

IBM DB2 Information Integrator
OmniFind Edition



发行说明

版本 8.2 修订包 1

IBM DB2 Information Integrator
OmniFind Edition



发行说明

版本 8.2 修订包 1

在使用本资料及其支持的产品之前，请务必阅读『声明』中的一般信息。

本文档包含 IBM 的专利信息。它是根据许可协议提供的，并受版权法保护。本出版物包含的信息不包括任何产品保证，且本手册提供的任何声明不应作如此解释。

可以在线方式或通过您当地的 IBM 代表订购 IBM 出版物。

- 要以在线方式订购出版物，可访问“IBM 出版物中心”（IBM Publications Center），网址为 www.ibm.com/shop/publications/order
- 要查找您当地的 IBM 代表，可访问“IBM 全球联系人目录”（IBM Directory of Worldwide Contacts），网址为 www.ibm.com/planetwide

当您发送信息给 IBM 后，即授予 IBM 非专有权，IBM 可以它认为合适的任何方式使用或分发此信息，而无须对您承担任何责任。

© Copyright International Business Machines Corporation 2004, 2005. All rights reserved.

目录

1 关于发行说明	v	1 将 HTML 元数据映射到搜索字段	20
1 Web 上的技术支持信息	v	1 HTML 元数据字段映射页面	20
1 安装 DB2 II OmniFind Edition V8.2.1	1	1 从日语和中文文档中的空格中除去换行字符	21
1 卸载 DB2 II OmniFind Edition	3	1 为集合解析器和 Stellent 会话配置文档类型	21
1 本发行版新增内容	5	1 为集合解析器配置文档类型	22
1 新的受支持操作系统和数据源类型	5	1 为 Stellent 会话配置文档类型	24
1 指定定制集合标识	5	1 对查询处理所做的增强	28
1 搜寻进行了密码保护的 Web 站点	5	1 折叠来自同一 Web 站点的结果	28
1 搜寻 Web 页面中的框架集	8	1 搜索结果中的扩展突出显示	28
1 从 Web 文档的 META 标记中配置安全性令牌	8	1 搜索应用程序中的超时支持	29
1 对指向被禁止文档的链接中的锚点文本建立索引	9		
1 覆盖 Web 页面中的 no-index 和 no-follow 伪指令	9	已知局限性、问题和变通方法	31
1 为 Web 搜寻器配置全局搜寻空间	10	1 将 DB2 UDB AIX 版 tar 文件解包	31
1 配置多台 Lotus Notes 服务器	11	1 AIX 中的共享内存段	31
1 “指定要搜寻的 Notes 服务器”页面	12	1 在 Windows 上卸载之后重新启动系统	31
1 VeniceBridge 搜寻器的性能增强	13	1 Linux 和 AIX 的孤立 URLFetcher 进程	32
1 为 Windows 文件系统指定登录信息	14	1 服务器上的索引位置	32
1 外部名现在变为显示名	14	1 堆大小限制	32
1 配置和监视数据侦听器	14	1 具有大对象的 DB2 UDB 表	33
1 “系统”视图: “数据侦听器”页面	15	1 索引状态不显示	33
1 “监视系统”视图: “数据侦听器”页面	15	1 Linux 和 AIX 的用户标识许可权	33
1 “数据侦听器详细信息”页面	16	1 为 FileNet Content Services 启用直接文档访问	33
1 停止并启动数据侦听器组件	16	1 搜寻昵称表时的局限性	33
1 估计集合中的文档数	17		
1 在使用集合向导时自动检查系统资源	17	声明	35
1 监视解析器活动	17	1 商标	36
1 “解析器详细信息”页面	18		
		与 IBM 联系	39
		1 获取产品信息	39
		1 提供对文档的意见	39

1 关于发行说明

1 发行说明文档描述了 IBM® DB2® Information Integrator OmniFind™ Edition V8.2.1 (V8.2
1 加修订包 1) 的本发行版中的新增内容、已知局限性、问题、变通方法和文档更新。

1 IBM DB2 Information Integrator OmniFind Edition V8.2.1 尚未经过翻译。对该产品的
1 界面和消息所做的更改只提供英语版本。

1 Web 上的技术支持信息

1 DB2 II OmniFind Edition 产品最近已被重命名为 WebSphere® Information Integrator
1 OmniFind Edition。您在与产品相关的 Web 页面上可能会看到对 WebSphere Information
1 Integrator OmniFind Edition 的引用，但是产品界面、文档和故障诊断信息（技术说明）
1 仍反映 DB2 品牌。

1 您可以在下列 Web 站点上找到技术支持信息：

1 **IBM DB2 Information Integrator OmniFind Edition**

1 www.ibm.com/software/data/integration/db2ii/supportomni.html

1 **IBM WebSphere Information Integrator**

1 <http://www.ibm.com/software/data/integration/db2ii/support.html>

1 **IBM WebSphere Application Server**

1 <http://www.ibm.com/software/webservers/appserv/was/support/>

1 **IBM DB2 通用数据库™ Linux™ 版、UNIX® 版和 Windows® 版版本 8 产品**

1 www.ibm.com/software/data/db2/udb/winos2unix/support

1 安装 DB2 II OmniFind Edition V8.2.1

1 使用 DB2 II OmniFind Edition 安装启动板，您既可以安装完整的产品（V8.2.1），也
1 可以仅仅在现有安装上安装 V8.2.1 升级。

1 在安装 V8.2.1 升级之前，应该对现有安装进行备份。请参阅《管理企业搜索》以了解
1 有关备份和恢复系统的信息。在安装期间，请指定现有数据目录的路径。安装启动板
1 将使用现有安装信息来安装升级。

1 如果要安装完整的版本 8.2.1 产品，并且先前已安装了 DB2 II OmniFind Edition V8.2，
1 则必须先卸载 DB2 II OmniFind Edition。既不能在现有安装上安装完整的产品，也不
1 能先备份现有系统并安装 V8.2.1，然后将 V8.2 配置设置复原到 V8.2.1 系统中。

1 要安装 V8.2.1:

1 1. 启动安装启动板。

选项	描述
Linux	作为 root 用户登录并输入 ./install-linux.bin
AIX	作为 root 用户登录并输入 ./install-aix.bin
Windows	您必须具有管理员权限。双击文件 install-win32.exe，或者输入 install-win32.exe

1 2. 在安装启动板打开后，单击**安装产品**。

1 3. 要安装 V8.2.1 升级:

- 1 a. 在升级安装窗口中选择**升级现有的 DB2 II OmniFind Edition 安装**。如果选择
1 了此选项，则现有的 V8.2 系统配置将被迁移到 V8.2.1 系统。
- 1 b. 输入现有数据目录的路径。单击**下一步**。
- 1 c. 在“总结”窗口中查看设置。单击**下一步**以开始进行安装。

1 4. 要安装完整的 V8.2.1 产品:

- 1 a. 选择**安装新版本的 DB2 II OmniFind Edition**。如果先前已安装了 DB2 II
1 OmniFind Edition V8.2，则选择此选项前务必将其卸载。
- 1 b. 遵循《企业搜索安装指南》中的过程来安装产品。

1 卸载 DB2 II OmniFind Edition

1

卸载程序将除去企业搜索组件。也可以除去所有数据系统配置。

1

在卸载 DB2 II OmniFind Edition 之前，您应该对系统进行备份。

1

对于多服务器配置，需在每台服务器上运行卸载程序。要卸载 DB2 II OmniFind Edition:

1

1

1. 对于 Linux 和 AIX，作为 root 用户登录。

1

2. 浏览或切换到安装目录 (ES_INSTALL_ROOT) 中的 _uninst 目录并启动卸载程序:

1

选项	描述
Linux 或 AIX	输入 ./uninstall.bin
Windows	双击文件或输入 uninstall.exe

1

1

1

3. 可选: 要除去所有企业搜索管理员配置信息，请选择“除去全部数据和配置文件”复选框。

1

1

警告: 如果选择了此复选框，则将除去系统上所有集合的所有数据。

1

在 Windows 上，必须先重新启动计算机，然后才能重新安装 DB2 II OmniFind Edition。

1

1 本发行版新增内容

DB2 II OmniFind Edition 的新增功能部件和功能包括改进了搜索管理、提高了搜索质量和性能以及支持 Microsoft® Windows 2003。

1 新的受支持操作系统和数据源类型

DB2 II OmniFind Edition 支持 32 位 Windows Server 2003 操作系统 (Standard Server、Advanced Server 和 Datacenter Server)，并且可以搜寻 Microsoft SQL Server 数据库。

要安装 Windows Server 2003，请遵循《企业搜索安装指南》中提供的有关 Windows Server 2000 的那些指示信息。

要搜寻 SQL Server 数据库，请使用企业搜索管理控制台来创建 DB2 搜寻器并配置要搜寻的数据源。

1 指定定制集合标识

在创建集合时，您可以选择使用系统创建的内部标识，也可以指定定制集合标识。

如果您指定定制集合标识，则搜索应用程序可以使用企业搜索应用程序编程接口来通过此标识调用该集合，而不是使用系统创建的有可能含义模糊的标识。

如果您指定了定制集合标识，则创建集合时也必须这样做。以后，您不能编辑该集合并更改此标识。

要指定您想要使用的集合标识的类型：

1. 启动企业搜索管理控制台。如果已启用安全性，则需作为具有企业搜索管理员权限的用户登录。

2. 根据您是否想要使用向导来创建集合，单击**创建集合**或**集合向导**。

3. 在**集合标识**字段中选择下列其中一个选项：

缺省集合标识

选择此选项以允许系统为该集合创建内部标识。

定制集合标识

如果您要为该集合指定自己的标识，则选择此选项。您指定的值只能包含字母数字字符和下划线字符。

1 搜寻进行了密码保护的 Web 站点

DB2 II OmniFind Edition Web 搜寻器可以通过使用 HTTP 基本认证并通过将信息保留在 cookie 中来搜寻进行了密码保护的 Web 站点。

HTTP 基本认证支持

作为企业搜索管理员，在开始进行搜寻前，您可以在名为 `httpauth.ini` 的配置文件中标识要搜寻的进行了密码保护的 Web 站点并指定必需的凭证。

`httpauth.ini` 文件必须遵循下列格式化规则：

- `httpauth.ini` 文件的内容是文本，它不能包含除空格以外的非打印字符。数据条目由换行符分隔。
- 此文件可以包含任意顺序的权限条目、注释和空白行。
- 注释是第一个非空格字符是 `#` 的行。
- 注释不能出现在带有非注释文本的行上。
- 每个权限条目都必须在一行上按以下顺序包含以下信息：
 - 目标 URL（需要权限的服务器文件系统的根）
 - 对该服务器有意义的域名
 - 用于登录的用户名
 - 用于登录的密码
 - 重新认证 URL
 - 分号分隔的安全性令牌列表，此列表可以是空的，但此字段仍必须存在，并且在第 5 个字段后面有一个逗号分隔符
- 同一行上的各个字段由逗号分隔。逗号两旁允许有空格。
- 每个条目中的密码的加密方法都与 DB2 II OmniFind Edition 其余部分中使用的加密方法相同。
- 对于服务器文件系统中位于指定 URL 下方的所有 URL，假定该 URL 的权限有效。

