

XPDFJによるレポート・論文PDF作成方法

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

2004年10月23日

関西オープンソース2004

2005年2月17日加筆修正

本日の内容

XPDFJ の標準マクロファイルと論文用マクロファイルを使って、レポートや論文の PDF を作成する方法を解説

- PDFJ と XPDFJ の概要
- 標準マクロの使い方
- 論文用マクロの使い方と実例
- 今後の課題

あらかじめお断り：

論文作成の方法としてはいくつかの大きな弱点があり、LaTeX にはまだまだ及びません

本発表は PDFJ0.81 にもとづいています **現在はさらにバージョンが上がっています**

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) アウトライン、ハイパーリンク、暗号化

PDFLib、FOP、OpenOffice と比べてどう？

XPDFJ とは～ 1

PDFJ の XML フロントエンド

PDFJ を用いたスクリプトと同等の処理を XML 形式で書けるようにする

- PDFJ のサブルーチンやメソッド呼び出しを XML のタグの形で記述
- Perl の変数や式を利用できる

1. XML 形式で原稿ファイルを書き

2. 「perl xpdfj.pl 原稿ファイル」と実行すると

3. PDF ファイルができる

```
use PDFJ 'EUC';
$Doc = Doc({version => 1.3,
  pagewidth => 595, pageheight =>
  842});
$font = $Doc->new_font({basefont
  => "Ryumin-Light", encoding =>
  "EUC-H"});
$page = $Doc->new_page;
Text({texts => ' こんにちは ', style =>
  TStyle({font => $font, fontsize =>
  40}}))
->show({page => $page, x => 50, y
  => 750});
$Doc->print({file => "sample2.pdf"});
```

```
<?xml version="1.0"
  encoding="x-euc-jp-jisx0221"?>
<XPDFJ version="0.2">
<Doc version="1.3" pagewidth="595"
  pageheight="842"/>
<new_font setvar="$font"
  basefont="Ryumin-Light"
  encoding="UniJIS-UCS2-HW-H"/>
<new_page setvar="$page"/>
<Text>
<texts> こんにちは </texts>
<style><TStyle font="$font"
  fontsize="40"/></style>
<call><show page="$page" x="50"
  y="750"/></call>
</Text>
<print file="sample2.pdf"/>
</XPDFJ>
```

XPDFJ とは～ 2

...しかし、スクリプトを単に XML 化するだけなら面倒くさいだけ

命令を拡張できるマクロ機能によって、簡潔な記述が可能に

```
<?xml version="1.0"
  encoding="x-euc-jp-jisx0221"?>
<XPDFJ version="0.2">
<do file="stddefs.inc"/>
<BODY>
<P><S fontsize="40">
こんにちは
</S></P>
</BODY>
<print file="sample2.pdf"/>
</XPDFJ>
```

標準マクロ stddefs.inc

HTML ライクなマクロ命令を定義

あくまで HTML「ライク」です。属性などかなり違いがあり、既存の HTML をそのまま使えるわけではありません

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, MLI	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

標準マクロの使い方～基本パターン

標準マクロ stddefs.inc を使った原稿の基本パターン

```
<?xml version="1.0" encoding="x-euc-jp-jisx0221"?> ... XML 宣言
<XPDFJ version="0.2"> ... 全体を <XPDFJ> で囲む
<do file="stddefs.inc"/> ... 標準マクロファイル stddefs.inc の読み込み
<BODY> ... 内容の全体を <BODY> で囲む
...ここに内容を書く...
</BODY>
<print file="$Args{outfile}"/> ... 出力する PDF ファイル名を指定
</XPDFJ>
```

XML 宣言では正しいエンコーディングを指定すること(XML::Parser で解釈される : x-sjis-cp932、 x-euc-jp-jisx0221)

