set by *setterm()*

Type	Name	Pad	Description
char *	AL	P*	Add new blank Line
bool	AM		Automatic Margins
char *	BC		Back Cursor movement
bool	BS		BackSpace works
char *	BT	P	Back Tab
bool	CA		Cursor Addressable
char *	CD	P*	Clear to end of Display
char *	CE	P	Clear to End of line
char *	CL	P*	CLear screen
char *	CM	P	Cursor Motion
char *	DC	P*	Delete Character
char *	DL	P*	Delete Line sequence
char *	DM		Delete Mode (enter)
char *	DO		DOwn line sequence
char *	ED		End Delete mode
bool	EO		can Erase Overstrikes with ' '
char *	EI		End Insert mode
char *	HO		HOme cursor
bool	HZ		HaZeltine ˜ braindamage
char *	IC	P	Insert Character
bool	IN		Insert-Null blessing
char *	IM		enter Insert Mode (IC usually set, too)
char *	IP	P*	Pad after char Inserted using IM+IE
char *	LL		quick to Last Line, column 0
char *	MA		ctrl character MAp for cmd mode
bool	MI		can Move in Insert mode
bool	NC		No Cr: \r sends \r\n then eats \n

variables set by *setterm()*

Type	Name	Pad	Description
char *	ND		Non-Destructive space
bool	OS		OverStrike works
char	PC		Pad Character
char *	SE		Standout End (may leave space)
char *	SF	P	Scroll Forwards
char *	SO		Stand Out begin (may leave space)
char *	SR	P	Scroll in Reverse
char *	TA	P	TAb (not ^I or with padding)
char *	TE		Terminal address enable Ending sequence
char *	TI		Terminal address enable Initialization
char *	UC		Underline a single Character
char *	UE		Underline Ending sequence
bool	UL		UnderLining works even though !OS
char *	UP		UPline
char *	US		Underline Starting sequence[10]
char *	VB		Visible Bell
char *	VE		Visual End sequence
char *	VS		Visual Start sequence
bool	XN		a Newline gets eaten after wrap

Names starting with *X* are reserved for severely nauseous glitches

1.4. Variables Set By gettmode()

variables set by *gettmode()*

type	name	description
bool	NONL	Term can't hack linefeeds doing a CR
bool	GT	Gtty indicates Tabs
bool	UPPERCASE	Terminal generates only uppercase letters

[10] US and UE, if they do not exist in the termcap entry, are copied from SO and SE in *setterm()*

1. The WINDOW structure

The WINDOW structure is defined as follows:

```
# define        WINDOW struct _win_st

struct _win_st {
        short     _cury, _curx;
        short     _maxy, _maxx;
        short     _begy, _begx;
        short     _flags;
        bool      _clear;
        bool      _leave;
        bool      _scroll;
        char      **_y;
        short     *_firstch;
        short     *_lastch;
};

# define        _SUBWIN       01
# define        _ENDLINE      02
# define        _FULLWIN      04
# define        _SCROLLWIN    010
# define        _STANDOUT     0200
```

_cury and *_curx* are the current (y, x) co-ordinates for the window. New characters added to the screen are added at this point. *_maxy* and *_maxx* are the maximum values allowed for (*_cury*, *_curx*). *_begy* and *_begx* are the starting (y, x) co-ordinates on the terminal for the window, i.e., the window's home. *_cury*, *_curx*, *_maxy*, and *_maxx* are measured relative to (*_begy*, *_begx*), not the terminal's home.

_clear tells if a clear-screen sequence is to be generated on the next *refresh()* call. This is only meaningful for screens. The initial clear-screen for the first *refresh()* call is generated by initially setting clear to be TRUE for *curscr*, which always generates a clear-screen if set, irrelevant of the dimensions of the window involved. *_leave* is TRUE if the current (y, x) co-ordinates and the cursor are to be left after the last character changed on the terminal, or not moved if there is no change. *_scroll* is TRUE if scrolling is allowed.

_y is a pointer to an array of lines which describe the terminal. Thus:

_y[i]

is a pointer to the *i*th line, and

_y[i][j]

is the *j*th character on the *i*th line.

_flags can have one or more values or'd into it. **_SUBWIN** means that the window is a subwindow, which indicates to *delwin()* that the space for the lines is not to be freed. **_ENDLINE** says that the end of the line for this window is also the end of a screen. **_FULLWIN** says that this window is a screen. **_SCROLLWIN** indicates that the last character of this screen is at the lower right-hand corner of the terminal; *i.e.*, if a character was put there, the terminal would scroll. **_STANDOUT** says that all characters added to the screen are in standout mode.

---

[11] All variables not normally accessed directly by the user are named with an initial "_" to avoid conflicts with the user's variables.

1. Examples

Here we present a few examples of how to use the package. They attempt to be representative, though not comprehensive.

2. Screen Updating

The following examples are intended to demonstrate the basic structure of a program using the screen updating sections of the package. Several of the programs require calculational sections which are irrelevant of to the example, and are therefore usually not included. It is hoped that the data structure definitions give enough of an idea to allow understanding of what the relevant portions do. The rest is left as an exercise to the reader, and will not be on the final.

2.1. Twinkle

This is a moderately simple program which prints pretty patterns on the screen that might even hold your interest for 30 seconds or more. It switches between patterns of asterisks, putting them on one by one in random order, and then taking them off in the same fashion. It is more efficient to write this using only the motion optimization, as is demonstrated below.

```
# include        <curses.h>
# include        <signal.h>

/*
 * the idea for this program was a product of the imagination of
 * Kurt Schoens.  Not responsible for minds lost or stolen.
 */

# define         NCOLS   80
# define         NLINES  24
# define         MAXPATTERNS     4

struct locs {
        char    y, x;
};

typedef struct locs             LOCS;

LOCS    Layout[NCOLS * NLINES];         /* current board layout */

int     Pattern,                        /* current pattern number */
        Numstars;                       /* number of stars in pattern */

main() {

        char            *getenv();
        int             die();

        srand(getpid());                        /* initialize random sequence */

        initscr();
        signal(SIGINT, die);
        noecho();
        nonl();
        leaveok(stdscr, TRUE);
        scrollok(stdscr, FALSE);

        for (;;) {
```

```
                        makeboard();              /* make the board setup */
                        puton('*');               /* put on '*'s */
                        puton(' ');               /* cover up with ' 's */
                }
}

/*
 * On program exit, move the cursor to the lower left corner by
 * direct addressing, since current location is not guaranteed.
 * We lie and say we used to be at the upper right corner to guarantee
 * absolute addressing.
 */
die() {

                signal(SIGINT, SIG_IGN);
                mvcur(0, COLS-1, LINES-1, 0);
                endwin();
                exit(0);
}


/*
 * Make the current board setup.  It picks a random pattern and
 * calls ison() to determine if the character is on that pattern
 * or not.
 */
makeboard() {

                reg int               y, x;
                reg LOCS              *lp;

                Pattern = rand() % MAXPATTERNS;
                lp = Layout;
                for (y = 0; y < NLINES; y++)
                        for (x = 0; x < NCOLS; x++)
                                if (ison(y, x)) {
                                        lp->y = y;
                                        lp++->x = x;
                                }
                Numstars = lp - Layout;
}

/*
 * Return TRUE if (y, x) is on the current pattern.
 */
ison(y, x)
reg int         y, x; {

                switch (Pattern) {
                  case 0:               /* alternating lines */
                        return !(y & 01);
                  case 1:               /* box */
                        if (x >= LINES && y >= NCOLS)
                                return FALSE;
                        if (y < 3 || y >= NLINES - 3)
```

```
                        return TRUE;
                return (x < 3 || x >= NCOLS - 3);
          case 2:               /* holy pattern! */
                return ((x + y) & 01);
          case 3:               /* bar across center */
                return (y >= 9 && y <= 15);
        }
        /* NOTREACHED */
}

puton(ch)
reg char                ch; {

        reg LOCS                *lp;
        reg int                 r;
        reg LOCS                *end;
        LOCS                    temp;

        end = &Layout[Numstars];
        for (lp = Layout; lp < end; lp++) {
                r = rand() % Numstars;
                temp = *lp;
                *lp = Layout[r];
                Layout[r] = temp;
        }

        for (lp = Layout; lp < end; lp++) {
                mvaddch(lp->y, lp->x, ch);
                refresh();
        }
}
```

2.2. Life

This program plays the famous computer pattern game of life (Scientific American, May, 1974). The calculational routines create a linked list of structures defining where each piece is. Nothing here claims to be optimal, merely demonstrative. This program, however, is a very good place to use the screen updating routines, as it allows them to worry about what the last position looked like, so you don't have to. It also demonstrates some of the input routines.

```
# include             <curses.h>
# include             <signal.h>

/*
 *              Run a life game.  This is a demonstration program for
 * the Screen Updating section of the -lcurses cursor package.
 */

struct lst_st {                                 /* linked list element */
        int             y, x;                   /* (y, x) position of piece */
        struct lst_st   *next, *last;           /* doubly linked */
};

typedef struct lst_st           LIST;
```

```
LIST        *Head;                          /* head of linked list */

main(ac, av)
int         ac;
char        *av[]; {

            int         die();

            evalargs(ac, av);                       /* evaluate arguments */

            initscr();                              /* initialize screen package */
            signal(SIGINT, die);                    /* set to restore tty stats */
            crmode();                               /* set for char—by—char */
            noecho();                               /*
            nonl();                                 /* for optimization */

            getstart();                             /* get starting position */
            for (;;) {
                        prboard();                  /* print out current board */
                        update();                   /* update board position */
            }
}

/*
 * This is the routine which is called when rubout is hit.
 * It resets the tty stats to their original values. This
 * is the normal way of leaving the program.
 */
die() {

            signal(SIGINT, SIG_IGN);                /* ignore rubouts */
            mvcur(0, COLS—1, LINES—1, 0);           /* go to bottom of screen */
            endwin();                               /* set terminal to initial state */
            exit(0);
}

/*
 * Get the starting position from the user. They keys u, i, o, j, l,
 * m, ·,, and . are used for moving their relative directions from the
 * k key. Thus, u move diagonally up to the left, , moves directly down,
 * etc. x places a piece at the current position, " " takes it away.
 * The input can also be from a file. The list is built after the
 * board setup is ready.
 */
getstart() {

            reg char            c;
            reg int             x, y;

            box(stdscr, '|', '_');                  /* box in the screen */
            move(1, 1);                             /* move to upper left corner */

            do {
```

```
                refresh();                              /* print current position */
                if ((c=getch()) == 'q')
                        break;
                switch (c) {
                  case 'u':
                  case 'i':
                  case 'o':
                  case 'j':
                  case 'l':
                  case 'm':
                  case ',':
                  case '.':
                        adjustyx(c);
                        break;
                  case 'f':
                        mvaddstr(0, 0, "File name: ");
                        getstr(buf);
                        readfile(buf);
                        break;
                  case 'x':
                        addch('X');
                        break;
                  case ' ':
                        addch(' ');
                        break;
                }
        }

        if (Head != NULL)                                       /* start new list */
                dellist(Head);
        Head = malloc(sizeof (LIST));

        /*
         * loop through the screen looking for 'x's, and add a list
         * element for each one
         */
        for (y = 1; y < LINES - 1; y++)
                for (x = 1; x < COLS - 1; x++) {
                        move(y, x);
                        if (inch() == 'x')
                                addlist(y, x);
                }
}

/*
 * Print out the current board position from the linked list
 */
prboard() {

        reg LIST                *hp;

        erase();                                        /* clear out last position */
```

```
		box(stdscr, '|', '_');				/* box in the screen */

		/*
		 * go through the list adding each piece to the newly
		 * blank board
		 */
		for (hp = Head; hp; hp = hp->next)
			mvaddch(hp->y, hp->x, 'X');

		refresh();
	}
```

3. Motion optimization

The following example shows how motion optimization is written on its own. Programs which flit from one place to another without regard for what is already there usually do not need the overhead of both space and time associated with screen updating. They should instead use motion optimization.

3.1. Twinkle

The **twinkle** program is a good candidate for simple motion optimization. Here is how it could be written (only the routines that have been changed are shown):

```
	main() {

		reg char		*sp;
		char			*getenv();
		int			_putchar(), die();

		srand(getpid());				/* initialize random sequence */

		if (isatty(0)) {
		      gettmode();
		      if (sp=getenv("TERM"))
				setterm(sp);
			signal(SIGINT, die);
		}
		else {
			printf("Need a terminal on %d\n", _tty_ch);
			exit(1);
		}
		_puts(TI);
		_puts(VS);

		noecho();
		nonl();
		tputs(CL, NLINES, _putchar);
		for (;;) {
			makeboard();			/* make the board setup */
			puton('*');			/* put on '*'s */
			puton(' ');			/* cover up with ' 's */
		}
	}
```

```
/*
 * _putchar defined for tputs() (and _puts())
 */
_putchar(c)
reg char            c; {

        putchar(c);
}

puton(ch)
char    ch; {

        static int          lasty, lastx;
        reg LOCS            *lp;
        reg int             r;
        reg LOCS            *end;
        LOCS                temp;

        end = &Layout[Numstars];
        for (lp = Layout; lp < end; lp++) {
                r = rand() % Numstars;
                temp = *lp;
                *lp = Layout[r];
                Layout[r] = temp;
        }

        for (lp = Layout; lp < end; lp++)
                        /* prevent scrolling */
                if (!AM || (lp->y < NLINES - 1 || lp->x < NCOLS - 1)) {
                        mvcur(lasty, lastx, lp->y, lp->x);
                        putchar(ch);
                        lasty = lp->y;
                        if ((lastx = lp->x + 1) >= NCOLS)
                                if (AM) {
                                        lastx = 0;
                                        lasty++;
                                }
                                else
                                        lastx = NCOLS - 1;
                }
}
```

A Tutorial Introduction to ADB

J. F. Maranzano

S. R. Bourne

ABSTRACT

Debugging tools generally provide a wealth of information about the inner workings of programs. These tools have been available on UNIX† to allow users to examine "core" files that result from aborted programs. A new debugging program, ADB, provides enhanced capabilities to examine "core" and other program files in a variety of formats, run programs with embedded breakpoints and patch files.

ADB is an indispensable but complex tool for debugging crashed systems and/or programs. This document provides an introduction to ADB with examples of its use. It explains the various formatting options, techniques for debugging C programs, examples of printing file system information and patching.

1. Introduction

ADB is a new debugging program that is available on UNIX. It provides capabilities to look at "core" files resulting from aborted programs, print output in a variety of formats, patch files, and run programs with embedded breakpoints. This document provides examples of the more useful features of ADB. The reader is expected to be familiar with the basic commands on UNIX with the C language, and with References 1, 2 and 3.

2. A Quick Survey

2.1. Invocation

ADB is invoked as:

adb objfile corefile

where *objfile* is an executable UNIX file and *corefile* is a core image file. Many times this will look like:

adb a.out core

or more simply:

adb

where the defaults are *a.out* and *core* respectively. The filename minus (–) means ignore this argument as in:

adb – core

ADB has requests for examining locations in either file. The **?** request examines the contents of *objfile,* the / request examines the *corefile.* The general form of these requests is:

† UNIX is a trademark of Bell Laboratories.

address ? format

or

address / format

2.2. Current Address

ADB maintains a current address, called dot, similar in function to the current pointer in the UNIX editor. When an address is entered, the current address is set to that location, so that:

0126?i

sets dot to octal 126 and prints the instruction at that address. The request:

.,10/d

prints 10 decimal numbers starting at dot. Dot ends up referring to the address of the last item printed. When used with the ? or / requests, the current address can be advanced by typing newline; it can be decremented by typing ^.

Addresses are represented by expressions. Expressions are made up from decimal, octal, and hexadecimal integers, and symbols from the program under test. These may be combined with the operators +, −, *, % (integer division), & (bitwise and), | (bitwise inclusive or), # (round up to the next multiple), and ~ (not). (All arithmetic within ADB is 32 bits.) When typing a symbolic address for a C program, the user can type *name* or *_name;* ADB will recognize both forms.

2.3. Formats

To print data, a user specifies a collection of letters and characters that describe the format of the printout. Formats are "remembered" in the sense that typing a request without one will cause the new printout to appear in the previous format. The following are the most commonly used format letters.

b	**one byte in octal**
c	**one byte as a character**
o	**one word in octal**
d	**one word in decimal**
f	**two words in floating point**
i	**PDP 11 instruction**
s	**a null terminated character string**
a	**the value of dot**
u	**one word as unsigned integer**
n	**print a newline**
r	**print a blank space**
^	**backup dot**

(Format letters are also available for "long" values, for example, '**D**' for long decimal, and '**F**' for double floating point.) For other formats see the ADB manual.

2.4. General Request Meanings

The general form of a request is:

address,count command modifier

which sets 'dot' to *address* and executes the command *count* times.

The following table illustrates some general ADB command meanings:

Command	**Meaning**
?	**Print contents from *a.out* file**
/	**Print contents from *core* file**

=	**Print value of "dot"**
:	**Breakpoint control**
$	**Miscellaneous requests**
;	**Request separator**
!	**Escape to shell**

ADB catches signals, so a user cannot use a quit signal to exit from ADB. The request $q or $Q (or cntl-D) must be used to exit from ADB.

3. Debugging C Programs

3.1. Debugging A Core Image

Consider the C program in Figure 1. The program is used to illustrate a common error made by C programmers. The object of the program is to change the lower case "t" to upper case in the string pointed to by *charp* and then write the character string to the file indicated by argument 1. The bug shown is that the character "T" is stored in the pointer *charp* instead of the string pointed to by *charp*. Executing the program produces a core file because of an out of bounds memory reference.

ADB is invoked by:

adb a.out core

The first debugging request:

$c

is used to give a C backtrace through the subroutines called. As shown in Figure 2 only one function (*main*) was called and the arguments *argc* and *argv* have octal values 02 and 0177762 respectively. Both of these values look reasonable; 02 = two arguments, 0177762 = address on stack of parameter vector.
The next request:

$C

is used to give a C backtrace plus an interpretation of all the local variables in each function and their values in octal. The value of the variable *cc* looks incorrect since *cc* was declared as a character.

The next request:

$r

prints out the registers including the program counter and an interpretation of the instruction at that location.

The request:

$e

prints out the values of all external variables.

A map exists for each file handled by ADB. The map for the *a.out* file is referenced by ? whereas the map for *core* file is referenced by /. Furthermore, a good rule of thumb is to use ? for instructions and / for data when looking at programs. To print out information about the maps type:

$m

This produces a report of the contents of the maps. More about these maps later.

In our example, it is useful to see the contents of the string pointed to by *charp*. This is done by:

***charp/s**

which says use *charp* as a pointer in the *core* file and print the information as a character string. This printout clearly shows that the character buffer was incorrectly overwritten and helps identify the error. Printing the locations around *charp* shows that the buffer is unchanged but that the pointer is destroyed. Using ADB similarly, we could print information about the arguments to a function. The request:

main.argc/d

prints the decimal *core* image value of the argument *argc* in the function *main*.
The request:

***main.argv,3/o**

prints the octal values of the three consecutive cells pointed to by *argv* in the function *main*. Note that these values are the addresses of the arguments to main. Therefore:

0177770/s

prints the ASCII value of the first argument. Another way to print this value would have been

***"/s**

The " means ditto which remembers the last address typed, in this case *main.argc* ; the * instructs ADB to use the address field of the *core* file as a pointer.

The request:

.=o

prints the current address (not its contents) in octal which has been set to the address of the first argument. The current address, dot, is used by ADB to "remember" its current location. It allows the user to reference locations relative to the current address, for example:

.–10/d

3.2. Multiple Functions

Consider the C program illustrated in Figure 3. This program calls functions *f, g,* and *h* until the stack is exhausted and a core image is produced.

Again you can enter the debugger via:

adb

which assumes the names *a.out* and *core* for the executable file and core image file respectively. The request:

$c

will fill a page of backtrace references to *f, g,* and *h*. Figure 4 shows an abbreviated list (typing *DEL* will terminate the output and bring you back to ADB request level).

The request:

,5$C

prints the five most recent activations.

Notice that each function (*f,g,h*) has a counter of the number of times it was called.

The request:

fcnt/d

prints the decimal value of the counter for the function *f*. Similarly *gcnt* and *hcnt* could be printed. To print the value of an automatic variable, for example the decimal value of *x* in the last call of the function *h*, type:

h.x/d

It is currently not possible in the exported version to print stack frames other than the most recent activation of a function. Therefore, a user can print everything with **$C** or the occurrence of a variable in the most recent call of a function. It is possible with the **$C** request, however, to print the stack frame starting at some address as **address$C.**

3.3. Setting Breakpoints

Consider the C program in Figure 5. This program, which changes tabs into blanks, is adapted from *Software Tools* by Kernighan and Plauger, pp. 18-27.

We will run this program under the control of ADB (see Figure 6a) by:

adb a.out –

Breakpoints are set in the program as:

address:b [**request**]

The requests:

settab+4:b
fopen+4:b
getc+4:b
tabpos+4:b

set breakpoints at the start of these functions. C does not generate statement labels. Therefore it is currently not possible to plant breakpoints at locations other than function entry points without a knowledge of the code generated by the C compiler. The above addresses are entered as **symbol+4** so that they will appear in any C backtrace since the first instruction of each function is a call to the C save routine (*csv*). Note that some of the functions are from the C library.

To print the location of breakpoints one types:

$b

The display indicates a *count* field. A breakpoint is bypassed *count –1* times before causing a stop. The *command* field indicates the ADB requests to be executed each time the breakpoint is encountered. In our example no *command* fields are present.

By displaying the original instructions at the function *settab* we see that the breakpoint is set after the jsr to the C save routine. We can display the instructions using the ADB request:

settab,5?ia

This request displays five instructions starting at *settab* with the addresses of each location displayed. Another variation is:

settab,5?i

which displays the instructions with only the starting address.

Notice that we accessed the addresses from the *a.out* file with the **?** command. In general when asking for a printout of multiple items, ADB will advance the current address the number of bytes necessary to satisfy the request; in the above example five instructions were displayed and the current address was advanced 18 (decimal) bytes.

To run the program one simply types:

:r

To delete a breakpoint, for instance the entry to the function *settab*, one types:

settab+4:d

To continue execution of the program from the breakpoint type:

:c

Once the program has stopped (in this case at the breakpoint for *fopen),* ADB requests can be used to display the contents of memory. For example:

$C

to display a stack trace, or:

tabs,3/8o

to print three lines of 8 locations each from the array called *tabs*. By this time (at location *fopen)* in the C program, *settab* has been called and should have set a one in every eighth location of *tabs*.

3.4. Advanced Breakpoint Usage

We continue execution of the program with:

:c

See Figure 6b. *Getc* is called three times and the contents of the variable *c* in the function *main* are displayed each time. The single character on the left hand edge is the output from the C program. On the third occurrence of *getc* the program stops. We can look at the full buffer of characters by typing:

ibuf+6/20c

When we continue the program with:

:c

we hit our first breakpoint at *tabpos* since there is a tab following the "This" word of the data.

Several breakpoints of *tabpos* will occur until the program has changed the tab into equivalent blanks. Since we feel that *tabpos* is working, we can remove the breakpoint at that location by:

tabpos+4:d

If the program is continued with:

:c

it resumes normal execution after ADB prints the message

a.out:running

The UNIX quit and interrupt signals act on ADB itself rather than on the program being debugged. If such a signal occurs then the program being debugged is stopped and control is returned to ADB. The signal is saved by ADB and is passed on to the test program if:

:c

is typed. This can be useful when testing interrupt handling routines. The signal is not passed on to the test program if:

:c 0

is typed.

