

ユーザ定義可能な仮想電子辞書システム

今井 健一* 近藤 幹大* 都司 達夫** 寶珍 輝尚** 樋口 健**
*福井大学大学院工学研究科 **福井大学工学部

あらまし 近年、記憶メディアの大容量化と低廉化にますます拍車がかかり、マルチメディア技術の進歩も相俟って、各種電子辞書の出版はその多様性を増してきている。電子辞書は、人手で引く時とは異なり、辞書の構造や配列にとらわれずに辞書の持つ豊富で多角的な情報をユーザが利用することが可能である。本論文では、ユーザの必要とする情報がユーザの望む配列で検索することができる「仮想電子辞書システム」の提案とその設計について述べる。今回は辞書データの構文解析により、辞書の各項目を切り出すために作成された構造化辞書を用いて辞書データの XML 化を行った。これにより、辞書の種類に依存しない汎用的な検索関数を提供することが可能である。ここでは、このような複数の XML 辞書を実体辞書としている。

A User Definable Virtual Electronic Dictionary System

Ken'ichi IMAI* Mikihiro KONDOU* Tatsuo TSUJI** Teruhisa HOCHIN** Ken HIGUCHI**
*Graduate School of Engineering, Fukui University
**Faculty of Engineering, Fukui University

Recently large scaled and low cost storage medias become available, and along with the progress of multimedia processing techniques, various kinds of electronic dictionary are being published. Unlike a paper dictionary, one can extract various information involved in an electronic dictionary regardless of its inherent structure. This paper concerns a proposal and a design of a virtual electronic dictionary system for providing a user definable dictionary by which one can collect and arrange his own information in more than two real dictionaries. In this research, real dictionaries are described by XML. By employing these XML dictionaries our system can provide general retrieving functions for any real dictionaries.

1 はじめに

現在、電子辞書の出版は、記憶メディアの大容量化と低廉化により、その多様性を増してきている[7]。また、インターネットを介したオンライン辞書の検索サービスも充実してきている。本来、辞書には極めて豊富で多角的な情報が内包されており[1]、これらの情報は「辞書の文法」に従って構造的に記述されている。しかし、人

手で辞書を引く時、その構造や配列が固定であるために辞書が潜在的に有する豊富な情報を十分利用することができない。そこで、辞書の持つ情報を有効に利用するためには、辞書の固定構造による強制から開放して、使用者のニーズにかなう情報を提示できる手段が必要である。単に手引きの代替ではなく、この手段の提供こそが、電子辞書の存在意義であると言える。

今回は辞書データの構文解析により、辞書の

各項目を切り出すために作成された構造化辞書を用いて辞書データの XML 化[5]を行った。これにより、辞書の種類に依存しない汎用的な検索関数を提供することが可能である。ここでは、このような複数の XML 辞書を実体辞書としている。

電子辞書を含めた電子出版物の構造化には、たとえば EPWING[6]などの規格が存在しており、それに従った電子辞書も出回っている。また、既存辞書のハイパーテキスト化の方式なども提案されている[2]。しかし、辞書の論理構造(スキーマ)を前面に押し出しユーザーレベルで既存辞書からユーザーニーズに合った辞書を定義し、利用するためのシステムは見当たらない。

2 仮想辞書

個々の辞書データの実体を持っている辞書(英和・和英辞書など)を実体辞書としたとき、仮想辞書は、実体辞書の各項目に対するユーザー定義のビューであると考えられ、データの実体を持たない。ビューは単に、(複数の)実体辞書のどの部分をどのような配列で再配置するかを指定するもので、ユーザーの定義により実体辞書の一部を参照したり、複数辞書を特定の項目を介して、結合させたりすることができる。その際、複数辞書の切り替えを意識させることなく、論理的に閉じた単一の辞書として操作できることが最大の特徴である。ここでは、構造記述力に優れ、また、柔軟な構造解析が行える XML を用いて実体辞書群を構造化している。

