

Perl 言語による日本語 PDF 生成モジュール - PDFJ

中島 靖[†]

PDF 形式の文書ファイルを Perl 言語によって生成するためのライブラリモジュールを作成した。JIS 規格の日本語の組版ルールを組み込んでおり、文字列を折り返して行を構成する際の禁則処理や文字間隔の調整など、日本語に固有の組版処理をサポートする。文字段落や画像や図形といった構成要素を並べたブロックを、入れ子にして表を作成したり、指定の大きさに分割してページ毎に配置したりすることができる。また、XML 形式の原稿から PDF を生成する、XPDFJ というモジュールも付属している。

Perl Japanese PDF Generation Module - PDFJ

Yasushi Nakajima[†]

PDFJ is a Perl language library module to generate PDF document files. It conforms to the JIS standard of Japanese document layout rules. Then it supports the inhibitions on line folding or letter spacing rules. It can make blocks with paragraphs, images and shapes and it can make tables by nested blocks. It can break a block by specified size and layout each broken block to a page. It includes XPDFJ module which generates PDF files from XML source texts.

1. はじめに

1.1 PDFJ 作成の動機

PDFJ を作成した動機は次の二つである。

(1) Web アプリケーションにおけるサーバー側での PDF 生成

Web アプリケーションに帳票などの印刷機能を持たせようとすると、Web ブラウザの印刷機能ではページ制御ができないなど力不足である。この一つの解決方法として、普及した文書フォーマットである PDF 形式で帳票等の内容を出力することが考えられる。そのためにサーバー側で動的に PDF を生成しなければならず、プログラム言語に PDF 生成機能が必要となる。サーバー側で採用する言語が Java である場合は FOP¹⁾ というライブラリが PDF 生成機能を持っているが、Perl 言語を使用する場合、従来は PDFLib²⁾ というライブラリを使用するのが一般的であった。しかし PDFLib の機能はプリミティブなもので、日本語の禁則処理や文字間隔の調整などのルールを意識した組版処理をおこなうにはプログラムの負担が大きい。罫線の入った表の作成や、大きなレイアウト要素を分割してページをまたがって配置する、といった処理についても同様である。

そこで、こういった機能を盛り込んだライブラリモジュールとして作成したのが PDFJ である。

(2) 書籍や論文の作成

書籍や論文を自分でレイアウトしようとする、LaTeX を使用するのが一般的である。筆者も過去に書籍を執筆した際に、レイアウトのための LaTeX の複雑なマクロを作成しようとしてデバッグに多大の時間を費やした経験を持つ。LaTeX のマクロプログラミングの作業は、少なくとも筆者にとって

は容易なものではない。できることなら使いなれば Perl 言語でレイアウトがプログラミングできないものかと考えたのが、PDFJ を作成したもう一つの動機である。

1.2 PDFJ の到達点

PDFJ は 2002 年に開発を始めて、現在も開発途上にある。

帳票作成の用途についてはある程度実用になるところまで来ており、実際の業務での使用例もある。書籍や論文作成の用途については、まだ実際に使用する段階には至っていない。

2. PDFJ の概要

2.1 PDFJ の機能

PDFJ は次のような機能を持っている。

- ・ JIS X 4051「日本語文書の行組版方法」にほぼ準拠。禁則、行の詰め伸ばし、ルビ、添え字、縦書き中の欧文、縦中横、欧文のハイフネーション、下線・傍線、圈点、網掛け
- ・ 使用可能な Type1 フォントは、和文に Ryumin-Light と GothicBBB-Medium、欧文に Times, Helvetica, Courier の各ファミリ。これらのフォントは埋め込まれない
- ・ TrueType フォントは任意のものが使用可能。TrueType フォントは埋め込まれる（和文についてはサブセットで）
- ・ 欧文に、固定ピッチの半角フォントと、プロポーション的な欧文フォントが可能
- ・ シフト JIS、日本語 EUC、UTF8、Unicode に対応
- ・ JPEG 画像(ファイルおよび URL 指定)と線画図形。画像や図形の行内配置。線画図形中のテキストや画像
- ・ 行長と行送りの指定による段落。箇条書きのためのラベル可
- ・ 段落、画像、図形などを並べたブロック。ブロックには、内容の配置、周囲の余白、枠線、塗りつぶし色などを指定可。ブロック内の並び方向は、上 下、左 右、右 左。入れ子のブロックによる表作成

