

# 日本語教育におけるコンピュータ利用

— WWWブラウザを利用した漢和辞典検索 —

波 平 八 郎

## 要 旨

本論文では、留学生に対する日本語学習方法としてインターネットの閲覧ソフトであるWWWブラウザを利用した漢和辞典検索システム構築の方法を提案する。

### 1. OS

漢和辞典検索システムの構築におけるOS(オペレーティング・システム)には、FreeBSDを使用する。FreeBSDは、カリフォルニア大学バークレイ校がリリースした4.4BSD-liteをもとにしたコンピュータ用のBSD UNIXオペレーティングシステムである<sup>1</sup>。

### 2. WWWサーバー

データをインターネット上で公開するためのWWWサーバーには、Apacheを使用する<sup>2</sup>。

### 3. perl

データを処理する言語にはperlを使用する。これはLaly Wallが開発した、テキスト処理に優れた能力を発揮するコンピュータ言語である<sup>3</sup>。オリジナルのperlに日本語処理のためのパッチをあてる<sup>4</sup>。

また、処理速度の向上のため、Apacheにperlをモジュールとして組み込んだmod\_perlを使用する<sup>5</sup>。

データベースの形式は、UNIXシステムで標準的に使われているDBM(データ・ベース・マネジメント)を利用する。perlのDBMには現在五種類のDBMモジュールがある。本論文では、その中でももっとも柔軟性が高いDB\_Fileモジュールを使用する。これは、Berkeley DBにアクセスする、キーと値の長さに制限がない高機能なモジュールである。

### 4. データ

データはYDIC<sup>6</sup>および、Ydicbus<sup>7</sup>を利用する。これらのデータは、ネットワーク上で広く配布されている<sup>8</sup>。

YDICのデータの構造は次のようになっている。

一、1,0,1,1,1

七、1,1,2,3,6

丁、1,1,2,4,2

(以下略)

コマンドで区切られたフィールドはそれぞれ、次のような内容である。

1 漢字

2 所属部首の部首番号

3 部首内画数

4 総画数

5 『新字源』(角川書店、常用漢字版)番号

6 諸橋轍次『大漢和辞典』(大修館書店、旧版、索引篇所載の補遺は除く)番号

Ydicbusの形式は次のようになっている。

(前略)

ノ 4/1

乙 5/1

ノ 6/1

(以下略)

それぞれ、「1 部首をあらわす漢字・2 部首番号・(区切りのスラッシュ)・3 部首の画数」である。

YDICについては漢字をキーにした連想配列

を作成する。Ydicbusについては部首番号をキーにした連想配列を作成する。部首番号とは、『康熙字典』（『新字源』、『大漢和辞典』も踏襲）の214部首を1から順に214までの番号で示したものである。

## 5. 手書き認識

パーソナルコンピュータの事実上の標準OSであるWindows95（NT）に添付されているMS-IME97を併用すると手書き文字の認識が行える。読めない漢字の形をマウスでなぞることにより、『大漢和辞典』『新字源』の漢字番号を知ることができる。また、部首、部首内画数により、両辞典以外の辞典の検索も容易に行える。

## 6. プログラム

- (1) YDICデータのDBMデータベース作成プログラム
- (2) YdicbusのDBMデータベース作成プログラム
- (3) 漢字検索プログラム

- (1) YDICデータのDBMデータベース作成プログラム

```
-----
1: use I18N::Japanese;
2: use Fcntl;
3: use DB_File;
4: tie %KANJI , DB_File, "ydic_db.
    dat" , O_RDWR | O_CREAT, 0644;
5: while (<>) {
6:   chomp;
7:   @data = split(/./);
8:   $kan = shift(@data);
9:   $KANJI{"$kan"} = "@data";
10: }
11: untie %KANJI
-----
```

1行目では、日本語を扱う指定をする。3行目ではBerkeley DBファイルにアクセスするよう指定する。4行目では、「ydic\_db.dat」という

Berkeley DB形式のファイルを、「KANJI」という名前の連想配列に結びつけるという指定をする。9行目では、YDICデータの第一番目のフィールドである漢字をキーにした連想配列を作成することを指定している。

このプログラムの実行により、漢字をキーにした連想配列のデータベース「ydic\_db.dat」が作成される。

- (2) YdicbusのDBMデータベース作成プログラム

```
-----
1: use I18N::Japanese;
2: use Fcntl;
3: use DB_File;
4: tie %BUSHU , DB_File, "bushu_db.
    dat" , O_RDWR | O_CREAT, 0644;
5: while (<>) {
6:   chomp;
7:   @data = split(/#/);
8:   $bus = shift(@data);
9:   @data2 = split(/(¥D+)/, $bus);
10:  $BUSHU{"$data2[2]"} =
    "$data2[1]" . " " . "$data[0]";
11: }
12: untie %BUSHU
-----
```

4行目で「bushu\_db.dat」というBerkeley DB形式のファイルを、「BUSHU」という名前の連想配列に結びつけるという指定をする。7行目では、部首データをまずスラッシュの部分で分割するという指定をする。9行目で部首をあらわす漢字と、部首番号に分割する。10行目で部首番号をキー、部首をあらわす漢字と部首の画数を値とした連想配列を作成する。