2.1 仮想辞書の定義

ここでは、仮想辞書の結合の例として「和訳付英語シソーラス辞書」を紹介する。この辞書は、英語シソーラス辞書で引いた類義語についてその和訳を英和辞書で付記するも

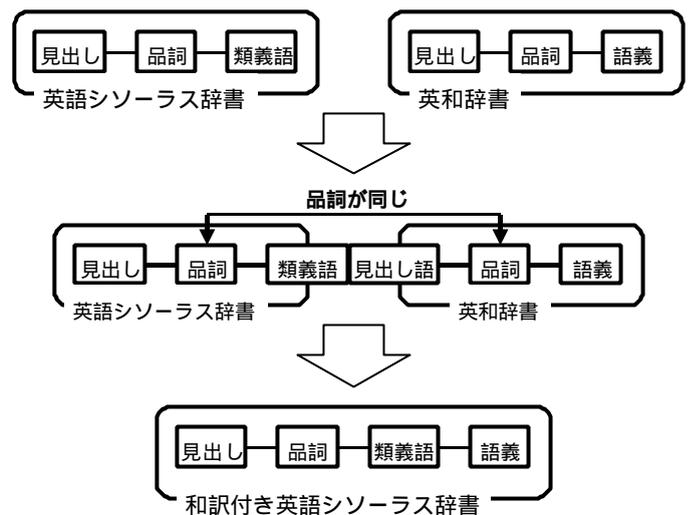


図 1. 和訳付き英語シソーラス辞書の定義

のである(図1)。

ここで、図1のような仮想辞書をユーザが理解する構文で記述する必要がある。仮想辞書の論理構造を定義する **DEFINE 句**は、項目名と括弧の入れ子で表現され、この例では、

和訳付英語シソーラス辞書 (見出し語,
見出し語項目 (品詞, 品詞項目 (類義語,
類義語項目 (語義)))

となる。母体辞書(この例では、**英語シソーラス辞書**《ET》と**英和辞書**《EJ》)を指定する **FROM 句**、仮想辞書の論理構造の各項目と該当する母体辞書の項目との対応付けを行う **ASSIGN 句**、この例では、

見出し語=ET.見出し語
見出し語項目.品詞=ET.見出し語項目.品詞
見出し語項目.品詞項目.類義語=
ET.見出し語項目.品詞項目.語義項目.類義語
見出し語項目.品詞項目.類義語項目.語義=
EJ.見出し語項目.品詞項目.語義

となる。

母体辞書の該当項目の結合情報（この例では、英語シソーラス辞書の類義語と英和辞書の見出し語）を保持する **CONNECTED BY** 句は、

EJ.見出し語=
ET.見出し語項目.品詞項目.語義項目.類義語

となる。

そして、選択条件（この例では「品詞が同じ」）を表す **WHERE** 句

ET.見出し語項目.品詞=EJ.見出し語項目.品詞

が必要である。これらの句をまとめて **DEFINE** 文と呼ぶ。

2.2 仮想辞書の検索

仮想辞書の検索には単に見出しを指定するだけではなく、仮想辞書の論理構造を検索したい項目名と検索条件となる定数で記述する **FIND** 文を用いる。

この例では、和訳付英語シソーラス辞書の見出し語が“test”、品詞が“名詞”のものについて品詞項目以下をすべて検索することを示している。

3 仮想辞書システム

現在、本研究において従来構築してきた仮想辞書システムは、辞書の検索をその辞書固有の検索関数に頼っているため、新規に辞書が追加されると、検索関数を新たに作る必要があり、このことに多大な労力が費やされた。そこで、本研究では辞書データを XML 化して、既存のアプリケーションに大幅な変更を加える必要の無い仮想辞書システムの設計・構築を目指した。

3.1 辞書データの XML 化

前述したように、辞書は「辞書の文法」に従っ

た論理構造をもっている（図 2）。

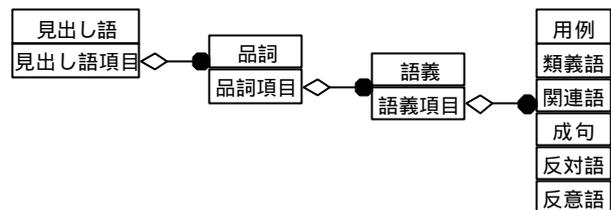


図 2. 英語シソーラス辞書の論理構造

この構造に従い、XML の文章構造定義（DTD）を記述して、辞書データの XML 化を行った。

図 2 の構造を DTD にしたものを以下に示す。

```
<!ELEMENT thesaurus (見出し語,見出し項目)>
<!ELEMENT 見出し語 (#PCDATA)>
<!ELEMENT 見出し項目 (品詞,品詞項目)*>
<!ELEMENT 品詞 (#PCDATA)>
<!ELEMENT 品詞項目 (語義,語義項目)*>
<!ELEMENT 語義 (#PCDATA)>
<!ELEMENT 語義項目 (用例*,類義語*,関連語*,
成句*,反意語*,反対語*)>
<!ELEMENT 用例 (#PCDATA)>
<!ELEMENT 類義語 (#PCDATA)>
<!ELEMENT 関連語 (#PCDATA)>
<!ELEMENT 成句 (#PCDATA)>
<!ELEMENT 反意語 (#PCDATA)>
<!ELEMENT 反対語 (#PCDATA)>
```