出力するファイル名を \$Args{outfile} とすれば原稿ファイル名に .pdf を付けたものに

参考：XML の書き方

- 先頭には XML 宣言を書いて原稿の文字コードと合致した encoding を指定する。
- 大文字小文字は区別される。<BODY> を <Body> と書いたりしてはダメ。特に HTML に慣れた方は注意。
- <XXX...> が開始タグ </XXX> が終了タグで、入れ子になってもよいが、開始タグと終了タグがきちんと対応すること。<XXX...></XXX> と内容がない場合は <XXX.../> と書ける。HTML のように <P> だけ書いて </P> を書かないのはダメ、
 は
 と書かねばならない。
- 開始タグの中に「名前 = " 値 "」あるいは「名前 = ' 値 '」と属性を指定できる。
- タグを書くため以外に < や > を書くときは、< > とする。その他、& は &、' は '、" は " と書ける。
- < や > をたくさん含むようなテキストを書くときに一つ一つ < などと書くのが面倒なら、<![CDATA[と]]> で囲むとそのまま書ける。
- <!-- から --> まではコメント。コメント中には -- という文字列は入れられないので、コメントは入れ子にはできない。

標準マクロの使い方～ stddefs.inc の読み込み

<do file="stddefs.inc"/> で読み込むと、次のように設定される。

属性名	省略時値	意味
pagewidth	595	ページ幅
pageheight	842	ページ高さ
leftmargin, rightmargin, topmargin, bottommargin	72	左右上下の余白
fontsize	10	本文フォントサイズ
linefeed	150%	本文行送り

B5 の用紙で余白 1.5cm、フォントは 9 ポイントとしたければ、
<do file="stddefs.inc" pagewidth="516" pageheight="728"
margin="42" fontsize="9"/>

1 ポイントは 1/72 インチ。1 インチは 2.54cm。

標準マクロの使い方～見出しと本文段落

見出しは <H1> ~ <H4> で囲む（<H1> は中央寄せ）

本文段落は <P> で囲む

<P><TD> での改行と空白の扱い（HTML と異なる）

- ・ タグの前後の改行は無視される
- ・ テキスト中の改行は生きる（
 で改行することもできる）
- ・ 行頭のタブや半角空白は無視される（入れ子を字下げして原稿を見やすくするため）

<H1> 表題用の見出し </H1>

<H2> 大見出し </H2>

<H3> 中見出し </H3>

<H4> 小見出し </H4>

<P>

本文用の段落

改行は生きる
BR での改行も
可能

</P>

表題用の見出し

大見出し

中見出し

小見出し

本文用の段落

改行は生きる

BR での改行も可能

実際のフォントは見出しはゴシック、本文は明朝です

標準マクロの使い方～テキストスタイル

<I><U><SUP><SUB> で太字、斜体、下線、上付き、下付き
<S> で次のテキストスタイルを指定できる

スタイル名	値と機能
fontsize	フォントサイズ
italic	1: 斜体
bold	1: 太字
ruby	ルビ文字列
shapestyle	文字描画の図形スタイル
withdot	1: 圏点あり
withline	1: 下線あり
withlinestyle	下線の図形スタイル

withbox	囲み枠。s: 枠線、f: 塗りつぶし、sf: 枠線 + 塗りつぶし。s の代わりに l: 左辺、r: 右辺、t: 上辺、b: 下辺の組み合わせも可。rX: (X は数値) 角を半径 X で丸く
withboxstyle	囲み枠の図形スタイル

図形スタイル

スタイル名	値と機能
fillcolor	塗りつぶし色
strokecolor	線色
linewidth	線幅
linedash	破線パターン (オン幅, オフ幅, 開始オフセット)

図形スタイルを省略すると枠線色塗りつぶし色とも黒、線幅 1、実線色は「#RRGGBB」(16進)または 0 (黒) から 1 (白) の間の数値
破線パターンを指定しなければ実線
破線パターンの開始オフセットは省略可

 太字 <I> 斜体 </I><U> 下線
</U><SUP> 上付き
</SUP>_{下付き}

<S shapestyle="fillcolor:#FF0000">
赤文字 </S>

<S shapestyle="fillcolor:1"
withbox="f"
withboxstyle="fillcolor:0.5"> 網掛け
白抜き </S>