您可以使用名为 `httpauthedit` 的命令行实用程序来修改 `httpauth.ini` 文件。使用此实用程序，您可以在创建或编辑条目时输入明文（未加密的）密码，但是此实用程序在将密码写至磁盘前将对它们进行加密。要更改密码，必须删除条目并再次添加该条目。

要启动 `httpauthedit` 实用程序，请在一行上输入以下命令：

```
java -classpath ES_INSTALL_ROOT/lib/URLFetcher.jar:esinstall/lib/es.oss.jar  
com.ibm.es.wc.uf.HTTPAuthEdit configdir
```

`ES_INSTALL_ROOT` 是 DB2 II OmniFind Edition 安装目录，`configdir` 是搜寻器实例的主配置目录，将在该目录中创建（或修改）`httpauth.ini` 文件。

主配置目录位于 `ES_NODE_ROOT/master_config/CR_INST_DIR` 中，其中 `ES_NODE_ROOT` 是 `ES_CFG` 文件中的 `NodeDirectory` 值，`CR_INST_DIR` 是为新的 Web 搜寻器实例创建的目录（管理软件对此目录指定了任意的名称）。您务必对正确的实例配置执行操作。在搜寻器下次启动之前，您所做的更改不会对其产生影响。

`httpauthedit` 实用程序接受下列命令：

表 1. `HTTPAUTHEDIT` 实用程序接受的命令

命令	描述
?	帮助：描述命令集。
b	浏览：打印文件内容。

表 1. HTTPAUTHEDIT 实用程序接受的命令 (续)

命令	描述
a	<p>添加: 将新的权限条目添加到文件中。此命令需要 5 个或 6 个自变量, 这些自变量位于一行上, 并且用逗号分隔。例如: <code>http://www.ibm.com/solutions/, solutions, wwwsearch, foobar, xyz123, http://www.ibm.com/solutions/, group1;group2;admin3</code></p> <ul style="list-style-type: none"> • 目标 URL: <code>http://www.ibm.com/solutions/</code> • 域名: <code>solutions</code> • 用户名: <code>wwwsearch, foobar</code> • 密码: <code>xyz123</code> • 重新认证 URL: <code>http://www.ibm.com/solutions/</code> • 可选: 安全性令牌: <code>group1;group2;admin3</code>
d	<p>删除: 从文件中除去权限条目。此命令需要一个自变量: 要从文件中除去的认证条目的目标 URL, 如 <code>http://www.ibm.com/solutions/</code>。</p>
q	<p>退出: 退出程序。</p>

重新认证 URL 可以与目标 URL 相同:

- 重新认证 URL 必须是这样的 URL: 如果在没有权限的情况下请求此 URL, 它将从服务器触发 401 响应并进行 HTTP 基本提问, 以获取客户机想要提交权限的域。
- 重新认证 URL 必须在不进行重定向的情况下触发此响应 (重新认证 URL 必须是真实的目标页面, 而不能是诸如 301 和 302 之类的页面, 那些页面将客户机重定向至另一个页面)。这个重新认证 URL 有时与它所授权的域的根 URL 不相同。重新认证 URL 是否与根 URL 相同取决于 Web 站点的内部逻辑。Web 站点管理员可以提供详细信息。

字段值中的前导空格和尾部空格将被除去。如果您输入了无效的输入, `httpauthedit` 将显示错误消息。

当您输入 `quit` 命令时, 您所做的更改将被保存下来。没有显式的 `save` 命令。要取消更改, 可以通过在不使用 `quit` 的情况下退出来杀死该程序。

Cookie

Web 搜寻器将保留从服务器接收到的 cookie 并在该搜寻器实例的生命期内使用它们。当搜寻器关闭时, 它将把所有未到期的 cookie 保存到磁盘并在下次搜寻会话开始时重新装入它们。cookie 文件名为 `cookies.ini`, 它位于搜寻器数据目录中。

搜寻器中的 cookie 处理与 HTTP 认证 (`httpauth.ini` 文件) 无关, 但是, 如果 Web 服务器使用 cookie 来保护内容, 则也可能需要进行认证, 甚至对于手工提供的 cookie 亦如此。Web 站点管理员应该能够提供有关搜寻他们的站点时需要提供的内容的信息。

作为企业搜索管理员, 您除了可以添加搜寻器找到的 cookie 以外, 还可以添加别的 cookie, 方法是使用文本编辑器将它们追加到 `cookies.ini` 文件中。由于搜寻器每次关闭时都会重写此文件, 所以, 您应该将手工输入的 cookie 保存在另一个文件中, 并在开始进行搜寻前根据需要添加它们。(搜寻器不会废弃未到期的 cookie, 但是, 当由于发生问题而导致无法写整个 cookie 集合时, 您不会希望丢失手工指定的 cookie。)

1 要为 Web 搜寻器指定 cookie, 请编辑 cookies.ini 文件。在 cookies.ini 文件中, 每一
1 行都是一个条目。允许使用空白行和注释, 但不会将它们保留下来。每个条目都具有
1 以下格式。每个条目都必须位于一行上:

```
1 Cookie(53,40)ASPSESSIONIDQSQTACSD=SLNSIDFNLSIDNFLSNDNLSNL;path=/  
1 https://www.ibm.com:443/help/solutions/
```

1 *Cookie* 是必需的关键字, 它指示条目开始。数字 53 是 cookie 名-值对以及要随该
1 cookie 一起发送的任何其它内容(路径和安全设置等等)的长度(以字符计)。数字 40
1 是用于进行验证的相关 URL 的长度。

1 对于 ASPSESSION. . . , 这 53 个字符是要发送至始发服务器的 cookie 的内容。此字
1 符串后面跟着逗号(,)分隔符。

1 URL https://www.ibm.com:443/help/solutions/ 用来通过提供一些信息(例如域名)
1 来验证该 cookie。对于手工添加的 cookie, 您应该提供能够满足 cookie 安全性和隐私
1 限制的验证 URL。

1 搜寻 Web 页面中的框架集

1 Web 搜寻器现在能够搜寻使用 HTML 框架创建的 Web 页面。

1 在 DB2 II OmniFind Edition V8.2 中, Web 搜寻器无法解析框架集或者从框架中抽取
1 指向其它 HTML 页面的链接。

1 从 Web 文档的 META 标记中配置安全性令牌

1 新的 Web 搜寻器安全性选项使您能够在配置文件中指定 META 标记的名称。那些标
1 记的内容被用作安全性令牌, 这些令牌用来控制对包含那些标记的文档进行的访问。

1 在 DB2 II OmniFind Edition V8.2 中, 安全性令牌是以全局方式定义的, 根据是否为
1 Web 搜寻器启用了文档级别的安全性的不同, 这些安全性令牌将应用于所有文档或者不
1 应用于任何文档。

1 现在, 您可以通过指定 META 标记作为安全性令牌来控制对文档进行的访问。要使用
1 基于 META 标记的安全性, 您必须为 Web 搜寻器启用文档级别安全性, 并且必须编
1 辑一个配置文件以指定 META 标记名(无法在管理控制台中指定 META 标记名)。

1 要指定作为安全性令牌使用的 META 标记:

- 1 1. 作为企业搜索管理员(此用户标识和密码是在 DB2 II OmniFind Edition 的安装期
1 间配置的)登录到搜寻器服务器。
- 1 2. 使用文本编辑器来编辑 master_config 目录树中的 crawl.properties 文件。
- 1 3. 如果有必要的话, 插入以下参数: security_meta_tags。(如果 security_meta_tags
1 参数已存在, 则缺省值为 nil。)
- 1 4. 在逗号定界的列表中指定 META 标记名。例如:

```
1 security_meta_tags=security,acl,access
```

1 标记名的规则如下:

- 1 • 从配置文件读取的标记名将被转换为小写。

- 在搜寻到的文档中找到的标记名在与配置文件中的名称作比较之前将被转换为小写。
 - 在名称列表中，唯一的分隔符是逗号（空格不对名称进行定界）。
 - 名称中的嵌入空白将被保留下来；前导和尾部空白将被除去。
5. 保存文件并退出。

如果为 `security_meta_tags` 参数配置了标记名，则在所下载每个 HTML 页面的处理期间，将针对 `security_meta_tags` 列表来检查文档的 META 标记。

要了解有关使用安全性令牌来控制文档访问的完整信息，请参阅《管理企业搜索》中的『文档级别安全性』主题。

对指向被禁止文档的链接中的锚点文本建立索引

如果文档包含的链接指向不允许 Web 搜寻器搜寻的文档，则您可以指定是否要在索引中为那些链接保留锚点文本。

`robots.txt` 文件中的设置或所指定的元标记中的设置可以使 Web 搜寻器无法访问 Web 站点上的某些文档。如果允许 Web 搜寻器搜索的文档包含指向被禁止文档的链接，则您可以在“高级 Web 搜寻器属性”页面上指定如何处理那些链接的锚点文本：

1. 如果要将指向被禁止文档的链接的锚点文本包括在索引中，则选取**对指向被禁止文档的链接中的锚点文本建立索引**复选框。（锚点文本是 Web 文档中出现在 `<A>` 与 `` 标记之间的描述性文本。如果选取此复选框，则通过搜索指向被禁止页面的链接的锚点文本中的文本可以找到那些页面。
2. 如果要将被禁止链接的锚点文本排除在索引之外（并排除被禁止文档），则清除**对指向被禁止文档的链接中的锚点文本建立索引**复选框。

您可以在创建 Web 搜寻器时指定高级 Web 搜寻器属性，也可以在编辑现有 Web 搜寻器的搜寻空间时指定这些属性。

覆盖 Web 页面中的 no-index 和 no-follow 伪指令

您可以在配置文件中指定规则以便控制 Web 搜寻器对于包含 `no-follow` 或 `no-index` 伪指令的页面是否遵循指向那些页面的链接或者对那些页面建立索引。在企业搜索管理控制台中，未提供对此功能的支持。

某些 Web 页面指定了 `no-index` 或 `no-follow` 伪指令，这些伪指令指示机器人（如 Web 搜寻器）不要将那些页面的内容包括在索引中、不要遵循那些页面中的链接或者不要执行这些操作中的任何一项操作。