この DTD を元にして既存電子辞書のデータを XML 化した。見出し語「test」に対して XML 化したものの一部分を以下に併せて載せておく。図 2 において、例えば、「見出し語」や「品詞」等は辞書のテキストとして現れる文字列であり、以下ではこれらの項目を具体項目という。また、「見出し語項目」や「品詞項目」はそれ以下の項目（の集合）に対して意味を与えるメタ項目であり、抽象項目と呼ぶ。

ここで、前提として、実体辞書、仮想辞書に関わらず、抽象項目のオカーレンス（インスタンス）を一意に決定する具体項目のオカーレンスが必ず存在するものとして、この具体項目を

抽象項目に対するセレクトと呼ぶこととする。

```

<見出し>test</見出し>
<見出し項目>
  <品詞>名詞</品詞>
  <品詞項目>
    <語義/>
    <語義項目>
      <類義語>EXPERIMENT</類義語>
      <類義語>experimentation</類義語>
      <類義語>trial</類義語>
      <類義語>trial and error</類義語>
      <類義語>trial run</類義語>
      <関連語>inspection</関連語>
      <関連語>scrutiny</関連語>
      <関連語>confirmation</関連語>
      <関連語>corroboration</関連語>
      <関連語>substantiation</関連語>
      <関連語>verification</関連語>
    </語義項目>
  </品詞項目>
  <品詞>動詞</品詞>
  <品詞項目>
    <語義/>
    <語義項目>
      <類義語>TRY 1</類義語>
      <類義語>check</類義語>
      <類義語>examine</類義語>
      <類義語>prove</類義語>
      <関連語>assay</関連語>
      <関連語>essay</関連語>
      <関連語>confirm</関連語>
  </品詞項目>
</見出し項目>
  
```

3.2 XML による仮想辞書システム

図 3 は、本研究で設計した仮想辞書システムである。

本システムは、ユーザインタフェース・仮想辞書サーバ・実体辞書サーバの 3 つから構成される。ユーザインタフェースからの要求に応じて仮想辞書サーバで処理を行い、必要に応じて実体辞書サーバに辞書データの問い合わせを行う。この仮想辞書サーバと実体辞書サーバの詳細は以降で説明する。

4 仮想辞書サーバ

仮想辞書サーバは、ユーザの要求を受け取り、実体辞書サーバに対して辞書データの検索要求を出す。図 3 に示したように、

以下の処理を行う。

- 実体辞書の登録・削除
- 仮想辞書の定義・削除
- 辞書検索

まず、実体辞書の登録と削除は、実体辞書サーバに置かれている実体辞書の登録と登録の削除を行う。仮想辞書サーバが利用可能な実体辞書の情報を操作する。次に、仮想辞書の定義と削除では、前述した『和訳付き英語シソーラス辞書』のような仮想辞書の定義の登録と抹消を行う。最後に、辞書検索は、定義された仮想辞書を検索するものである。

4.1 仮想辞書の定義

ここではユーザにより新しく定義された仮想辞書定義を仮想辞書サーバが保存する際の処理に関して述べる。前述したように、仮想辞書は DEFINE 文によって定義される。DEFINE 文は、ユーザによってユーザインタフェースを介して生成される。生成された