Now let us reset the breakpoint at *settab* and display the instructions located there when we reach the breakpoint. This is accomplished by:

settab+4:b settab,5?ia *

It is also possible to execute the ADB requests for each occurrence of the breakpoint but only stop

* Owing to a bug in early versions of ADB (including the version distributed in Generic 3 UNIX) these statements must be written as:

after the third occurrence by typing:

getc+4,3:b main.c?C *

This request will print the local variable *c* in the function *main* at each occurrence of the breakpoint. The semicolon is used to separate multiple ADB requests on a single line.

Warning: setting a breakpoint causes the value of dot to be changed; executing the program under ADB does not change dot. Therefore:

settab+4:b .,5?ia
fopen+4:b

will print the last thing dot was set to (in the example *fopen+4*) *not* the current location (*settab+4*) at which the program is executing.

A breakpoint can be overwritten without first deleting the old breakpoint. For example:

settab+4:b settab,5?ia; ptab/o *

could be entered after typing the above requests.

Now the display of breakpoints:

$b

shows the above request for the *settab* breakpoint. When the breakpoint at *settab* is encountered the ADB requests are executed. Note that the location at *settab+4* has been changed to plant the breakpoint; all the other locations match their original value.

Using the functions, *f*, *g* and *h* shown in Figure 3, we can follow the execution of each function by planting non-stopping breakpoints. We call ADB with the executable program of Figure 3 as follows:

adb ex3 –

Suppose we enter the following breakpoints:

```
h+4:b    hcnt/d; h.hi/; h.hr/
g+4:b    gcnt/d; g.gi/; g.gr/
f+4:b    fcnt/d; f.fi/; f.fr/
:r
```

Each request line indicates that the variables are printed in decimal (by the specification **d**). Since the format is not changed, the **d** can be left off all but the first request.

The output in Figure 7 illustrates two points. First, the ADB requests in the breakpoint line are not examined until the program under test is run. That means any errors in those ADB requests is not detected until run time. At the location of the error ADB stops running the program.

The second point is the way ADB handles register variables. ADB uses the symbol table to address variables. Register variables, like *f.fr* above, have pointers to uninitialized places on the stack. Therefore the message "symbol not found".

Another way of getting at the data in this example is to print the variables used in the call as:

```
f+4:b    fcnt/d; f.a/; f.b/; f.fi/
g+4:b    gcnt/d; g.p/; g.q/; g.gi/
:c
```

```
settab+4:b    settab,5?ia;0
getc+4,3:b    main.c?C;0
settab+4:b    settab,5?ia; ptab/o;0
```

Note that ;0 will set dot to zero and stop at the breakpoint.

The operator / was used instead of ? to read values from the *core* file. The output for each function, as shown in Figure 7, has the same format. For the function *f*, for example, it shows the name and value of the *external* variable *fcnt*. It also shows the address on the stack and value of the variables *a*, *b* and *fi*.

Notice that the addresses on the stack will continue to decrease until no address space is left for program execution at which time (after many pages of output) the program under test aborts. A display with names would be produced by requests like the following:

f+4:b   fcnt/d; f.a/"a="d; f.b/"b="d; f.fi/"fi="d

In this format the quoted string is printed literally and the **d** produces a decimal display of the variables. The results are shown in Figure 7.

3.5. Other Breakpoint Facilities

- Arguments and change of standard input and output are passed to a program as:

  :r arg1 arg2 ... <infile >outfile

  This request kills any existing program under test and starts the *a.out* afresh.

- The program being debugged can be single stepped by:

  :s

  If necessary, this request will start up the program being debugged and stop after executing the first instruction.

- ADB allows a program to be entered at a specific address by typing:

  address:r

- The count field can be used to skip the first *n* breakpoints as:

  ,n:r

  The request:

  ,n:c

  may also be used for skipping the first *n* breakpoints when continuing a program.

- A program can be continued at an address different from the breakpoint by:

  address:c

- The program being debugged runs as a separate process and can be killed by:

  :k

4. Maps

UNIX supports several executable file formats. These are used to tell the loader how to load the program file. File type 407 is the most common and is generated by a C compiler invocation such as **cc pgm.c**. A 410 file is produced by a C compiler command of the form **cc -n pgm.c**, whereas a 411 file is produced by **cc -i pgm.c**. ADB interprets these different file formats and provides access to the different segments through a set of maps (see Figure 8). To print the maps type:

$m

In 407 files, both text (instructions) and data are intermixed. This makes it impossible for ADB to differentiate data from instructions and some of the printed symbolic addresses look incorrect; for example, printing data addresses as offsets from routines.

In 410 files (shared text), the instructions are separated from data and **?*** accesses the data part of the *a.out* file. The **?*** request tells ADB to use the second part of the map in the *a.out* file. Accessing data in the *core* file shows the data after it was modified by the execution of the program. Notice also that the data segment may have grown during program execution.

In 411 files (separated I & D space), the instructions and data are also separated. However, in this case, since data is mapped through a separate set of segmentation registers, the base of the data segment is also relative to address zero. In this case since the addresses overlap it is necessary to use the **?*** operator to access the data space of the *a.out* file. In both 410 and 411 files the corresponding core file does not contain the program text.

Figure 9 shows the display of three maps for the same program linked as a 407, 410, 411 respectively. The b, e, and f fields are used by ADB to map addresses into file addresses. The "f1" field is the length of the header at the beginning of the file (020 bytes for an *a.out* file and 02000 bytes for a *core* file). The "f2" field is the displacement from the beginning of the file to the data. For a 407 file with mixed text and data this is the same as the length of the header; for 410 and 411 files this is the length of the header plus the size of the text portion.

The "b" and "e" fields are the starting and ending locations for a segment. Given an address, A, the location in the file (either *a.out* or *core*) is calculated as:

$$\mathbf{b1 \leq A \leq e1 \Longrightarrow file\ address = (A{-}b1){+}f1}$$
$$\mathbf{b2 \leq A \leq e2 \Longrightarrow file\ address = (A{-}b2){+}f2}$$

A user can access locations by using the ADB defined variables. The **$v** request prints the variables initialized by ADB:

b	**base address of data segment**
d	**length of the data segment**
s	**length of the stack**
t	**length of the text**
m	**execution type (407,410,411)**

In Figure 9 those variables not present are zero. Use can be made of these variables by expressions such as:

<b

in the address field. Similarly the value of the variable can be changed by an assignment request such as:

02000>b

that sets **b** to octal 2000. These variables are useful to know if the file under examination is an executable or *core* image file.

ADB reads the header of the *core* image file to find the values for these variables. If the second file specified does not seem to be a *core* file, or if it is missing then the header of the executable file is used instead.

5. Advanced Usage

It is possible with ADB to combine formatting requests to provide elaborate displays. Below are several examples.

5.1. Formatted dump

The line:

<b,-1/4o4^8Cn

prints 4 octal words followed by their ASCII interpretation from the data space of the core image file. Broken down, the various request pieces mean:

<b The base address of the data segment.

<b,–1 Print from the base address to the end of file. A negative count is used here and elsewhere to loop indefinitely or until some error condition (like end of file) is detected.

The format **4o4^8Cn** is broken down as follows:

4o Print 4 octal locations.

4^ Backup the current address 4 locations (to the original start of the field).

8C Print 8 consecutive characters using an escape convention; each character in the range 0 to 037 is printed as @ followed by the corresponding character in the range 0140 to 0177. An @ is printed as @@.

n Print a newline.

The request:

```
<b,<d/4o4^8Cn
```

could have been used instead to allow the printing to stop at the end of the data segment (<d provides the data segment size in bytes).

The formatting requests can be combined with ADB's ability to read in a script to produce a core image dump script. ADB is invoked as:

```
adb a.out core < dump
```

to read in a script file, *dump*, of requests. An example of such a script is:

```
120$w
4095$s
$v
=3n
$m
=3n"C Stack Backtrace"
$C
=3n"C External Variables"
$e
=3n"Registers"
$r
0$s
=3n"Data Segment"
<b,-1/8ona
```

The request **120$w** sets the width of the output to 120 characters (normally, the width is 80 characters). ADB attempts to print addresses as:

symbol + offset

The request **4095$s** increases the maximum permissible offset to the nearest symbolic address from 255 (default) to 4095. The request = can be used to print literal strings. Thus, headings are provided in this *dump* program with requests of the form:

=3n"C Stack Backtrace"

that spaces three lines and prints the literal string. The request **$v** prints all non-zero ADB variables (see Figure 8). The request **0$s** sets the maximum offset for symbol matches to zero thus suppressing the printing of symbolic labels in favor of octal values. Note that this is only done for the printing of the data segment. The request:

<b,-1/8ona

prints a dump from the base of the data segment to the end of file with an octal address field and eight octal numbers per line.

Figure 11 shows the results of some formatting requests on the C program of Figure 10.

5.2. Directory Dump

As another illustration (Figure 12) consider a set of requests to dump the contents of a directory (which is made up of an integer *inumber* followed by a 14 character name):

```
adb dir -
=n8t"Inum"8t"Name"
0,-1? u8t14cn
```

In this example, the **u** prints the *inumber* as an unsigned decimal integer, the **8t** means that ADB will space to the next multiple of 8 on the output line, and the **14c** prints the 14 character file name.

5.3. Ilist Dump

Similarly the contents of the *ilist* of a file system, (e.g. /dev/src, on UNIX systems distributed by the UNIX Support Group; see UNIX Programmer's Manual Section V) could be dumped with the following set of requests:

```
adb /dev/src -
02000>b
?m <b
<b,-1?"flags"8ton"links,uid,gid"8t3bn",size"8tbrdn"addr"8t8un"times"8t2Y2na
```

In this example the value of the base for the map was changed to 02000 (by saying **?m<b**) since that is the start of an *ilist* within a file system. An artifice (**brd** above) was used to print the 24 bit size field as a byte, a space, and a decimal integer. The last access time and last modify time are printed with the **2Y** operator. Figure 12 shows portions of these requests as applied to a directory and file system.

5.4. Converting values

ADB may be used to convert values from one representation to another. For example:

072 = odx

will print

072 58 #3a

which is the octal, decimal and hexadecimal representations of 072 (octal). The format is remembered so that typing subsequent numbers will print them in the given formats. Character values may be converted similarly, for example:

'a' = co

prints

a 0141

It may also be used to evaluate expressions but be warned that all binary operators have the same precedence which is lower than that for unary operators.

6. Patching

Patching files with ADB is accomplished with the *write*, **w** or **W**, request (which is not like the *ed* editor write command). This is often used in conjunction with the *locate*, **l** or **L** request. In general, the request syntax for **l** and **w** are similar as follows:

?l value

The request **l** is used to match on two bytes, **L** is used for four bytes. The request **w** is used to write two bytes, whereas **W** writes four bytes. The **value** field in either *locate* or *write* requests is an expression. Therefore, decimal and octal numbers, or character strings are supported.

In order to modify a file, ADB must be called as:

adb –w file1 file2

When called with this option, *file1* and *file2* are created if necessary and opened for both reading and writing.

For example, consider the C program shown in Figure 10. We can change the word "This" to "The " in the executable file for this program, *ex7*, by using the following requests:

```
adb –w ex7 –
?l 'Th'
?W 'The '
```

The request **?l** starts at dot and stops at the first match of "Th" having set dot to the address of the location found. Note the use of **?** to write to the *a.out* file. The form **?*** would have been used for a 411 file.

More frequently the request will be typed as:

?l 'Th'; ?s

and locates the first occurrence of "Th" and print the entire string. Execution of this ADB request will set dot to the address of the "Th" characters.

As another example of the utility of the patching facility, consider a C program that has an internal logic flag. The flag could be set by the user through ADB and the program run. For example:

```
adb a.out –
:s arg1 arg2
flag/w 1
:c
```

The **:s** request is normally used to single step through a process or start a process in single step mode. In this case it starts *a.out* as a subprocess with arguments **arg1** and **arg2**. If there is a subprocess running ADB writes to it rather than to the file so the **w** request causes *flag* to be changed in the memory of the subprocess.

7. Anomalies

Below is a list of some strange things that users should be aware of.

1. Function calls and arguments are put on the stack by the C save routine. Putting breakpoints at the entry point to routines means that the function appears not to have been called when the breakpoint occurs.
2. When printing addresses, ADB uses either text or data symbols from the *a.out* file. This sometimes causes unexpected symbol names to be printed with data (e.g. *savr5+022*). This does not happen if **?** is used for text (instructions) and / for data.

3. ADB cannot handle C register variables in the most recently activated function.

8. Acknowledgements

The authors are grateful for the thoughtful comments on how to organize this document from R. B. Brandt, E. N. Pinson and B. A. Tague. D. M. Ritchie made the system changes necessary to accommodate tracing within ADB. He also participated in discussions during the writing of ADB. His earlier work with DB and CDB led to many of the features found in ADB.

9. References

1. D. M. Ritchie and K. Thompson, "The UNIX Time-Sharing System," CACM, July, 1974.
2. B. W. Kernighan and D. M. Ritchie, *The C Programming Language,* Prentice-Hall, 1978.
3. K. Thompson and D. M. Ritchie, UNIX Programmer's Manual - 7th Edition, 1978.
4. B. W. Kernighan and P. J. Plauger, *Software Tools,* Addison-Wesley, 1976.

Figure 1: C program with pointer bug

```
struct buf {
        int fildes;
        int nleft;
        char *nextp;
        char buff[512];
        }bb;
struct buf *obuf;

char *charp "this is a sentence.";

main(argc,argv)
int argc;
char **argv;
{
        char    cc;

        if(argc < 2) {
                printf("Input file missing\n");
                exit(8);
        }

        if((fcreat(argv[1],obuf)) < 0){
                printf("%s : not found\n", argv[1]);
                exit(8);
        }
        charp = 'T';
printf("debug 1 %s\n",charp);
        while(cc= *charp++)
                putc(cc,obuf);
        fflush(obuf);
}
```

Figure 2: ADB output for C program of Figure 1

```
adb a.out core
$c
~main(02,0177762)
$C
~main(02,0177762)
        argc:       02
        argv:       0177762
        cc:         02124
$r
ps      0170010
pc      0204     ~main+0152
sp      0177740
r5      0177752
r4      01
r3      0
r2      0
r1      0
r0      0124
~main+0152:     mov     _obuf,(sp)
$e
savr5:    0
_obuf:    0
_charp:   0124
_errno:   0
_fout:    0
$m
text map   `ex1´
b1 = 0              e1   = 02360          f1 = 020
b2 = 0              e2   = 02360          f2 = 020
data map   `core1´
b1 = 0              e1   = 03500          f1 = 02000
b2 = 0175400        e2   = 0200000                f2 = 05500
*charp/s
0124:           TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTLx    Nh@x &_
~

charp/s
_charp:         T

_charp+02:      this is a sentence.

_charp+026:     Input file missing
main.argc/d
0177756: 2
*main.argv/3o
0177762: 0177770 0177776 0177777
0177770/s
0177770: a.out
*main.argv/3o
0177762: 0177770 0177776 0177777
*"/s
0177770: a.out
 .=o
                0177770
 .-10/d
0177756: 2
$q
```

Figure 3: Multiple function C program for stack trace illustration

```
int     fcnt,gcnt,hcnt;
h(x,y)
{
        int hi; register int hr;
        hi = x+1;
        hr = x-y+1;
        hcnt++ ;
        hj:
        f(hr,hi);
}

g(p,q)
{
        int gi; register int gr;
        gi = q-p;
        gr = q-p+1;
        gcnt++ ;
        gj:
        h(gr,gi);
}

f(a,b)
{
        int fi; register int fr;
        fi = a+2*b;
        fr = a+b;
        fcnt++ ;
        fj:
        g(fr,fi);
}

main()
{
        f(1,1);
}
```

Figure 4: ADB output for C program of Figure 3

```
adb
$c
~h(04452,04451)
~g(04453,011124)
~f(02,04451)
~h(04450,04447)
~g(04451,011120)
~f(02,04447)
~h(04446,04445)
~g(04447,011114)
~f(02,04445)
~h(04444,04443)
HIT DEL KEY
adb
,5$C
~h(04452,04451)
        x:      04452
        y:      04451
        hi:     ?
~g(04453,011124)
        p:      04453
        q:      011124
        gi:     04451
        gr:     ?
~f(02,04451)
        a:      02
        b:      04451
        fi:     011124
        fr:     04453
~h(04450,04447)
        x:      04450
        y:      04447
        hi:     04451
        hr:     02
~g(04451,011120)
        p:      04451
        q:      011120
        gi:     04447
        gr:     04450
fcnt/d
_fcnt:          1173
gcnt/d
_gcnt:          1173
hcnt/d
_hcnt:          1172
h.x/d
022004:         2346
$q
```

Figure 5: C program to decode tabs

```
#define MAXLINE     80
#define YES         1
#define NO          0
#define TABSP       8

char    input[] "data";
char    ibuf[518];
int     tabs[MAXLINE];

main()
{
        int col, *ptab;
        char c;

        ptab = tabs;
        settab(ptab);       /*Set initial tab stops */
        col = 1;
        if(fopen(input,ibuf) < 0) {
                printf("%s : not found\n",input);
                exit(8);
        }
        while((c = getc(ibuf)) != -1) {
                switch(c) {
                        case '\t' :         /* TAB */
                                while(tabpos(col) != YES) {
                                        putchar(' ');       /* put BLANK */
                                        col++ ;
                                }
                                break;
                        case '\n' :         /*NEWLINE */
                                putchar('\n');
                                col = 1;
                                break;
                        default:
                                putchar(c);
                                col++ ;
                }
        }
}
/* Tabpos return YES if col is a tab stop */
tabpos(col)
int col;
{
        if(col > MAXLINE)
                return(YES);
        else
                return(tabs[col]);
}
/* Settab - Set initial tab stops */
settab(tabp)
int *tabp;
{
        int i;

        for(i = 0; i<= MAXLINE; i++)
                (i%TABSP) ? (tabs[i] = NO) : (tabs[i] = YES);
}
```

Figure 6a: ADB output for C program of Figure 5

```
adb a.out -
settab+4:b
fopen+4:b
getc+4:b
tabpos+4:b
$b
breakpoints
count   bkpt            command
1       ~tabpos+04
1       _getc+04
1       _fopen+04
1       ~settab+04
settab,5?ia
~settab:        jsr     r5,csv
~settab+04:     tst     -(sp)
~settab+06:     clr     0177770(r5)
~settab+012:    cmp     $0120,0177770(r5)
~settab+020:    blt     ~settab+076
~settab+022:
settab,5?i
~settab:        jsr     r5,csv
                tst     -(sp)
                clr     0177770(r5)
                cmp     $0120,0177770(r5)
                blt     ~settab+076
:r
a.out: running
breakpoint      ~settab+04:     tst     -(sp)
settab+4:d
:c
a.out: running
breakpoint      _fopen+04:      mov     04(r5),nulstr+012
$C
_fopen(02302,02472)
~main(01,0177770)
        col:    01
        c:      0
        ptab:   03500
tabs,3/8o
03500:          01      0       0       0       0       0       0       0
                01      0       0       0       0       0       0       0
                01      0       0       0       0       0       0       0
```

Figure 6b: ADB output for C program of Figure 5

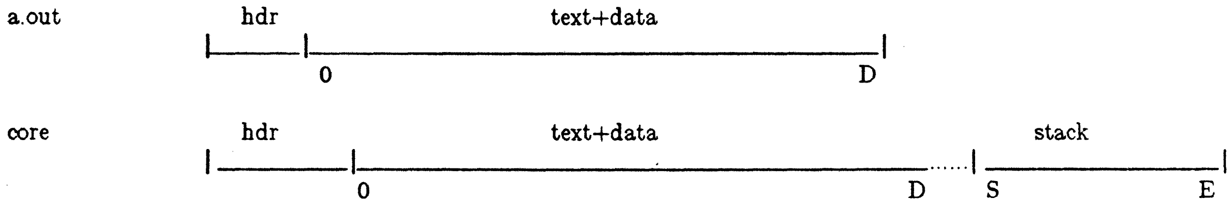
```
:c
a.out: running
breakpoint      _getc+04:       mov     04(r5),r1
ibuf+6/20c
__cleanu+0202:          This    is      a test  of
:c
a.out: running
breakpoint      ~tabpos+04:     cmp     $0120,04(r5)
tabpos+4:d
settab+4:b settab,5?ia
settab+4:b settab,5?ia; 0
getc+4,3:b main.c?C; 0
settab+4:b settab,5?ia; ptab/o; 0
$b
breakpoints
count   bkpt            command
1       ~tabpos+04
3       _getc+04        main.c?C;0
1       _fopen+04
1       ~settab+04      settab,5?ia;ptab?o;0
~settab:        jsr     r5,csv
~settab+04:     bpt
~settab+06:     clr     0177770(r5)
~settab+012:    cmp     $0120,0177770(r5)
~settab+020:    blt     ~settab+076
~settab+022:
0177766: 0177770
0177744: @`
T0177744:       T
h0177744:       h
i0177744:       i
s0177744:       s
```

Figure 7: ADB output for C program with breakpoints

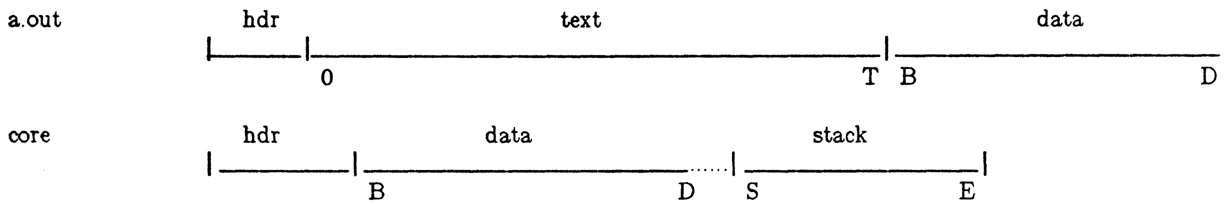
```
adb ex3 -
h+4:b hcnt/d; h.hi/; h.hr/
g+4:b gcnt/d; g.gi/; g.gr/
f+4:b fcnt/d; f.fi/; f.fr/
:r
ex3: running
_fcnt:      0
0177732:    214
symbol not found
f+4:b fcnt/d; f.a/; f.b/; f.fi/
g+4:b gcnt/d; g.p/; g.q/; g.gi/
h+4:b hcnt/d; h.x/; h.y/; h.hi/
:c
ex3: running
_fcnt:      0
0177746:    1
0177750:    1
0177732:    214
_gcnt:      0
0177726:    2
0177730:    3
0177712:    214
_hcnt:      0
0177706:    2
0177710:    1
0177672:    214
_fcnt:      1
0177666:    2
0177670:    3
0177652:    214
_gcnt:      1
0177646:    5
0177650:    8
0177632:    214
HIT DEL
f+4:b fcnt/d; f.a/"a = "d; f.b/"b = "d; f.fi/"fi = "d
g+4:b gcnt/d; g.p/"p = "d; g.q/"q = "d; g.gi/"gi = "d
h+4:b hcnt/d; h.x/"x = "d; h.y/"h = "d; h.hi/"hi = "d
:r
ex3: running
_fcnt:      0
0177746:    a = 1
0177750:    b = 1
0177732:    fi = 214
_gcnt:      0
0177726:    p = 2
0177730:    q = 3
0177712:    gi = 214
_hcnt:      0
0177706:    x = 2
0177710:    y = 1
0177672:    hi = 214
_fcnt:      1
0177666:    a = 2
0177670:    b = 3
0177652:    fi = 214
HIT DEL
$q
```

Figure 8: ADB address maps

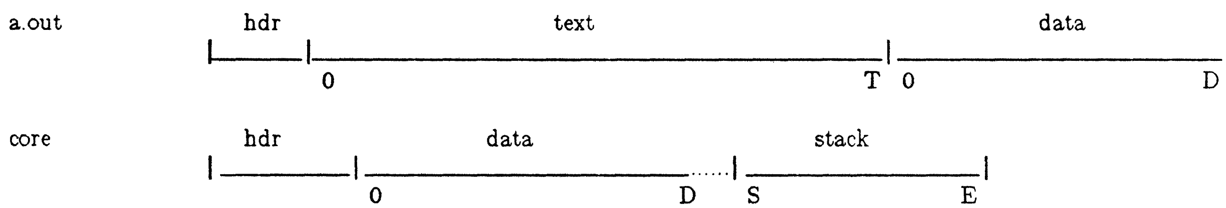
407 files



410 files (shared text)



411 files (separated I and D space)



The following *adb* variables are set.