在 DB2 II OmniFind Edition V8.2 中，这些限制是由 Web 站点管理员或内容创建程序强制实施的。但是，通过对这些设置进行控制，可以提高搜寻质量。例如，对于包含数以千计链接但未包含有用内容的目录页面集合，应该对那些页面执行搜寻并遵循它们的链接，但不应该对那些页面本身建立索引。

并且，搜寻器管理员有时候不想搜寻层次结构低于某个级别的内容，但所需的叶子页包含链接并且未包含 `no-follow` 伪指令。由于这些页面中的某些页面是自动生成的并且没有所有者，所以可能能够也可能无法让此类页面的内容所有者插入必需的伪指令。

V8.2.1 搜寻器在它的配置目录中查找名为 `followindex.rules` 的新配置文件。此文件可以包含任意数目的 `follow` 和 `index` 规则:

```
forbid follow URLprefix
allow follow URLprefix
forbid index URLprefix
allow index URLprefix
```

不能在 `follow` 和 `index` 规则中指定 IP 地址和 DNS 主机名; 只能指定 URL 前缀 (包括通配符)。正如指定搜寻规则时顺序十分重要一样, 这里的顺序也十分重要。如果定义的规则明确允许或禁止遵循链接或建立索引, 则它将覆盖其它设置, 包括目标文档中的那些设置。

为 Web 搜寻器配置全局搜寻空间

您可以在配置文件中指定规则以便为 Web 搜寻器创建全局搜寻空间并且更好地控制从索引中除去 URL 的行为。在企业搜索管理控制台中, 未提供对此功能的支持。

每个 Web 搜寻器都配置有搜寻空间, 搜寻空间定义要搜寻的或不要搜寻的 URL。搜寻空间中已被发现的 URL 将被保留下来 (保留在数据库中) 以用于将来的搜寻; 那些不在搜寻空间中的 URL 将被废弃。如果搜寻器在开始时使用的是空数据库, 则只要搜寻器保持运行, 搜寻空间定义与数据库就会保持一致。

有时, 搜寻器会停止, 并且它的搜寻空间会被缩小 (通过添加用于禁止页面的新规则)。当搜寻器重新启动时, 它的搜寻空间定义将与数据库不一致: 数据库将包含不存在于新的并且更小的搜寻空间中的 URL - 在这些 URL 中, 某些已被搜寻, 某些尚未被搜寻。在 DB2 II OmniFind Edition V8.2 中, Web 搜寻器通过将这些 URL 的 HTTP 返回码更改为 760 (搜寻空间排除) 并编写用于从索引中除去现在已被排除的页面的指令来恢复一致性。它这样做的原因是: 它假定它是集合中的唯一 Web 搜寻器, 并且它的搜寻空间标识了所有已被允许包括在索引中的 Web 文档。

有时候, 在搜寻器对大型搜寻空间操作一段时间以后, 管理员决定将该搜寻空间划分给两个或更多个搜寻器使用 (这有可能是为了确保某些部分的搜寻频率高于其余部分)。将创建新的搜寻器, 每个搜寻器都有自己的独立数据库表 (这些表最初是空的), 并且每个搜寻器都使用原始搜寻空间的不同部分。于是, 原始搜寻器的搜寻空间缩小为在除去被新搜寻器所搜寻的那些部分后剩下的那部分空间。

如果原始搜寻器是 V8.2 搜寻器, 则它将按照上面描述的方式恢复与其数据库保持一致性。这并不是管理员所希望的情况: 不从索引中除去已移动的 URL, 而仅仅是由其它搜寻器搜寻它们。

DB2 II OmniFind Edition V8.2.1 提供了具有更高级别的第二个搜寻空间, 后者表示不会被当前搜寻器搜寻但也不会被删除的 URL。(在任何一个空间中都不存在的 URL 在发现期间仍将被拒绝, 并且在重新扫描时将被从索引中除去, 这与以前完全一样。)

第二个搜寻空间是全局搜寻空间, 它是由名为 `global.rules` 的文件的内容定义的, 该文件位于搜寻器的配置目录中。如果此文件存在, 则搜寻器初始化期间将读取此文件。如果没有 `global.rules` 文件, 则搜寻器将对单级别搜寻空间执行操作 (与 V8.2 相同)。如果定义了全局空间, 则搜寻器将象以前那样通过规则包括或排除 URL, 但仅发送用于当 URL 在任何一个空间中都不存在时从索引中除去该 URL 的指令。

1 global.rules 文件的语法与局部 crawl.rules 文件相同，但只有域名规则是有意义的。（因
1 此，只能通过使用 DNS 主机名来在搜寻器之间对搜寻空间进行分区，而不能使用 IP 地
1 址或 URL 前缀模式来执行此操作。）

1 只要 global.rules 文件存在，就会启用全局搜寻空间功能（只要它不存在，就会禁用该功
1 能）。局部搜寻空间中（crawl.rules 文件中定义的）URL 前缀规则或 IP 地址规则排
1 除的 URL 不受全局搜寻空间影响；它们仍将被排除。全局搜寻空间只能用来防止从索
1 引中除去被局部域规则从一个搜寻器的域空间中排除的 URL。

1 规则的工作方式如下所示：

- 1 1. 如果局部前缀规则或地址规则排除了搜寻器数据库中的一个 URL，则对其指定返回
1 码 760 并将其从索引中除去。它的搜寻空间处理到此为止。该 URL 将不再被搜寻。
- 1 2. 在步骤 1 之后：如果局部域规则排除了搜寻器数据库中的一个 URL，并且没有全局
1 搜寻空间，则对其指定返回码 760 并将其从索引中除去。它的搜寻空间处理到此为
1 止。该 URL 将不再被搜寻。
- 1 3. 在步骤 1 之后：如果局部域规则排除了搜寻器数据库中的一个 URL，但全局搜寻空
1 间中的规则明确允许该 URL，则对其指定返回码 761。搜寻器不会再次搜寻该
1 URL，但也不会从索引中除去它（假定它存在于另外某个搜寻器的局部搜寻空间
1 中）。
- 1 4. 在步骤 1 之后：如果局部域规则排除了搜寻器数据库中的一个 URL，并且全局搜寻
1 空间中的规则未明确允许该 URL，则对其指定返回码 760，并从索引中除去该
1 URL。它的搜寻空间处理到此为止。

1 下面这一点很重要：由于查询全局搜寻空间的目的仅仅是为了防止删除已被局部搜
1 寻空间排除的 URL，因此，当没有规则适用于候选 URL 时，全局搜寻空间的缺省
1 结果是禁止。这与局部搜寻空间的缺省结果相反。

1 注意，global.rules 文件需存在于共享全局搜寻空间的每个搜寻器的 master_config 目录
1 中，您需要谨慎地编辑全局文件的所有副本以及各个局部 crawl.rules 文件以使它们相互
1 一致。

1 配置多台 Lotus Notes 服务器

1 您可以将来自多台 Lotus Notes® 服务器的数据库添加到同一个搜寻空间中，而不必配
1 置其它搜寻器来搜寻它们。在 DB2 II OmniFind Edition V8.2 中，每个 Notes 搜寻器
1 都限于只能搜寻一台 Lotus Notes 服务器。

1 在配置 Notes 搜寻器以搜寻一台 Lotus Notes 服务器上的数据库之后，执行下列操作来
1 将来自其它 Lotus Notes 服务器的数据库添加到同一个搜寻空间中：

- 1 1. 在“搜寻”页面上，选择您想要配置的 Notes 搜寻器并单击**搜寻空间**图标。
- 1 2. 在“Notes 搜寻空间”页面上，单击**添加数据库**。
- 1 3. 使用以下窗口来指定有关要搜寻的 Lotus Notes 服务器的信息：第 12 页的『“指定要
1 搜寻的 Notes 服务器”页面』

1 如果您需要搜寻器向导所显示的后续页面的帮助，请单击**帮助**。

1 **提示：**要更改搜寻器在访问 Lotus Notes 服务器时使用的密码，请在“Notes 搜寻空间”
1 页面上选择服务器，然后单击**编辑服务器**。（以前，当 Notes 搜寻器限于只能搜
1 寻一台 Lotus Notes 服务器时，此操作是在“Notes 搜寻空间”页面顶部提供的。）

“指定要搜寻的 Notes 服务器”页面

使用“指定要搜寻的 Notes 服务器”页面来标识您要为现有搜寻空间配置的 Lotus Notes 服务器。如果您正在将服务器添加到搜寻空间中，则指定一些信息以使搜寻器能够访问该服务器上的数据库。

要了解有关设置 Notes® 环境以使其可以被 Notes 搜寻器搜寻的信息，请参阅《管理企业搜索》。

要点: 在您单击下一步后，系统将验证它是否能够访问您指定的 Notes 服务器。如果它无法访问该服务器，则将显示错误消息。验证并更正服务器名和协议信息，然后再次单击下一步。

字段和控件

现有服务器

指定您想要从搜寻空间中已存在的 Lotus Notes 服务器添加数据库。

如果选中此选项，则请选择要从中添加数据库的 Lotus Notes 服务器。

新服务器

指定您想要将关于新的 Lotus Notes 服务器的信息添加到搜寻空间中并从该服务器中选择要搜寻的数据库。

如果选中此选项，则指定一些选项以使 Notes 搜寻器能够访问该服务器上的数据库。

Lotus Notes 服务器名

输入 Domino® 服务器的标准主机名称，如 server1.ibm.com，该服务器负责主管您要让此搜寻器搜寻的 Notes 数据库 (.nsf 文件)。

使用 NRPC 协议的 Notes 服务器的缺省端口是 1352。使用 DIIOP 协议的 Notes 服务器的缺省端口是 80。如果服务器未侦听缺省端口，则指定端口号（例如 server1.ibm.com:1353 或 server1.ibm.com:81）。

协议 指定关于服务器所使用的通信协议的信息：

Notes 远程过程调用 (NRPC)

如果 Domino 服务器使用 Notes 远程过程调用协议 (NRPC) 来在服务器与 Web 浏览器之间交换信息，则选中此选项。

要点: 如果您现在已经为此集合配置了任何使用 NRPC 协议的 Notes 搜寻器，则先停止它们，然后再配置有关此 Notes 搜寻器的信息。使用 NRPC 的企业搜索发现进程与使用 NRPC 的搜寻器进程不能并发运行。（发现进程为搜寻器提供有关可供搜寻的来源的信息。）

如果选中了 NRPC 选项，则指定下列信息：

Lotus Notes 用户标识文件

输入一个 Notes 用户标识文件的绝对路径，此用户标识有权访问此服务器上的数据库。例如，在 UNIX 机器上，可以输入 /home/username/username.id。

每个 Notes 用户都有一个标识文件，该文件唯一地标识该用户。请确保您在此处指定的标识文件的副本存在于搜寻器机器上。企业搜索管理员标识对该 Notes

1 用户标识文件以及该文件的存储目录必须具有读访问
1 权限。（企业搜索管理员的用户标识是在 DB2 II
1 OmniFind Edition 的安装期间指定的。）

