

XMLを利用した日本古典史料の英日全文連携検索システムの構築 －日米共同研究について－

桶谷猪久夫^{*1} Delmer Brown^{*2} 大久保祐子^{*3} 山尾正之^{*4}

A Construction of the Full Text Coordinated Retrieval System of Japanese Historical Resources using XML (Extensible Markup Language) －A Collaborative Research between Japan and the U.S.－

Ikuko Oketani^{*1} Delmer Brown^{*2} Yuko Okubo^{*3} Masayuki Yamao^{*4}

Abstract

The goals of this full text coordinated retrieval system are to study ancient Japanese culture such as Shinto and Japanese spiritual life, to study the historical development of ancient Japanese state and its structure, and to conduct ethnological/ethnographic research. It further aims to send out the information regarding Japanese culture to the world and to promote the international collaboration. This system will provide the means for making a database of texts available to scholars not only who are Japanese speakers but English speakers, thus reaping synergetic research results and promoting collaboration between English-speaking researchers and Japanese researchers.

The texts we have chosen to work with are 1) Japanese ancient chronologies such as *Kojiki*, *Nihon Shoki* and *Shoku Nihongi*, 2) *Engishiki*, a collection of laws and regulations on shrines, 3) *Izumo Fudoki*, a local document of a certain geographic area, and 4) *Gukansho*, the representative history text of the Middle Ages. We intend to digitize the twenty-five volumes of Japanese classic texts and to facilitate an internet-based coordinated search, browse and reuse between the texts using both English and Japanese.

The purpose of this full text coordinated retrieval system is to assist historical research that requires a large volume of data, without any special processing, i.e.

* 1 : おけたに いくお : Faculty of Human Sciences, Osaka International University

* 2 : デルマー ブラウン : University of California, Berkeley 〈「投稿者の資格基準」①により承認〉

* 3 : おおくぼ ゆうこ : University of California, Berkeley 〈「投稿者の資格基準」①により承認〉

* 4 : やまお まさゆき : Osaka International Junior College 〈2005. 6. 22受理〉

Japanese historical documents that do not have fixed data forms. As these documents require a search function using a certain program, we used the CGI (Common Gateway Interface) function that provides the contents retrieved from the Web server upon request of the browser with a dynamic and interactive environment. As the number of stored historical documents increases, the efficiency rate of the search function will decrease if we depend on the technique of matching simple letter patterns. The difficulty in extracting certain subjects from the texts that have sentence structures (logic structures) also results in significant limitations and exposes the weakness of the system. In order to solve these problems, it is important to develop a search method focusing on the document structures and descriptions of Japanese classic texts and to design and implement a more robust system.

XML (Extensible Markup Language), which could define the logic structure of the documents and attributes in addition to texts, has been receiving attention, and certain standards of the language have been set, which we are using for our standard for managing, circulating, and providing documents on the Web. XML enables us to record information about each text such as notes, titles, extracts, important insertions, and interactive references independently from the text. In order to take full advantage of these features of the language we employed XML to define the logic structure of the texts, and designed the full text coordinated retrieval system. In this paper, we explain the processes and problems we encountered in developing the full text coordinated retrieval system, describing our efforts to increase efficiency and the addition of various search functions.

Key Words

Full Text Retrieval System, Japanese Historical Resources, Internet, XML
Kojiki (the Imperial Records of Ancient Matters),
Nihon-Shoki (the Imperial Chronicle of Japan),
Shoku-Nihongi (Chronicles of Japan, continued, from 697-791 A.D.)

1. はじめに

近年、人文科学分野の研究でも手元の史料、新たに発掘した史料や各種文献が、手軽で安価なパソコンを利用し電子テキストや画像ファイルとして入力・格納・蓄積されてきている。また、インターネット利用の急激な普及は、それを利用した電子化された情報の公開を一般化させ、コンピュータ利用による研究者間の共同研究の促進、研究への新しい視点、新しい研究課題と研究方法を生み出せないかが模索されてきている。

これは歴史学研究分野においても例外でなく、古典文献を電子化し研究に活用しようとする動きが盛んになりつつある。このような状況下で、我々は日本古典史料を対象に、その

文書構造や歴史的記述方法に着目し設計した英日全文連携検索システムを開発し、インターネット上に公開することにより、歴史学研究を援用し、さらに、国際的なコラボレーションを促進することを目的とする。