		407	410	411
b	base of data	0	B	0
d	length of data	D	D−B	D
s	length of stack	S	S	S
t	length of text	0	T	T

Figure 9: ADB output for maps

```
adb map407 core407
$m
text map    `map407'
b1 = 0          e1    = 0256          f1 = 020
b2 = 0          e2    = 0256          f2 = 020
data map    `core407'
b1 = 0          e1    = 0300          f1 = 02000
b2 = 0175400    e2    = 0200000  f2 = 02300
$v
variables
d = 0300
m = 0407
s = 02400
$q
```

```
adb map410 core410
$m
text map    `map410'
b1 = 0          e1    = 0200          f1 = 020
b2 = 020000     e2    = 020116   f2 = 0220
data map    `core410'
b1 = 020000     e1    = 020200   f1 = 02000
b2 = 0175400    e2    = 0200000  f2 = 02200
$v
variables
b = 020000
d = 0200
m = 0410
s = 02400
t = 0200
$q
```

```
adb map411 core411
$m
text map    `map411'
b1 = 0          e1    = 0200          f1 = 020
b2 = 0          e2    = 0116          f2 = 0220
data map    `core411'
b1 = 0          e1    = 0200          f1 = 02000
b2 = 0175400    e2    = 0200000  f2 = 02200
$v
variables
d = 0200
m = 0411
s = 02400
t = 0200
$q
```

Figure 10: Simple C program for illustrating formatting and patching

```
char    str1[]   "This is a character string";
int     one      1;
int     number   456;
long    lnum     1234;
float   fpt      1.25;
char    str2[]   "This is the second character string";
main()
{
        one = 2;
}
```

Figure 11: ADB output illustrating fancy formats

```
adb map410 core410
<b,-1/8ona
020000:        0      064124    071551    064440    020163    020141    064143071141

_str1+016: 061541     062564    020162    072163    064562    063556    0     02

_number:
_number:   0710 0     02322     040240    0    064124    071551    064440

_str2+06:  020163     064164    020145    062563    067543    062156    061440060550

_str2+026: 060562     072143    071145    071440    071164    067151    0147 0

savr5+02:  0     0    0    0    0    0    0    0

<b,20/4o4^8Cn
020000:        0      064124    071551    064440    @`@`This i
           020163     020141    064143    071141    s a char
           061541     062564    020162    072163    acter st
           064562     063556    0    02   ring@`@`@b@`

_number:   0710 0     02322     040240    H@a@`@`R@d @@
           0     064124    071551    064440    @`@`This i
           020163     064164    020145    062563    s the se
           067543     062156    061440    060550    cond cha
           060562     072143    071145    071440    racter s
           071164     067151    0147 0    tring@`@`@`
           0     0    0    0    @`@`@`@`@`@`@`@`
           0     0    0    0    @`@`@`@`@`@`@`@`
data address not found
<b,20/4o4^8t8cna
020000:        0      064124    071551    064440        This i
_str1+06:  020163     020141    064143    071141        s a char
_str1+016: 061541     062564    020162    072163        acter st
_str1+026: 064562     063556    0    02        ring
_number:
_number:   0710 0     02322     040240        HR
_fpt+02:   0     064124    071551    064440        This i
_str2+06:  020163     064164    020145    062563        s the se
_str2+016: 067543     062156    061440    060550        cond cha
_str2+026: 060562     072143    071145    071440        racter      s
_str2+036: 071164     067151    0147 0        tring
savr5+02:  0     0    0    0
savr5+012: 0     0    0    0
data address not found
<b,10/2b8t^2cn
020000:        0    0

_str1:     0124 0150        Th
           0151 0163        is
           040  0151         i
           0163 040         s
           0141 040         a
           0143 0150        ch
           0141 0162        ar
           0141 0143        ac
           0164 0145        te
$Q
```

Figure 12: Directory and inode dumps

```
adb dir -
=nt"Inode"t"Name"
0,-1?ut14cn

        Inode Name
0:      652   .
        82    ..
        5971  cap.c
        5323  cap
        0     pp
```

```
adb /dev/src -
02000>b
?m<b
new map        `/dev/src'
b1 = 02000          e1      = 0100000000    f1 = 0
b2 = 0              e2      = 0             f2 = 0
$v
variables
b = 02000
<b,-1?"flags"8ton"links,uid,gid"8t3bn"size"8tbrdn"addr"8t8un"times"8t2Y2na
02000:          flags 073145
        links,uid,gid     0163 0164 0141
        size   0162 10356
        addr 28770      8236 25956      27766      25455      8236 25956      25206
        times 1976 Feb 5 08:34:56    1975 Dec 28 10:55:15

02040:          flags 024555
        links,uid,gid     012   0163 0164
        size   0162 25461
        addr 8308 30050      8294 25130      15216      26890      29806      10784
        times 1976 Aug 17 12:16:51   1976 Aug 17 12:16:51

02100:          flags 05173
        links,uid,gid     011   0162 0145
        size   0147 29545
        addr 25972      8306 28265      8308 25642      15216      2314 25970
        times 1977 Apr 2 08:58:01    1977 Feb 5 10:21:44
```

ADB Summary

Command Summary

a) formatted printing

? *format*	print from *a.out* file according to *format*
/ *format*	print from *core* file according to *format*
= *format*	print the value of *dot*
?w expr	write expression into *a.out* file
/w expr	write expression into *core* file
?l expr	locate expression in *a.out* file

b) breakpoint and program control

:b	set breakpoint at *dot*
:c	continue running program
:d	delete breakpoint
:k	kill the program being debugged
:r	run *a.out* file under ADB control
:s	single step

c) miscellaneous printing

$b	print current breakpoints
$c	C stack trace
$e	external variables
$f	floating registers
$m	print ADB segment maps
$q	exit from ADB
$r	general registers
$s	set offset for symbol match
$v	print ADB variables
$w	set output line width

d) calling the shell

!	call *shell* to read rest of line

e) assignment to variables

>*name*	assign dot to variable or register *name*

Format Summary

a	the value of dot
b	one byte in octal
c	one byte as a character
d	one word in decimal
f	two words in floating point
i	PDP 11 instruction
o	one word in octal
n	print a newline
r	print a blank space
s	a null terminated character string
*n***t**	move to next *n* space tab
u	one word as unsigned integer
x	hexadecimal
Y	date
^	backup dot
"..."	print string

Expression Summary

a) expression components

decimal integer	e.g. 256
octal integer	e.g. 0277
hexadecimal	e.g. #ff
symbols	e.g. flag _main main.argc
variables	e.g. <b
registers	e.g. <pc <r0
(expression)	expression grouping

b) dyadic operators

+	add
−	subtract
*	multiply
%	integer division
&	bitwise and
|	bitwise or
#	round up to the next multiple

c) monadic operators

~	not
*	contents of location
−	integer negate

An Introduction to the Source Code Control System

Eric Allman
Project Ingres
University of California at Berkeley

This document gives a quick introduction to using the Source Code Control System (SCCS). The presentation is geared to programmers who are more concerned with what to do to get a task done rather than how it works; for this reason some of the examples are not well explained. For details of what the magic options do, see the section on "Further Information".

> This is a working document. Please send any comments or suggestions to csvax:eric.

1. Introduction

SCCS is a source management system. Such a system maintains a record of versions of a system; a record is kept with each set of changes of what the changes are, why they were made, and who made them and when. Old versions can be recovered, and different versions can be maintained simultaneously. In projects with more than one person, SCCS will insure that two people are not editing the same file at the same time.

All versions of your program, plus the log and other information, is kept in a file called the "s-file". There are three major operations that can be performed on the s-file:

(1) Get a file for compilation (not for editing). This operation retrieves a version of the file from the s-file. By default, the latest version is retrieved. This file is intended for compilation, printing, or whatever; it is specifically NOT intended to be edited or changed in any way; any changes made to a file retrieved in this way will probably be lost.

(2) Get a file for editing. This operation also retrieves a version of the file from the s-file, but this file is intended to be edited and then incorporated back into the s-file. Only one person may be editing a file at one time.

(3) Merge a file back into the s-file. This is the companion operation to (2). A new version number is assigned, and comments are saved explaining why this change was made.

2. Learning the Lingo

There are a number of terms that are worth learning before we go any farther.

2.1. S-file

The s-file is a single file that holds all the different versions of your file. The s-file is stored in differential format; *i.e.*, only the differences between versions are stored, rather than the entire text of the new version. This saves disk space and allows selective changes to be removed later. Also included in the s-file is some header information for each version, including the comments given by the person who created the version explaining why the changes were made.

This is version 1.21 of this document. It was last modified on 12/5/80.

2.2. Deltas

Each set of changes to the s-file (which is approximately [but not exactly!] equivalent to a version of the file) is called a *delta*. Although technically a delta only includes the *changes* made, in practice it is usual for each delta to be made with respect to all the deltas that have occurred before[1]. However, it is possible to get a version of the file that has selected deltas removed out of the middle of the list of changes – equivalent to removing your changes later.

2.3. SID's (or, version numbers)

A SID (SCCS Id) is a number that represents a delta. This is normally a two-part number consisting of a "release" number and a "level" number. Normally the release number stays the same, however, it is possible to move into a new release if some major change is being made.

Since all past deltas are normally applied, the SID of the final delta applied can be used to represent a version number of the file as a whole.

2.4. Id keywords

When you get a version of a file with intent to compile and install it (*i.e.*, something other than edit it), some special keywords are expanded inline by SCCS. These *Id Keywords* can be used to include the current version number or other information into the file. All id keywords are of the form %*x*%, where *x* is an upper case letter. For example, %I% is the SID of the latest delta applied, %W% includes the module name, SID, and a mark that makes it findable by a program, and %G% is the date of the latest delta applied. There are many others, most of which are of dubious usefulness.

When you get a file for editing, the id keywords are not expanded; this is so that after you put them back in to the s-file, they will be expanded automatically on each new version. But notice: if you were to get them expanded accidently, then your file would appear to be the same version forever more, which would of course defeat the purpose. Also, if you should install a version of the program without expanding the id keywords, it will be impossible to tell what version it is (since all it will have is "%W%" or whatever).

3. Creating SCCS Files

To put source files into SCCS format, run the following shell script from csh:

```
mkdir SCCS save
foreach i (*.[ch])
        sccs admin -i$i $i
        mv $i save/$i
end
```

This will put the named files into s-files in the subdirectory "SCCS" The files will be removed from the current directory and hidden away in the directory "save", so the next thing you will probably want to do is to get all the files (described below). When you are convinced that SCCS has correctly created the s-files, you should remove the directory "save".

If you want to have id keywords in the files, it is best to put them in before you create the s-files. If you do not, *admin* will print "No Id Keywords (cm7)", which is a warning message only.

4. Getting Files for Compilation

To get a copy of the latest version of a file, run

```
sccs get prog.c
```

SCCS will respond:

---

[1]This matches normal usage, where the previous changes are not saved at all, so all changes are automatically based on all other changes that have happened through history.

1.1
87 lines

meaning that version 1.1 was retrieved[2] and that it has 87 lines. The file *prog.c* will be created in the current directory. The file will be read-only to remind you that you are not supposed to change it.

This copy of the file should not be changed, since SCCS is unable to merge the changes back into the s-file. If you do make changes, they will be lost the next time someone does a *get*.

5. Changing Files (or, Creating Deltas)

5.1. Getting a copy to edit

To edit a source file, you must first get it, requesting permission to edit it[3]:

```
sccs edit prog.c
```

The response will be the same as with *get* except that it will also say:

```
New delta 1.2
```

You then edit it, using a standard text editor:

```
vi prog.c
```

5.2. Merging the changes back into the s-file

When the desired changes are made, you can put your changes into the SCCS file using the *delta* command:

```
sccs delta prog.c
```

Delta will prompt you for "comments?" before it merges the changes in. At this prompt you should type a one-line description of what the changes mean (more lines can be entered by ending each line except the last with a backslash[4]). *Delta* will then type:

```
1.2
5 inserted
3 deleted
84 unchanged
```

saying that delta 1.2 was created, and it inserted five lines, removed three lines, and left 84 lines unchanged[5]. The *prog.c* file will be removed; it can be retrieved using *get*.

5.3. When to make deltas

It is probably unwise to make a delta before every recompilation or test; otherwise, you tend to get a lot of deltas with comments like "fixed compilation problem in previous delta" or "fixed botch in 1.3". However, it is very important to delta everything before installing a module for general use. A good technique is to edit the files you need, make all necessary changes and tests, compiling and editing as often as necessary without making deltas. When you are satisfied that you have a working version, delta everything being edited, re-get them, and recompile everything.

---

[2]Actually, the SID of the final delta applied was 1.1.

[3]The "edit" command is equivalent to using the –e flag to *get*, as:

```
sccs get –e prog.c
```

Keep *this* in mind when reading other documentation.

[4]Yes, this is a stupid default.

[5]Changes to a line are counted as a line deleted and a line inserted.

5.4. What's going on: the info command

To find out what files where being edited, you can use:

```
sccs info
```

to print out all the files being edited and other information such as the name of the user who did the edit. Also, the command:

```
sccs check
```

is nearly equivalent to the *info* command, except that it is silent if nothing is being edited, and returns non-zero exit status if anything is being edited; it can be used in an "install" entry in a makefile to abort the install if anything has not been properly deltaed.

If you know that everything being edited should be deltaed, you can use:

```
sccs delta `sccs tell`
```

The *tell* command is similar to *info* except that only the names of files being edited are output, one per line.

All of these commands take a **–b** flag to ignore "branches" (alternate versions, described later) and the **–u** flag to only give files being edited by you. The **–u** flag takes an optional *user* argument, giving only files being edited by that user. For example,

```
sccs info –ujohn
```

gives a listing of files being edited by john.

5.5. ID keywords

Id keywords can be inserted into your file that will be expanded automatically by *get*. For example, a line such as:

```
static char SccsId[] = "%W%\t%G%";
```

will be replaced with something like:

```
static char SccsId[] = "@(#)prog.c	1.2	08/29/80";
```

This tells you the name and version of the source file and the time the delta was created. The string "@(#)" is a special string which signals the beginning of an SCCS Id keyword.

5.5.1. The what command

To find out what version of a program is being run, use:

```
sccs what prog.c /usr/bin/prog
```

which will print all strings it finds that begin with "@(#)". This works on all types of files, including binaries and libraries. For example, the above command will output something like:

```
prog.c:
	prog.c	1.2	08/29/80
/usr/bin/prog:
	prog.c	1.1	02/05/79
```

From this I can see that the source that I have in prog.c will not compile into the same version as the binary in /usr/bin/prog.

5.5.2. Where to put id keywords

ID keywords can be inserted anywhere, including in comments, but Id Keywords that are compiled into the object module are especially useful, since it lets you find out what version of the object is being run, as well as the source. However, there is a cost: data space is used up to store the keywords, and on small address space machines this may be prohibitive.

When you put id keywords into header files, it is important that you assign them to different variables. For example, you might use:

static char AccessSid[] = "%W%   %G%";

in the file *access.h* and:

static char OpsysSid[] = "%W%   %G%";

in the file *opsys.h*. Otherwise, you will get compilation errors because "SccsId" is redefined. The problem with this is that if the header file is included by many modules that are loaded together, the version number of that header file is included in the object module many times; you may find it more to your taste to put id keywords in header files in comments.

5.6. Keeping SID's consistent across files

With some care, it is possible to keep the SID's consistent in multi-file systems. The trick here is to always *edit* all files at once. The changes can then be made to whatever files are necessary and then all files (even those not changed) are redeltaed. This can be done fairly easily by just specifying the name of the directory that the SCCS files are in:

sccs edit SCCS

which will *edit* all files in that directory. To make the delta, use:

sccs delta SCCS

You will be prompted for comments only once.

5.7. Creating new releases

When you want to create a new release of a program, you can specify the release number you want to create on the *edit* command. For example:

sccs edit –r2 prog.c

will cause the next delta to be in release two (that is, it will be numbered 2.1). Future deltas will automatically be in release two. To change the release number of an entire system, use:

sccs edit –r2 SCCS

6. Restoring Old Versions

6.1. Reverting to old versions

Suppose that after delta 1.2 was stable you made and released a delta 1.3. But this introduced a bug, so you made a delta 1.4 to correct it. But 1.4 was still buggy, and you decided you wanted to go back to the old version. You could revert to delta 1.2 by choosing the SID in a get:

sccs get –r1.2 prog.c

This will produce a version of *prog.c* that is delta 1.2 that can be reinstalled so that work can proceed.

In some cases you don't know what the SID of the delta you want is. However, you can revert to the version of the program that was running as of a certain date by using the **–c** (cutoff) flag. For example,

sccs get –c800722120000 prog.c

will retrieve whatever version was current as of July 22, 1980 at 12:00 noon. Trailing components can be stripped off (defaulting to their highest legal value), and punctuation can be inserted in the obvious places; for example, the above line could be equivalently stated:

sccs get –c"80/07/22 12:00:00" prog.c

6.2. Selectively deleting old deltas

Suppose that you later decided that you liked the changes in delta 1.4, but that delta 1.3 should be removed. You could do this by *excluding* delta 1.3:

```
sces edit -x1.3 prog.c
```

When delta 1.5 is made, it will include the changes made in delta 1.4, but will exclude the changes made in delta 1.3. You can exclude a range of deltas using a dash. For example, if you want to get rid of 1.3 and 1.4 you can use:

```
sccs edit -x1.3-1.4 prog.c
```

which will exclude all deltas from 1.3 to 1.4. Alternatively,

```
sccs edit -x1.3-1 prog.c
```

will exclude a range of deltas from 1.3 to the current highest delta in release 1.

In certain cases when using –x (or –i; see below) there will be conflicts between versions; for example, it may be necessary to both include and delete a particular line. If this happens, SCCS always prints out a message telling the range of lines effected; these lines should then be examined very carefully to see if the version SCCS got is ok.

Since each delta (in the sense of "a set of changes") can be excluded at will, that this makes it most useful to put each semantically distinct change into its own delta.

7. Auditing Changes

7.1. The prt command

When you created a delta, you presumably gave a reason for the delta to the "comments?" prompt. To print out these comments later, use:

```
sccs prt prog.c
```

This will produce a report for each delta of the SID, time and date of creation, user who created the delta, number of lines inserted, deleted, and unchanged, and the comments associated with the delta. For example, the output of the above command might be:

```
D 1.2   80/08/29 12:35:31   bill   2   1   00005/00003/00084
removed "-q" option
D 1.1   79/02/05 00:19:31   eric   1   0   00087/00000/00000
date and time created 80/06/10 00:19:31 by eric
```

7.2. Finding why lines were inserted

To find out why you inserted lines, you can get a copy of the file with each line preceded by the SID that created it:

```
sccs get -m prog.c
```

You can then find out what this delta did by printing the comments using *prt*.

To find out what lines are associated with a particular delta (*e.g.*, 1.3), use:

```
sccs get -m -p prog.c | grep '^1.3'
```

The **–p** flag causes SCCS to output the generated source to the standard output rather than to a file.

7.3. Finding what changes you have made

When you are editing a file, you can find out what changes you have made using:

```
sccs diffs prog.c
```

Most of the "diff" flags can be used. To pass the **–c** flag, use **–C**.

To compare two versions that are in deltas, use:

```
sccs sccsdiff -r1.3 -r1.6 prog.c
```

to see the differences between delta 1.3 and delta 1.6.

8. Shorthand Notations

There are several sequences of commands that get executed frequently. *Sccs* tries to make it easy to do these.

8.1. Delget

A frequent requirement is to make a delta of some file and then get that file. This can be done by using:

```
sccs delget prog.c
```

which is entirely equivalent to using:

```
sccs delta prog.c
sccs get prog.c
```

The "deledit" command is equivalent to "delget" except that the "edit" command is used instead of the "get" command.

8.2. Fix

Frequently, there are small bugs in deltas, e.g., compilation errors, for which there is no reason to maintain an audit trail. To *replace* a delta, use:

```
sccs fix -r1.4 prog.c
```

This will get a copy of delta 1.4 of prog.c for you to edit and then delete delta 1.4 from the SCCS file. When you do a delta of prog.c, it will be delta 1.4 again. The –r flag must be specified, and the delta that is specified must be a leaf delta, i.e., no other deltas may have been made subsequent to the creation of that delta.

8.3. Unedit

If you found you edited a file that you did not want to edit, you can back out by using:

```
sccs unedit prog.c
```

8.4. The –d flag

If you are working on a project where the SCCS code is in a directory somewhere, you may be able to simplify things by using a shell alias. For example, the alias:

```
alias syssccs sccs -d/usr/src
```

will allow you to issue commands such as:

```
syssccs edit cmd/who.c
```

which will look for the file "/usr/src/cmd/SCCS/who.c". The file "who.c" will always be created in your current directory regardless of the value of the –d flag.

9. Using SCCS on a Project

Working on a project with several people has its own set of special problems. The main problem occurs when two people modify a file at the same time. SCCS prevents this by locking an s-file while it is being edited.

As a result, files should not be reserved for editing unless they are actually being edited at the time, since this will prevent other people on the project from making necessary changes. For example, a good scenario for working might be:

```
sccs edit a.c g.c t.c
vi a.c g.c t.c
# do testing of the (experimental) version
sccs delget a.c g.c t.c
sccs info
# should respond "Nothing being edited"
make install
```

As a general rule, all source files should be deltaed before installing the program for general use. This will insure that it is possible to restore any version in use at any time.

10. Saving Yourself

10.1. Recovering a munged edit file

Sometimes you may find that you have destroyed or trashed a file that you were trying to edit[6]. Unfortunately, you can't just remove it and re-*edit* it; SCCS keeps track of the fact that someone is trying to edit it, so it won't let you do it again. Neither can you just get it using *get*, since that would expand the Id keywords. Instead, you can say:

```
sccs get -k prog.c
```

This will not expand the Id keywords, so it is safe to do a delta with it.

Alternately, you can *unedit* and *edit* the file.