1 **密码** 输入指定的 Notes 用户标识文件的密码。

1 **Domino 因特网 ORB 间协议 (DIIOIP)**

1 如果 Domino 服务器使用 Domino 因特网 ORB 间协议
1 (DIIOIP) 来在服务器与 Web 浏览器之间交换信息，则选中此
1 选项。要使用此协议，必须在 Domino 服务器上启用 HTTP 和
1 DIIOIP 任务。

1 如果选中了此选项，则指定下列信息：

1 **Lotus Notes 用户标识**

1 输入有权访问此服务器上的数据库的 Notes 用户标
1 识，如 user1/Hometown/IBM。

1 **密码** 输入在 Domino 目录中为此用户标识指定的 Notes 因
1 特网密码。（在为 Notes 服务器配置 DIIOIP 时，其中
1 一个步骤就是为有权访问该服务器的用户创建因特网
1 密码。）

1 **VeniceBridge 搜寻器的性能增强**

1 Ventica VeniceBridge 现在是名为 DB2 Information Integrator Content Edition (DB2 II
1 Content Edition) 的 IBM 产品。由于 DB2 II OmniFind Edition 产品界面和文档仍将
1 其称为 VeniceBridge，所以，这些发行说明也使用术语 VeniceBridge 来描述对用户界面
1 进行的增强。

1 为了提高重新搜寻性能，现在，您在为各个 VeniceBridge 项类配置规则时可以选择数据
1 图。数据图是一个 DB2 II Content Edition 构造，它提供有关项类的 VeniceBridge 搜
1 寻器信息，如创建日期、修订日期、排序键和内容大小等等。通过使数据图与搜寻器
1 配置中的项类相关联，可以帮助确保搜寻器能够快速地重新搜寻该项类中的内容。

1 对于 Documentum 和 FileNet Content Services，由于 VeniceBridge 搜寻器有缺省数据
1 图，所以您可以在不指定数据图的情况下利用此功能。

1 **要点:**

1 要利用这些性能增强功能，必须运行 DB2 II Content Edition 修订包 1。

1 要使数据图与现有 VeniceBridge 搜寻空间中的项类相关联:

- 1 1. 编辑 VeniceBridge 搜寻器所属的集合。
- 1 2. 单击“搜寻”页面。
- 1 3. 找到 VeniceBridge 搜寻器名并单击**搜寻空间**按钮。
- 1 4. 在“VeniceBridge 搜寻空间”页面上，找到您想要配置的项类并单击**编辑**。
- 1 5. 在“为 VeniceBridge 项类配置规则”页面上的**数据图**字段中，选择要与此项类配合使
1 用的数据图的名称。
- 1 6. 单击**确定**。

如果您正在创建 VeniceBridge 搜寻器，或者正在将项类添加到现有搜寻空间中，则您在“为 VeniceBridge 项类配置规则”页面上指定其它选项时选择数据图名。

为 Windows 文件系统指定登录信息

在为 Windows 子目录配置搜寻规则时，您可以指定用户标识和密码，此信息使搜寻器能够访问该子目录中的数据。

要为现有 Windows 文件系统搜寻空间中的子目录配置登录凭证：

1. 编辑 Windows 文件系统搜寻器所属的集合。
2. 单击“搜寻”页面。
3. 找到 Windows 文件系统搜寻器名并单击**搜寻空间**按钮。
4. 在“Windows 搜寻空间”页面上，找到您想要配置的子目录并单击**编辑**。
5. 在“为 Windows 子目录配置规则”页面上，在**用户标识和密码**字段中为此子目录输入有效用户标识和密码。
6. 单击**确定**。

如果您正在创建 Windows 文件系统搜寻器或将子目录添加到现有搜寻空间中，则在“为 Windows 子目录配置规则”页面上指定其它选项时指定登录凭证。

外部名现在变为显示名

在企业搜索管理控制台中，搜寻器配置向导中的标签**外部名**已更改为**显示名**，从而更好地表达此字段中指定的值用于显示用途这一事实。

在指定显示名时，并不会更改对象的内部名。例如，可以对名为 lastUpdateDateTime 的数据库字段指定更易于理解的显示名“上次更新时间”。

使用有意义的显示名使用户能够方便地识别他们要搜索的字段。这还使用户能够对一个字段（如“作者”）进行搜索并找到在文档的相应数据源中使用了不同内部字段名（如“作者”、“创建者”和“生产者”等等）的文档。

配置和监视数据侦听器

您可以使用企业搜索管理控制台来配置对定制数据侦听器应用程序的支持。

在 DB2 II OmniFind Edition V8.2 中，您必须先手工编辑若干个文件，然后应用程序才能更新企业搜索集合。以下信息替换《管理企业搜索》中『配置对外部搜寻器的支持』中的指示信息。

数据侦听器应用程序可以对集合添加数据、从集合中除去 URI 或者指示 Web 搜寻器访问或重新访问 URL。当用户（客户机）连接至数据侦听器组件时，它需要传入客户机标识、密码以及要更新的集合的标识。数据侦听器组件对客户机标识和密码进行认证，如果它们有效，则授权该应用程序更新指定的集合。

要配置对您的数据侦听器应用程序的支持，请在管理控制台工具栏上选择**系统**选项。使用下列窗口来指定有关数据侦听器应用程序的信息以及监视数据侦听器活动：

- 第 15 页的『“系统”视图：“数据侦听器”页面』
- 第 15 页的『“监视系统”视图：“数据侦听器”页面』

• 第 16 页的『“数据侦听器详细信息”页面』

“系统”视图：“数据侦听器”页面

使用“系统”视图的“数据侦听器”页面来指定一些信息，这些信息使定制数据侦听器应用程序能够更新集合。

数据侦听器应用程序使您能够搜寻 DB2 Information Integrator OmniFind Edition 附带提供的缺省搜寻器无法搜寻的数据源类型。您的应用程序可以对集合添加数据、从集合中除去数据或者指示 Web 搜寻器访问或重新访问 URL。

字段和控件

线程数 输入数据侦听器组件为了处理来自数据侦听器应用程序的请求而创建的工作线程的最大数目。缺省值是 6。

数据侦听器端口

输入数据侦听器组件的端口号。缺省值是 6668。

要点：如果更改了端口号，则必须重新启动数据侦听器组件才能使更改生效。

添加数据侦听器客户机标识

在数据侦听器客户机标识列表中添加一个空白行，以便您可以指定关于新的数据侦听器应用程序的信息。

数据侦听器客户机标识

输入数据侦听器应用程序的客户机标识，该应用程序将有权访问企业搜索集合。每个标识在企业搜索系统中都必须是唯一的。

密码 输入数据侦听器应用程序的密码。要访问企业搜索集合，数据侦听器应用程序必须使用您指定的凭证（标识和密码）。

集合名 选择要让这个数据侦听器应用程序更新的集合。每个数据侦听器应用程序都只能更新一个集合中的内容。



除去

从企业搜索系统中除去选择的数据侦听器应用程序。此应用程序将不会被用来对任何企业搜索集合添加数据或除去数据。

“监视系统”视图：“数据侦听器”页面

当您监视系统时，可以使用“数据侦听器”页面来查看有关数据侦听器组件和数据侦听器应用程序活动的详细信息。

字段和控件

刷新 刷新此页面上的信息。将有一条状态消息指示此信息的上次刷新时间。

名称 显示组件名（数据侦听器）。



详细信息

显示有关数据侦听器活动的详细信息。（如果数据侦听器已停止，则不会提供详细信息。）

状态 下列图标指示了数据侦听器的当前状态：



正在运行

指示数据侦听器组件处于活动状态。数据侦听器组件是在企业搜索系统启动时自动启动的。通常，除非您更改了为数据侦听器组件配置的端口号，否则不需要停止它或重新启动它。



已停止

指示数据侦听器组件已停止。

“数据侦听器详细信息”页面

使用“数据侦听器详细信息”页面来监视有关当前数据侦听器活动的详细信息，如有关线程状态和数据侦听器应用程序请求的信息。

字段和控件

刷新 刷新此页面上的信息。将有一条状态消息指示此信息的上次刷新时间。

状态 下列图标指示了数据侦听器的当前状态:



正在运行

指示数据侦听器组件处于活动状态。数据侦听器组件是在企业搜索系统启动时自动启动的。通常，您不需要停止它或重新启动它。



已停止

指示数据侦听器组件已停止。

要处理的请求数

显示已排队并等待处理的数据侦听器应用程序请求数。

线程总数

显示数据侦听器组件为了处理来自数据侦听器应用程序的请求而可以创建的线程的总数。

线程状态

显示可能的线程状态:

- 线程已初始化
- 线程正在读取客户机请求
- 线程正在处理客户机请求
- 线程正在将响应发送至客户机
- 线程已完成将响应发送至客户机
- 线程正在等待下一个客户机请求
- 线程处于暂停状态
- 线程正在停止

线程数 显示当前具有给定线程状态的线程数（例如，5 个线程当前正在访存数据）。

停止并启动数据侦听器组件

如果您更改了为数据侦听器组件配置的端口号，则必须停止该组件并重新启动它。

过程

要停止并重新启动数据侦听器组件:

1. 在索引服务器上, 作为企业搜索管理员登录。此用户标识是在 DB2 II OmniFind Edition 的安装期间指定的。
2. 切换到 ES_NODE_ROOT/master_config/datalistener 目录。
3. 输入下列命令以停止并重新启动数据侦听器:

```
esadmin datalistener stop
esadmin datalistener start
```

估计集合中的文档数

在创建集合时, 您可以提供期望该集合包含的文档数估计。DB2 II OmniFind Edition 中的资源管理器使用此数目来估计集合所需的内存量和磁盘资源量。

当估计的内存和磁盘资源需求量超出系统中当前可用的数量时, 资源管理器将发出警告。这些警告使您能够预防将来发生资源不足的问题。

监视器使用估计的文档数以及您为了接收警报而配置的阈值百分比来确定何时发送有关集合将要达到最大文档数的通知。

要提供估计的可能集合大小, 当您创建集合时, 可以在**估计的文档数**字段中输入一个数目。缺省值是 100 万个文档 (1000 000)。

在创建集合之后, 可以通过编辑集合并在“常规”页面上指定新值来更改这个估计值。

即使集合增大到您估计的大小, 系统也不会停止将文档添加到索引中。但是, 如果为集合配置了警报, 并且启用了当索引中的文档数超出限制时发送通知的选项, 则当集合大小达到您指定的通知百分比时, 您将接收到通知。

注: 当您在 V8.2 中配置此警报时, 您通过输入索引可以包含的文档数来指定索引中的文档数限制。对于 V8.2.1, 您在**估计的文档数**字段中指定的数目将自动显示在**限制**字段中, 并且, 您通过指定当索引中的文档数达到此限制的某个百分比时要得到通知来配置通知阈值。

在使用集合向导时自动检查系统资源

如果您使用集合向导, 则系统将自动确定系统资源是否足以支持对企业搜索系统添加新集合或搜寻器。在 DB2 II OmniFind Edition V8.2 中, 您只能通过明确地单击**检查系统资源**按钮来检查系统资源。

如果您在创建集合或搜寻器之后但在开始搜索文档之前检查系统资源, 则可以根据当前配置设置来确定可用的系统资源是否足以支持全力运行 DB2 II OmniFind Edition。

要了解有关此功能的完整信息, 请参阅 DB2 II OmniFind Edition 支持 Web 站点上的以下技术说明 (参考号为 1193761): Checking DB2 Information Integrator OmniFind Edition system resources.