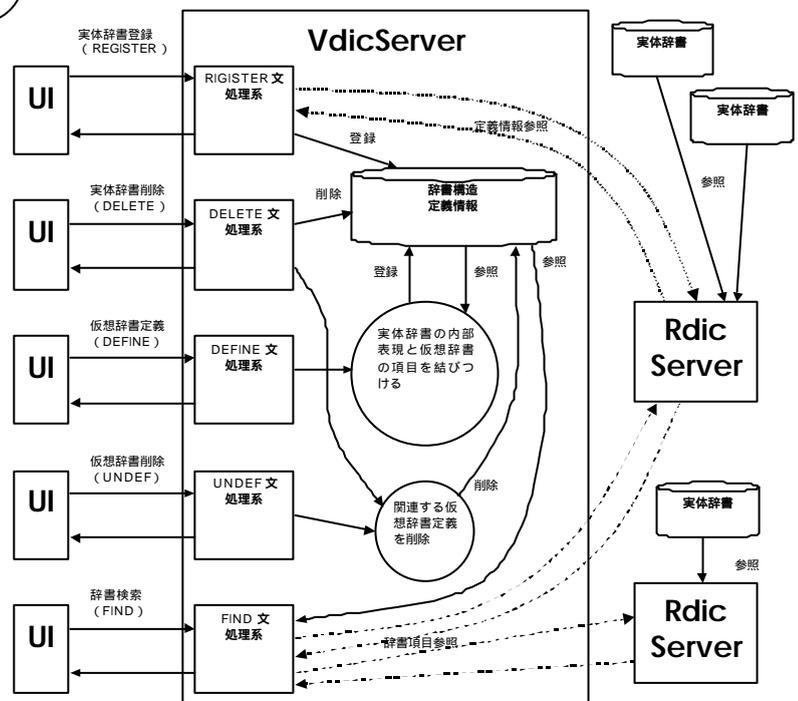


図 3. 仮想辞書システム

DEFINE 文は仮想辞書サーバに送られ、格納される。DEFINE 文の各句の各項目の取り出しには DOM ツリーの検索ライブラリを使用する。そのため、DEFINE 文は XML 文に変換される。ここで DOM とは XML 文書のタグとテキストを 1 つのノードとして構築されたオブジェクトの

ツリーを動的に操作するためのインタフェースのことである。前述した「和訳付英語シソーラス辞書」の DEFINE 文を XML 文に変換したものを示す。仮想辞書の具体項目として参照される母体辞書のパス (ASSIGN 句の右辺) は名前空間を使用した XML で表現し、仮想辞書の具体項目の下位に展開している。FROM 句・CONNECTED BY 句・WHERE 句はルート要素 (和訳付英語シソーラス辞書) の直下に置く。

```

<和訳付英語シソーラス辞書>
  <FROM xmlns:ET="/data/dtd/ET">
    英語シソーラス辞書
  </FROM>
  <FROM xmlns:EJ="/data/dtd/EJ">
    英和辞書
  </FROM>

  <CONNECTED_BY>
    EJ.midashi=
      ET.midashi_lv.part_lv.sense_lv.syn
  </CONNECTED_BY>

  <WHERE>
    ET.midashi_lv.part=EJ.midashi_lv.part
  </WHERE>

  <見出し語>
    <ET:midashi/>
  </見出し語>
  <見出し項目>
    <品詞>
      <ET:midashi_lv>
        <ET:part/>
      </ET:midashi_lv>
    </品詞>
    <品詞項目>
      <類義語>
        <ET:midashi_lv>
          <ET:part_lv>
            <ET:sense_lv>
              <ET:syn/>
            </ET:sense_lv>
          </ET:part_lv>
        </ET:midashi_lv>
      </類義語>
      <類義語項目>
        <語義>
          <EJ:midashi_lv>
            <EJ:part_lv>
              <EJ:sense/>
            </EJ:part_lv>
          </EJ:midashi_lv>
        </語義>
      </類義語項目>
    </品詞項目>
  </見出し項目>
</和訳付英語シソーラス辞書>

```

4.2 仮想辞書の検索

母体辞書である実体辞書に対する検索文 (FIND 文) の生成には、この XML による仮想辞書の DEFINE 文の DOM ツリーを使用して、逐次、仮想辞書の項目定義を取り出すことによって行う。以下では、先に示した FIND 文、

```

FIND 和訳付英語シソーラス辞書("test",
    見出し項目("名詞", 品詞項目))

```

を使って処理の流れを追う。

まず、FIND 文の解析結果から仮想辞書名が得られる。仮想辞書サーバでは該当する辞書の DEFINE 文が読み込まれる。この場合は、和訳付英語シソーラス辞書である。このとき、この DEFINE 文の DOM ツリーが作成される。次に、FROM 句 (FROM タグの部分) が参照され、検索を行う母体辞書名が決定される。また、CONNECTED BY 句 (CONNECTED_BY タグ) や WHERE 句 (WHERE タグ) も参照され、結合情報や結合条件が参照できるように保持する。