10.2. Restoring the s-file

In particularly bad circumstances, the SCCS file itself may get munged. The most common way this happens is that it gets edited. Since SCCS keeps a checksum, you will get errors every time you read the file. To fix this checksum, use:

```
sccs admin -z prog.c
```

11. Using the Admin Command

There are a number of parameters that can be set using the *admin* command. The most interesting of these are flags. Flags can be added by using the **-f** flag. For example:

```
sccs admin -fd1 prog.c
```

sets the "d" flag to the value "1". This flag can be deleted by using:

```
sccs admin -dd prog.c
```

The most useful flags are:

b — Allow branches to be made using the -b flag to *edit*.

d*SID* — Default SID to be used on a *get* or *edit*. If this is just a release number it constrains the version to a particular release only.

i — Give a fatal error if there are no Id Keywords in a file. This is useful to guarantee that a version of the file does not get merged into the s-file that has the Id Keywords inserted as constants instead of internal forms.

y — The "type" of the module. Actually, the value of this flag is unused by SCCS except that it replaces the %Y% keyword.

The **-t***file* flag can be used to store descriptive text from *file*. This descriptive text might be the documentation or a design and implementation document. Using the **-t** flag insures that if the SCCS file is sent, the documentation will be sent also. If *file* is omitted, the descriptive text is deleted. To see the descriptive text, use "prt -t".

[6] Or given up and decided to start over.

The *admin* command can be used safely any number of times on files. A file need not be gotten for *admin* to work.

12. Maintaining Different Versions (Branches)

Sometimes it is convenient to maintain an experimental version of a program for an extended period while normal maintenance continues on the version in production. This can be done using a "branch." Normally deltas continue in a straight line, each depending on the delta before. Creating a branch "forks off" a version of the program.

The ability to create branches must be enabled in advance using:

```
sccs admin –fb prog.c
```

The **–fb** flag can be specified when the SCCS file is first created.

12.1. Creating a branch

To create a branch, use:

```
sccs edit –b prog.c
```

This will create a branch with (for example) SID 1.5.1.1. The deltas for this version will be numbered 1.5.1.*n*.

12.2. Getting from a branch

Deltas in a branch are normally not included when you do a get. To get these versions, you will have to say:

```
sccs get –r1.5.1 prog.c
```

12.3. Merging a branch back into the main trunk

At some point you will have finished the experiment, and if it was successful you will want to incorporate it into the release version. But in the meantime someone may have created a delta 1.6 that you don't want to lose. The commands:

```
sccs edit –i1.5.1.1–1.5.1 prog.c
sccs delta prog.c
```

will merge all of your changes into the release system. If some of the changes conflict, get will print an error; the generated result should be carefully examined before the delta is made.

12.4. A more detailed example

The following technique might be used to maintain a different version of a program. First, create a directory to contain the new version:

```
mkdir ../newxyz
cd ../newxyz
```

Edit a copy of the program on a branch:

```
sccs –d../xyz edit prog.c
```

When using the old version, be sure to use the **–b** flag to info, check, tell, and clean to avoid confusion. For example, use:

```
sccs info –b
```

when in the directory "xyz".

If you want to save a copy of the program (still on the branch) back in the s-file, you can use:

```
sccs -d../xyz deledit prog.c
```

which will do a delta on the branch and reedit it for you.

When the experiment is complete, merge it back into the s-file using delta:

```
sccs -d../xyz delta prog.c
```

At this point you must decide whether this version should be merged back into the trunk (*i.e.* the default version), which may have undergone changes. If so, it can be merged using the –i flag to *edit* as described above.

12.5. A warning

Branches should be kept to a minimum. After the first branch from the trunk, SID's are assigned rather haphazardly, and the structure gets complex fast.

13. Using SCCS with Make

SCCS and make can be made to work together with a little care. A few sample makefiles for common applications are shown.

There are a few basic entries that every makefile ought to have. These are:

a.out — (or whatever the makefile generates.) This entry regenerates whatever this makefile is supposed to regenerate. If the makefile regenerates many things, this should be called "all" and should in turn have dependencies on everything the makefile can generate.

install — Moves the objects to the final resting place, doing any special *chmod*'s or *ranlib*'s as appropriate.

sources — Creates all the source files from SCCS files.

clean — Removes all cruft from the directory.

print — Prints the contents of the directory.

The examples shown below are only partial examples, and may omit some of these entries when they are deemed to be obvious.

The *clean* entry should not remove files that can be regenerated from the SCCS files. It is sufficiently important to have the source files around at all times that the only time they should be removed is when the directory is being mothballed. To do this, the command:

```
sccs clean
```

can be used. This will remove all files for which an s-file exists, but which is not being edited.

13.1. To maintain single programs

Frequently there are directories with several largely unrelated programs (such as simple commands). These can be put into a single makefile:

```
LDFLAGS= -i -s

prog: prog.o
	$(CC) $(LDFLAGS) -o prog prog.o
prog.o: prog.c prog.h

example: example.o
	$(CC) $(LDFLAGS) -o example example.o
example.o: example.c

.DEFAULT:
	sccs get $<
```

The trick here is that the .DEFAULT rule is called every time something is needed that does not exist, and no other rule exists to make it. The explicit dependency of the .o file on the .c file is important. Another way of doing the same thing is:

```
SRCS=prog.c prog.h example.c
LDFLAGS= -i -s
prog: prog.o
	$(CC) $(LDFLAGS) -o prog prog.o
prog.o: prog.h
example: example.o
	$(CC) $(LDFLAGS) -o example example.o
sources: $(SRCS)
$(SRCS):
	sccs get $@
```

There are a couple of advantages to this approach: (1) the explicit dependencies of the .o on the .c files are not needed, (2) there is an entry called "sources" so if you want to get all the sources you can just say "make sources", and (3) the makefile is less likely to do confusing things since it won't try to *get* things that do not exist.

13.2. To maintain a library

Libraries that are largely static are best updated using explicit commands, since *make* doesn't know about updating them properly. However, libraries that are in the process of being developed can be handled quite adequately. The problem is that the .o files have to be kept out of the library as well as in the library.

```
# configuration information
OBJS=a.o b.o c.o d.o
SRCS=a.c b.c c.c d.s x.h y.h z.h
TARG=	/usr/lib
# programs
GET= sccs get
REL=
AR= -ar
RANLIB=	ranlib
lib.a: $(OBJS)
	$(AR) rvu lib.a $(OBJS)
	$(RANLIB) lib.a
install: lib.a
	sccs check
	cp lib.a $(TARG)/lib.a
	$(RANLIB) $(TARG)/lib.a
sources: $(SRCS)
$(SRCS):
	$(GET) $(REL) $@
print: sources
	pr *.h *.[cs]
clean:
	rm -f *.o
	rm -f core a.out $(LIB)
```

The "$(REL)" in the get can be used to get old versions easily; for example:

```
make b.o REL=-r1.3
```

The *install* entry includes the line "sccs check" before anything else. This guarantees that all the s-files are up to date (*i.e.*, nothing is being edited), and will abort the *make* if this condition is not met.

13.3. To maintain a large program

```
OBJS=a.o b.o c.o d.o
SRCS=a.c b.c c.y d.s x.h y.h z.h

GET= sccs get
REL=

a.out: $(OBJS)
        $(CC) $(LDFLAGS) $(OBJS) $(LIBS)

sources: $(SRCS)
$(SRCS):
        $(GET) $(REL) $@
```

(The *print* and *clean* entries are identical to the previous case.) This makefile requires copies of the source and object files to be kept during development. It is probably also wise to include lines of the form:

```
a.o: x.h y.h
b.o: z.h
c.o: x.h y.h z.h
z.h: x.h
```

so that modules will be recompiled if header files change.

Since *make* does not do transitive closure on dependencies, you may find in some makefiles lines like:

```
z.h: x.h
        touch z.h
```

This would be used in cases where file z.h has a line:

```
#include "x.h"
```

in order to bring the mod date of z.h in line with the mod date of x.h. When you have a makefile such as above, the *touch* command can be removed completely; the equivalent effect will be achieved by doing an automatic *get* on z.h.

14. Further Information

The *SCCS/PWB User's Manual* gives a deeper description of how to use SCCS. Of particular interest are the numbering of branches, the l-file, which gives a description of what deltas were used on a get, and certain other SCCS commands.

The SCCS manual pages are a good last resort. These should be read by software managers and by people who want to know everything about everything.

Both of these documents were written without the *sccs* front end in mind, so most of the examples are slightly different from those in this document.

Quick Reference

1. Commands

The following commands should all be preceded with "sccs". This list is not exhaustive; for more options see *Further Information*.

get Gets files for compilation (not for editing). Id keywords are expanded.

- –r*SID* Version to get.
- –p Send to standard output rather than to the actual file.
- –k Don't expand id keywords.
- –i*list* List of deltas to include.
- –x*list* List of deltas to exclude.
- –m Precede each line with SID of creating delta.
- –c*date* Don't apply any deltas created after *date*.

edit Gets files for editing. Id keywords are not expanded. Should be matched with a *delta* command.

- –r*SID* Same as *get*. If *SID* specifies a release that does not yet exist, the highest numbered delta is retrieved and the new delta is numbered with *SID*.
- –b Create a branch.
- –i*list* Same as *get*.
- –x*list* Same as *get*.

delta Merge a file gotten using *edit* back into the s-file. Collect comments about why this delta was made.

unedit Remove a file that has been edited previously without merging the changes into the s-file.

prt Produce a report of changes.

- –t Print the descriptive text.
- –e Print (nearly) everything.

info Give a list of all files being edited.

- –b Ignore branches.
- –u[*user*] Ignore files not being edited by *user*.

check Same as *info*, except that nothing is printed if nothing is being edited and exit status is returned.

tell Same as *info*, except that one line is produced per file being edited containing only the file name.

clean Remove all files that can be regenerated from the s-file.

what Find and print id keywords.

admin Create or set parameters on s-files.

- –i*file* Create, using *file* as the initial contents.
- –z Rebuild the checksum in case the file has been trashed.
- –f*flag* Turn on the *flag*.

–d*flag*	Turn off (delete) the *flag*.
–t*file*	Replace the descriptive text in the s-file with the contents of *file*. If *file* is omitted, the text is deleted. Useful for storing documentation or "design & implementation" documents to insure they get distributed with the s-file.

Useful flags are:

b	Allow branches to be made using the –b flag to *edit*.
d*SID*	Default SID to be used on a *get* or *edit*.
i	Cause "No Id Keywords" error message to be a fatal error rather than a warning.
t	The module "type"; the value of this flag replaces the %Y% keyword.

fix	Remove a delta and reedit it.
delget	Do a *delta* followed by a *get*.
deledit	Do a *delta* followed by an *edit*.

2. Id Keywords

%Z% Expands to "@(#)" for the *what* command to find.

%M% The current module name, *e.g.*, "prog.c".

%I% The highest SID applied.

%W% A shorthand for "%Z%%M% <tab> %I%".

%G% The date of the delta corresponding to the "%I%" keyword.

%R% The current release number, *i.e.*, the first component of the "%I%" keyword.

%Y% Replaced by the value of the t flag (set by *admin*).

A Guide to the Dungeons of Doom

Michael C. Toy
Kenneth C. R. C. Arnold

Computer Systems Research Group
Department of Electrical Engineering and Computer Science
University of California
Berkeley, California 94720

ABSTRACT

Rogue is a visual CRT based fantasy game which runs under the UNIX† timesharing system. This paper describes how to play rogue, and gives a few hints for those who might otherwise get lost in the Dungeons of Doom.

†UNIX is a trademark of Bell Laboratories

1. Introduction

You have just finished your years as a student at the local fighter's guild. After much practice and sweat you have finally completed your training and are ready to embark upon a perilous adventure. As a test of your skills, the local guildmasters have sent you into the Dungeons of Doom. Your task is to return with the Amulet of Yendor. Your reward for the completion of this task will be a full membership in the local guild. In addition, you are allowed to keep all the loot you bring back from the dungeons.

In preparation for your journey, you are given an enchanted mace, a bow, and a quiver of arrows taken from a dragon's hoard in the far off Dark Mountains. You are also outfitted with elf-crafted armor and given enough food to reach the dungeons. You say goodbye to family and friends for what may be the last time and head up the road.

You set out on your way to the dungeons and after several days of uneventful travel, you see the ancient ruins that mark the entrance to the Dungeons of Doom. It is late at night, so you make camp at the entrance and spend the night sleeping under the open skies. In the morning you gather your weapons, put on your armor, eat what is almost your last food, and enter the dungeons.

2. What is going on here?

You have just begun a game of rogue. Your goal is to grab as much treasure as you can, find the Amulet of Yendor, and get out of the Dungeons of Doom alive. On the screen, a map of where you have been and what you have seen on the current dungeon level is kept. As you explore more of the level, it appears on the screen in front of you.

Rogue differs from most computer fantasy games in that it is screen oriented. Commands are all one or two keystrokes[1] and the results of your commands are displayed graphically on the screen rather than being explained in words.[2]

Another major difference between rogue and other computer fantasy games is that once you have solved all the puzzles in a standard fantasy game, it has lost most of its excitement and it ceases to be fun. Rogue, on the other hand, generates a new dungeon every time you play it and even the author finds it an entertaining and exciting game.

3. What do all those things on the screen mean?

In order to understand what is going on in rogue you have to first get some grasp of what rogue is doing with the screen. The rogue screen is intended to replace the "You can see ..." descriptions of standard fantasy games. Figure 1 is a sample of what a rogue screen might look like.

3.1. The bottom line

At the bottom line of the screen are a few pieces of cryptic information describing your current status. Here is an explanation of what these things mean:

Level This number indicates how deep you have gone in the dungeon. It starts at one and goes up as you go deeper into the dungeon.

Gold The number of gold pieces you have managed to find and keep with you so far.

Hp Your current and maximum hit points. Hit points indicate how much damage you can take before you die. The more you get hit in a fight, the lower they get. You can regain hit points by resting. The number in parentheses is the maximum number your hit points can reach.

[1] As opposed to pseudo English sentences.

[2] A minimum screen size of 24 lines by 80 columns is required. If the screen is larger, only the 24x80 section will be used for the map.

```
 - - - - - - - - - - - -
|. . . . . . . . . . . +
|. . @ . . . . ] . . . |
|. . . . B . . . . . . |
|. . . . . . . . . . . |
 - - - - - + - - - - - -
```

Level: 1 Gold: 0   Hp: 12(12) Str: 16(16) Ac: 6 Exp: 1/0

Figure 1

Str — Your current strength and maximum ever strength. This can be any integer less than or equal to 31, or greater than or equal to three. The higher the number, the stronger you are. The number in the parentheses is the maximum strength you have attained so far this game.

Ac — Your current armor class. This number indicates how effective your armor is in stopping blows from unfriendly creatures. The lower this number is, the more effective the armor.

Exp — These two numbers give your current experience level and experience points. As you do things, you gain experience points. At certain experience point totals, you gain an experience level. The more experienced you are, the better you are able to fight and to withstand magical attacks.

3.2. The top line

The top line of the screen is reserved for printing messages that describe things that are impossible to represent visually. If you see a "--More--" on the top line, this means that rogue wants to print another message on the screen, but it wants to make certain that you have read the one that is there first. To read the next message, just type a space.

3.3. The rest of the screen

The rest of the screen is the map of the level as you have explored it so far. Each symbol on the screen represents something. Here is a list of what the various symbols mean:

@ — This symbol represents you, the adventurer.

-| — These symbols represent the walls of rooms.

+ — A door to/from a room.

. — The floor of a room.

— The floor of a passage between rooms.

* — A pile or pot of gold.

) — A weapon of some sort.

] — A piece of armor.

! — A flask containing a magic potion.

? — A piece of paper, usually a magic scroll.

= — A ring with magic properties

/ A magical staff or wand

^ A trap, watch out for these.

% A staircase to other levels

: A piece of food.

A-Z The uppercase letters represent the various inhabitants of the Dungeons of Doom. Watch out, they can be nasty and vicious.

4. Commands

Commands are given to rogue by typing one or two characters. Most commands can be preceded by a count to repeat them (e.g. typing "10s" will do ten searches). Commands for which counts make no sense have the count ignored. To cancel a count or a prefix, type <ESCAPE>. The list of commands is rather long, but it can be read at any time during the game with the "?" command. Here it is for reference, with a short explanation of each command.

? The help command. Asks for a character to give help on. If you type a "*", it will list all the commands, otherwise it will explain what the character you typed does.

/ This is the "What is that on the screen?" command. A "/" followed by any character that you see on the level, will tell you what that character is. For instance, typing "/@" will tell you that the "@" symbol represents you, the player.

h, H, ^H
Move left. You move one space to the left. If you use upper case "h", you will continue to move left until you run into something. This works for all movement commands (e.g. "L" means run in direction "l") If you use the "control" "h", you will continue moving in the specified direction until you pass something interesting or run into a wall. You should experiment with this, since it is a very useful command, but very difficult to describe. This also works for all movement commands.

j Move down.

k Move up.

l Move right.

y Move diagonally up and left.

u Move diagonally up and right.

b Move diagonally down and left.

n Move diagonally down and right.

t Throw an object. This is a prefix command. When followed with a direction it throws an object in the specified direction. (e.g. type "th" to throw something to the left.)

f Fight until someone dies. When followed with a direction this will force you to fight the creature in that direction until either you or it bites the big one.

m Move onto something without picking it up. This will move you one space in the direction you specify and, if there is an object there you can pick up, it won't do it.

z Zap prefix. Point a staff or wand in a given direction and fire it. Even non-directional staves must be pointed in some direction to be used.

^ Identify trap command. If a trap is on your map and you can't remember what type it is, you can get rogue to remind you by getting next to it and typing "^" followed by the direction that would move you on top of it.

s Search for traps and secret doors. Examine each space immediately adjacent to you for the existence of a trap or secret door. There is a large chance that even if there is something there, you won't find it, so you might have to search a while before you find something.

> Climb down a staircase to the next level. Not surprisingly, this can only be done if you are standing on staircase.

< Climb up a staircase to the level above. This can't be done without the Amulet of Yendor in your possession.

. Rest. This is the "do nothing" command. This is good for waiting and healing.

* Inventory. List what you are carrying in your pack.

I Selective inventory. Tells you what a single item in your pack is.

q Quaff one of the potions you are carrying.

r Read one of the scrolls in your pack.

e Eat food from your pack.

w Wield a weapon. Take a weapon out of your pack and carry it for use in combat, replacing the one you are currently using (if any).

W Wear armor. You can only wear one suit of armor at a time. This takes extra time.

T Take armor off. You can't remove armor that is cursed. This takes extra time.

P Put on a ring. You can wear only two rings at a time (one on each hand). If you aren't wearing any rings, this command will ask you which hand you want to wear it on, otherwise, it will place it on the unused hand. The program assumes that you wield your sword in your right hand.

R Remove a ring. If you are only wearing one ring, this command takes it off. If you are wearing two, it will ask you which one you wish to remove,

d Drop an object. Take something out of your pack and leave it lying on the floor. Only one object can occupy each space. You cannot drop a cursed object at all if you are wielding or wearing it.

c Call an object something. If you have a type of object in your pack which you wish to remember something about, you can use the call command to give a name to that type of object. This is usually used when you figure out what a potion, scroll, ring, or staff is after you pick it up, or when you want to remember which of those swords in your pack you were wielding.

D Print out which things you've discovered something about. This command will ask you what type of thing you are interested in. If you type the character for a given type of object (*e.g.* "!" for potion) it will tell you which kinds of that type of object you've discovered (*i.e.*, figured out what they are). This command works for potions, scrolls, rings, and staves and wands.

o Examine and set options. This command is further explained in the section on options.

^R Redraws the screen. Useful if spurious messages or transmission errors have messed up the display.

^P Print last message. Useful when a message disappears before you can read it. This only repeats the last message that was not a mistyped command so that you don't loose anything by accidentally typing the wrong character instead of ^P.

<ESCAPE>
Cancel a command, prefix, or count.

! Escape to a shell for some commands.

Q Quit. Leave the game.

S Save the current game in a file. It will ask you whether you wish to use the default save file. *Caveat*: Rogue won't let you start up a copy of a saved game, and it removes the save file as soon as you start up a restored game. This is to prevent people from saving a game just before a dangerous position and then restarting it if they die. To restore a saved game, give the file name as an argument to rogue. As in

% rogue *save_file*

To restart from the default save file (see below), run

% rogue -r

v Prints the program version number.

) Print the weapon you are currently wielding

] Print the armor you are currently wearing

= Print the rings you are currently wearing

@ Reprint the status line on the message line

5. Rooms

Rooms in the dungeons are either lit or dark. If you walk into a lit room, the entire room will be drawn on the screen as soon as you enter. If you walk into a dark room, it will only be displayed as you explore it. Upon leaving a room, all monsters inside the room are erased from the screen. In the darkness you can only see one space in all directions around you. A corridor is always dark.

6. Fighting

If you see a monster and you wish to fight it, just attempt to run into it. Many times a monster you find will mind its own business unless you attack it. It is often the case that discretion is the better part of valor.

7. Objects you can find

When you find something in the dungeon, it is common to want to pick the object up. This is accomplished in rogue by walking over the object (unless you use the "m" prefix, see above). If you are carrying too many things, the program will tell you and it won't pick up the object, otherwise it will add it to your pack and tell you what you just picked up.

Many of the commands that operate on objects must prompt you to find out which object you want to use. If you change your mind and don't want to do that command after all, just type an <ESCAPE> and the command will be aborted.

Some objects, like armor and weapons, are easily differentiated. Others, like scrolls and potions, are given labels which vary according to type. During a game, any two of the same kind of object with the same label are the same type. However, the labels will vary from game to game.

When you use one of these labeled objects, if its effect is obvious, rogue will remember what it is for you. If it's effect isn't extremely obvious you will be asked what you want to scribble on it so you will recognize it later, or you can use the "call" command (see above).

7.1. Weapons

Some weapons, like arrows, come in bunches, but most come one at a time. In order to use a weapon, you must wield it. To fire an arrow out of a bow, you must first wield the bow, then throw the arrow. You can only wield one weapon at a time, but you can't change weapons if the one you are currently wielding is cursed. The commands to use weapons are "w" (wield) and "t" (throw).

7.2. Armor

There are various sorts of armor lying around in the dungeon. Some of it is enchanted, some is cursed, and some is just normal. Different armor types have different armor classes. The lower the armor class, the more protection the armor affords against the blows of monsters. Here is a list of the various armor types and their normal armor class:

Type	*Class*
None	10
Leather armor	8
Studded leather / Ring mail	7
Scale mail	6
Chain mail	5
Banded mail / Splint mail	4
Plate mail	3

If a piece of armor is enchanted, its armor class will be lower than normal. If a suit of armor is cursed, its armor class will be higher, and you will not be able to remove it. However, not all armor with a class that is higher than normal is cursed.

The commands to use weapons are "W" (wear) and "T" (take off).

7.3. Scrolls

Scrolls come with titles in an unknown tongue[3]. After you read a scroll, it disappears from your pack. The command to use a scroll is "r" (read).

7.4. Potions

Potions are labeled by the color of the liquid inside the flask. They disappear after being quaffed. The command to use a scroll is "q" (quaff).

7.5. Staves and Wands

Staves and wands do the same kinds of things. Staves are identified by a type of wood; wands by a type of metal or bone. They are generally things you want to do to something over a long distance, so you must point them at what you wish to affect to use them. Some staves are not affected by the direction they are pointed, though. Staves come with multiple magic charges, the number being random, and when they are used up, the staff is just a piece of wood or metal.