监视解析器活动

在监视集合时, 您可以查看有关解析器活动的详细信息。

在 DB2 II OmniFind Edition V8.2 中，您可以监视解析器以开始或停止处理，但您无法查看详细状态信息或者获取有关解析器所解析的文档数的信息。

要监视集合的解析器以及查看详细状态信息：

1. 在企业搜索管理控制台中，对您想要监视的集合单击**监视**图标。
2. 单击“解析”页面。
3. 如果解析器的状态是**正在运行**，则单击**详细信息**图标。
4. 使用以下窗口来查看有关解析器活动的详细信息：
 - 『“解析器详细信息”页面』

“解析器详细信息”页面

使用“解析器详细信息”页面来查看有关为集合解析的文档的详细状态信息。提供的选项使您能够查看统计信息以及管理解析器活动。

字段和控件

刷新 刷新此页面上的信息。将有一条状态消息指示此信息的上次刷新时间。

解析器状态总结

此区域中的选项显示解析器的状态并提供有关已解析但尚未建立索引的文档的统计信息。

状态 状态行上的第一个图标指示搜寻器的当前状态：



正在运行

指示解析器正在运行。

根据解析器实际状态的不同，您可能会看到一条消息，该消息将提供有关当前解析活动的更多信息。例如，解析器当前可能正在解析文档、处于空闲状态（休眠到有更多文档可供解析为止、由于错误条件而正在等待被重新启动或者正在等待解析器服务重新启动）或者已暂停（正在等待，例如，正在等待索引重组完成）。



已停止

指示解析器已停止。

根据解析器当前状态以及您的管理角色的不同，下列图标使您能够管理解析器的活动：



启动

启动解析器。

在解析器启动后，它通常会对已搜寻到的文档进行解析，直到没有更多文档可供解析为止。然后，解析器进入休眠方式，并且，它在下次检查到有新搜寻到的文档可供解析之前不运行。



停止

停止解析器。

您可能需要停止并重新启动解析器，例如，如果更改了分类类型或添加了分类规则，就需要执行该操作。

数据存储中需要建立索引的文档数

显示解析器已解析并写至数据存储的文档数。这些文档尚未被添加到企业搜索索引。

对于集合中的每个搜寻器，提供了下列统计信息：

快照统计信息

此区域中的选项显示解析器活动的当前状态并提供有关各个搜寻器所收集的文档的统计信息。为了获得最佳的搜索结果，在解析器对集合中各个搜寻器搜寻到的所有文档进行解析之前，请不要构建索引。

快照起始

显示当前解析器状态视图中的统计信息的收集日期和时间。

解析速率

显示解析器每秒钟解析的页面数。

对于集合中的每个搜寻器，提供了下列统计信息：

搜寻器名

显示集合中每个搜寻器的名称。

搜寻器类型

对于每个搜寻器，显示该搜寻器所搜寻的数据的类型。

解析的文档数

对于每个搜寻器，显示可供解析的文档数以及迄今为止已解析的文档数。

状态 对于每个搜寻器，显示文档解析活动的状态：



正在运行

指示解析器当前正在解析此搜寻器搜寻到的文档。



已停止

指示解析器当前未解析此搜寻器的任何文档（可能是因为没有文档可供解析）。



已暂停

指示此搜寻器搜寻到的文档的解析操作已暂停。



错误

指示解析此搜寻器搜寻到的文档时发生了错误。要查看错误，请在监视集合时打开“日志”页面并查看解析器的日志文件。



完成

指示解析器已完成解析此搜寻器搜寻到的文档。

1 将 HTML 元数据映射到搜索字段

1 解析器可以将 HTML 元数据元素映射到字段名。在为集合配置解析器时，您可以指定
1 要将哪些 HTML 元数据元素作为搜索字段来建立索引。然后，用户可以通过在查询中
1 指定这些字段名来搜索 HTML 文档的特定部分。

1 在将 HTML 元数据映射到字段时，您指定是要让用户能够搜索该字段、使用参数查询
1 来搜索该字段还是在搜索结果中查看该字段。

1 要将 HTML 元数据元素映射到搜索字段，请执行下列操作：

- 1 在管理控制台中，对您想要更改的集合单击**编辑**图标。
- 1 选择“解析”页面。
- 1 单击将 **HTML 元数据映射到字段**。
- 1 使用以下窗口来指定有关您想要搜索的元数据字段的信息：
 - 1 • 『HTML 元数据字段映射页面』

1 HTML 元数据字段映射页面

1 使用“HTML 元数据字段映射”页面来将 HTML 元数据元素映射到字段名。通过在查询
1 中指定映射的字段名，用户可以搜索 HTML 文档的特定部分并获得更精确的搜索结果。

1 当您创建 HTML 字段映射时，所做的更改将在您下次停止并重新启动解析器时生效。
1 新映射不会影响已被解析并建立了索引的数据。

1 字段和控件

1 添加字段

1 添加一个空白行，您可以在该行中将另一个 HTML 元数据元素映射到搜索字段
1 名。

1 对于每个要映射到搜索字段名的 HTML 元数据元素，提供下列信息。

1 **字段名** 输入要与 HTML 元数据元素相关联的名称。当用户查询此集合中的 HTML 文
1 档时，他们可以指定此字段名。例如，要映射名为 publisher 的 HTML 元数
1 据标记，可以输入 publisher 来作为搜索字段名。

1 此值必须是单个字，并且只能包含字母数字字符。

1 HTML 元数据字段

1 输入一个 HTML 元数据元素的名称，这个 HTML 元数据元素将映射到您**在字
1 段名字段中指定的字段名**。使用同一个示例，由于正在将 HTML 元数据元素
1 publisher 映射到名为 publisher 的搜索字段，所以，请在此字段中输入
1 publisher。

1 此值必须是单个字，并且只能包含字母数字字符。

1 按字段名搜索

1 如果要让用户能够通过指定您在**字段名字段中指定的字段名**来搜索 HTML 文
1 档，则选中此复选框。如果清除了此复选框，则用户在搜索此集合中的 HTML
1 文档时将无法指定字段名。

参数搜索

如果要在搜索此字段时能够设计参数查询，则选中此复选框。参数查询使用户能够对数字值和日期值进行求值。例如，用户可能想确定此字段的值是否小于或大于另一个值。

如果选中此复选框，则确保 `HTML` 元数据元素的数据类型是 `DECIMAL`、`DOUBLE`、`INTEGER`、`SHORT`、`TIME` 或 `TIMESTAMP`。

显示在搜索结果中

如果要在搜索结果中显示此字段，则选中此复选框。如果清除了此复选框，则此字段不会包括在搜索结果中。



删除

删除选择的 `HTML` 元数据字段映射。

从日语和中文文档中的空格中除去换行字符

在不使用空格来对字边界进行定界的语言（如日语和中文）中，可以将解析器配置为除去某些将会导致换行的空格字符。

要对集合启用此功能：

1. 作为企业搜索管理员（此用户标识和密码是在 `DB2 II OmniFind Edition` 的安装期间设置的）登录。
2. 使用文本编辑器来编辑 `collection_ID.parserdriver/collection.properties` 文件，其中 `collection_ID` 是系统在您创建集合时对该集合指定的标识。
3. 将 `removeCjNewlineChars` 属性值由 `false` 更改为 `true`。

如果启用了此功能，并且指定集合中的文档是日语文档或中文文档，则解析器将从非 `ASCII` 字符范围内除去用于将两个字母字符分隔开的任何换行符序列或跳进字符序列。将除去下列字符：`Tab (0x09)`、`LF` 或换行 (`0x0A`) 以及 `CR` 或回车符 (`0x0D`)。

为集合解析器和 `Stellent` 会话配置文档类型

在 `DB2 II OmniFind Edition V8.2` 中，文件类型和内容类型已预先与特定集合解析器和 `Stellent` 文档过滤会话相关联。现在，您可以创建配置文件以指定特定文档类型的解析方式。

在 `V8.2` 中，某些类型的文档无法被解析（如 `PostScript` 文档），并且某些类型的文档已被硬编码为发送给 `Stellent` 会话以进行处理（`Microsoft Word`、`Microsoft Excel`、`Microsoft PowerPoint`、`Lotus Freelance`、`Lotus 123`、`PDF`、`RT` 和 `Ichitaro`）。所有其它文档格式都由内置的 `HTML` 或 `XML` 解析器处理。由于元数据可能有误，所以纯文本和 `HTML` 文档可能会被错误地发送给 `Stellent` 会话，然后被发送回给其中一个内置解析器（这种情况会影响性能）。

为了确保准确并高效地对搜寻空间中的文档进行解析，您可以创建配置文件来指定哪些类型的文档将由集合解析器解析以及哪些类型的文档将由 `Stellent` 会话解析。

使文档类型与集合解析器和 `Stellent` 会话相关联的工作涉及下列任务：

- 1 为集合解析器配置文档类型。此步骤涉及创建一个配置文件，该配置文件将文档类型映射至集合所使用的解析器。您可以为每个集合创建这些配置文件中的一个。
- 1 为 Stellent 会话配置文档类型。此步骤涉及创建一个配置文件，该配置文件将文档类型映射至集合所使用的 Stellent 文档过滤器。您可以为每个集合创建这些配置文件中的一个。
- 1 要使更改生效，请使用企业搜索管理控制台来监视已配置了文档类型的集合，然后停止并重新启动解析器。

1 为集合解析器配置文档类型

1 要使特定类型的文档与集合解析器相关联，您应该创建配置文件。在企业搜索管理控制台中，未提供对此功能的支持。

1 配置文件指定了以下内容：

- 1 • 要根据文件扩展名或内容类型将哪些文档发送给 Stellent 会话。
- 1 • 如何对由于元数据不完整而具有未知类型的文档进行解析。

1 如果配置文件不存在，集合解析器就会使用缺省配置设置，这些设置与 DB2 II OmniFind Edition V8.2 的行为完全对应。要了解这些缺省设置的描述，请参阅第 23 页的『缺省解析器服务规则』。

