

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 <nakajima@netstock.co.jp>

2002/5/11 Kansan.pm 2周年イベント

2002/10/18 改訂

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">* サーバー側にたくさんのソフトをインストールしないといけない* 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">* 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 かあ...(^;;;)* 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)

目標

- * 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）
- * 日本語組版ルール（JIS X 4051）を組み込む
- * 縦書きも OK
- * 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
- * アウトラインやハイパーリンクといった PDF 機能も利用できるようにする
- * 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のうち次のものをサポート

1. 約物の幅と間隔
2. 禁則処理と分離禁止処理
3. 縦中横
4. ルビ
5. 添え字
6. 圏点
7. 下線・傍線
8. 行長揃えのための間隔調整方法
9. 欧文のハイフネーション

日本語組版ルール～ 2

JIS X 4051 と異なる点

1. 文字間の間隔はすべて後ろの文字のサイズに従う
2. 欧文スペースは三分空きでなくフォントに従う
3. 割注は未サポート
4. タブ処理は未サポート

入力はどうする？

Perl スクリプトで直接書くのでない限り、なんらかのマークアップ言語を入力とする必要

独自のマークアップ言語	<ul style="list-style-type: none">* 必要なものだけ実装できる* 自分で本を書くためならこれがラク
LaTeX のサブセット	<ul style="list-style-type: none">* LaTeX からの移行がラク
XSL	<ul style="list-style-type: none">* 将来性として有望* フルサポートは大変、サブセットか
HTML+CSS	<ul style="list-style-type: none">* 既存文書をそのまま使える* フルサポートは大変、互換性の泥沼が...
その他	<ul style="list-style-type: none">* pod...pod2pdf はそれなりに便利かも

現状と計画

一応実用になるレベルまでできてきました

主な課題

1. ページ分割の際のフロート要素
2. 段落が指定領域を避ける
3. マークアップ言語による入力
4. チュートリアル的な説明書

まだまだバグがあると思います

デモ

「約物の、(幅と間隔)」

ゴシック Gothic 明朝 Minchou

シツク Hankaku 明朝

Hankaku 色文字 傍線 傍破線

圈点、網掛け 日本語斜体 (1) 縦中横

添え字⁽¹⁾ 中島 文字位置 曙 大親分

大親分^{boss} fifteen^{フィフティーン}

「約物の、(幅と間隔)」

ゴシック Gothic 明朝

Minchou ゴシック Hankaku

明朝 Hankaku 色文字 下線

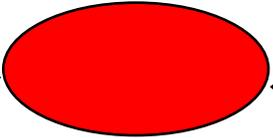
下破線 圈点 網掛け 日本語

斜体 添え字¹⁾ 中島^{なかじま}

文字位置^{もじいち} 曙^{あけぼの} 大親分^{ボス}

大親分^{boss} fifteen^{フィフティーン}

図形と画像のデモ

テキスト中の画像、テキスト中の図形、テキスト中の図形中のテキスト