The command to use a wand or staff is "z" (zap)

7.6. Rings

Rings are very useful items, since they are relatively permanent magic, unlike the usually fleeting effects of potions, scrolls, and staves. Of course, the bad rings are also more powerful. Most rings also cause you to use up food more rapidly, the rate varying with the type of ring. Rings are differentiated by their stone settings. The commands to use rings are "P" (put on) and "R" (remove).

7.7. Food

Food is necessary to keep you going. If you go too long without eating you will faint, and eventually die of starvation. The command to use food is "e" (eat).

8. Options

Due to variations in personal tastes and conceptions of the way rogue should do things, there are a set of options you can set that cause rogue to behave in various different ways.

8.1. Setting the options

There are two ways to set the options. The first is with the "o" command of rogue; the

[3] Actually, it's a dialect spoken only by the twenty-seven members of a tribe in Outer Mongolia, but you're not supposed to *know* that.

second is with the "ROGUEOPTS" environment variable[4].

8.1.1. Using the 'o' command

When you type "o" in rogue, it clears the screen and displays the current settings for all the options. It then places the cursor by the value of the first option and waits for you to type. You can type a <RETURN> which means to go to the next option, a "–" which means to go to the previous option, an <ESCAPE> which means to return to the game, or you can give the option a value. For boolean options this merely involves typing "t" for true or "f" for false. For string options, type the new value followed by a <RETURN>.

8.1.2. Using the ROGUEOPTS variable

The ROGUEOPTS variable is a string containing a comma separated list of initial values for the various options. Boolean variables can be turned on by listing their name or turned off by putting a "no" in front of the name. Thus to set up an environment variable so that **jump** is on, **terse** is off, and the **name** is set to "Blue Meanie", use the command

% setenv ROGUEOPTS "jump,noterse,name=Blue Meanie"[5]

8.2. Option list

Here is a list of the options and an explanation of what each one is for. The default value for each is enclosed in square brackets. For character string options, input over fifty characters will be ignored.

terse [*noterse*]
: Useful for those who are tired of the sometimes lengthy messages of rogue. This is a useful option for playing on slow terminals, so this option defaults to *terse* if you are on a slow (1200 baud or under) terminal.

jump [*nojump*]
: If this option is set, running moves will not be displayed until you reach the end of the move. This saves considerable cpu and display time. This option defaults to *jump* if you are using a slow terminal.

flush [*noflush*]
: All typeahead is thrown away after each round of battle. This is useful for those who type far ahead and then watch in dismay as a Bat kills them.

seefloor [*seefloor*]
: Display the floor around you on the screen as you move through dark rooms. Due to the amount of characters generated, this option defaults to *noseefloor* if you are using a slow terminal.

passgo [*nopassgo*]
: Follow turnings in passageways. If you run in a passage and you run into stone or a wall, rogue will see if it can turn to the right or left. If it can only turn one way, it will turn that way. If it can turn either or neither, it will stop. This is followed strictly, which can sometimes lead to slightly confusing occurrences (which is why it defaults to *nopassgo*).

tombstone [*tombstone*]
: Print out the tombstone at the end if you get killed. This is nice but slow, so you can turn it off if you like.

inven [*overwrite*]
: Inventory type. This can have one of three values: *overwrite*, *slow*, or *clear*. With *overwrite*

[4] On Version 6 systems, there is no equivalent of the ROGUEOPTS feature.

[5] For those of you who use the bourne shell, the commands would be
$ ROGUEOPTS="jump,noterse,name=Blue Meanie"
$ export ROGUEOPTS

the top lines of the map are overwritten with the list when inventory is requested or when "Which item do you wish to . . .? " questions are answered with a "*". However, if the list is longer than a screenful, the screen is cleared. With *slow*, lists are displayed one item at a time on the top of the screen, and with *clear*, the screen is cleared, the list is displayed, and then the dungeon level is re-displayed. Due to speed considerations, *clear* is the default for terminals without clear-to-end-of-line capabilities.

name [account name]
This is the name of your character. It is used if you get on the top ten scorer's list.

fruit [*slime-mold*]
This should hold the name of a fruit that you enjoy eating. It is basically a whimsey that rogue uses in a couple of places.

file [*~/rogue.save*]
The default file name for saving the game. If your phone is hung up by accident, rogue will automatically save the game in this file. The file name may start with the special character "~" which expands to be your home directory.

9. Scoring

Rogue usually maintains a list of the top scoring people or scores on your machine. Depending on how it is set up, it can post either the top scores or the top players. In the latter case, each account on the machine can post only one non-winning score on this list. If you score higher than someone else on this list, or better your previous score on the list, you will be inserted in the proper place under your current name. How many scores are kept can also be set up by whoever installs it on your machine.

If you quit the game, you get out with all of your gold intact. If, however, you get killed in the Dungeons of Doom, your body is forwarded to your next-of-kin, along with 90% of your gold; ten percent of your gold is kept by the Dungeons' wizard as a fee[6]. This should make you consider whether you want to take one last hit at that monster and possibly live, or quit and thus stop with whatever you have. If you quit, you do get all your gold, but if you swing and live, you might find more.

If you just want to see what the current top players/games list is, you can type

% rogue -s

10.
Acknowledgements

Rogue was originally conceived of by Glenn Wichman and Michael Toy. Ken Arnold and Michael Toy then smoothed out the user interface, and added jillions of new features. We would like to thank Bob Arnold, Michelle Busch, Andy Hatcher, Kipp Hickman, Mark Horton, Daniel Jensen, Bill Joy, Joe Kalash, Steve Maurer, Marty McNary, Jan Miller, and Scott Nelson for their ideas and assistance; and also the teeming multitudes who graciously ignored work, school, and social life to play rogue and send us bugs, complaints, suggestions, and just plain flames. And also Mom.

[6] The Dungeon's wizard is named Wally the Wonder Badger. Invocations should be accompanied by a sizable donative.

STAR

TREK

by

Eric Allman
University of California
Berkeley

INTRODUCTION

Well, the federation is once again at war with the Klingon empire. It is up to you, as captain of the U.S.S. Enterprise, to wipe out the invasion fleet and save the Federation.

For the purposes of the game the galaxy is divided into 64 quadrants on an eight by eight grid, with quadrant 0,0 in the upper left hand corner. Each quadrant is divided into 100 sectors on a ten by ten grid. Each sector contains one object (e.g., the Enterprise, a Klingon, or a star).

Navigation is handled in degrees, with zero being straight up and ninty being to the right. Distances are measured in quadrants. One tenth quadrant is one sector.

The galaxy contains starbases, at which you can dock to refuel, repair damages, etc. The galaxy also contains stars. Stars usually have a knack for getting in your way, but they can be triggered into going nova by shooting a photon torpedo at one, thereby (hopefully) destroying any adjacent Klingons. This is not a good practice however, because you are penalized for destroying stars. Also, a star will sometimes go supernova, which obliterates an entire quadrant. You must never stop in a supernova quadrant, although you may "jump over" one.

Some starsystems have inhabited planets. Klingons can attack inhabited planets and enslave the populace, which they then put to work building more Klingon battle cruisers.

STARTING UP THE GAME

To request the game, issue the command

/usr/games/trek

from the shell. If a filename is stated, a log of the game is written onto that file. If omitted, the file is not written. If the "**-a**" flag is stated before the filename, thata file is appended to rather than created.

The game will ask you what length game you would like. Valid responses are "short", "medium", and "long". You may also type "restart", which restarts a previously saved game. Ideally, the length of the game does not affect the difficulty, but currently the shorter games tend to be harder than the longer ones.

You will then be prompted for the skill, to which you must respond "novice", "fair", "good", "expert", "commadore", or "impossible". You should start out with a novice and work up, but if you really want to see how fast you can be slaughtered, start out with an impossible game.

In general, throughout the game, if you forget what is appropriate the game will tell you what it expects if you just type in a question mark.

ISSUING COMMANDS

If the game expects you to enter a command, it will say "Command: " and wait for your response. Most commands can be abbreviated.

At almost any time you can type more than one thing on a line. For example, to move straight up one quadrant, you can type

move 0 1

or you could just type

move

and the game would prompt you with

Course:

to which you could type

0 1

The "1" is the distance, which could be put on still another line. Also, the "move" command could have been abbreviated "mov", "mo", or just "m".

If you are partway through a command and you change your mind, you can usually type "-1" to cancel the command.

Klingons generally cannot hit you if you don't consume anything (e.g., time or energy), so some commands are considered "free". As soon as you consume anything though -- POW!

THE COMMANDS

Short Range Scan

Mnemonic: srscan
Shortest Appreviation: s
Full Commands: srscan
srscan yes/no
Consumes: nothing

The short range scan gives you a picture of the quadrant you are in, and (if you say "yes") a status report which tells you a whole bunch of interesting stuff. You can get a status report alone by using the *status* command. An example follows:

```
Short range sensor scan
    0 1 2 3 4 5 6 7 8 9
0   . . . . . . . * . *  0    stardate        3702.16
1   . . E . . . . . . .  1    condition       RED
2   . . . . . . . . . *  2    position        0,3/1,2
3   * . . . . # . . . .  3    warp factor     5.0
4   . . . . . . . . . .  4    total energy    4376
5   . . * . * . . . . .  5    torpedoes       9
6   . . . @ . .   . . .  6    shields         down, 78%
7   . . . . . . . . . .  7    Klingons left   3
8   . . . K . . . . . .  8    time left       6.43
9   . . . . . . * . . .  9    life support    damaged, reserves = 2.4
    0 1 2 3 4 5 6 7 8 9

Distressed Starsystem Marcus XII
```

The cast of characters is as follows:

E	the hero
K	the villain
#	the starbase
*	stars
@	inhabited starsystem
.	empty space
	a black hole

The name of the starsystem is listed underneath the short range scan. The word "distressed", if present, means that the starsystem is under attack.

Short range scans are absolutely free. They use no time, no energy, and they don't give the Klingons another chance to hit you.

Status Report

Mnemonic: status
Shortest Abbreviation: st
Consumes: nothing

This command gives you information about the current status of the game and your ship, as follows:

Stardate -- The current stardate.

Condition -- as follows:
RED -- in battle
YELLOW -- low on energy
GREEN -- normal state
DOCKED -- docked at starbase
CLOAKED -- the cloaking device is activated

Position -- Your current quadrant and sector.

Warp Factor -- The speed you will move at when you move under warp power (with the *move* command).

Total Energy -- Your energy reserves. If they drop to zero, you die. Energy regenerates, but the higher the skill of the game, the slower it regenerates.

Torpedoes -- How many photon torpedoes you have left.

Shields -- Whether your shields are up or down, and how effective they are if up (what percentage of a hit they will absorb).

Klingons Left -- Guess.

Time Left -- How long the Federation can hold out if you sit on your fat ass and do nothing. If you kill Klingons quickly, this number goes up, otherwise, it goes down. If it hits zero, the Federation is conquered.

Life Support -- If "active", everything is fine. If "damaged", your reserves tell you how long you have to repair your life support or get to a starbase before you starve, suffocate, or something equally unpleasant.

Current Crew -- The number of crew members left. This figures does not include officers.

Brig Space -- The space left in your brig for Klingon captives.

Klingon Power -- The number of units needed to kill a Klingon. Remember, as Klingons fire at you they use up their own energy, so you probably need somewhat less than this.

Skill, Length -- The skill and length of the game you are playing.

Status information is absolutely free.

Long Range Scan

Mnemonic: lrscan
Shortest Abbreviation: l
Consumes: nothing

Long range scan gives you information about the eight quadrants that surround the quadrant you're in. A sample long range scan follows:

```
Long range scan for quadrant 0,3

     2     3     4
  -----------------
  !  *  !  *  !  *  !
  -----------------
0 ! 108 !   6 !  19 !
  -----------------
1 !   9 ! /// !   8 !
  -----------------
```

The three digit numbers tell the number of objects in the quadrants. The units digit tells the number of stars, the tens digit the number of starbases, and the hundreds digit is the number of Klingons. "*" indicates the negative energy barrier at the edge of the galaxy, which you cannot enter. "///" means that that is a supernova quadrant and must not be entered.

Damage Report

Mnemonic: damages
Shortest Abbreviation: da
Consumes: nothing

A damage report tells you what devices are damaged and how long it will take to repair them. Repairs proceed faster when you are docked at a starbase.

Set Warp Factor

Mnemonic: warp
Shortest Abbreviation: w
Full Command: warp factor
Consumes: nothing

The warp factor tells the speed of your starship when you move under warp power (with the *move* command). The higher the warp factor, the faster you go, and the more energy you use.

The minimum warp factor is 1.0 and the maximum is 10.0. At speeds above warp 6 there is danger of the warp engines being damaged. The probability of this increases at higher warp speeds. Above warp 9.0 there is a chance of entering a time warp.

Move Under Warp Power

Mnemonic: move
Shortest Abbreviation: m
Full Command: move course distance
Consumes: time and energy

This is the usual way of moving. The course is in degrees and the distance is in quadrants. To move one sector specify a distance of 0.1.

Time is consumed proportionately to the inverse of the warp factor squared, and directly to the distance. Energy is consumed as the warp factor cubed, and directly to the distance. If you move with your shields up it doubles the amount of energy consumed.

When you move in a quadrant containing Klingons, they get a chance to attack you.

The computer detects navigation errors. If the computer is out, you run the risk of running into things.

The course is determined by the Space Inertial Navigation System [SINS]. As described in Star Fleet Technical Order TO:02:06:12, the SINS is calibrated, after which it becomes the base for navigation. If damaged, navigation becomes inaccurate. When it is fixed, Spock recalibrates it, however, it cannot be calibrated extremely accurately until you dock at starbase.

Move Under Impulse Power

Mnemonic: impulse
Shortest Abbreviation: i
Full Command: impulse course distance
Consumes: time and energy

The impulse engines give you a chance to maneuver when your warp engines are damaged; however, they are incredibly slow (0.095 quadrants/stardate). They require 20 units of energy to engage, and ten units per sector to move.

The same comments about the computer and the SINS apply as above.

There is no penalty to move under impulse power with shields up.

Deflector Shields

Mnemonic: shields
Shortest Abbreviation: sh
Full Command: shields up/down
Consumes: energy

Shields protect you from Klingon attack and nearby novas. As they protect you, they weaken. A shield which is 78% effective will absorb 78% of a hit and let 22% in to hurt you.

The Klingons have a chance to attack you every time you raise or lower shields. Shields do not rise and lower instantaneously, so the hit you receive will be computed with the shields at an intermediate effectiveness.

It takes energy to raise shields, but not to drop them.

Cloaking Device

Mnemonic: cloak
Shortest Abbreviation: cl
Full Command: cloak up/down
Consumes: energy

When you are cloaked, Klingons cannot see you, and hence they do not fire at you. They are useful for entering a quadrant and selecting a good position, however, weapons cannot be fired through the cloak due to the huge energy drain that it requires.

The cloak up command only starts the cloaking process; Klingons will continue to fire at you until you do something which consumes time.

Fire Phasers

Mnmemonic: phasers
Shortest Abbreviation: p
Full Commands: phasers automatic amount
phasers manual amt1 course1 spread1 ...
Consumes: energy

Phasers are energy weapons; the energy comes from your ship's reserves ("total energy" on a srscan). It takes about 250 units of hits to kill a Klingon. Hits are cumulative as long as you stay in the quadrant.

Phasers become less effective the further from a Klingon you are. Adjacent Klingons receive about 90% of what you fire, at five sectors about 60%, and at ten sectors about 35%. They have no effect outside of the quadrant.

Phasers cannot be fired while shields are up; to do so would fry you. They have no effect on starbases or stars.

In automatic mode the computer decides how to divide up the energy among the Klingons present; in manual mode you do that yourself.

In manual mode firing you specify a direction, amount (number of units to fire) and spread (0 -> 1.0) for each of the six phaser banks. A zero amount terminates the manual input.

Fire Photon Torpedoes

Mnemonic: torpedo
Shortest Abbreviation: t
Full Command: torpedo course [yes/no] [burst angle]
Consumes: torpedoes

Torpedoes are projectile weapons -- there are no partial hits. You either hit your target or you don't. A hit on a Klingon destroys him. A hit on a starbase destroys that starbase (woops!). Hitting a star usually causes it to go nova, and occasionally supernova.

Photon torpedoes cannot be aimed precisely. They can be fired with shields up, but they get even more random as they pass through the shields.

Torpedoes may be fired in bursts of three. If this is desired, the burst angle is the angle between the three shots, which may vary from one to fifteen. The word "no" says that a burst is not wanted; the word "yes" (which may be omitted if stated on the same line as the course) says that a burst is wanted.

Photon torpedoes have no effect outside the quadrant.

Onboard Computer Request

Mnemonic: computer
Shortest Abbreviation: c
Full Command: computer request; request;...
Consumes: nothing

The computer command gives you access to the facilities of the onboard computer, which allows you to do all sorts of fascinating stuff. Computer requests are:

score -- Shows your current score.

course quad/sect -- Computes the course and distance from whereever you are to the given location. If you type "course /x,y" you will be given the course to sector x,y in the current quadrant.

move quad/sect -- Identical to the course request, except that the move is executed.

chart -- prints a chart of the known galaxy, i.e., everything that you have seen with a long range scan. The format is the same as on a long range scan, except that "..." means that you don't yet know what is there, and ".1." means that you know that a starbase exists, but you don't know anything else. "$$$" mans the quadrant that you are currently in.

trajectory -- prints the course and distance to all the Klingons in the quadrant.

warpcost dist warp_factor -- computes the cost in time and energy to move 'dist' quadrants at warp 'warp_factor'.

impcost dist -- same as warpcost for impulse engines.

pheff range -- tells how effective your phasers are at a given range.

distresslist -- gives a list of currently distressed starbases and starsystems.

More than one request may be stated on a line by seperating them with semicolons.

Dock at Starbase

Mnemonic: dock
Shortest Abbreviation: do

Consumes: nothing

You may dock at a starbase when you are in one of the eight adjacent sectors.

When you dock you are resupplied with energy, photon torpedoes, and life support reserves. Repairs are also done faster at starbase. Any prisoners you have taken are unloaded. You do not recieve points for taking prisoners until this time.

Starbases have their own deflector shields, so you are safe from attack while docked.

Undock from Starbase

Mnemonic: undock
Shortest Abbreviation: u
Consumes: nothing

This just allows you to leave starbase so that you may proceed on your way.

Rest

Mnemonic: rest
Shortest Abbreviation: r
Full Command: rest time
Consumes: time

This command allows you to rest to repair damages. It is not advisable to rest while under attack.

Call Starbase For Help

Mnemonic: help
Shortest Abbreviation: help
Consumes: nothing

You may call starbase for help via your subspace radio. Starbase has long range transporter beams to get you. Problem is, they can't always rematerialize you.

You should avoid using this command unless absolutely necessary, for the above reason and because it counts heavily against you in the scoring.

Capture Klingon

Mnemonic: capture
Shortest Abbreviation: ca
Consumes: time

You may request that a Klingon surrender to you. If he accepts, you get to take captives (but only as many as your brig can hold). It is good if you do this, because you get points for captives. Also, if you ever get captured, you want to be sure that the Federation has prisoners to exchange for you.

You must go to a starbase to turn over your prisoners to Federation authorities.

Visual Scan

Mnemonic: visual
Shortest Abbreviation: v
Full Command: visual course
Consumes: time

When your short range scanners are out, you can still see what is out "there" by doing a visual scan. Unfortunately, you can only see three sectors at one time, and it takes 0.005 stardates to perform.

The three sectors in the general direction of the course specified are examined and displayed.

Abandon Ship

Mnemonic: abandon
Shortest Abbreviation: abandon
Consumes: nothing

The officers escape the Enterprise in the shuttlecraft. If the transporter is working and there is an inhabitable starsystem in the area, the crew beams down, otherwise you leave them to die. You are given an old but still usable ship, the Faire Queene.

Ram

Mnemonic: ram
Shortest Abbreviation: ram
Full Command: ram course distance
Consumes: time and energy

This command is identical to "move", except that the computer doesn't stop you from making navigation errors.

You get very nearly slaughtered if you ram anything.

Self Destruct

Mnemonic: destruct
Shortest Abbreviation: destruct
Consumes: everything

Your starship is self-destructed. Chances are you will destroy any Klingons (and stars, and starbases) left in your quadrant.

Terminate the Game

Mnemonic: terminate
Shortest Abbreviation: terminate
Full Command: terminate yes/no

Cancels the current game. No score is computed. If you answer yes, a new game will be started, otherwise trek exits.

Call the Shell

Mnemonic: shell
Shortest Abbreviation: shell

Temporarily escapes to the shell. When you log out of the shell you will return to the game.

SCORING

The scoring algorithm is rather complicated. Basically, you get points for each Klingon you kill, for your Klingon per stardate kill rate, and a bonus if you win the game. You lose points for the number of Klingons left in the galaxy at the end of the game, for getting killed, for each star, starbase, or inhabited starsystem you destroy, for calling for help, and for each casualty you incur.

You will be promoted if you play very well. You will never get a promotion if you call for help, abandon the Enterprise, get killed, destroy a starbase or inhabited starsystem, or destroy too many stars.

COMMAND SUMMARY

Command	Requires	Consumes
abandon	shuttlecraft, transporter	-
capture	subspace radio	time
cloak **up/d**own	cloaking device	energy
computer request; ...	computer	-
damages	-	-
destruct	computer	-
dock	-	-
help	subspace radio	-
impulse course distance	impulse engines, computer, SINS	time, energy
lrscan	L.R. sensors	-
move course distance	warp engines, computer, SINS	time, energy
phasers **a**utomatic amount	phasers, computer	energy **p**hasers **m**anual
amt1 course1 spread1 ...	phasers	energy
torpedo course [**y**es] angle/**n**o	torpedo tubes	torpedoes
ram course distance	warp engines, computer, SINS	time, energy
rest time	-	time
shell	-	-
shields **u**p/**d**own	shields	energy
srscan [**y**es/**n**o]	S.R. sensors	-
status	-	-
terminate **y**es/**n**o	-	-
undock	-	-
visual course	-	time
warp **w**arp_factor	-	-

A 4.2BSD Interprocess Communication Primer
DRAFT of September 15, 1986

Samuel J. Leffler

Robert S. Fabry

William N. Joy

Computer Systems Research Group
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, California 94720
(415) 642-7780

ABSTRACT

This document provides an introduction to the interprocess communication facilities included in the 4.2BSD release of the VAX* UNIX** system.

It discusses the overall model for interprocess communication and introduces the interprocess communication primitives which have been added to the system. The majority of the document considers the use of these primitives in developing applications. The reader is expected to be familiar with the C programming language as all examples are written in C.

---

1. INTRODUCTION

One of the most important parts of 4.2BSD is the interprocess communication facilities. These facilities are the result of more than two years of discussion and research. The facilities provided in 4.2BSD incorporate many of the ideas from current research, while trying to maintain the UNIX philosophy of simplicity and conciseness. It is hoped that the interprocess communication facilities included in 4.2BSD will establish a standard for UNIX. From the response to the design, it appears many organizations carrying out work with UNIX are adopting it.

UNIX has previously been very weak in the area of interprocess communication. Prior to the 4.2BSD facilities, the only standard mechanism which allowed two processes to communicate were pipes (the mpx files which were part of Version 7 were experimental). Unfortunately, pipes are very restrictive in that the two communicating processes must be related through a common ancestor. Further, the semantics of pipes makes them almost impossible to maintain in a distributed environment.