<S withline="1"
withlinestyle="linedash:3,3"> 点線
下線 </S>

<S ruby=" あじさい "> 紫陽花 </S>

太字斜体下線^{上付き}_{下付き}

赤文字

網掛け白抜き

点線下線

あじさい
紫陽花

標準マクロの使い方～段落スタイル

<P> の属性

属性名	値と機能
pstyle	追加の段落スタイル（文書の本文設定を元に）
tstyle	追加のテキストスタイル（文書の本文設定を元に）
align	b,left: 左寄せ、m,center: 中央寄せ、e,right: 右寄せ、w: 両端寄せ

主な段落スタイル

スタイル名	値と機能
align	b: 左寄せ、m: 中央寄せ、e: 右寄せ、w: 両端寄せ
linefeed	行送り（ポイントまたは％）
preskip	段落前の間隔

postskip	段落後の間隔
beginindent	行頭インデント
endindent	行末インデント

<pre><P align="center"> 中央寄せ </P></pre>	中央寄せ
<pre><P pstyle="beginindent:40" tstyle="fontsize:15"> フォントサイ ズ 15 ポイント、行頭を 40 ポイン ト下げる </P></pre>	フォントサイズ 15 ポイント、行頭を 40 ポ イント下げる
<pre><P pstyle="align:e; linefeed:200%;"> 右寄せで 行間を広く </P></pre>	右寄せで 行間を広く

標準マクロの使い方～箇条書き

、、<DL><DT><DD> で箇条書き
各行を で囲むのが面倒なら <MLI> で

	属性名	値と機能
UL	label	ラベル文字列（省略すると「・」）
OL	label	ラベル初期値（半角英数字、省略すると「1」）「-」を指定すると前回の OL から引き継ぐ
	prelabel	ラベルの前置文字列
	postlabel	ラベルの後置文字列（省略すると「.」）
DL	labelskip	語句と本体の間隔（省略すると本文フォントサイズに同じ）
共通	labelsize	二行目以降の字下げ（省略すると本文フォントサイズに同じ）

```
<UL><MLI>
各行が箇条書きに
各行が箇条書きに
</MLI></UL>
```

```
<OL>
<LI> 外側の箇条書き 1 </LI>
<P> 外側の箇条書き 1 の続き </P>
<OL label="a">
  <LI> 内側の箇条書き 1 </LI>
</OL>
<LI> 外側の箇条書き 2 </LI>
</OL>
```

```
<DL labelsizes="40" labelskip="20">
  <DT> 語句説明 </DT><DD> 語句
と説明の間が 20 で二行目以降が 40
下げ </DD>
</DL>
```

- 各行が箇条書きに
- 各行が箇条書きに

1. 外側の箇条書き 1
外側の箇条書き 1 の続き
 - a. 内側の箇条書き 1
2. 外側の箇条書き 2

語句説明 語句と説明の間が 20 で二
行目以降が 40 下げ

標準マクロの使い方～表

<TABLE><TR><TH><TD> で表が作れる

内容に応じてセル幅を自動調節する機能はなく、cellwidth 属性ですべてのセル幅（列幅）を指定する必要がある

<TD> の中は小さな BODY と考えることができ、その中にさらに <P><DL><TABLE> を入れることができる

	属性名	値と機能
TABLE	cellwidth	セル幅（カンマ区切りで列毎に指定、省略すると 200）
	tablebox	表枠線。指定方法は withbox と同じ
	tableboxstyle	表枠線の図形スタイル（省略すると「linewidth:0.25」）
	preskip	前間隔
	postskip	後間隔

共通	cellpadding	セル内余白
	cellbox	セル枠線。指定方法は withbox と同じ
	cellboxstyle	セル枠線の図形スタイル (省略すると「linewidth:0.25」)
	align	セル内配置。left,center,right
	pstyle	追加の段落スタイル
	tstyle	追加のテキストスタイル
TD	colspan	セルがまたがる列数
	rowspan	セルがまたがる行数

cellbox と align を TABLE や TR で指定するときはカンマ区切りで列毎に指定できる。このとき列数の方が多ければ最後の要素が繰り返し使われる

```
<TABLE cellwidth="40%, 30%" cellpadding="5" cellbox=",l" align="center,
right" tablebox="sr10" tableboxstyle="linewidth:1">
<TR cellbox="b,lb"><TH> 年月 </TH><TH>A</TH><TH>B</TH></TR>
<TR><TD>2004 年 1 月 </TD><TD>15,000</TD><TD>25,000</TD></TR>
<TR><TD>2004 年 2 月 </TD><TD>12,000</TD><TD>21,000</TD></TR>
</TABLE>
```