1 要使文档类型与集合解析器相关联：

- 1 在索引服务器上，作为企业搜索管理员登录。此用户标识是在 DB2 II OmniFind Edition 的安装期间指定的。
- 1 创建如下配置文件，其中 *collection_ID* 标识您想要配置的集合：

```
1 ES_NODE_ROOT/master_config/collection_ID.parserdriver/parserTypes.cfg
```

1 此文件列示要发送到 Stellent 会话的文件扩展名和内容类型。文件的格式是行序列，每一行都是下列其中一个规则：

1 **EXTENSION** *extension parser*

1 以指定扩展名作为 URL 结尾的所有文档都将由指定的解析器处理。不要在扩展名中包括句点；比较操作不区分大小写。

1 **CONTENTTYPE** *type/subtype parser*

1 内容类型与指定的类型 / 子类型相匹配的所有文档都将由指定的解析器处理。假定文档的内容类型为 *t/s*，如果 *t* 等于类型，并且 *s* 等于子类型或者子类型是通配符（星号 *），则将发生匹配。

1 **UNKNOWN** *parser*

1 所有具有未知扩展名和内容类型的文档（即，不是由搜寻器提供的文档）都将由指定的解析器处理。

1 **DEFAULT** *parser*

1 所有未被任何其它规则包括的文档都将由指定的解析器处理。

1 在所有情况下，*parser* 必须是 *html*、*xml*、*stellent* 或 *none*，其中 *none* 表示不对文档进行解析。

1 **规则优先级：**如果有多个规则与一个文档相匹配，则最具特定性的规则优先，而不考虑规则的出现顺序。规则的特定性按照规则在以上列表中的出现顺序递减。

- 1 • EXTENSION 规则比 CONTENTTYPE 规则更具特定性。
- 1 • 包含子类型的 CONTENTTYPE 规则比包含通配符的 CONTENTTYPE 规则更具
- 1 特定性。例如，内容类型 application/postscript 的规则优先于 application/*
- 1 的规则。
- 1 • 不应该存在两个用于相同扩展名和内容类型的规则。在那种情况下，由实现负责
- 1 确定哪个规则具有优先权。

1 示例

1 在以下示例中，内置的 HTML 解析器处理所有具有扩展名 txt、htm 或 html、内容类

1 型以 text/ 开头或者具有未知扩展名和内容类型的文档。内置的 XML 解析器处理所

1 有具有扩展名 xml 或者具有内容类型 text/xml 的文档。所有其它文档（包括内容类型

1 以 application/ 开头的文档）都将被发送给 Stellent 会话。

```
1           EXTENSION doc stellent
1           EXTENSION txt html
1           EXTENSION htm html
1           EXTENSION html html
1           EXTENSION xml xml
1           EXTENSION ps none
1           CONTENTTYPE text/xml xml
1           CONTENTTYPE text/* html
1           CONTENTTYPE application/* stellent
1           UNKNOWN html
1           DEFAULT stellent
```

1 缺省解析器服务规则

1 如果您未创建配置文件来将文件类型和内容类型映射至集合的解析器，则将使用缺省

1 规则来解析文档。

1 集合解析器使用的缺省规则如下所示：

```
1           EXTENSION pdf stellent
1           EXTENSION ppt stellent
1           EXTENSION prz stellent
1           EXTENSION lwp stellent
1           EXTENSION doc stellent
1           EXTENSION rtf stellent
1           EXTENSION xls stellent
1           EXTENSION 123 stellent
1           EXTENSION vsd stellent
1           EXTENSION vdx stellent
1           EXTENSION jxw stellent
1           EXTENSION jsw stellent
1           EXTENSION jtw stellent
1           EXTENSION jaw stellent
1           EXTENSION juw stellent
1           EXTENSION jbw stellent
1           EXTENSION jvw stellent
1           EXTENSION jfw stellent
1           EXTENSION jtt stellent
1           EXTENSION jtd stellent
1           EXTENSION jttc stellent
1           EXTENSION jtdc stellent
1           EXTENSION jtdx stellent
1           EXTENSION ps none
1           EXTENSION xml xml
1           EXTENSION txt text
1           EXTENSION htm html
1           EXTENSION html html
1           EXTENSION shtml html
```

```

1      EXTENSION xhtml html
1      EXTENSION asp html
1
1      CONTENTTYPE application/postscript none
1      CONTENTTYPE application/* stellent
1      CONTENTTYPE text/rtf stellent
1      CONTENTTYPE text/richtext stellent
1      CONTENTTYPE text/xml xml
1      CONTENTTYPE text/html html
1      CONTENTTYPE text/plain text
1
1      UNKNOWN html
1      DEFAULT html

```

1 为 Stellent 会话配置文档类型

1 要指定 Stellent 文档过滤器要解析的文档类型，您应创建配置文件。在企业搜索管理控制台中，未提供对此功能的支持。

1 配置文件指定了以下内容：

- 1 • Stellent 会话将要对哪些文件类型进行解析。文件类型与 Stellent 库所识别的其中一种文件类型相对应。
- 1 • 要将哪些文件类型发送回给集合解析器以便使用其中一个内置解析器进行处理。（由于集合解析器可能会由于元数据不正确而错误地将文档发送给 Stellent 会话，所以需要执行此操作。）
- 1 • 哪些文件类型由于不受 DB2 II OmniFind Edition 支持而将被拒绝。

1 如果指定了配置文件，但它不存在，则解析器将无法启动。如果未对 `stellent.properties` 文件中的 `OutsideInSupportedTypes` 属性指定配置文件，则将使用缺省配置设置。要了解这些缺省设置的描述（这些设置与 DB2 II OmniFind Edition V8.2 的行为相对应），请参阅第 26 页的『Stellent 会话的缺省解析规则』。

1 要为 Stellent 会话配置文档类型：

- 1 1. 在索引服务器上，作为企业搜索管理员登录。此用户标识是在 DB2 II OmniFind Edition 的安装期间指定的。
- 1 2. 编辑 `ES_NODE_ROOT/master_config/collection_ID.stellent/stellent.properties` 文件，其中 `collection_ID` 标识您想要配置的集合。
- 1 3. 对于 `OutsideInSupportedTypes` 属性，指定您正在创建的配置文件的绝对路径。

1 例如，可以为单个集合创建以下配置文件并将其与其它特定于集合的文件存储在一起：

```
1 ES_NODE_ROOT/master_config/collection_ID.stellent/stellenttypes.cfg
```

1 要对所有集合使用相同的设置，可以创建以下配置文件并将其与其它系统级别的文件存储在一起。（如果使用此方法，则务必在每个集合的 `stellent.properties` 文件中指定此路径，如步骤 2 所述。）

```
1 ES_INSTALL_ROOT/default_config/stellent/stellenttypes.cfg
```

- 1 4. 创建配置文件。此文件列示文档类型以及它们的处理方式。文件的格式是行序列，每一行都是下列其中一个规则：


```
1 accept DEFAULT
1 accept ALL doctype
1 accept type doctype
1 native DEFAULT
1 native type doctype
1 reject type
```

1 其中:

1 *doctype*

1 是用于 *doctype* 查询令牌的值。例如, 通过在查询中指定 *doctype:pdf*, 可
1 以对 *doctype* 令牌执行搜索。

1 *type* 是 Stellent 库中的其中一个 FI_ 值, *doctype* 是当规则被应用时用于 *doctype*
1 令牌的值。

1 **DEFAULT**

1 表示接受类型列表或本机类型列表(这取决于规则的类型)包括缺省列
1 表。此选项使您能够扩展缺省配置, 而不是替换缺省配置。

1 **ALL** 表示指定的 *doctype* 令牌接受所有未明确列示的类型。

1 **规则处理:**

- 1 • 如果存在 *type* 的 *accept* 规则, 包括缺省列表(如果指定了 *accept DEFAULT* 的
1 话), 则接受该类型。
- 1 • 否则, 如果存在 *type* 的 *reject* 规则, 则不接受该类型。
- 1 • 否则, 如果指定了 *accept ALL*, 则接受该类型。
- 1 • 否则, 不接受该类型。

1 如果接受文档类型, 则使用已应用的规则中指定的 *doctype* 值。将把此值与所解析
1 的内容一起发送回给集合解析器。如果不接受文档类型, 则发生下列行为:

- 1 • 如果存在 *type* 的本机规则, 包括缺省列表(如果指定了 *native DEFAULT* 的
1 话), 则将文档发送回给内置解析器, 并对 *doctype* 令牌指定由此规则指定的
1 值。*doctype* 值必须是 *txt*、*htm* 或 *xml*, 它们分别指示纯文本、HTML 或
1 XML。
- 1 • 否则, 拒绝该文档, 并且不对该文档进行解析。

1 **示例**

1 给定以下配置文件, Stellent 会话除接受缺省列表以外还接受 Microsoft Visio 格式。

```
1 accept DEFAULT
1 accept FI_VISIO3 visio
1 accept FI_VISIO4 visio
1 accept FI_VISIO5 visio
1 accept FI_VISIO6 visio
```

1 给定以下配置文件, 将支持 Postscript 文档并可以对它们执行搜索, 这些文档的文档类
1 型为 *ps*; 将把 X 像素图格式(XPM)发送回给内置文本解析器, 将拒绝 PNG 图像格
1 式; 接受所有其它文件类型并可以对它们执行搜索, 它们的文档类型为 *other*。

```
1 accept DEFAULT
1 accept FI_POSTSCRIPT ps
1 native FI_XPIXMAP txt
1 accept ALL other
1 reject FI_PNG
```