Earlier attempts at extending the ipc facilities of UNIX have met with mixed reaction. The majority of the problems have been related to the fact these facilities have been tied to the UNIX file system; either through naming, or implementation. Consequently, the ipc facilities provided in 4.2BSD have been designed as a totally independent subsystem. The 4.2BSD ipc allows processes to rendezvous in many ways. Processes may rendezvous through a UNIX file system-like name space (a space where all names are path names) as well as through a network name space. In fact, new name spaces may be added at a future time with only minor changes visible to users. Further, the communication facilities have been extended to included more than the simple byte stream provided by a pipe-like entity. These extensions have resulted in a completely new part of the system which users will need time to familiarize themselves with. It is likely that as more use is made of these facilities they will be refined; only time will tell.

The remainder of this document is organized in four sections. Section 2 introduces the new system calls and the basic model of communication. Section 3 describes some of the supporting library routines users may find useful in constructing distributed applications. Section 4 is concerned with the client/server model used in developing applications and includes examples of the two major types of servers. Section 5 delves into advanced topics which sophisticated users are likely to encounter when using the ipc facilities.

2. BASICS

The basic building block for communication is the *socket*. A socket is an endpoint of communication to which a name may be *bound*. Each socket in use has a *type* and one or more associated processes. Sockets exist within *communication domains*. A communication domain is an abstraction introduced to bundle common properties of processes communicating through sockets. One such property is the scheme used to name sockets. For example, in the UNIX communication domain sockets are named with UNIX path names; e.g. a socket may be named "/dev/foo". Sockets normally exchange data only with sockets in the same domain (it may be possible to cross domain boundaries, but only if some translation process is performed). The 4.2BSD ipc supports two separate communication domains: the UNIX domain, and the Internet domain is used by processes which communicate using the the DARPA standard communication protocols. The underlying communication facilities provided by these domains have a significant influence on the internal system implementation as well as the interface to socket facilities available to a user. An example of the latter is that a socket "operating" in the UNIX domain sees a subset of the possible error conditions which are possible when operating in the Internet domain.

2.1. Socket types

Sockets are typed according to the communication properties visible to a user. Processes are presumed to communicate only between sockets of the same type, although there is nothing that prevents communication between sockets of different types should the underlying communication protocols support this.

Three types of sockets currently are available to a user. A *stream* socket provides for the bidirectional, reliable, sequenced, and unduplicated flow of data without record boundaries. Aside from the bidirectionality of data flow, a pair of connected stream sockets provides an interface nearly identical to that of pipes*.

A *datagram* socket supports bidirectional flow of data which is not promised to be sequenced, reliable, or unduplicated. That is, a process receiving messages on a datagram socket may find messages duplicated, and, possibly, in an order different from the order in which it was sent. An important characteristic of a datagram socket is that record boundaries in data are preserved. Datagram sockets closely model the facilities found in many contemporary packet switched networks such as the Ethernet.

A *raw* socket provides users access to the underlying communication protocols which support socket abstractions. These sockets are normally datagram oriented, though their exact characteristics are dependent on the interface provided by the protocol. Raw sockets are not intended for the general user; they have been provided mainly for those interested in developing new communication protocols, or for gaining access to some of the more esoteric facilities of an existing protocol. The use of raw sockets is considered in section 5.

Two potential socket types which have interesting properties are the *sequenced packet* socket and the *reliably delivered message* socket. A sequenced packet socket is identical to a stream socket with the exception that record boundaries are preserved. This interface is very similar to that provided by the Xerox NS Sequenced Packet protocol. The reliably delivered message socket has similar properties to a datagram socket, but with reliable delivery. While these two socket types have been loosely defined, they are currently unimplemented in 4.2BSD. As such, in this document we will concern ourselves only with the three socket types for which support exists.

* In the UNIX domain, in fact, the semantics are identical and, as one might expect, pipes have been implemented internally as simply a pair of connected stream sockets.

2.2. Socket creation

To create a socket the *socket* system call is used:

```
s = socket(domain, type, protocol);
```

This call requests that the system create a socket in the specified *domain* and of the specified *type*. A particular protocol may also be requested. If the protocol is left unspecified (a value of 0), the system will select an appropriate protocol from those protocols which comprise the communication domain and which may be used to support the requested socket type. The user is returned a descriptor (a small integer number) which may be used in later system calls which operate on sockets. The domain is specified as one of the manifest constants defined in the file *<sys/socket.h>*. For the UNIX domain the constant is AF_UNIX*; for the Internet domain AF_INET. The socket types are also defined in this file and one of SOCK_STREAM, SOCK_DGRAM, or SOCK_RAW must be specified. To create a stream socket in the Internet domain the following call might be used:

```
s = socket(AF_INET, SOCK_STREAM, 0);
```

This call would result in a stream socket being created with the TCP protocol providing the underlying communication support. To create a datagram socket for on-machine use a sample call might be:

```
s = socket(AF_UNIX, SOCK_DGRAM, 0);
```

To obtain a particular protocol one selects the protocol number, as defined within the communication domain. For the Internet domain the available protocols are defined in *<netinet/in.h>* or, better yet, one may use one of the library routines discussed in section 3, such as *getprotobyname*:

```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <netdb.h>
...
pp = getprotobyname("tcp");
s = socket(AF_INET, SOCK_STREAM, pp->p_proto);
```

There are several reasons a socket call may fail. Aside from the rare occurrence of lack of memory (ENOBUFS), a socket request may fail due to a request for an unknown protocol (EPROTONOSUPPORT), or a request for a type of socket for which there is no supporting protocol (EPROTOTYPE).

2.3. Binding names

A socket is created without a name. Until a name is bound to a socket, processes have no way to reference it and, consequently, no messages may be received on it. The *bind* call is used to assign a name to a socket:

```
bind(s, name, namelen);
```

The bound name is a variable length byte string which is interpreted by the supporting protocol(s). Its interpretation may vary from communication domain to communication domain (this is one of the properties which comprise the "domain"). In the UNIX domain names are path names while in the Internet domain names contain an Internet address and port number. If one wanted to bind the name "/dev/foo" to a UNIX domain socket, the following would be used:

* The manifest constants are named AF_whatever as they indicate the "address format" to use in interpreting names.

```
bind(s, "/dev/foo", sizeof ("/dev/foo") - 1);
```

(Note how the null byte in the name is not counted as part of the name.) In binding an Internet address things become more complicated. The actual call is simple,

```
#include <sys/types.h>
#include <netinet/in.h>
 ...
struct sockaddr_in sin;
 ...
bind(s, &sin, sizeof (sin));
```

but the selection of what to place in the address *sin* requires some discussion. We will come back to the problem of formulating Internet addresses in section 3 when the library routines used in name resolution are discussed.

2.4. Connection establishment

With a bound socket it is possible to rendezvous with an unrelated process. This operation is usually asymmetric with one process a "client" and the other a "server". The client requests services from the server by initiating a "connection" to the server's socket. The server, when willing to offer its advertised services, passively "listens" on its socket. On the client side the *connect* call is used to initiate a connection. Using the UNIX domain, this might appear as,

```
connect(s, "server-name", sizeof ("server-name"));
```

while in the Internet domain,

```
struct sockaddr_in server;
connect(s, &server, sizeof (server));
```

If the client process's socket is unbound at the time of the connect call, the system will automatically select and bind a name to the socket; c.f. section 5.4. An error is returned when the connection was unsuccessful (any name automatically bound by the system, however, remains). Otherwise, the socket is associated with the server and data transfer may begin.

Many errors can be returned when a connection attempt fails. The most common are:

ETIMEDOUT
: After failing to establish a connection for a period of time, the system decided there was no point in retrying the connection attempt any more. This usually occurs because the destination host is down, or because problems in the network resulted in transmissions being lost.

ECONNREFUSED
: The host refused service for some reason. When connecting to a host running 4.2BSD this is usually due to a server process not being present at the requested name.

ENETDOWN or EHOSTDOWN
: These operational errors are returned based on status information delivered to the client host by the underlying communication services.

ENETUNREACH or EHOSTUNREACH
: These operational errors can occur either because the network or host is unknown (no route to the network or host is present), or because of status information returned by intermediate gateways or switching nodes. Many times the status returned is not sufficient to distinguish a network being down from a host being down. In these cases the system is conservative and indicates the entire network is unreachable.

For the server to receive a client's connection it must perform two steps after binding its socket. The first is to indicate a willingness to listen for incoming connection requests:

```
listen(s, 5);
```

The second parameter to the *listen* call specifies the maximum number of outstanding connections which may be queued awaiting acceptance by the server process. Should a connection be requested while the queue is full, the connection will not be refused, but rather the individual messages which comprise the request will be ignored. This gives a harried server time to make room in its pending connection queue while the client retries the connection request. Had the connection been returned with the ECONNREFUSED error, the client would be unable to tell if the server was up or not. As it is now it is still possible to get the ETIMEDOUT error back, though this is unlikely. The backlog figure supplied with the listen call is limited by the system to a maximum of 5 pending connections on any one queue. This avoids the problem of processes hogging system resources by setting an infinite backlog, then ignoring all connection requests.

With a socket marked as listening, a server may *accept* a connection:

```
fromlen = sizeof (from);
snew = accept(s, &from, &fromlen);
```

A new descriptor is returned on receipt of a connection (along with a new socket). If the server wishes to find out who its client is, it may supply a buffer for the client socket's name. The value-result parameter *fromlen* is initialized by the server to indicate how much space is associated with *from*, then modified on return to reflect the true size of the name. If the client's name is not of interest, the second parameter may be zero.

Accept normally blocks. That is, the call to accept will not return until a connection is available or the system call is interrupted by a signal to the process. Further, there is no way for a process to indicate it will accept connections from only a specific individual, or individuals. It is up to the user process to consider who the connection is from and close down the connection if it does not wish to speak to the process. If the server process wants to accept connections on more than one socket, or not block on the accept call there are alternatives; they will be considered in section 5.

2.5. Data transfer

With a connection established, data may begin to flow. To send and receive data there are a number of possible calls. With the peer entity at each end of a connection anchored, a user can send or receive a message without specifying the peer. As one might expect, in this case, then the normal *read* and *write* system calls are useable,

```
write(s, buf, sizeof (buf));
read(s, buf, sizeof (buf));
```

In addition to *read* and *write*, the new calls *send* and *recv* may be used:

```
send(s, buf, sizeof (buf), flags);
recv(s, buf, sizeof (buf), flags);
```

While *send* and *recv* are virtually identical to *read* and *write*, the extra *flags* argument is important. The flags may be specified as a non-zero value if one or more of the following is required:

SOF_OOB	send/receive out of band data
SOF_PREVIEW	look at data without reading
SOF_DONTROUTE	send data without routing packets

Out of band data is a notion specific to stream sockets, and one which we will not immediately consider. The option to have data sent without routing applied to the outgoing packets is currently used only by the routing table management process, and is unlikely to be of interest to the casual user. The ability to preview data is, however, of interest. When SOF_PREVIEW is specified with a *recv* call, any data present is returned to the user, but treated as still "unread". That is, the next *read* or *recv* call applied to the socket will return the data previously previewed.

2.6. Discarding sockets

Once a socket is no longer of interest, it may be discarded by applying a *close* to the descriptor,

```
close(s);
```

If data is associated with a socket which promises reliable delivery (e.g. a stream socket) when a close takes place, the system will continue to attempt to transfer the data. However, after a fairly long period of time, if the data is still undelivered, it will be discarded. Should a user have no use for any pending data, it may perform a *shutdown* on the socket prior to closing it. This call is of the form:

```
shutdown(s, how);
```

where *how* is 0 if the user is no longer interested in reading data, 1 if no more data will be sent, or 2 if no data is to be sent or received. Applying shutdown to a socket causes any data queued to be immediately discarded.

2.7. Connectionless sockets

To this point we have been concerned mostly with sockets which follow a connection oriented model. However, there is also support for connectionless interactions typical of the datagram facilities found in contemporary packet switched networks. A datagram socket provides a symmetric interface to data exchange. While processes are still likely to be client and server, there is no requirement for connection establishment. Instead, each message includes the destination address.

Datagram sockets are created as before, and each should have a name bound to it in order that the recipient of a message may identify the sender. To send data, the *sendto* primitive is used,

```
sendto(s, buf, buflen, flags, &to, tolen);
```

The *s*, *buf*, *buflen*, and *flags* parameters are used as before. The *to* and *tolen* values are used to indicate the intended recipient of the message. When using an unreliable datagram interface, it is unlikely any errors will be reported to the sender. Where information is present locally to recognize a message which may never be delivered (for instance when a network is unreachable), the call will return −1 and the global value *errno* will contain an error number.

To receive messages on an unconnected datagram socket, the *recvfrom* primitive is provided:

```
recvfrom(s, buf, buflen, flags, &from, &fromlen);
```

Once again, the *fromlen* parameter is handled in a value-result fashion, initially containing the size of the *from* buffer.

In addition to the two calls mentioned above, datagram sockets may also use the *connect* call to associate a socket with a specific address. In this case, any data sent on the socket will automatically be addressed to the connected peer, and only data received from that peer will be delivered to the user. Only one connected address is permitted for each socket (i.e. no multicasting). Connect requests on datagram sockets return immediately, as this simply results in the system recording the peer's address (as compared to a stream socket where a connect request initiates establishment of an end to end connection). Other of the less important details of datagram sockets are described in section 5.

2.8. Input/Output multiplexing

One last facility often used in developing applications is the ability to multiplex i/o requests among multiple sockets and/or files. This is done using the *select* call:

```
select(nfds, &readfds, &writefds, &execptfds, &timeout);
```

Select takes as arguments three bit masks, one for the set of file descriptors for which the caller wishes to be able to read data on, one for those descriptors to which data is to be written, and one for which exceptional conditions are pending. Bit masks are created by or-ing bits of the form "1 << fd". That is, a descriptor *fd* is selected if a 1 is present in the *fd*'th bit of the mask. The parameter *nfds* specifies the range of file descriptors (i.e. one plus the value of the largest descriptor) specified in a mask.

A timeout value may be specified if the selection is not to last more than a predetermined period of time. If *timeout* is set to 0, the selection takes the form of a *poll*, returning immediately. If the last parameter is a null pointer, the selection will block indefinitely*. *Select* normally returns the number of file descriptors selected. If the *select* call returns due to the timeout expiring, then a value of –1 is returned along with the error number EINTR.

Select provides a synchronous multiplexing scheme. Asynchronous notification of output completion, input availability, and exceptional conditions is possible through use of the SIGIO and SIGURG signals described in section 5.

* To be more specific, a return takes place only when a descriptor is selectable, or when a signal is received by the caller, interrupting the system call.

3. NETWORK LIBRARY ROUTINES

The discussion in section 2 indicated the possible need to locate and construct network addresses when using the interprocess communication facilities in a distributed environment. To aid in this task a number of routines have been added to the standard C run-time library. In this section we will consider the new routines provided to manipulate network addresses. While the 4.2BSD networking facilities support only the DARPA standard Internet protocols, these routines have been designed with flexibility in mind. As more communication protocols become available, we hope the same user interface will be maintained in accessing network-related address data bases. The only difference should be the values returned to the user. Since these values are normally supplied the system, users should not need to be directly aware of the communication protocol and/or naming conventions in use.

Locating a service on a remote host requires many levels of mapping before client and server may communicate. A service is assigned a name which is intended for human consumption; e.g. "the *login server* on host monet". This name, and the name of the peer host, must then be translated into network *addresses* which are not necessarily suitable for human consumption. Finally, the address must then used in locating a physical *location* and *route* to the service. The specifics of these three mappings is likely to vary between network architectures. For instance, it is desirable for a network to not require hosts be named in such a way that their physical location is known by the client host. Instead, underlying services in the network may discover the actual location of the host at the time a client host wishes to communicate. This ability to have hosts named in a location independent manner may induce overhead in connection establishment, as a discovery process must take place, but allows a host to be physically mobile without requiring it to notify its clientele of its current location.

Standard routines are provided for: mapping host names to network addresses, network names to network numbers, protocol names to protocol numbers, and service names to port numbers and the appropriate protocol to use in communicating with the server process. The file <*netdb.h*> must be included when using any of these routines.

3.1. Host names

A host name to address mapping is represented by the *hostent* structure:

```
struct  hostent {
        char     *h_name;       /* official name of host */
        char     **h_aliases;   /* alias list */
        int      h_addrtype;    /* host address type */
        int      h_length;      /* length of address */
        char     *h_addr;       /* address */
};
```

The official name of the host and its public aliases are returned, along with a variable length address and address type. The routine *gethostbyname*(3N) takes a host name and returns a *hostent* structure, while the routine *gethostbyaddr*(3N) maps host addresses into a *hostent* structure. It is possible for a host to have many addresses, all having the same name. *Gethostybyname* returns the first matching entry in the data base file */etc/hosts*; if this is unsuitable, the lower level routine *gethostent*(3N) may be used. For example, to obtain a *hostent* structure for a host on a particular network the following routine might be used (for simplicity, only Internet addresses are considered):

```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <netdb.h>
 ...
struct hostent *
gethostbynameandnet(name, net)
        char *name;
        int net;
{
        register struct hostent *hp;
        register char **cp;

        sethostent(0);
        while ((hp = gethostent()) != NULL) {
                if (hp->h_addrtype != AF_INET)
                        continue;
                if (strcmp(name, hp->h_name)) {
                        for (cp = hp->h_aliases; cp && *cp != NULL; cp++)
                                if (strcmp(name, *cp) == 0)
                                goto found;
                        continue;
                }
        found:
                if (in_netof(*(struct in_addr *)hp->h_addr)) == net)
                        break;
        }
        endhostent(0);
        return (hp);
}
```

(*in_netof*(3N) is a standard routine which returns the network portion of an Internet address.)

3.2. Network names

As for host names, routines for mapping network names to numbers, and back, are provided. These routines return a *netent* structure:

```
/*
 * Assumption here is that a network number
 * fits in 32 bits -- probably a poor one.
 */
struct  netent {
        char        *n_name;          /* official name of net */
        char        **n_aliases;      /* alias list */
        int         n_addrtype;       /* net address type */
        int         n_net;            /* network # */
};
```

The routines *getnetbyname*(3N), *getnetbynumber*(3N), and *getnetent*(3N) are the network counterparts to the host routines described above.

3.3. Protocol names

For protocols the *protoent* structure defines the protocol-name mapping used with the routines *getprotobyname*(3N), *getprotobynumber*(3N), and *getprotoent*(3N):

```
struct  protoent {
        char      *p_name;         /* official protocol name */
        char      **p_aliases;     /* alias list */
        int       p_proto;         /* protocol # */
};
```

3.4. Service names

Information regarding services is a bit more complicated. A service is expected to reside at a specific "port" and employ a particular communication protocol. This view is consistent with the Internet domain, but inconsistent with other network architectures. Further, a service may reside on multiple ports or support multiple protocols. If either of these occurs, the higher level library routines will have to be bypassed in favor of homegrown routines similar in spirit to the "gethostbynameandnet" routine described above. A service mapping is described by the *servent* structure,

```
struct  servent {
        char      *s_name;         /* official service name */
        char      **s_aliases;     /* alias list */
        int       s_port;          /* port # */
        char      *s_proto;        /* protocol to use */
};
```

The routine *getservbyname*(3N) maps service names to a servent structure by specifying a service name and, optionally, a qualifying protocol. Thus the call

```
sp = getservbyname("telnet", (char *)0);
```

returns the service specification for a telnet server using any protocol, while the call

```
sp = getservbyname("telnet", "tcp");
```

returns only that telnet server which uses the TCP protocol. The routines *getservbyport*(3N) and *getservent*(3N) are also provided. The *getservbyport* routine has an interface similar to that provided by *getservbyname*; an optional protocol name may be specified to qualify lookups.

3.5. Miscellaneous

With the support routines described above, an application program should rarely have to deal directly with addresses. This allows services to be developed as much as possible in a network independent fashion. It is clear, however, that purging all network dependencies is very difficult. So long as the user is required to supply network addresses when naming services and sockets there will always some network dependency in a program. For example, the normal code included in client programs, such as the remote login program, is of the form shown in Figure 1. (This example will be considered in more detail in section 4.)

If we wanted to make the remote login program independent of the Internet protocols and addressing scheme we would be forced to add a layer of routines which masked the network dependent aspects from the mainstream login code. For the current facilities available in the system this does not appear to be worthwhile. Perhaps when the system is adapted to different network architectures the utilities will be reorganized more cleanly.

Aside from the address-related data base routines, there are several other routines available in the run-time library which are of interest to users. These are intended mostly to simplify manipulation of names and addresses. Table 1 summarizes the routines for manipulating variable length byte strings and handling byte swapping of network addresses and values.

The byte swapping routines are provided because the operating system expects addresses to be supplied in network order. On a VAX, or machine with similar architecture, this is usually reversed. Consequently, programs are sometimes required to byte swap quantities. The library routines which return network addresses provide them in network order so that they may simply

```
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <stdio.h>
#include <netdb.h>
 ...
main(argc, argv)
        char *argv[];
{
        struct sockaddr_in sin;
        struct servent *sp;
        struct hostent *hp;
        int s;
        ...
        sp = getservbyname("login", "tcp");
        if (sp == NULL) {
                fprintf(stderr, "rlogin: tcp/login: unknown service\n");
                exit(1);
        }
        hp = gethostbyname(argv[1]);
        if (hp == NULL) {
                fprintf(stderr, "rlogin: %s: unknown host\n", argv[1]);
                exit(2);
        }
        bzero((char *)&sin, sizeof (sin));
        bcopy(hp->h_addr, (char *)&sin.sin_addr, hp->h_length);
        sin.sin_family = hp->h_addrtype;
        sin.sin_port = sp->s_port;
        s = socket(AF_INET, SOCK_STREAM, 0);
        if (s < 0) {
                perror("rlogin: socket");
                exit(3);
        }
        ...
        if (connect(s, (char *)&sin, sizeof (sin)) < 0) {
                perror("rlogin: connect");
                exit(5);
        }
        ...
}
```

Figure 1. Remote login client code.

Call	Synopsis
bcmp(s1, s2, n)	compare byte-strings; 0 if same, not 0 otherwise
bcopy(s1, s2, n)	copy n bytes from s1 to s2
bzero(base, n)	zero-fill n bytes starting at base
htonl(val)	convert 32-bit quantity from host to network byte order
htons(val)	convert 16-bit quantity from host to network byte order
ntohl(val)	convert 32-bit quantity from network to host byte order
ntohs(val)	convert 16-bit quantity from network to host byte order

Table 1. C run-time routines.

be copied into the structures provided to the system. This implies users should encounter the byte swapping problem only when *interpreting* network addresses. For example, if an Internet port is to be printed out the following code would be required:

```
printf("port number %d\n", ntohs(sp->s_port));
```

On machines other than the VAX these routines are defined as null macros.

4. CLIENT/SERVER MODEL

The most commonly used paradigm in constructing distributed applications is the client/server model. In this scheme client applications request services from a server process. This implies an asymmetry in establishing communication between the client and server which has been examined in section 2. In this section we will look more closely at the interactions between client and server, and consider some of the problems in developing client and server applications.