年月	A	B
2004 年 1 月	15,000	25,000
2004 年 2 月	12,000	21,000

標準マクロの使い方～その他

その他に次のようなマクロが用意されている

マクロ名	機能
HEADER, FOOTER	ページヘッダ、フッタを指定する
HR	横罫線を引く
IMG	画像を挿入する
DIV	ブロックを作って囲み枠を付けたり配置を調整したりする
SKIP	間隔を調整する
NEWPAGE	改ページする

線画図形を描くマクロを定義した `shape.inc` というマクロファイルも用意されている

簡単なマクロ定義

よく使う属性がある場合...alias を使う

<S shapestyle="fillcolor:#FF0000"> をよく使うなら、
<alias tag="RED" aliasof="S" shapestyle="fillcolor:#FF0000"/> と定義すれば、
<RED> 赤文字 </RED> と書ける

一連のタグのまとまりをマクロ化する...def を使う

<P align="right">
 ネットストック
 中島 靖 </P> をよく使うなら、

<def tag="MYSIGN"><return><P align="right">
 ネットストック

 中島 靖 </P></return></def> と定義すれば、
<MYSIGN/> と書くだけでよい

マクロ定義を別ファイルにして読み込むこともできる

上記のようなマクロ定義を、<?xml version="1.0" encoding="..."?><do> と
</do> で挟んだ内容のファイル mymacro.inc を作る
<do file="stddefs.inc"/> の後に <do file="mymacro.inc"/> と書けば OK

論文用マクロ article.inc

二段組みの論文用のマクロ

stddefs.inc も自動的に読み込まれるので、標準マクロも使える

マクロ名	機能
ARTICLE	論文全体を囲む
HEAD	ヘッダ部を囲む (ARTICLE の中)
TITLE	表題
ETITLE	英文表題
AUTHOR	著者名
EAUTHOR	英文著者名
SUMMARY	要約
ESUMMARY	英文要約

BODY	本文部を囲む (ARTICLE の中)
SECTITLE	章見出し。「1.」の連番が自動的に付く。label 属性で指定も可
SUBSECTITLE	節見出し。「1.1」の連番が自動的に付く。label 属性で指定も可
FOOTNOTE	脚注。label 属性でラベルを指定
FIG	図表を囲む。ページ上部に中央寄せで配置。float="e" でページ下部に配置
CAPTION	図表キャプション (FIG の中)
ECAPTION	英文図表キャプション (FIG の中)
REFERENCE	参考文献。label 属性でラベルを指定

今のところ脚注や参考文献のラベルは自動連番されない

論文用マクロの使い方

論文全体を <ARTICLE> で囲み、ヘッダ部を <HEAD> で、本文部を <BODY> で囲む

ヘッダ部には <TITLE> <ETITLE> <AUTHOR> <EAUTHOR> <SUMMARY> <ESUMMARY> から必要なものを並べる

本文部には標準マクロの他、<SECTITLE> <SUBSECTITLE> <FOOTNOTE> <FIG> <REFERENCE> から必要なものを並べる

実例として `articledemo.xp` が付属しています（別紙資料）

おわりに

課題

- 段をまたがった図表がレイアウトできない 近い将来に改善したい
- 線画図形を位置や大きさを指定して配置することはできるが、GUI で作成できない 将来は SVG に対応したい
- 数式のレイアウト機能がない 将来は MathML に対応したい

発展途上のソフトウェアです。是非ご意見をお寄せください

PDFJ は次の場所にあります (XPDFJ は PDFJ に付属しています)

<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>