Stellent 会话的缺省解析规则

如果您未创建配置文件来将文件类型映射至 Stellent 会话文档过滤器，则将使用缺省规则来解析文档。

Stellent 会话使用的缺省规则如下所示：

```
ACCEPT FI_WORD4 doc
ACCEPT FI_WORD5 doc
ACCEPT FI_RTF rtf
ACCEPT FI_WINWORD1 doc
ACCEPT FI_WINWORD1COMPLEX doc
ACCEPT FI_WINWORD2 doc
ACCEPT FI_WORD6 doc
ACCEPT FI_WINWORD6 doc
ACCEPT FI_ICHITAR03 jxw
ACCEPT FI_ICHITAR04 jsw
ACCEPT FI_WINWORD1J doc
ACCEPT FI_WINWORD5J doc
ACCEPT FI_RTFJ rtf
ACCEPT FI_WINWORD7 doc
ACCEPT FI_WORDPRO lwp
ACCEPT FI_WINWORD97 doc
ACCEPT FI_ICHITAR08 jtd
ACCEPT FI_WORDPRO97 lwp
ACCEPT FI_WINWORD2000 doc
ACCEPT FI_WINWORD2002 doc
ACCEPT FI_WINWORD2003 doc
ACCEPT FI_123R1 123
ACCEPT FI_123R2 123
ACCEPT FI_123R3 123
ACCEPT FI_EXCEL xls
ACCEPT FI_EXCEL3 xls
ACCEPT FI_EXCEL4 xls
ACCEPT FI_123R4 123
ACCEPT FI_EXCEL5 xls
ACCEPT FI_123R6 123
ACCEPT FI_EXCEL97 xls
ACCEPT FI_123R9 123
ACCEPT FI_EXCEL2000 xls
ACCEPT FI_EXCEL2002 xls
ACCEPT FI_EXCEL2003 xls
ACCEPT FI_FREELANCE prz
ACCEPT FI_POWERPOINT4 ppt
ACCEPT FI_POWERPOINT3 ppt
ACCEPT FI_POWERPOINT7 ppt
ACCEPT FI_FREELANCE3 prz
ACCEPT FI_POWERPOINTMAC3 ppt
ACCEPT FI_POWERPOINTMAC4 ppt
ACCEPT FI_PDF pdf
ACCEPT FI_EXTPOWERPOINT4 ppt
ACCEPT FI_EXTPOWERPOINTMAC4 ppt
ACCEPT FI_POWERPOINTMACB3 ppt
ACCEPT FI_POWERPOINTMACB4 ppt
ACCEPT FI_POWERPOINT97 ppt
ACCEPT FI_PDFMACBIN pdf
ACCEPT FI_POWERPOINT9597 ppt
ACCEPT FI_POWERPOINT2000 ppt
ACCEPT FI_POWERPOINT2 ppt

NATIVE FI_HTML htm
NATIVE FI_HTML_LATIN2 htm
NATIVE FI_HTML_JAPANESESJIS htm
NATIVE FI_HTML_JAPANESEEUC htm
NATIVE FI_HTML_CHINESEBIG5 htm
NATIVE FI_HTML_CHINESEEUC htm
NATIVE FI_HTML_CHINESEGB htm
```

1	NATIVE FI_HTML_KOREANHANGUL	htm
1	NATIVE FI_HTML_CYRILLIC1251	htm
1	NATIVE FI_HTML_CYRILLICKOI8	htm
1	NATIVE FI_CYRILLIC1251	txt
1	NATIVE FI_CYRILLICKOI8	txt
1	NATIVE FI_W2KHTML	htm
1	NATIVE FI_XL2KHTML	htm
1	NATIVE FI_PP2KHTML	htm
1	NATIVE FI_XML	xml
1	NATIVE FI_WML	xml
1	NATIVE FI_HTML_JAPANESEJIS	htm
1	NATIVE FI_WML_CHINESEBIG5	xml
1	NATIVE FI_WML_CHINESEEEUC	xml
1	NATIVE FI_WML_CHINESEGB	xml
1	NATIVE FI_WML_CYRILLIC1251	xml
1	NATIVE FI_WML_CYRILLICKOI8	xml
1	NATIVE FI_WML_JAPANESEJIS	xml
1	NATIVE FI_WML_JAPANESEJIS	xml
1	NATIVE FI_WML_JAPANESEEEUC	xml
1	NATIVE FI_WML_KOREANHANGUL	xml
1	NATIVE FI_WML_LATIN2	xml
1	NATIVE FI_HTMLUNICODE	htm
1	NATIVE FI_XML_DOCTYPE_HTML	htm
1	NATIVE FI_XHTML	htm
1	NATIVE FI_ASCII	txt
1	NATIVE FI_ANSI	txt
1	NATIVE FI_UNICODE	txt
1	NATIVE FI_ASCII8	txt
1	NATIVE FI_ANSI8	txt
1	NATIVE FI_MAC	txt
1	NATIVE FI_MAC8	txt
1	NATIVE FI_SHIFTJIS	txt
1	NATIVE FI_CHINESEGB	txt
1	NATIVE FI_HANGEUL	txt
1	NATIVE FI_CHINESEBIG5	txt
1	NATIVE FI_LATIN2	txt
1	NATIVE FI_JAPANESE_EUC	txt
1	NATIVE FI_HEBREW_OLDCODE	txt
1	NATIVE FI_HEBREW_PC8	txt
1	NATIVE FI_HEBREW_E0	txt
1	NATIVE FI_HEBREW_WINDOWS	txt
1	NATIVE FI_ARABIC_710	txt
1	NATIVE FI_ARABIC_720	txt
1	NATIVE FI_ARABIC_WINDOWS	txt
1	NATIVE FI_7BITTEXT	txt
1	NATIVE FI_JAPANESE_JIS	txt
1	NATIVE FI_CENTRALEU_1250	txt
1	NATIVE FI_UTF8	txt
1	NATIVE FI_EBCDIC_37	txt
1	NATIVE FI_EBCDIC_273	txt
1	NATIVE FI_EBCDIC_277	txt
1	NATIVE FI_EBCDIC_278	txt
1	NATIVE FI_EBCDIC_280	txt
1	NATIVE FI_EBCDIC_284	txt
1	NATIVE FI_EBCDIC_285	txt
1	NATIVE FI_EBCDIC_297	txt
1	NATIVE FI_EBCDIC_500	txt
1	NATIVE FI_EBCDIC_870	txt
1	NATIVE FI_EBCDIC_871	txt
1	NATIVE FI_EBCDIC_1026	txt

1 对查询处理所做的增强

1 在多服务器配置中，在集合级别提供了故障转移保护，而不是仅仅在服务器级别提供
1 此功能。如果一台搜索服务器上的集合由于任何原因而变得不可用，则对该集合进行
1 的查询将自动地被传递至另一台搜索服务器。

1 折叠来自同一 Web 站点的结果

1 当您编辑 SiteDefs.txt 文件来指定选项以便折叠来自同一 Web 站点的搜索结果时，
1 可以随 URL 前缀一起指定一个字符串来作为站点标识。

1 在 DB2 II OmniFind Edition V8.2 中，SiteDefs.txt 文件仅包含 URL 前缀列表。现
1 在，您可以为每个站点指定一个字符串来作为标识。如以下示例所示，您在行首输入
1 URL 前缀，并可以选择在右边输入字符串标识：

```
1 http://mycompany.server1.com/hr/          hr  
1 http://mycompany.server2.com/hr/          hr  
1 http://mycompany.server3.com/hr/          hr  
1 http://mycompany.server1.com/finance/     finance  
1 http://mycompany.server1.com/news         news  
1 http://mycompany.server1.com/USA
```

1 搜索应用程序可以使用 URL 前缀或标识来对搜索结果中的文档进行分组和折叠。例如，
1 hr 标识使来自主管人力资源文档的三台服务器的搜索结果能够在全部搜索结果中折叠在
1 一起。

1 企业搜索查询语法已扩展为能够支持此功能。例如，要检索属于同一组的文档，请使
1 用以下语法，其中 *group* 是组标识或属于该组的某个文档的标准 URL：

```
1 samegroupas:group
```

1 例如：

```
1 samegroupas:hr  
1 samegroupas:http://mycompany.server3.com/hr/
```

1 要了解有关更新 SiteDefs.txt 文件的完整信息，请参阅《管理企业搜索》中的『折叠
1 来自同一 Web 站点的结果』主题。

1 搜索结果中的扩展突出显示

1 Query 接口中的一个属性使查询项能够在搜索结果详细信息的数个区域中突出显示。在
1 DB2 II OmniFind Edition V8.2 中，只能在总结字段中进行突出显示。

1 在创建搜索应用程序时，您可以使用 Query 接口的 getProperty 和 setProperty 方法
1 来处理查询处理。可以在这些方法中使用的新属性 HighlightingMode 使您能够控制查
1 询项在搜索结果中的突出显示方式。此属性具有下列值：

1 DefaultHighlighting

1 仅在总结字段中突出显示查询项。当搜索应用程序省略了 HighlightingMode 属性
1 时，这就是缺省操作。

1 ExtendedHighlighting

1 将查询项的突出显示扩展到搜索结果的其它区域，如标题、URL 和其它字段。

1 您可以使用 `getProperty` 或 `getProperties` 方法来从 `Result` 对象的属性中检索突出
1 显示的 URL 字段。属性名是 `HighlightedDocumentID`，它的值就是突出显示的
1 URL。`Result` 方法 `getDocumentID` 仍返回未突出显示的 URL。

1 搜索应用程序中的超时支持

1 `Remote` 应用程序编程接口 (API) 中的一个属性使您能够为远程请求指定超时值。

1 搜索应用程序在调用下列方法时可以通过 `Properties` 对象传递 `timeout` 属性:

```
1 SearchFactory.getSearchService(Properties)  
1 BrowseFactory.getBrowseService(Properties)
```

1 例如:

```
1 // create a new Properties object.  
1 Properties config = new Properties();  
1 config.setProperty("hostname", hostname);  
1 config.setProperty("port", portNumber);  
1 config.setProperty("locale", "en_US");  
1 config.setProperty("timeout", "60");  
1 SearchService searchService = factory.getSearchService(config);
```

1 超时值表示在远程请求超时前要等待的秒数，它必须是整数（如 60，而不能是 60.5 或
1 sixty）。如果您未指定超时值，`Remote API` 将使用缺省值（30 秒）。

已知局限性、问题和变通方法

本信息描述 DB2 Information Integrator OmniFind Edition V8.2 的已知局限性、问题和变通方法。

任何局限性和限制都可能适用于也可能不适用于产品的其它发行版。

将 DB2 UDB AIX 版 tar 文件解包

对于 DB2 通用数据库企业服务器版版本 8.2 AIX 版，必须先将 CD 映像解包，然后才能安装 DB2 UDB。

在开始安装企业搜索之前，将 CD 解包。

要了解有关将 CD 解包的更多信息，请访问 DB2 信息中心版本 8.2: <http://publib.boulder.ibm.com/infocenter/db2help/index.jsp>。在 DB2 信息中心中，单击安装 → 数据库系统 → **DB2 通用数据库 Linux 版、UNIX 版和 Windows 版** → **DB2 服务器** → **DB2 UDB 企业服务器版（非分区）** → **AIX** → 启动 **DB2 安装向导**。

必须先将此文件解包，然后企业搜索安装程序才能安装 DB2 UDB。

要将 CD 映像解包：

1. 将 product.tar.Z（其中，product 表示许可您安装的产品）复制到临时文件系统。
2. 输入以下命令以启动 DB2 安装向导：

```
zcat product.tar.Z | tar -xf - ;./product/db2setup
```

例如，如果 DB2 UDB 企业服务器版的产品名是 ese，则输入以下命令：

```
zcat ese.tar.Z | tar -xf - ;./ese/db2setup
```