Client and server require a well known set of conventions before service may be rendered (and accepted). This set of conventions comprises a protocol which must be implemented at both ends of a connection. Depending on the situation, the protocol may be symmetric or asymmetric. In a symmetric protocol, either side may play the master or slave roles. In an asymmetric protocol, one side is immutably recognized as the master, with the other the slave. An example of a symmetric protocol is the TELNET protocol used in the Internet for remote terminal emulation. An example of an asymmetric protocol is the Internet file transfer protocol, FTP. No matter whether the specific protocol used in obtaining a service is symmetric or asymmetric, when accessing a service there is a "client process" and a "server process". We will first consider the properties of server processes, then client processes.

A server process normally listens at a well know address for service requests. Alternative schemes which use a service server may be used to eliminate a flock of server processes clogging the system while remaining dormant most of the time. The Xerox Courier protocol uses the latter scheme. When using Courier, a Courier client process contacts a Courier server at the remote host and identifies the service it requires. The Courier server process then creates the appropriate server process based on a data base and "splices" the client and server together, voiding its part in the transaction. This scheme is attractive in that the Courier server process may provide a single contact point for all services, as well as carrying out the initial steps in authentication. However, while this is an attractive possibility for standardizing access to services, it does introduce a certain amount of overhead due to the intermediate process involved. Implementations which provide this type of service within the system can minimize the cost of client server rendezvous. The *portal* notion described in the "4.2BSD System Manual" embodies many of the ideas found in Courier, with the rendezvous mechanism implemented internal to the system.

4.1. Servers

In 4.2BSD most servers are accessed at well known Internet addresses or UNIX domain names. When a server is started at boot time it advertises it services by listening at a well know location. For example, the remote login server's main loop is of the form shown in Figure 2.

The first step taken by the server is look up its service definition:

```
sp = getservbyname("login", "tcp");
if (sp == NULL) {
        fprintf(stderr, "rlogind: tcp/login: unknown service\n");
        exit(1);
}
```

This definition is used in later portions of the code to define the Internet port at which it listens for service requests (indicated by a connection).

Step two is to disassociate the server from the controlling terminal of its invoker. This is important as the server will likely not want to receive signals delivered to the process group of the controlling terminal.

Once a server has established a pristine environment, it creates a socket and begins accepting service requests. The *bind* call is required to insure the server listens at its expected location. The main body of the loop is fairly simple:

```
main(argc, argv)
	int argc;
	char **argv;
{
	int f;
	struct sockaddr_in from;
	struct servent *sp;

	sp = getservbyname("login", "tcp");
	if (sp == NULL) {
		fprintf(stderr, "rlogind: tcp/login: unknown service\n");
		exit(1);
	}
	...
#ifndef DEBUG
	<<disassociate server from controlling terminal>>
#endif
	...
	sin.sin_port = sp->s_port;
	...
	f = socket(AF_INET, SOCK_STREAM, 0);
	...
	if (bind(f, (caddr_t)&sin, sizeof (sin)) < 0) {
		...
	}
	...
	listen(f, 5);
	for (;;) {
		int g, len = sizeof (from);

		g = accept(f, &from, &len);
		if (g < 0) {
			if (errno != EINTR)
				perror("rlogind: accept");
			continue;
		}
		if (fork() == 0) {
			close(f);
			doit(g, &from);
		}
		close(g);
	}
}
```

Figure 2. Remote login server.

```
for (;;) {
	int g, len = sizeof (from);

	g = accept(f, &from, &len);
	if (g < 0) {
		if (errno != EINTR)
			perror("rlogind: accept");
		continue;
```

```
                }
                if (fork() == 0) {
                        close(f);
                        doit(g, &from);
                }
                close(g);
        }
```

An *accept* call blocks the server until a client requests service. This call could return a failure status if the call is interrupted by a signal such as SIGCHLD (to be discussed in section 5). Therefore, the return value from *accept* is checked to insure a connection has actually been established. With a connection in hand, the server then forks a child process and invokes the main body of the remote login protocol processing. Note how the socket used by the parent for queueing connection requests is closed in the child, while the socket created as a result of the accept is closed in the parent. The address of the client is also handed the *doit* routine because it requires it in authenticating clients.

4.2. Clients

The client side of the remote login service was shown earlier in Figure 1. One can see the separate, asymmetric roles of the client and server clearly in the code. The server is a passive entity, listening for client connections, while the client process is an active entity, initiating a connection when invoked.

Let us consider more closely the steps taken by the client remote login process. As in the server process the first step is to locate the service definition for a remote login:

```
sp = getservbyname("login", "tcp");
if (sp == NULL) {
        fprintf(stderr, "rlogin: tcp/login: unknown service\n");
        exit(1);
}
```

Next the destination host is looked up with a *gethostbyname* call:

```
hp = gethostbyname(argv[1]);
if (hp == NULL) {
        fprintf(stderr, "rlogin: %s: unknown host\n", argv[1]);
        exit(2);
}
```

With this accomplished, all that is required is to establish a connection to the server at the requested host and start up the remote login protocol. The address buffer is cleared, then filled in with the Internet address of the foreign host and the port number at which the login process resides:

```
bzero((char *)&sin, sizeof (sin));
bcopy(hp->h_addr, (char *)sin.sin_addr, hp->h_length);
sin.sin_family = hp->h_addrtype;
sin.sin_port = sp->s_port;
```

A socket is created, and a connection initiated.

```
s = socket(hp->h_addrtype, SOCK_STREAM, 0);
if (s < 0) {
      perror("rlogin: socket");
      exit(3);
}
 ...
if (connect(s, (char *)&sin, sizeof (sin)) < 0) {
      perror("rlogin: connect");
      exit(4);
}
```

The details of the remote login protocol will not be considered here.

4.3. Connectionless servers

While connection-based services are the norm, some services are based on the use of datagram sockets. One, in particular, is the "rwho" service which provides users with status information for hosts connected to a local area network. This service, while predicated on the ability to *broadcast* information to all hosts connected to a particular network, is of interest as an example usage of datagram sockets.

A user on any machine running the rwho server may find out the current status of a machine with the *ruptime*(1) program. The output generated is illustrated in Figure 3.

arpa	up	9:45,	5 users, load	1.15,	1.39,	1.31
cad	up	2+12:04,	8 users, load	4.67,	5.13,	4.59
calder	up	10:10,	0 users, load	0.27,	0.15,	0.14
dali	up	2+06:28,	9 users, load	1.04,	1.20,	1.65
degas	up	25+09:48,	0 users, load	1.49,	1.43,	1.41
ear	up	5+00:05,	0 users, load	1.51,	1.54,	1.56
ernie	down	0:24				
esvax	down	17:04				
ingres	down	0:26				
kim	up	3+09:16,	8 users, load	2.03,	2.46,	3.11
matisse	up	3+06:18,	0 users, load	0.03,	0.03,	0.05
medea	up	3+09:39,	2 users, load	0.35,	0.37,	0.50
merlin	down	19+15:37				
miro	up	1+07:20,	7 users, load	4.59,	3.28,	2.12
monet	up	1+00:43,	2 users, load	0.22,	0.09,	0.07
oz	down	16:09				
statvax	up	2+15:57,	3 users, load	1.52,	1.81,	1.86
ucbvax	up	9:34,	2 users, load	6.08,	5.16,	3.28

Figure 3. ruptime output.

Status information for each host is periodically broadcast by rwho server processes on each machine. The same server process also receives the status information and uses it to update a database. This database is then interpreted to generate the status information for each host. Servers operate autonomously, coupled only by the local network and its broadcast capabilities.

The rwho server, in a simplified form, is pictured in Figure 4. There are two separate tasks performed by the server. The first task is to act as a receiver of status information broadcast by other hosts on the network. This job is carried out in the main loop of the program. Packets received at the rwho port are interrogated to insure they've been sent by another rwho server process, then are time stamped with their arrival time and used to update a file indicating the status of the host. When a host has not been heard from for an extended period of time, the database interpretation routines assume the host is down and indicate such on the status reports. This algorithm is prone to error as a server may be down while a host is actually up, but serves our current needs.

```
main()
{
        ...
        sp = getservbyname("who", "udp");
        net = getnetbyname("localnet");
        sin.sin_addr = inet_makeaddr(INADDR_ANY, net);
        sin.sin_port = sp->s_port;
        ...
        s = socket(AF_INET, SOCK_DGRAM, 0);
        ...
        bind(s, &sin, sizeof (sin));
        ...
        sigset(SIGALRM, onalrm);
        onalrm();
        for (;;) {
                struct whod wd;
                int cc, whod, len = sizeof (from);

                cc = recvfrom(s, (char *)&wd, sizeof (struct whod), 0, &from, &len);
                if (cc <= 0) {
                        if (cc < 0 && errno != EINTR)
                                perror("rwhod: recv");
                        continue;
                }
                if (from.sin_port != sp->s_port) {
                        fprintf(stderr, "rwhod: %d: bad from port\n",
                                ntohs(from.sin_port));
                        continue;
                }
                ...
                if (!verify(wd.wd_hostname)) {
                        fprintf(stderr, "rwhod: malformed host name from %x\n",
                                ntohl(from.sin_addr.s_addr));
                        continue;
                }
                (void) sprintf(path, "%s/whod.%s", RWHODIR, wd.wd_hostname);
                whod = open(path, FWRONLY|FCREATE|FTRUNCATE, 0666);
                ...
                (void) time(&wd.wd_recvtime);
                (void) write(whod, (char *)&wd, cc);
                (void) close(whod);
        }
}
```

Figure 4. rwho server.

The second task performed by the server is to supply information regarding the status of its host. This involves periodically acquiring system status information, packaging it up in a message and broadcasting it on the local network for other rwho servers to hear. The supply function is triggered by a timer and runs off a signal. Locating the system status information is somewhat involved, but uninteresting. Deciding where to transmit the resultant packet does, however, indicates some problems with the current protocol.

Status information is broadcast on the local network. For networks which do not support the notion of broadcast another scheme must be used to simulate or replace broadcasting. One possibility is to enumerate the known neighbors (based on the status received). This, unfortunately, requires some bootstrapping information, as a server started up on a quiet network will have no known neighbors and thus never receive, or send, any status information. This is the identical problem faced by the routing table management process in propagating routing status information. The standard solution, unsatisfactory as it may be, is to inform one or more servers of known neighbors and request that they always communicate with these neighbors. If each server has at least one neighbor supplied it, status information may then propagate through a neighbor to hosts which are not (possibly) directly neighbors. If the server is able to support networks which provide a broadcast capability, as well as those which do not, then networks with an arbitrary topology may share status information*.

The second problem with the current scheme is that the rwho process services only a single local network, and this network is found by reading a file. It is important that software operating in a distributed environment not have any site-dependent information compiled into it. This would require a separate copy of the server at each host and make maintenance a severe headache. 4.2BSD attempts to isolate host-specific information from applications by providing system calls which return the necessary information†. Unfortunately, no straightforward mechanism currently exists for finding the collection of networks to which a host is directly connected. Thus the rwho server performs a lookup in a file to find its local network. A better, though still unsatisfactory, scheme used by the routing process is to interrogate the system data structures to locate those directly connected networks. A mechanism to acquire this information from the system would be a useful addition.

* One must, however, be concerned about "loops". That is, if a host is connected to multiple networks, it will receive status information from itself. This can lead to an endless, wasteful, exchange of information.
† An example of such a system call is the *gethostname*(2) call which returns the host's "official" name.

5. ADVANCED TOPICS

A number of facilities have yet to be discussed. For most users of the ipc the mechanisms already described will suffice in constructing distributed applications. However, others will find need to utilize some of the features which we consider in this section.

5.1. Out of band data

The stream socket abstraction includes the notion of "out of band" data. Out of band data is a logically independent transmission channel associated with each pair of connected stream sockets. Out of band data is delivered to the user independently of normal data along with the SIGURG signal. In addition to the information passed, a logical mark is placed in the data stream to indicate the point at which the out of band data was sent. The remote login and remote shell applications use this facility to propagate signals from between client and server processes. When a signal is expected to flush any pending output from the remote process(es), all data up to the mark in the data stream is discarded.

The stream abstraction defines that the out of band data facilities must support the reliable delivery of at least one out of band message at a time. This message may contain at least one byte of data, and at least one message may be pending delivery to the user at any one time. For communications protocols which support only in-band signaling (i.e. the urgent data is delivered in sequence with the normal data) the system extracts the data from the normal data stream and stores it separately. This allows users to choose between receiving the urgent data in order and receiving it out of sequence without having to buffer all the intervening data.

To send an out of band message the SOF_OOB flag is supplied to a *send* or *sendto* calls, while to receive out of band data SOF_OOB should be indicated when performing a *recvfrom* or *recv* call. To find out if the read pointer is currently pointing at the mark in the data stream, the SIOCATMARK ioctl is provided:

```
ioctl(s, SIOCATMARK, &yes);
```

If *yes* is a 1 on return, the next read will return data after the mark. Otherwise (assuming out of band data has arrived), the next read will provide data sent by the client prior to transmission of the out of band signal. The routine used in the remote login process to flush output on receipt of an interrupt or quit signal is shown in Figure 5.

5.2. Signals and process groups

Due to the existence of the SIGURG and SIGIO signals each socket has an associated process group (just as is done for terminals). This process group is initialized to the process group of its creator, but may be redefined at a later time with the SIOCSPGRP ioctl:

```
oob()
{
	int out = 1+1;
	char waste[BUFSIZ], mark;

	signal(SIGURG, oob);
	/* flush local terminal input and output */
	ioctl(1, TIOCFLUSH, (char *)&out);
	for (;;) {
		if (ioctl(rem, SIOCATMARK, &mark) < 0) {
			perror("ioctl");
			break;
		}
		if (mark)
			break;
		(void) read(rem, waste, sizeof (waste));
	}
	recv(rem, &mark, 1, SOF_OOB);
	...
}
```

Figure 5. Flushing terminal i/o on receipt of out of band data.

```
ioctl(s, SIOCSPGRP, &pgrp);
```

A similar ioctl, SIOCGPGRP, is available for determining the current process group of a socket.

5.3. Pseudo terminals

Many programs will not function properly without a terminal for standard input and output. Since a socket is not a terminal, it is often necessary to have a process communicating over the network do so through a *pseudo terminal.* A pseudo terminal is actually a pair of devices, master and slave, which allow a process to serve as an active agent in communication between processes and users. Data written on the slave side of a pseudo terminal is supplied as input to a process reading from the master side. Data written on the master side is given the slave as input. In this way, the process manipulating the master side of the pseudo terminal has control over the information read and written on the slave side. The remote login server uses pseudo terminals for remote login sessions. A user logging in to a machine across the network is provided a shell with a slave pseudo terminal as standard input, output, and error. The server process then handles the communication between the programs invoked by the remote shell and the user's local client process. When a user sends an interrupt or quit signal to a process executing on a remote machine, the client login program traps the signal, sends an out of band message to the server process who then uses the signal number, sent as the data value in the out of band message, to perform a *killpg*(2) on the appropriate process group.

5.4. Internet address binding

Binding addresses to sockets in the Internet domain can be fairly complex. Communicating processes are bound by an *association.* An association is composed of local and foreign addresses, and local and foreign ports. Port numbers are allocated out of separate spaces, one for each Internet protocol. Associations are always unique. That is, there may never be duplicate <protocol, local address, local port, foreign address, foreign port> tuples.

The bind system call allows a process to specify half of an association, <local address, local port>, while the connect and accept primitives are used to complete a socket's association. Since the association is created in two steps the association uniqueness requirement indicated above could be violated unless care is taken. Further, it is unrealistic to expect user programs to always know proper values to use for the local address and local port since a host may reside on multiple networks and the set of allocated port numbers is not directly accessible to a user.

To simplify local address binding the notion of a "wildcard" address has been provided. When an address is specified as INADDR_ANY (a manifest constant defined in <netinet/in.h>), the system interprets the address as "any valid address". For example, to bind a specific port number to a socket, but leave the local address unspecified, the following code might be used:

```
#include <sys/types.h>
#include <netinet/in.h>
 ...
struct sockaddr_in sin;
 ...
s = socket(AF_INET, SOCK_STREAM, 0);
sin.sin_family = AF_INET;
sin.sin_addr.s_addr = INADDR_ANY;
sin.sin_port = MYPORT;
bind(s, (char *)&sin, sizeof (sin));
```

Sockets with wildcarded local addresses may receive messages directed to the specified port number, and addressed to any of the possible addresses assigned a host. For example, if a host is on a networks 46 and 10 and a socket is bound as above, then an accept call is performed, the process will be able to accept connection requests which arrive either from network 46 or network 10.

In a similar fashion, a local port may be left unspecified (specified as zero), in which case the system will select an appropriate port number for it. For example:

```
sin.sin_addr.s_addr = MYADDRESS;
sin.sin_port = 0;
bind(s, (char *)&sin, sizeof (sin));
```

The system selects the port number based on two criteria. The first is that ports numbered 0 through 1023 are reserved for privileged users (i.e. the super user). The second is that the port number is not currently bound to some other socket. In order to find a free port number in the privileged range the following code is used by the remote shell server:

```
struct sockaddr_in sin;
 ...
lport = IPPORT_RESERVED - 1;
sin.sin_addr.s_addr = INADDR_ANY;
 ...
for (;;) {
        sin.sin_port = htons((u_short)lport);
        if (bind(s, (caddr_t)&sin, sizeof (sin)) >= 0)
                break;
        if (errno != EADDRINUSE && errno != EADDRNOTAVAIL) {
                perror("socket");
                break;
        }
        lport--;
        if (lport == IPPORT_RESERVED/2) {
                fprintf(stderr, "socket: All ports in use\n");
                break;
        }
}
```

The restriction on allocating ports was done to allow processes executing in a "secure" environment to perform authentication based on the originating address and port number.

In certain cases the algorithm used by the system in selecting port numbers is unsuitable for an application. This is due to associations being created in a two step process. For example, the Internet file transfer protocol, FTP, specifies that data connections must always originate from the same local port. However, duplicate associations are avoided by connecting to different foreign ports. In this situation the system would disallow binding the same local address and port number to a socket if a previous data connection's socket were around. To override the default port selection algorithm then an option call must be performed prior to address binding:

```
setsockopt(s, SOL_SOCKET, SO_REUSEADDR, (char *)0, 0);
bind(s, (char *)&sin, sizeof (sin));
```

With the above call, local addresses may be bound which are already in use. This does not violate the uniqueness requirement as the system still checks at connect time to be sure any other sockets with the same local address and port do not have the same foreign address and port (if an association already exists, the error EADDRINUSE is returned).

Local address binding by the system is currently done somewhat haphazardly when a host is on multiple networks. Logically, one would expect the system to bind the local address associated with the network through which a peer was communicating. For instance, if the local host is connected to networks 46 and 10 and the foreign host is on network 32, and traffic from network 32 were arriving via network 10, the local address to be bound would be the host's address on network 10, not network 46. This unfortunately, is not always the case. For reasons too complicated to discuss here, the local address bound may be appear to be chosen at random. This property of local address binding will normally be invisible to users unless the foreign host does not understand how to reach the address selected*.

5.5. Broadcasting and datagram sockets

By using a datagram socket it is possible to send broadcast packets on many networks supported by the system (the network itself must support the notion of broadcasting; the system provides no broadcast simulation in software). Broadcast messages can place a high load on a network since they force every host on the network to service them. Consequently, the ability to send broadcast packets has been limited to the super user.

To send a broadcast message, an Internet datagram socket should be created:

```
s = socket(AF_INET, SOCK_DGRAM, 0);
```

and at least a port number should be bound to the socket:

```
sin.sin_family = AF_INET;
sin.sin_addr.s_addr = INADDR_ANY;
sin.sin_port = MYPORT;
bind(s, (char *)&sin, sizeof (sin));
```

Then the message should be addressed as:

```
dst.sin_family = AF_INET;
dst.sin_addr.s_addr = INADDR_ANY;
dst.sin_port = DESTPORT;
```

and, finally, a sendto call may be used:

```
sendto(s, buf, buflen, 0, &dst, sizeof (dst));
```

Received broadcast messages contain the senders address and port (datagram sockets are anchored before a message is allowed to go out).

* For example, if network 46 were unknown to the host on network 32, and the local address were bound to that located on network 46, then even though a route between the two hosts existed through network 10, a connection would fail.

5.6. Signals

Two new signals have been added to the system which may be used in conjunction with the interprocess communication facilities. The SIGURG signal is associated with the existence of an "urgent condition". The SIGIO signal is used with "interrupt driven i/o" (not presently implemented). SIGURG is currently supplied a process when out of band data is present at a socket. If multiple sockets have out of band data awaiting delivery, a select call may be used to determine those sockets with such data.

An old signal which is useful when constructing server processes is SIGCHLD. This signal is delivered to a process when any children processes have changed state. Normally servers use the signal to "reap" child processes after exiting. For example, the remote login server loop shown in Figure 2 may be augmented as follows:

```
int reaper();
 ...
sigset(SIGCHLD, reaper);
listen(f, 10);
for (;;) {
	int g, len = sizeof (from);

	g = accept(f, &from, &len, 0);
	if (g < 0) {
		if (errno != EINTR)
			perror("rlogind: accept");
		continue;
	}
	...
}
 ...
#include <wait.h>
reaper()
{
	union wait status;

	while (wait3(&status, WNOHANG, 0) > 0)
		;
}
```

If the parent server process fails to reap its children, a large number of "zombie" processes may be created.

Reprinted from
Proceedings of the ACM SIGPLAN '82 Symposium on Compiler Construction
SIGPLAN Notices 17, 6 (June 1982), pp 120-126†

gprof: a Call Graph Execution Profiler[1]

by
Susan L. Graham
Peter B. Kessler
Marshall K. McKusick

Computer Science Division
Electrical Engineering and Computer Science Department
University of California, Berkeley
Berkeley, California 94720

Abstract

Large complex programs are composed of many small routines that implement abstractions for the routines that call them. To be useful, an execution profiler must attribute execution time in a way that is significant for the logical structure of a program as well as for its textual decomposition. This data must then be displayed to the user in a convenient and informative way. The **gprof** profiler accounts for the running time of called routines in the running time of the routines that call them. The design and use of this profiler is described.

1. Programs to be Profiled

Software research environments normally include many large programs both for production use and for experimental investigation. These programs are typically modular, in accordance with generally accepted principles of good program design. Often they consist of numerous small routines that implement various abstractions. Sometimes such large programs are written by one programmer who has understood the requirements for these abstractions, and has programmed them appropriately. More frequently the program has had multiple authors and has evolved over time, changing the demands placed on the implementation of the abstractions without changing the implementation itself. Finally, the program may be assembled from a library of abstraction implementations unexamined by the programmer.

Once a large program is executable, it is often desirable to increase its speed, especially if small portions of the program are found to dominate its execution time. The purpose of the **gprof** profiling tool is to help the user evaluate alternative implementations of abstractions. We developed this tool in response to our efforts to improve a code generator we were writing [Graham82].

[1]This work was supported by grant MCS80-05144 from the National Science Foundation.

The **gprof** design takes advantage of the fact that the programs to be measured are large, structured and hierarchical. We provide a profile in which the execution time for a set of routines that implement an abstraction is collected and charged to that abstraction. The profile can be used to compare and assess the costs of various implementations.

The profiler can be linked into a program without special planning by the programmer. The overhead for using **gprof** is low; both in terms of added execution time and in the volume of profiling information recorded.

2. Types of Profiling

There are several different uses for program profiles, and each may require different information from the profiles, or different presentation of the information. We distinguish two broad categories of profiles: those that present counts of statement or routine invocations, and those that display timing information about statements or routines. Counts are typically presented in tabular form, often in parallel with a listing of the source code. Timing information could be similarly presented; but more than one measure of time might be associated with each statement or routine. For example, in the framework used by **gprof** each profiled segment would display two times: one for the time used by the segment itself, and another for the time inherited from code segments it invokes.

