

Perl だけで作る日本語 PDF

日本語組版ルールを組み込んだ PDF 生成モジュール PDFJ

中島 靖 (Kansai.pm、(株)ネットストック)
<nakajima@netstock.co.jp>

2002 年 12 月 6 日

関西オープンソース+フリーウェア 2002

2003 年 10 月 7 日、2004 年 8 月 20 日、2005 年 2 月 17 日加筆修正

PDFJ とは

日本語 PDF を作成するための Perl のライブラリモジュール

- 1) 日本語の組版ルール(禁則、行の詰め伸ばし、ルビ、添え字、縦書き中の欧文、縦中横、下線・傍線、圏点など)
- 2) フォントは標準 Type1 と TrueType (TrueType は埋め込み)
- 3) テキスト中の画像や線画図形、線画図形中のテキストや画像
- 4) 箇条書きのためのラベル付き段落
- 5) 段落、画像、図形などを並べたブロック。入れ子のブロックによる表
- 6) ページをまたぐ段落やブロックの分割配置
- 7) PDF のアウトラインやハイパーリンク
- 8) **XML で原稿が書けます**

2002 年 5 月に初公開の、発展途上のソフトウェアです

背景と目的～ 1

Web アプリケーションに印刷機能を持たせるためにサーバーで動的に日本語 PDF を生成したい

TeX+dvipdfm	<ul style="list-style-type: none">・ サーバー側の仕掛としてはちょっと大げさ・ 複雑な数式を扱うならこれしかないが、ビジネスアプリには不要だろう
PDFLib	<ul style="list-style-type: none">・ レイアウト機能が弱く、かなりのプログラミングが必要・ 日本語対応の細かなところまでは無理・ 商用の場合ライセンスが必要
FOP	<ul style="list-style-type: none">・ うーん、JAVA かなぁ...(^;;)・ XSL は魅力だけど XSLT を書くのは...(^;;)

Perl だけでちょいちょいっとやりたいなあ

背景と目的～ 2

本を書くのに LaTeX のマクロの代わりに Perl スクリプトでレイアウトできたらラクだなあ

LaTeX で複雑なマクロを書いたら不安定で困った...(すみません、TeXnician じゃないもので)

自作モジュールなら日本語組版ルールもきちんと組み込めるだろう

TeX では難しい、ルビの配置とか、追い込みの時の空白の詰め方とか...(すみません、TeXnician じゃないもので)

前提として

- 1) Perl だけで書く (C のコンパイルができないサーバーでも使えるように)
- 2) 日本語組版ルール (JIS X 4051) を組み込む (ただし厳密な準拠にはこだわらない)
- 3) 縦書きも OK
- 4) 段落、表、箇条書き、図といったレイアウト要素のページへの適切な配置
- 5) アウトラインやハイパーリンクなどの PDF の便利な機能はできるだけ利用できるようにする
- 6) できる限りフリーにする

日本語組版ルール～ 1

JIS X 4051 「日本語文書の行組版方法」(1995) を採用

JIS X 4051 のうち概ね次のものをサポート

- ・ 約物の基本的な用法
- ・ 行頭禁則処理
- ・ 行末禁則処理
- ・ 分離禁止処理
- ・ 和欧文の混植処理
- ・ 縦中横処理
- ・ ルビ処理
- ・ 添え字処理
- ・ 圈点処理
- ・ 下線・傍線処理
- ・ 段落整形処理
- ・ そろえ（中央、行頭、行末、均等割）
- ・ 行の調整処理
- ・ 表処理

日本語組版ルール～ 2

JIS X 4051 と異なる点

- ・ 文字間の間隔はすべて後ろの文字のサイズに従う
- ・ 欧文スペースは三分空きでなくフォントに従う

不完全または未サポートの処理

- ・ 連数字の配置法
- ・ 割注
- ・ 段落末尾処理（段落の最終行を所定の文字数以上に）
- ・ タブ処理
- ・ 表処理における行間隔の均等割

組版機能デモ

「約物の、(幅と間隔)」ゴシック
 ク Gothic 明朝 Minchou ゴシック
 ク Hankaku 明朝 Hankaku **色文字**
 下線 下破線 圏点 網掛け 日本
 語斜体 添え字 ¹⁾ 中島文字 位置
あけぼの ボ ス boss フィフティーン
 曙 大親分大親分 fifteen

「約物の、(幅と間隔)」ゴシック
 ク Gothic 明朝 Minchou ゴシック
 ク Hankaku 明 朝
 H a n k a k u **色文字** 傍線 傍
 破線 圏点、網掛け 日本語斜体
 (1) 縦中横 添え字 (1) なかじま も じ い ち 中島文字 位置
あけぼの ボ ス boss フィフティーン
 曙 大親分大親分 fifteen

帳票デモ

納品書

2002年9月1日

PDFJ社 御中

(株) ネットストック
東京都大田区蒲田 XX-YYY
03-XXXX-YYYY

名称	数量	摘要
ノートパソコン	20	Windows XP Professional、 Office XP Personal
スイッチングハブ	3	

XPDFJ とは

XML でマークアップされたテキストから PDFJ を用いて PDF を生成する Perl モジュール

- PDFJ を直接使うと自由度は高いがいちいちプログラミングが必要
- とりあえずこのスライドが作れる程度まで来たが、まだ仕様が動いている段階。近日公開 **かなり実用的になってきました**
- XML ライク？ 自動化のためは XML だが、今のところ書きやすさを優先 **結局 XML 準拠にしました**
- 最終的には本を丸ごと一冊書けるように... **目次と索引付きの小冊子程度は書けます。論文用マクロも**

さきほどの帳票デモを生成する XPDFJ のソーステキストは次のようになる **現時点のソースにあります**

帳票デモソース

```
<?xml version="1.0" encoding="x-sjis-cp932"?>
<XPDFJ version="0.1"><do file="of2002.inc"/><BODY>
<P align="center"> 納品書 </P>
<P align="right" pstyle="$PStyle{small}" tstyle="$TStyle{small}">2002 年 9 月
1 日 </P>
<P align="left"><T><S withline="1">PDFJ 社 御中 </S></T></P>
<P align="right" pstyle="$PStyle{small}" tstyle="$TStyle{small}"> (株) ネット
ストック <BR/> 東京都大田区蒲田 XX-YYY<BR/>03-XXXX-YYYY</P>
<TABLE cellwidth="250, 100, 350" cellpadding="5" tablebox="s"
tableboxstyle="linewidth:3" cellbox="s">
<TR><TH> 名称 </TH><TH> 数量 </TH><TH> 摘要 </TH></TR>
<TR><TD> ノートパソコン </TD><TD align="right">20</TD><TD>Windows
XP Professional、 Office XP Personal</TD></TR>
<TR><TD> スイッチングハブ </TD><TD align="right">3</TD><TD>
</TD></TR>
</TABLE>
</BODY></XPDFJ>
```

今後の展開

当面は、XPDFJ を含むバージョン 0.7 の公開 **しました**

今後の大きな展開としては

- 1) XPDFJ を充実させて本を一冊書いてみる
- 2) PDF のフォーム、スライドショー、注釈、暗号化、**電子署名**などへの対応 **暗号化とスライドショーは一応できます**
- 3) 既存の PDF の編集機能
- 4) **MathML** による数式記述
- 5) **SVG** による線画図形記述

PDFJ は次の場所にあります

<http://hp1.jonex.ne.jp/~nakajima.yasushi/>

メーリングリストがあります

<mailto:nakajima.yasushi-pdfj-subscribe@jonex.ne.jp>