

FreePWING で作る電子辞書

大浦 真@ Debian.org
[http://www.netfort.gr.jp/~ohura/
ohura@debian.org](http://www.netfort.gr.jp/~ohura/ohura@debian.org)

2006 年 3 月 19 日

目次

- ① はじめに
- ② FreePWING
- ③ 使い方
 - 変換スクリプトの用意
 - 辞書の生成
 - catalogs の生成
- ④ 終わりに

はじめに

- EPWING (イーピーウィング): 日本で広く使われている電子辞書の標準規格。
 - 準拠した電子辞書が多数存在する。

<http://www.epwing.or.jp/>
- 多くの検索ソフトが存在する。
 - DDwin (Windows)、lookup (Emacs)、電子辞書付属のソフトウェア
- 自分で EPWING の辞書を作成できないか。
→ EPWING の規格全体は公開されていないが、EPWING v1 のサブセットが JIS 規格になっている。(JIS X 4081)

FreePWING

- FreePWING: JIS X 4081 の形式の電子辞書を生成するための Perl のライブラリ。
- フリーソフトウェア。
- UNIX 互換の開発環境がないと動作しないが、生成された辞書はどの環境でも使用できる。
- Web ページ:
<http://www.sra.co.jp/people/m-kasahr/freepwing/>

FreePWING

- FreePWING が生成する辞書で使える機能
 - 前方一致検索、後方一致検索
 - 条件一致検索
 - 本文内の相互参照
 - 外字の利用
 - カラー図版の参照
- FreePWING が生成する辞書で使えない機能
 - 音声、動画の再生
 - カラー図版の表示

使い方

- FreePWING の役割は、単に JIS X 4081 形式のファイルを出力するだけ。
- 入力のテキスト、HTML、XML などの処理は自分で用意する。
- 日本語の文字コードには、EUC-JP を使う。

変換スクリプトの用意

- FreePWING のドキュメントのサンプルに基づき説明
- 変換前のデータ

たいさく 【大作】
大掛かりな作品。
たいさく 【対策】
状況に応じてとる手段や策略。

検索語 たいさく、対策
見出し たいさく 【対策】
本文 たいさく 【対策】
状況に応じてとる手段や策略。

変換スクリプトの用意

- モジュールの読み込み

```
use FreePWING::FPWUtils::FPWParser;
```

- 初期化 (インスタンスの生成と作業ファイルのオープン)

```
initialize_fpwparser('text' => ¥$fpwtext,  
'heading' => ¥$fpwheading,  
'word2' => ¥$fpword2);
```

- 'text': 本文
- 'heading': 見出し
- 'word2': 検索語

変換スクリプトの用意

- エントリの切替え

```
## 本文と見出しを新しいエントリに切り替える。  
$fpwtext->new_entry()  
|| die $fpwtext->error_message() . "\n";  
$fpwheading->new_entry()  
|| die $fpwheading->error_message() . "\n";
```

- 現在の項目の終了を宣言し、次の項目に切り替える。

変換スクリプトの用意

- 見出しを 'heading' と 'text' の一行目に入力。

```
## 見出しを書き込む。  
$fpwheading->add_text($_)  
|| die $fpwheading->error_message() . "\n";  
## 本文を書き込む。  
if (!$fpwtext->add_keyword_start()  
|| !$fpwtext->add_text($_)  
|| !$fpwtext->add_keyword_end()  
|| !$fpwtext->add_newline()) {  
die $fpwtext->error_message() . "\n";  
}
```

変換スクリプトの用意

- 検索語を word2 に登録

```
## かなの検索語を登録する。  
($kanaword = $_) = s/【.*$//;  
$heading_position =  
$fpwheading->entry_position();  
$text_position = $fpwtext->entry_position();  
if (!$fpword2->add_entry($kanaword,  
$heading_position, $text_position)) {  
die $fpword2->error_message() . "\n";  
}
```

変換スクリプトの用意

- 本文を書き込む。

```
## 次の一行（説明）を読み込む。  
last if (!defined($_ = <>));  
  
## 本文を書き込む。  
if (!$fpwtext->add_text($_)  
|| !$fpwtext->add_newline()) {  
die $fpwtext->error_message() . "\n"; }  
}
```

変換スクリプトの用意

- 終了処理

```
## 書き込み用の作業ファイルを閉じる。  
finalize_fpwpaser('text' => ¥$fpwtext,  
'heading' => ¥$fpwheading,  
'word2' => ¥$fpword2);
```

辞書の生成

- Makefile の用意

```
FPWPARSER = fpwkokugo
FPWPARSERFLAGS = kokugo.dat
DIR = kokugo
INSTALLDIR = /home/mohura/dict

include fpwutils.mk
```

- fpwutils.mk が GNU make 用に書かれているので、GNU make が必要
- fpwmake を実行すると honmon という名前で辞書ファイルが生成される。

catalogs の生成

- どんな辞書が含まれているかを示す catalogs ファイルが必要。catalogs.txt を用意。

```
[Catalog]
FileName   = catalogs
Type       = EPWING1
Books      = 1
[Book]
Title      = "国語辞典"
BookType   = 6001
Directory  = "KOKUGO"
```

- fpwmake catalogs を実行。
- fpwmake install でインストール。

終わりに

- 元のテキストデータがあれば、簡単に電子辞書に変換できる。
- 「FreePWING による各種辞書」
<http://openlab.ring.gr.jp/edict/fpw/>
- 「Wikipedia 日本語版」(<http://ja.wikipedia.org/>) を電子辞書に変換できないか模索中。
 - http://meta.wikimedia.org/wiki/Data_dumps から XML ファイルをダウンロードできる。