Execution counts are used in many different contexts. The exact number of times a routine or statement is activated can be used to determine if an algorithm is performing as expected. Cursory inspection of such counters may show algorithms whose complexity is unsuited to the task at hand. Careful interpretation of counters can often suggest improvements to acceptable algorithms. Precise examination can uncover subtle errors in an algorithm. At this level, profiling counters are similar to debugging statements whose purpose is to show the number of times a piece of code is executed. Another view of such counters is as boolean values. One may be interested that a portion of code has executed at all, for exhaustive testing, or to check that one implementation of an abstraction completely replaces a previous one.

Execution counts are not necessarily proportional to the amount of time required to execute the routine or statement. Further, the execution time of a routine will not be the same for all calls on the routine. The criteria for establishing execution time must be decided. If a routine implements an abstraction by invoking other abstractions, the time spent in the routine will not accurately reflect the time required by the abstraction it implements. Similarly, if an abstraction is implemented by several routines the time required by the abstraction will be distributed across those routines.

Given the execution time of individual routines, **gprof** accounts to each routine the time spent for it by the routines it invokes. This accounting is done by assembling a *call graph* with nodes that are the routines of the program and directed arcs that represent calls from call sites to routines. We distinguish among three different call graphs for a program. The *complete call graph* incorporates all routines and all potential arcs, including arcs that represent calls to functional parameters or functional variables. This graph contains the other two graphs as subgraphs. The *static call graph* includes all routines and all possible arcs that are not calls to functional parameters or variables. The *dynamic call graph* includes only those routines and arcs traversed by the profiled execution of the program. This graph need not include all routines, nor need it include all potential arcs between the routines it covers. It may, however, include arcs to functional parameters or variables that the static call graph may omit. The static call graph can be determined from the (static) program text. The dynamic call graph is determined only by profiling an execution of the program. The complete call graph for a monolithic program could be determined by data flow analysis techniques. The complete call graph for programs that change during execution, by modifying themselves or dynamically loading or overlaying code, may never be determinable. Both the static call graph and the dynamic call graph are used by **gprof**, but it does not search for the complete call graph.

3. Gathering Profile Data

Routine calls or statement executions can be measured by having a compiler augment the code at strategic points. The additions can be inline increments to counters [Knuth71] [Satterthwaite72] [Joy79] or calls to monitoring routines [Unix]. The counter increment overhead is low, and is suitable for profiling statements. A call of the monitoring routine has an overhead comparable with a call of a regular routine, and is therefore only suited to profiling on a routine by routine basis. However, the monitoring routine solution has certain advantages. Whatever counters are needed by the monitoring routine can be managed by the monitoring routine itself, rather than being distributed around the code. In particular, a monitoring routine can easily be called from separately compiled programs. In addition, different monitoring routines can be linked into the program being measured to assemble different profiling data without having to change the compiler or recompile the program. We have exploited this approach; our compilers for C, Fortran77, and Pascal can insert calls to a monitoring routine in the prologue for each routine. Use of the monitoring routine requires no planning on part of a programmer other than to request that augmented routine prologues be produced during compilation.

We are interested in gathering three pieces of information during program execution: call counts and execution times for each profiled routine, and the arcs of the dynamic call graph traversed by this execution of the program. By post-processing of this data we can build the dynamic call graph for this execution of the program and propagate times along the edges of this graph to attribute times for routines to the routines that invoke them.

Gathering of the profiling information should not greatly interfere with the running of the program. Thus, the monitoring routine must not produce trace output each time it is invoked. The volume of data thus produced would be unmanageably large, and the time required to record it would overwhelm the running time of most programs. Similarly, the monitoring routine can not do the analysis of the profiling data (e.g. assembling the call graph, propagating times around it, discovering cycles, etc.) during program execution. Our solution is to gather profiling data in memory during program execution and to condense it to a file as the profiled program exits. This file is then processed by a separate program to produce the listing of the profile data. An advantage of this approach is that the profile data for several executions of a program can be combined by the post-processing to provide a profile of many executions.

The execution time monitoring consists of three parts. The first part allocates and initializes the runtime monitoring data structures before the program begins execution. The second part is the monitoring routine invoked from the prologue of each profiled routine. The third part condenses the data structures and writes them to a file as the program terminates. The monitoring routine is discussed in detail in the following sections.

3.1. Execution Counts

The **gprof** monitoring routine counts the number of times each profiled routine is called. The monitoring routine also records the arc in the call graph that activated the profiled routine. The count is associated with the arc in the call graph rather than with the routine. Call counts for routines can then be determined by summing the counts on arcs directed into that routine. In a machine-dependent fashion, the monitoring routine notes its own return address. This address is in the prologue of some profiled routine that is the destination of an arc in the dynamic call graph. The monitoring routine also discovers the return address for that routine, thus identifying the call site, or source of the arc. The source of the arc is in the *caller*, and the destination is in the *callee*. For example, if a routine A calls a routine B, A is the caller, and B is the callee. The prologue of B will include a call to the monitoring routine that will note the arc from A to B and either initialize or increment a counter for that arc.

One can not afford to have the monitoring routine output tracing information as each arc is identified. Therefore, the monitoring routine maintains a table of all the arcs discovered, with counts of the numbers of times each is traversed during execution. This table is accessed once per routine call. Access to it must be as fast as possible so as not to overwhelm the time required to execute the program.

Our solution is to access the table through a hash table. We use the call site as the primary key with the callee address being the secondary key. Since each call site typically calls only one callee, we can reduce (usually to one) the number of minor lookups based on the callee. Another alternative would use the callee as the primary key and the call site as the secondary key. Such an organization has the advantage of associating callers with callees, at the expense of longer lookups in the monitoring routine. We are fortunate to be running in a virtual memory environment, and (for the sake of speed) were able to allocate enough space for the primary hash table to allow a one-to-one mapping from call site addresses to the primary hash table. Thus our hash function is trivial to calculate and collisions occur only for call sites that call multiple destinations (e.g. functional parameters and functional variables). A one level hash function using both call site and callee would result in an unreasonably large hash table. Further, the number of dynamic call sites and callees is not known during execution of the profiled program.

Not all callers and callees can be identified by the monitoring routine. Routines that were compiled without the profiling augmentations will not call the monitoring routine as part of their prologue, and thus no arcs will be recorded whose destinations are in these routines. One need not profile all the routines in a program. Routines that are not profiled run at full speed. Certain routines, notably exception handlers, are invoked by non-standard calling sequences. Thus the monitoring routine may know the destination of an arc (the callee), but find it difficult or impossible to determine the source of the arc (the caller). Often in these cases the apparent source of the arc is not a call site at all. Such anomalous invocations are declared "spontaneous".

3.2. Execution Times

The execution times for routines can be gathered in at least two ways. One method measures the execution time of a routine by measuring the elapsed time from routine entry to routine exit. Unfortunately, time measurement is complicated on time-sharing systems by the time-slicing of the program. A second method samples the value of the program counter at some interval, and infers execution time from the distribution of the samples within the program. This technique is particularly suited to time-sharing systems, where the time-slicing can serve as the basis for sampling the program counter. Notice that, whereas the first method could provide exact timings, the second is inherently a statistical approximation.

The sampling method need not require support from the operating system: all that is needed is the ability to set and respond to "alarm clock" interrupts that run relative to program time. It is imperative that the intervals be uniform since the sampling of the program counter rather than the duration of the interval is the basis of the distribution. If sampling is done too often, the interruptions to sample the program counter will overwhelm the running of the profiled program. On the other hand, the program must run for enough sampled intervals that the distribution of the samples accurately represents the distribution of time for the execution of the program. As with routine call tracing, the monitoring routine can not afford to output information for each program counter sample. In our computing environment, the operating system can provide a histogram of the location of the program counter at the end of each clock tick (1/60th of a second) in which a program runs. The histogram is assembled in memory as the program runs. This facility is enabled by our monitoring routine. We have adjusted the granularity of the histogram so that program counter values map one-to-one onto the histogram. We make the simplifying assumption that all calls to a specific routine require the same amount of time to execute. This assumption may disguise that some calls (or worse, some call sites) always invoke a routine such that its execution is faster (or slower) than the average time for that routine.

When the profiled program terminates, the arc table and the histogram of program counter samples is written to a file. The arc table is condensed to consist of the source and destination addresses of the arc and the count of the number of times the arc was traversed by this execution of the program. The recorded histogram consists of counters of the number of times the program counter was found to be in each of the ranges covered by the histogram. The ranges themselves are summarized as a lower and upper bound and a step size.

4. Post Processing

Having gathered the arcs of the call graph and timing information for an execution of the program, we are interested in attributing the time for each routine to the routines that call it. We build a dynamic call graph with arcs from caller to callee, and propagate time from descendants to ancestors by topologically sorting the call graph. Time propagation is performed from the leaves of the call graph toward the roots, according to the order assigned by a topological numbering algorithm. The topological numbering ensures that all edges in the graph go from higher numbered nodes to lower numbered nodes. An example is given in Figure 1. If we propagate time from nodes in the order assigned by the algorithm, execution time can be propagated from descendants to ancestors after a single traversal of each arc in the call graph. Each parent receives some fraction of a child's time. Thus time is charged to the caller in addition to being charged to the callee.

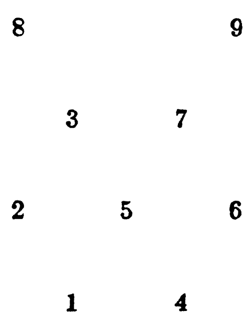
Let C_e be the number of calls to some routine, e, and C_e^r be the number of calls from a caller r to a callee e. Since we are assuming each call to a routine takes the average amount of time for all calls to that routine, the caller is accountable for C_e^r / C_e of the time spent by the callee. Let the S_e be the *selftime* of a routine, e. The selftime of a routine can be determined from the timing information gathered during profiled program execution. The total time, T_r, we wish to account to a routine r, is then given by the recurrence equation:

$$T_r = S_r + \sum_{r \text{ CALLS } e} T_e \times \frac{C_e^r}{C_e}$$

where r CALLS e is a relation showing all routines e called by a routine r. This relation is easily available from the call graph.

However, if the execution contains recursive calls, the call graph has cycles that cannot be topologically sorted. In these cases, we discover strongly-connected components in the call graph, treat each such component as a single node, and then sort the resulting graph. We use a variation of Tarjan's strongly-connected components algorithm that discovers strongly-connected components as it is assigning topological order numbers [Tarjan72].

Time propagation within strongly connected components is a problem. For example, a self-recursive routine (a trivial cycle in the call graph) is accountable for all the time it uses in all its recursive instantiations. In our scheme, this time should be shared among its call graph parents. The arcs from a routine to itself are of interest, but do not participate in time propagation. Thus the simple equation for time propagation does not work within strongly connected components. Time is not propagated from one member of a cycle to another, since, by definition, this involves propagating time from a routine to itself. In addition, children of one member of a cycle must be considered children of all members of the cycle. Similarly, parents of one member of the cycle



Topological ordering
Figure 1.

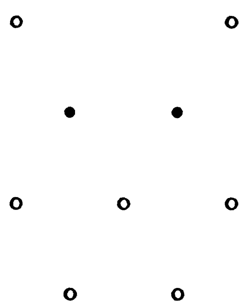
must inherit all members of the cycle as descendants. It is for these reasons that we collapse connected components. Our solution collects all members of a cycle together, summing the time and call counts for all members. All calls into the cycle are made to share the total time of the cycle, and all descendants of the cycle propagate time into the cycle as a whole. Calls among the members of the cycle do not propagate any time, though they are listed in the call graph profile.

Figure 2 shows a modified version of the call graph of Figure 1, in which the nodes labelled 3 and 7 in Figure 1 are mutually recursive. The topologically sorted graph after the cycle is collapsed is given in Figure 3.

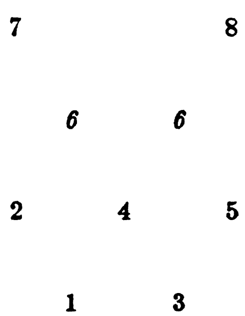
Since the technique described above only collects the dynamic call graph, and the program typically does not call every routine on each execution, different executions can introduce different cycles in the dynamic call graph. Since cycles often have a significant effect on time propagation, it is desirable to incorporate the static call graph so that cycles will have the same members regardless of how the program runs.

The static call graph can be constructed from the source text of the program. However, discovering the static call graph from the source text would require two moderately difficult steps: finding the source text for the program (which may not be available), and scanning and parsing that text, which may be in any one of several languages.

In our programming system, the static calling information is also contained in the executable version of the program, which we already have available, and which is in language-independent form. One can examine the instructions in the object program, looking for calls to routines, and note which routines can be called. This technique allows us to add arcs to those already in the



Cycle to be collapsed.
Figure 2.



Topological numbering after cycle collapsing.
Figure 3.

dynamic call graph. If a statically discovered arc already exists in the dynamic call graph, no action is required. Statically discovered arcs that do not exist in the dynamic call graph are added to the graph with a traversal count of zero. Thus they are never responsible for any time propagation. However, they may affect the structure of the graph. Since they may complete strongly connected components, the static call graph construction is done before topological ordering.

5. Data Presentation

The data is presented to the user in two different formats. The first presentation simply lists the routines without regard to the amount of time their descendants use. The second presentation incorporates the call graph of the program.

5.1. The Flat Profile

The flat profile consists of a list of all the routines that are called during execution of the program, with the count of the number of times they are called and the number of seconds of execution time for which they are themselves accountable. The routines are listed in decreasing order of execution time. A list of the routines that are never called during execution of the program is also available to verify that nothing important is omitted by this execution. The flat profile gives a quick overview of the routines that are used, and shows the routines that are themselves responsible for large fractions of the execution time. In practice, this profile usually shows that no single function is overwhelmingly responsible for the total time of the program. Notice that for this profile, the individual times sum to the total execution time.

5.2. The Call Graph Profile

Ideally, we would like to print the call graph of the program, but we are limited by the two-dimensional nature of our output devices. We cannot assume that a call graph is planar, and even if it is, that we can print a planar version of it. Instead, we choose to list each routine, together with information about the routines that are its direct parents and children. This listing presents a window into the call graph. Based on our experience, both parent information and child information is important, and should be available without searching through the output.

The major entries of the call graph profile are the entries from the flat profile, augmented by the time propagated to each routine from its descendants. This profile is sorted by the sum of the time for the routine itself plus the time inherited from its descendants. The profile shows which of the higher level routines spend large portions of the total execution time in the routines that they call. For each routine, we show the amount of time passed by each child to the routine, which includes time for the child itself and for the descendants of the child (and thus the descendants of the routine). We also show the percentage these times represent of the total time accounted to the child. Similarly, the parents of each routine are listed, along with time, and percentage of total routine time, propagated to each one.

index	%time	self	descendants	called/total called+self called/total	parents name children	index
		0.20	1.20	4/10	CALLER1	[7]
		0.30	1.80	6/10	CALLER2	[1]
[2]	41.5	0.50	3.00	10+4	EXAMPLE	[2]
		1.50	1.00	20/40	SUB1 <cycle1>	[4]
		0.00	0.50	1/5	SUB2	[9]
		0.00	0.00	0/5	SUB3	[11]

Profile entry for EXAMPLE.
Figure 4.

Cycles are handled as single entities. The cycle as a whole is shown as though it were a single routine, except that members of the cycle are listed in place of the children. Although the number of calls of each member from within the cycle are shown, they do not affect time propagation. When a child is a member of a cycle, the time shown is the appropriate fraction of the time for the whole cycle. Self-recursive routines have their calls broken down into calls from the outside and self-recursive calls. Only the outside calls affect the propagation of time.

The following example is a typical fragment of a call graph.

CALLER1 CALLER2

EXAMPLE

SUB1 SUB2 SUB3

The entry in the call graph profile listing for this example is shown in Figure 4.

The entry is for routine EXAMPLE, which has the Caller routines as its parents, and the Sub routines as its children. The reader should keep in mind that all information is given *with respect to EXAMPLE.* The index in the first column shows that EXAMPLE is the second entry in the profile listing. The EXAMPLE routine is called ten times, four times by CALLER1, and six times by CALLER2. Consequently 40% of EXAMPLE's time is propagated to CALLER1, and 60% of EXAMPLE's time is propagated to CALLER2. The self and descendant fields of the parents show the amount of self and descendant time EXAMPLE propagates to them (but not the time used by the parents directly). Note that EXAMPLE calls itself recursively four times. The routine EXAMPLE calls routine SUB1 twenty times, SUB2 once, and never calls SUB3. Since SUB2 is called a total of five times, 20% of its self and descendant time is propagated to EXAMPLE's descendant time field. Because SUB1 is a member of *cycle 1,* the self and descendant times and call count fraction are those for the cycle as a whole. Since cycle 1 is called a total of forty times (not counting calls among members of the cycle), it propagates 50% of the cycle's self and descendant time to EXAMPLE's descendant time field. Finally each name is followed by an index that shows where on the listing to find the entry for that routine.

6. Using the Profiles

The profiler is a useful tool for improving a set of routines that implement an abstraction. It can be helpful in identifying poorly coded routines, and in evaluating the new algorithms and code that replace them. Taking full advantage of the profiler requires a careful examination of the call graph profile, and a thorough knowledge of the abstractions underlying the program.

The easiest optimization that can be performed is a small change to a control construct or data structure that improves the running time of the program. An obvious starting point is a routine that is called many times. For example, suppose an output routine is the only parent of a routine that formats the data. If this format routine is expanded inline in the output routine, the overhead of a function call and return can be saved for each datum that needs to be formatted.

The drawback to inline expansion is that the data abstractions in the program may become less parameterized, hence less clearly defined. The profiling will also become less useful since the loss of routines will make its output more granular. For example, if the symbol table functions "lookup", "insert", and "delete" are all merged into a single parameterized routine, it will be impossible to determine the costs of any one of these individual functions from the profile.

Further potential for optimization lies in routines that implement data abstractions whose total execution time is long. For example, a lookup routine might be called only a few times, but use an inefficient linear search algorithm, that might be replaced with a binary search. Alternately, the discovery that a rehashing function is being called excessively, can lead to a different hash function or a larger hash table. If the data abstraction function cannot easily be speeded up, it may be advantageous to cache its results, and eliminate the need to rerun it for identical inputs. These and other ideas for program improvement are discussed in [Bentley81].

This tool is best used in an iterative approach: profiling the program, eliminating one bottleneck, then finding some other part of the program that begins to dominate execution time. For instance, we have used **gprof** on itself; eliminating, rewriting, and inline expanding routines, until reading data files (hardly a target for optimization!) represents the dominating factor in its execution time.

Certain types of programs are not easily analyzed by **gprof**. They are typified by programs that exhibit a large degree of recursion, such as recursive descent compilers. The problem is that most of the major routines are grouped into a single monolithic cycle. As in the symbol table abstraction that is placed in one routine, it is impossible to distinguish which members of the cycle are responsible for the execution time. Unfortunately there are no easy modifications to these programs that make them amenable to analysis.

A completely different use of the profiler is to analyze the control flow of an unfamiliar program. If you receive a program from another user that you need to modify in some small way, it is often unclear where the changes need to be made. By running the program on an example and then using **gprof**, you can get a view of the structure of the program.

Consider an example in which you need to change the output format of the program. For purposes of this example suppose that the call graph of the output portion of the program has the following structure:

CALC1 CALC2 CALC3

FORMAT1 FORMAT2

"WRITE"

Initially you look through the **gprof** output for the system call "WRITE". The format routine you will need to change is probably among the parents of the "WRITE" procedure. The next step is to look at the profile entry for each of parents of "WRITE", in this example either "FORMAT1" or "FORMAT2", to determine which one to change. Each format routine will have one or more parents, in this example "CALC1", "CALC2", and "CALC3". By inspecting the source code for each of these routines you can determine which format routine generates the output that you wish to modify. Since the **gprof** entry shows all the potential calls to the format routine you intend to change, you can determine if your modifications will affect output that should be left alone. If you desire to change the output of "CALC2", but not "CALC3", then formatting routine "FORMAT2" needs to be split into two separate routines, one of which implements the new format. You can then retarget just the call by "CALC2" that needs the new format. It should be noted that the static call information is particularly useful here since the test case you run probably will not exercise the entire program.

7. Conclusions

We have created a profiler that aids in the evaluation of modular programs. For each routine in the program, the profile shows the extent to which that routine helps support various abstractions, and how that routine uses other abstractions. The profile accurately assesses the cost of routines at all levels of the program decomposition. The profiler is easily used, and can be compiled into the program without any prior planning by the programmer. It adds only five to thirty percent execution overhead to the program being profiled, produces no additional output until after the program finishes, and allows the program to be measured in its actual environment. Finally, the profiler runs on a time-sharing system using only the normal services provided by the operating system and compilers.

8. References

[Bentley81]
Bentley, J. L., "Writing Efficient Code", Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania, CMU-CS-81-116, 1981.

[Graham82]
Graham, S. L., Henry, R. R., Schulman, R. A., "An Experiment in Table Driven Code Generation", SIGPLAN '82 Symposium on Compiler Construction, June, 1982.

[Joy79]
Joy, W. N., Graham, S. L., Haley, C. B. "Berkeley Pascal User's Manual", Version 1.1, Computer Science Division University of California, Berkeley, CA. April 1979.

[Knuth71]
Knuth, D. E. "An empirical study of FORTRAN programs", Software - Practice and Experience, 1, 105-133. 1971

[Satterthwaite72]
Satterthwaite, E. "Debugging Tools for High Level Languages", Software - Practice and Experience, 2, 197-217, 1972

[Tarjan72]
Tarjan, R. E., "Depth first search and linear graph algorithm," *SIAM J. Computing* **1**:2, 146-160, 1972.

[Unix]
Unix Programmer's Manual, "**prof** command", section 1, Bell Laboratories, Murray Hill, NJ. January 1979.

C O M M E N T S

ICON/UXB REFERENCE MANUAL Volume 2 **P/N 172-022-003**

Your comments and suggestions are appreciated and will help us to provide you with the very best in system and application documentation. Send your comments to the address at the bottom of this page. Users who respond will be entitled to free updates of this manual for one year.

1. How would you rate this manual for COMPLETENESS? (Please Circle)
   Excellent Poor
   5 ------------- 4 -------------- 3 ------------- 2 -------------- 1 ------------- 0

2. Is there any information that you feel should be included or removed?

   ____
   ____
   ____

3. How would you rate this manual for ACCURACY? (Please Circle)
   Excellent Poor
   5 ------------- 4 -------------- 3 ------------- 2 -------------- 1 ------------- 0

4. Indicate the page number and nature of any error(s) found in this manual.

   ____
   ____
   ____

5. How would you rate this manual for USABILITY? (Please Circle)
   Excellent Poor
   5 ------------- 4 -------------- 3 ------------- 2 -------------- 1 ------------- 0

6. Describe any format or packaging problems you have experienced with this manual and/or binder.

   ____
   ____
   ____
   ____

7. Do you have any general comments or suggestions regarding this publication or future publications?

   ____
   ____
   ____
   ____

Your Name ______________________________
Company ______________________________
Address ______________________________ Phone (____)__________
City & State ______________________________ Zip Code __________
Job Function ______________________________
Type of Equipment Installed: ______________________________