[†] (株) ネットストック
Netstock Corporation
Email: nakajima@netstock.co.jp

- ・段落やブロックを指定の大きさを超えないように分割して、ページ毎に配置。分割の際に先頭や末尾に移動するブロック要素

- ・PDF の文書情報、アウトライン情報、ハイパーリンク(文書内および URL) の付加。暗号化

2.2 動作環境

PDFJ は Perl 言語だけで書かれており、Perl5.005 以降で動作する。下請けとして、TeX::Hyphen、Compress::Zlib、Digest::MD5 の各 Perl モジュールを使用する。ただし Compress::Zlib は使用しないことも可能。Digest::MD5 は暗号化の時にのみ使用される。

3. XPDFJ の概要

XPDFJ は XML 形式で書かれた原稿を読み取って PDFJ を呼び出す命令に変換することによって PDF を生成するモジュールである。XML の読み取りには XML::Parser モジュールを使用する。

XPDFJ の原稿における XML ベースの書式では、PDFJ に用意されたサブルーチンやメソッドの呼び出しを XML 形式で書く規則が定められているとともに、Perl の式と変数によって PDFJ の各種オブジェクトを操作できる仕組みが盛り込まれている。これによって、PDFJ を使って PDF を生成している Perl プログラムを XML 形式に書き直すことができる。もちろんそれだけではかえって書き方が面倒なだけでメリットはない。XPDFJ では、新たな XML 要素名 = 命令とその働きを既存の命令から定義することのできるマクロ機能が用意されている。定義された命令には属性や内容によって引数を与えることができるので、上述の Perl の式や変数を操作する機能と組み合わせれば、複雑な機能を持たせることができる。XPDFJ に添付されている標準マクロ(stddefs.inc)では表 1 のような HTML ライクなマクロが用意されている。また、論文用マクロ(article.inc)では標準マクロに加えて表 2 のようなマクロが用意されている。本稿はこの論文用マクロの使用例となっている。ただし、これらのマクロファイルは開発途上であり、変更の可能性が大いにあることに注意していただきたい。

4. 今後の課題

今後の課題としては、XPDFJ のマクロなどの機能強化はもちろんであるが、チャレンジングな課題として次のものがあげられる。

- (1) 対話フォーム、スライドショー、注釈、電子署名といった PDF の機能への対応
特に対話フォームと電子署名は、電子文書ならではの PDF の活用の上で重要と考えている。
- (2) 数式のレイアウト機能
論文では数式のレイアウト機能が必要であり、LaTeX と比

XPDFJ の標準マクロで定義された HTML と同名の要素名は、そこから連想されるような機能を持っているが、属性は違いが大きく互換性はない。

表 1 XPDFJ の標準マクロにおけるマクロ命令

Fig.1 Macro commands defined in the XPDFJ standard macro file

S	テキストスタイル
B, I, U, SUP, SUB	太字、斜体、下線、上付、下付
H1 ~ 4, P, BR	見出し、段落、改行
UL, OL, LI	箇条書き
DL, DT, DD	語句説明
TABLE, TR, TD, TH	表
IMG	画像
DIV, HR, SKIP, NEWPAGE	ブロック、横罫線、間隔、改ページ
BODY	文書内容全体を囲む

表 2 XPDFJ の論文用マクロにおけるマクロ命令

Fig.2 Macro commands defined in the XPDFJ article macro file

ARTICLE	論文全体を囲む
HEAD	ヘッダ部を囲む
TITLE, ETITLE	表題、英文表題
AUTHOR, EAUTHOR	著者名、英文著者名
SUMMARY, ESUMMARY	概要、英文概要
BODY	本文部を囲む
SECTITLE, SUBSECTITLE	章見出し、節見出し
FOOTNOTE	脚注
FIG	図表ブロック
CAPTION, ECAPTION	図表のキャプション、同英文
REFERENCE	参考文献

較しての最大の弱点となっている。

(3) 図形エディタ

GUI 操作で図形を作成できないと、図形を活用しきれない。

(4) レイアウト機能の強化

段組の時の段をまたがる要素の配置、図形を自動的に避ける段落の作成など。

(5) 一冊本を書いてみる

書籍用マクロを作成し、一冊丸ごと本を書いてみたい。

(6) 既存 PDF の操作

この要望は非常に多い。PDFJ とは別のモジュールとして作成することになるだろう。

文 献

- 1) FOP: <http://xml.apache.org/fop/>
- 2) PDFLib: <http://www.pdflib.com/>
- 3) アドビシステムズ：PDF リファレンス第 2 版，ピアソンエデュケーション (2001)