さて、FIND 文によりこの仮想辞書の見出し語が "test" と指定されているので、DEFINE 文の見出し語ノード以下は、"test" で置き換えられる。FIND 文では、見出し語

項目という抽象項目が参照されるが、そのセレクトは"test"となる。このセレクトにより選択される"見出し項目"が参照される。この見出し項目には品詞として"名詞"が指定されているので、DEFINE 文の品詞の部分が"名詞"で置き換えられる。この"名詞"がセレクトとなり、対応する"品詞項目"が参照される。そして、FIND 文の品詞項目に関して処理が行われる。ここで、品詞項目は抽象項目であり、その下位に項目が指定されていないので、仮想辞書定義の品詞項目以下の全てが検索対象となる(類義語および類義語項目・語義)。検索項目となった類義語では、その下位の英語シソーラス辞書のsyn へのパス部分が参照されて、実体辞書サーバへ問い合わせをするためのFIND 文が生成される。midashi_lv や part_lv は抽象項目で、対応する具体項目が(midashi や part)がある。今、ユーザが問い合わせをしたFIND 文によりmidashi は"test"、part は"名詞"となっており、それぞれ、midashi_lv、part_lv のセレクトとなる。また、sense_lv に関しては、対応する具体項目 sense は存在するがこれまでの処理によって限定されていないので、任意のsense の下位にあるsyn を取り出すためのFIND 文が作成される。

```
FIND thesaurus (midashi["test"],
midashi_lv (part["名詞"], part_lv
(sense[ * ], sense_lv (syn[ * ]))))
```

このように、生成された母体辞書(実体辞書) thesaurus に対する検索文は thesaurus を管理する実体辞書サーバに送り出される。実体辞書サーバでは以下の類義語が検索される。

```
EXPERIMENT, experimentation, trial,
trial and error, trial run
```

続いて、仮想辞書サーバではこれらの類義語をセレクトとして、抽象項目である類義語項目が参照され、語義に処理が及ぶ。語義に対応するASSIGN 句の右辺はEJ:midashi_lv ノードで始まっている。このmidashi_lv にはセレクト項目midashi が存在しており、EJ:midashi はCONNECTED BY 句により英語シソーラス辞書の類義語に結合されるように定義されているので、先に調べた英語シソーラス辞書の類義語を見出し語に指定して、母体辞書である英和辞書を引くようにFIND 文が作成される。また、EJ:part_lv ノードではWHERE 句により選択条件は品詞が同じものを選ぶように指定されているので、今は名詞のものについて検索が行われる。よって、上記の5つの類義語ごとに作成されるFIND 文の1つは次のようになる。

```
FIND eiwa (midashi["trial"],midashi_lv (
part["名詞"], part_lv (sense[ * ])))
```

最終的にユーザの要求した検索文に対する結果は2つの母体辞書に対する上記の検索結果をXML 文に合成して、次のようになる。

```

<和訳付英語シソーラス辞書>
<見出し語>test</見出し語>
<見出し項目>
<品詞>名詞</品詞>
<品詞項目>
<類義語>EXPERIMENT</類義語>
<類義語項目>
<語義/>
</類義語項目>
<類義語>experimentation</類義語>
<類義語項目>
<語義>実験すること, 実験作業.</語義>
</類義語項目>
<類義語>trial</類義語>
<類義語項目>
<語義>1 試み, 企て, (attempt). 2
形容詞的に 試験的な, 試すもの.
3 試練; 困難; 厄介なこと, 困りもの.
4 審判 法律 裁判, 審理; 公判.</語義>
</類義語項目>
<類義語>trial and error</類義語>
<類義語項目>
<語義/>
</類義語項目>
<類義語>trial run</類義語>
<類義語項目>
<語義/>
</類義語項目>
</品詞項目>
</見出し項目>
</和訳付英語シソーラス辞書>

```

5 実体辞書サーバ

実体辞書サーバは、仮想辞書サーバからの問い合わせに応じて処理を行う。仮想辞書サーバから発行される **FIND 文**を受け取り、所望のものを検索して、結果を返す。