このプログラムの実行により部首番号をキーにした連想配列のデータベース「bushu\_db.dat」が作成される。

- (3) 漢字検索プログラム

```
-----
1: use CGI;
2: use I18N::Japanese;
3: use Fcntl;
```

```

4: use DB_File;
5: tie %KANJI, DB_File,
    "ydic_db.dat", O_RDONLY,0644;
6: tie %BUSHU, DB_File,
    "bushu_db.dat", O_RDONLY, 0644;
7: $query = new CGI;
8: print $query->header,
9:   $query->start_html(-title=>
    '漢字検索'),
10:  $query->startform,
11:  "<CENTER><H3> 漢字検索
    </H3></CENTER>",
12:  "漢字を一文字入力してください。",
13:  $query->textfield(-name=>
    'kanji', -size=>5, -override=>1),
14:  $query->submit(-value=> '検索'),
15:  $query->end_form;
16: if ($query->param) {
17:   $kanji = "";
18:   $kanji = $query->param
    ('kanji');
19:   $sonota = $KANJI{ "$kanji" };
20:   if ($sonota) {
21:     @row = split(/ /, $sonota);
22:     ($bushu, $bukaku, $sokaku,
    $jigen, $daikanwa) = @row;
23:     $bushu_sonota = $BUSHU
    {$bushu};
24:     ($bushu_moji, $bushu_
    kakuuu) = split
    (/ /, $bushu_sonota);
25:     $maki = 1;
26:     if ($daikanwa > 1449) {
27:       $maki++;
28:     }
29:     (中略)
30: print "<font size=+4 color=blue>
    $dbname</font><hr width=8%
    align=left>",
31: "【大漢和辞典】の番号: <font size=+2
    color=blue>$maki 卷 </font> の
    <font size=+2 color=blue>

```

```

    $daikanwa</font>",
32: "<p>【新字源】の番号: <font size=
    +2 color=blue>$jigen</font>",
33: "<P>部首: <font size=+2 color=
    blue>$bushu_moji</font>
    <br>(部首の画数: $bushu_
    kakuuu) (部首番号: $bushu)
    <font size=+2>(部首内画数:
    $bukaku)</font>",
34: "<br> 絵画 : <font size=+2>
    $sokaku</font>";
35: } else {
36:   print "<hr> 漢字を一文字入力してく
    ださい";
37: }
38: }
39: untie %KANJI;
40: untie %BUSHU;

```

1行目は、WWWサーバー上でperlを使用するためのインターフェイスを呼び込むための記述。5行目、6行目は、先に作成していた漢字データベースと部首データベースを開くための指定。16行目以降が、実際に漢字が入力された場合の処理プログラムである。18行目で、入力された漢字を変数 \$kanji に保持する。19行目から22行目までで、入力された漢字に関する「所属部首の部首番号・部首内画数・総画数・『新字源』の漢字番号・『大漢和辞典』の漢字番号」をそれぞれの変数に保持する。23行目と24行目で、入力された部首番号をキーにした部首に関する値（部首の漢字・部首の画数）をそれぞれの変数に保持する。

25行目以降は、表示に関わる処理である。26行目では、『大漢和辞典』の漢字番号をもとにして、巻数のインクリメントを行っている。35行目から37行目は、漢字以外の文字が入力されたときのエラー処理である。39行目と40行目でデータベースファイルを閉じる。

## 7. まとめ

WWWを利用した漢字検索の実現により、まったく読めない漢字の検索が容易となる。また、手書き漢字認識ソフトとの併用により、漢字の画数や部首に関して十分に知識を持たない日本語学習の初級者にも容易に漢和辞典を検索することが可能となる。

例えば「名」という漢字の検索結果は次のように表示される。

-----  
**名**

【大漢和辞典】の番号: 2巻の 3297

【新字源】の番号: 977

部首: 口

(部首の画数: 3) (部首番号: 30) (部首内画数: 3)

総画: 6  
-----

WWWを利用した漢字検索システムを次のURLで公開する<sup>9</sup>。

[http://bsd.ics.meio-u.ac.jp/  
perl/kanji4.cgi](http://bsd.ics.meio-u.ac.jp/perl/kanji4.cgi)

<sup>1</sup>FreeBSDは次のURLで保守されている。

<http://www.freebsd.org/>

<sup>2</sup>Apacheは次のURLで保守されている。

<http://www.apache.org/>

<sup>3</sup>Perlは次のURLで保守されている。

<http://www.perl.com/>

<sup>4</sup>Japanization patch 4 by Yasushi Saito, 1996  
Modified by Hirofumi Watanabe, 1996, 1997  
jperl5.004\_04-971016 EUC version

<sup>5</sup>mod\_perlは次のURLで保守されている。

<http://perl.apache.org/>

<sup>6</sup>YDICは、豊島正之氏らの制作になる漢字データベースである。

Ydic by S.Kinsui,K.Furuta,M.Toyoshima  
(1988), based on JIS X0208-1983  
revised according to the suggestion by

massangeana(1991)

<sup>7</sup>Ydicbusは、豊島正之氏の制作になる漢字の部首のデータベースである。

copyright (C) Masayuki TOYOSHIMA,  
1988. All rights reserved.

<sup>8</sup>配布の要領は次のとおり(データの添付文書Ydic.docより)。

「本字書は、営利・非営利を問わず一切の利用・複製・頒布は自由である。但し、複製物、又はそれを部分として含むデータ/ソフトウェアの再頒布を、如何なる形でも妨げてはならない。」

<sup>9</sup>期間は1998年9月までとする。

(なみひらはちろう 国際文化学科助教授)