在企业搜索安装期间，请提供解包文件路径中包含文件 db2setup 的目录。例如，解包文件的目录可能是 .\ese.sbcsaix1。

AIX 中的共享内存段

如果您在 AIX 上看到消息 SQL1224N，则表示可能超出了 AIX 所允许的最大共享内存段数。

要解决此问题，请参阅 IBM 支持 Web 站点上的主题 SQL1224N Shared Memory Segments。

在 Windows 上卸载之后重新启动系统

如果在 Windows 计算机上卸载 DB2 II OmniFind Edition，则必须重新启动系统以完成所有程序文件的删除。

您还必须手工删除企业搜索安装目录。如果不重新启动系统或删除程序文件，然后尝试再次安装企业搜索，就有可能发生问题。

Linux 和 AIX 的孤立 URLFetcher 进程

问题: Web 搜寻器使用两个进程。当一个进程退出时，它会发信号让另一个进程也退出。有时，一个进程可能会意外退出，使另一个进程无限期地运行下去。

如果从 UNIX ps 命令的输出中看到类似以下进程的孤立进程，就可以确定发生了这种情况：

```
F S UID      PID  PPID  C PRI  NI ADDR  SZ WCHAN  STIME TTY          TIME CMD
0 S user1    26222  1  0  75   0   - 152641 schedu 10:34 pts/8    00:00:00
```

```
/opt/IBMJava2-141/jre/bin/java -classpath /.../URLFetcher.jar:...
```

注意：

- PPID（父进程标识）值为 1 表示该进程是孤立的。
- 类路径中存在 URLFetcher.jar 表示这是 Web 搜寻器进程。

解决方案

查找孤立进程的进程标识（PID）。在上一个示例中，它是 26222，然后杀死该进程。在 Linux 和 AIX 上，使用不带自变量的 kill 命令。您可能还会在进程表中看到“子代”条目。子代条目表示由孤立进程创建的线程。当您杀死这些子代条目的祖先进程时，这些子代条目应该会被删除。

当您看到孤立 URLFetcher 进程时，应该杀死该进程，这是因为如果它存在于进程表中，操作系统就会为它保留资源。您应该定期检查是否存在这些孤立进程，当您怀疑 Web 搜寻器异常终止时尤其如此。如果存在这些进程，则杀死它们。

注: 在 DB2 II OmniFind Edition V8.2.1 中，由于 URL 访存器在发现 C++ 组件退出时会自我终止，所以很少会出现孤立进程。如果确实发生了此问题，则请注意不要意外地杀死另一用户的 httpauth.ini 编辑会话（HTTPAuthEdit 实用程序是一个 JVM，它从 URLFetcher JAR 文件运行，在进程表中的外观很相似）。

服务器上的索引位置

当您创建集合时，如果想要指定除缺省位置以外的索引位置，并且如果指定不是位于服务器数据根目录中的索引位置，这个非缺省索引位置的磁盘空间使用情况就不会被监视。

对于多服务器安装，当您创建集合时，如果想要指定除缺省位置以外的索引位置，该索引位置在全部四台服务器上首先必须已经存在。

堆大小限制

文档解析器 Java™ 会话（也称为 parserservice）的堆大小限制为 400 MB。

堆大小值是以如下方式存储的：

```
ES_NODE_ROOT/config/services.ini:session7.max_heap=400
```

如果在 parserservice 会话的系统错误日志中出现了 OutOfMemory 消息，则停止 parserservice 会话，在 services.ini 文件中增大 parserservice 的 max_heap 值，然后重新启动 parserservice 会话。

具有大对象的 DB2 UDB 表

不能搜寻包含 128 MB 或更大的 LOB 数据的 DB2 通用数据库表。不能为大于 8 MB 的文档建立索引。

索引状态不显示

如果未除去先前安装就重新安装了新版本的 DB2 II OmniFind Edition, 则索引监视状态可能不能正确显示。

当您启动新的索引构建时, 构建状态在管理控制台的“索引”页面上不能正确显示。但是, 索引构建会继续, 完成构建时, “索引”页面将显示正确的已建立索引文档数和索引构建的正确次数。

建议您在安装新版本之前除去 DB2 II OmniFind Edition 的先前版本。

Linux 和 AIX 的用户标识许可权

可能需要添加对新的企业搜索用户标识的写许可权。

如果使用不同于企业搜索管理员的 Linux 或 AIX 用户来重新安装 DB2 II OmniFind Edition, 则当添加 DB2 搜寻器并指定您想要搜寻远程未编目数据库 (使用 DB2 通用数据库 4 类 Java 数据库连接 (JDBC) 驱动程序的数据库) 时, 可能会发生错误。在这种情况下, 必须将新的企业搜索管理员用户的写许可权添加至 `/home/db2inst1/sqllib/tmp/0.spf` 文件。

1 为 FileNet Content Services 启用直接文档访问

1 当您使用 VeniceBridge 搜寻器来搜寻由 DB2 Information Integrator Content Edition (旧
1 称 VeniceBridge) 服务器管理的 FileNet Content Services 存储库时, 目标 FileNet Content
1 1 Services 库和 FileNet Web Services 使用 FileNet Content Services 连接器, 并且必须
1 1 被配置为使用直接文档访问。请参阅 *FileNet Web Services Administrator's Guide* 和
1 1 FileNet Content Services 管理文档来设置直接文档访问支持。

1 搜寻昵称表时的局限性

1 在搜寻 IBM DB2 通用数据库 z/OS® 版和 Microsoft SQL Server 的 IBM DB2
1 1 Information Integrator 昵称表时, 有一些局限性。

1 如果目标表是 DB2 通用数据库 z/OS 版或 Microsoft SQL Server 的 DB2 Information
1 1 Integrator 昵称表, 并且如果指定了文档内容字段, 则唯一标识不能包括任何具有下列数
1 1 据类型的列:

- 1 • LONG VARCHAR
- 1 • LONG VARCHAR FOR BIT DATA
- 1 • LONG VARGRAPHIC

1 如果目标表是 Microsoft SQL Server 的 DB2 Information Integrator 昵称表, 并且如果
1 1 该昵称表是在使用 UTF-8 代码集的数据库中创建的, 则具有 BLOB 数据类型的列中的
1 1 文档内容可能会损坏。

1 如果目标表是 DB2 通用数据库 z/OS 版或 Microsoft SQL Server 的 DB2 Information
1 Integrator 昵称表，并且如果指定了文档内容字段，则执行下列步骤来提高 DB2 搜寻器
1 的性能：
1 1. 创建服务器并将 COLLATING_SEQUENCE 选项设置为 Y。
1 2. 在此服务器中创建昵称。
1 如果该昵称表在您将该选项设置为 Y 之前已定义，则删除该昵称并再次创建它。

声明

本信息是为在美国提供的产品和服务编写的。IBM 可能未在所有国家或地区提供本文中讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：IBM Director of Licensing IBM Corporation North Castle Drive Armonk, NY 10504-1785 U.S.A.

有关双字节（DBCS）信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：IBM World Trade Asia Corporation Licensing 2-31 Roppongi 3-chome, Minato-ku Tokyo 106-0032, Japan

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本资料中描述的产品和 / 或程序进行改进和 / 或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：（i）允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及（ii）允许对已经交换的信息进行相互使用，请与下列地址联系：

IBM Corporation J46A/G4
555 Bailey Avenue
San Jose, CA 95141-1003 U.S.A.

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本资料中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际软件许可协议或任何同等协议中的条款提供。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能会有差异。本文档的用户应当验证其特定环境的适用数据。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

所有关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含日常商业运作所使用的数据和报表的示例。为了尽可能全面地说明这些数据和报表，这些示例包括个人、公司、品牌和产品的名称。所有这些名称都是虚构的，如与实际商业企业所使用的名称和地址有任何雷同，纯属巧合。

版权许可:

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口 (API) 进行应用程序的开发、使用、经销或分发为目的，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。用户如果是为了按照 IBM 应用程序编程接口开发、使用、经销或分发应用程序，则可以任何形式复制、修改和分发这些样本程序，而无须向 IBM 付费。

凡这些样本程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明:

Outside In (®) Viewer Technology, © 1992-2005 Stellent, Chicago, IL., Inc. All Rights Reserved.

IBM XSLT Processor Licensed Materials - Property of IBM © Copyright IBM Corp., 1999-2005. All Rights Reserved.

商标

本主题列示 IBM 商标和某些非 IBM 商标。

下列各项是 International Business Machines Corporation 在美国和 / 或其他国家或地区的商标:

IBM
AIX
AIX 5L
DB2
DB2 Universal Database
Domino
Informix
Lotus

Lotus Notes
Notes
OmniFind
WebSphere
xSeries
z/OS

下列各项是其他公司的商标或注册商标:

Java 和所有基于 Java 的商标和徽标是 Sun Microsystems, Inc. 在美国和 / 或其他国家或地区的商标或注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和 / 或其他国家或地区的商标。

Intel、Intel Inside (徽标)、MMX 和 Pentium 是 Intel Corporation 在美国和 / 或其他国家或地区的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Linux 是 Linus Torvalds 在美国和 / 或其他国家或地区的商标。

其他公司、产品或服务名称可能是其他公司的商标或服务标记。

与 IBM 联系

在中国，请致电下列其中一个号码以与 IBM 联系：

- 800-810-1818 或 (010) 84981188 分机 5151，可获得售前客户服务；
- 800-810-1818 或 (010) 84981188 分机 5200，可获得售后客户服务；
- 800-810-1818 或 (010) 84981188 分机 5017，可获得市场营销与销售的信息；

要查找您所在国家或地区的 IBM 营业处，可在网上查看 IBM 全球联系人目录 (Directory of Worldwide Contacts)，网址为：www.ibm.com/planetwide。

获取产品信息

关于 DB2 Information Integrator 的信息可通过万维网获取，网址为：<http://www-900.ibm.com/cn/software/db2/>。

此站点包含有关 DB2 产品家族、DB2 解决方案、技术前沿与趋势、DB2 服务、成功案例、市场活动、培训与认证、DB2 开发者园地、合作伙伴、下载中心、资料库、第三方分析报告、殊荣与奖项、DB2 新闻以及如何购买 DB2 的最新信息。

提供对文档的意见

请发送您对本书或其它 DB2 Information Integrator 文档的任何意见。

您的反馈有助于 IBM 提供高质量的信息。请发送您对本书或其它 DB2 Information Integrator 文档的任何意见。可以使用下列任何一种方法提出意见：

1. 使用 www.ibm.com/software/data/rcf 上的在线读者意见表发送您的意见。
2. 通过电子邮件将您的意见发送至 ctscrcf@cn.ibm.com。确保包括产品的名称、产品的版本号和书籍的名称及部件号（如果适用的话）。如果您对特定文本有意见，请包括此文本的位置（例如，标题、表号或页码）。

IBM

中国印刷



Java[™]
COMPATIBLE