実体辞書の検索には、DOM を利用しているが、メモリの消費量が大きく1つの辞書をメモリ内に保持することができないので、見出し語によるインデックスを作成している。そして、検索対象となる見出し語の辞書項目のみを辞書から抽出して、DOM ツリーを作成して、必要項目の取り出しを行っている。

```

FIND thesaurus (midashi["test"],
midashi_lv (part["名詞"], part_lv
(sense[ * ], sense_lv (syn[ * ])))

```

これは、前述したように英語シソーラス辞書から見出し語 “ test ”・品詞 “ 名詞 ” である類義語を抽出するために仮想辞書サーバで作られた実体辞書サーバへの問い合わせを行う FIND 文である。これを受け取った実体辞書サーバは、英語シソーラス辞書の見出し語 “ test ” の項目を辞書データから抽出して、DOM ツリーに変換する。この DOM ツリーから検索を行うが、midashi_lv に対するセレクト項目 midashi、part_lv に対するセレクト項目 part により、抽出すべき項目を絞っていき、前述したように5つの類義語が抽出し、仮想辞書サーバに結果を返す。

なお、ここでは、実体辞書の例として、三青銅の新クラウン和英辞典[3]、および入センチュリー英和辞典[4]を構文解析により構造化したものに対して XML 化を行い使用している。

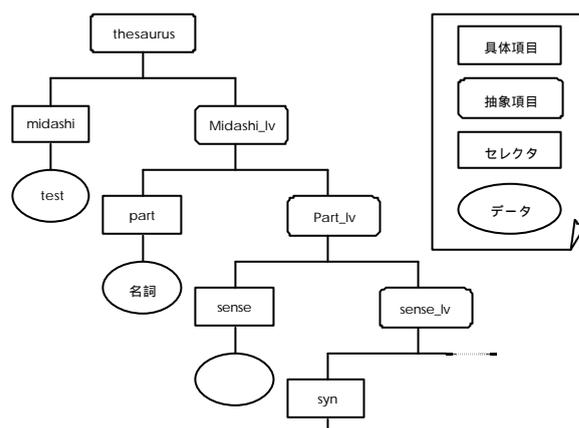


図 4. 実体辞書の DOM ツリー

6. おわりに

本論では、仮想辞書システム的设计について述べた。この設計に基づき現在構築中である。他

にも、本論では述べなかったが、検索結果の XML に対して GUI で表示させるための XSL の自動生成に関しても検討も行っている。

既存の実体辞書の問題点として、記述が不統一でスキーマに沿った構造化ができない項目説明があったり、項目として切り出す時の終端子として決めたマークが必ずしも一定でなかったり、スキーマとしてこちらが想定していない項目が出現したりする例外が出現し、構造化できないものがあった。可能な限り手作業で対応したが、なお十分ではない。これらのいくつかは、言葉が持つ多様性を十分に説明するためには、ある程度不可避であると思われる。電子辞書の規格化がこのあたりも含めて十分に検討され、標準化されたならば、逆に辞書の編集段階から、規格準拠を考慮でき、統一のとれた電子辞書を構築できる。本研究における仮想電子辞書の場合には、このような標準化された実体辞書が母体辞書として使用できるような環境が望まれる。

また、XML の世界は現在も発展中であり、その動向から目が離せない。現在は DOM を主に使っているが、今後の発展次第では様々な XML データを扱う技術 (XSL ・ XPath ・ XML-QL ・ DOM ・ SAX など) の中からどれを選ぶかを慎重に且つ迅速に決定し、それらをシステムに組み込みたいと考えている。

謝辞 和英辞書および英和辞書の元データは三省堂 (株) より提供していただいた。謝意を表す。

参考文献

- [1] 吉田 将, 辞書構築における諸問題, 情報処理, 27, 8, pp.933-939, 1986.¥¥
- [2] 内藤, 山下, 松山, 柵木, オンライン辞書のハイパーテキスト化手法, 情報処理学会論文誌, 34, 2, pp.320-329, 1993.
- [3] 山田和男 編: 新クラウン和英辞典, 三省

堂.

[4] 木原, 福村, 芦川 編: ニューセンチュリー英和辞典, 三省堂.

[5] W3C(XML) <http://www.w3c.org/>

[6] EPWING コンソシアム, <http://www.epwing.or.jp>

[7] <http://www.m-w.com/cgi-bin/netdict>