我々は、本システムの設計と開発をカリフォルニア大学バークレー校のブラウン教授が運営するJHTI (Japanese Historical Text Initiative) プロジェクトと共同で開発した。このプロジェクトの目的は、日本神道を中心に古代日本の文化と日本人の精神生活の研究、その当時の事物や社会の様相を研究する資料を提供することにより、日本文化の世界への発信と国際的なコラボレーションを促進する研究である。また、外国人研究者の古典入門や研究支援だけでなく日本に関する教育にも役立つと思われる。最終的には、日本古典文献25巻（後述）のデジタル化とデータベース化である。

大量で特別に加工されていない一次情報である歴史文献を対象に情報検索システムを実現するとき、プログラムを介した検索を必要とする。そのためCGI(Common Gateway Interface)と呼ばれる機能を利用した歴史史料検索システムを実現し、インターネット上のWWW(World Wide Web、以下、Webという)で公開してきた^[1,2,3]。しかし、データ量が膨大になったとき、全文からの単純なパターンマッチング技法だけでは検索効率を考慮したとき問題があり、また検索条件を適切に指定できず効率的な検索には大きな制約がある。大量のデータから利用者の所望のデータを高速にかつ効率的に検索するには、全文検索システムが必要になる。

英日連携検索システムを実現するには、日本古典史料の日本語、英訳文とページイメージ画像ファイルに対して、連携して検索を可能にすることが必要である。そのため、それら各文書に対して文献間連携用と属性のデータ記述の定義が必要である。現在、文書の実体と共にその論理構造(階層構造)や属性(XML Schema)を定義可能なマークアップ言語XML(Extensible Markup Language)が注目され、一連の規格も制定されてきており、Web上での文書管理・流通・提供の実質的な標準になりつつある。XMLは文書構造を記述するだけでなく、我々が取り扱う各文献に出現する注釈、解題、抄録、貴重な書き込み、相互参照などの情報や知識を文書本体と独立して表現することが可能である。

本稿では、全文検索システムとしてパトリシア(Patricia)ツリー方式を採用しているデータベース管理システム(カナダ、OpenText社)を格納構造として用い、XML化された各文献の文書に対しての検索手法を開発した。まず、Webから検索・表示が可能になるように、高速検索エンジン(PAT70)を利用し、UNIXシステム固有のバッファであるパイプ(PIPE)を利用して検索要求とその検索結果がやり取りできるように設計した。さらに、利用者と高速検索エンジン間のユーザーインターフェースをアプリケーションプログラムで作成した。最後に、我々が設計し実現した英日全文連携検索システムの目的と概要、各種機能の特徴と問題点、今後の検討課題について述べる。

2. 英日全文連携検索システムの目的と対象文献

2-1. 本検索システムの目的

本英日全文連携検索システムの目的は、特に日本神道を中心に古代日本の文化と日本人

の精神生活の研究、日本古代国家の成立史や構造の研究、民俗(民族)学的研究であり、日本文化の世界への発信と国際的なコラボレーションを促進する研究である。つまり、英語圏の研究者や学生の日本史・国文学の研究に貢献することであり、また、日本の研究者との共同研究を促進することで研究の相乗的な効果を追求することである。さらに、外国人研究者の古典入門や研究支援だけでなく日本に関する教育にも役立つと思われる。本研究は、Web上で日本古典史料の文書構造を解析し、また歴史的記述方法に着目した検索手法を開発し、英語と日本語・漢文(または、両言語)を利用した文献内検索、文献ページイメージ画像との連携、文献間連携検索、閲覧、再利用、語彙検索を目標にしている。

2-2. 本検索システムの対象文献の概要

本英日全文連携検索システムが直接対象とする文献は、日本の記紀である「古事記」、「日本書紀」や「続日本紀」、神祇関係の法令である「延喜式」、特定の地方誌の文書である「出雲国風土記」、歌集「万葉集」、天台座主の慈円の作で承久2年(1220年)ごろ成立した中世の代表的歴史書(全7巻から構成)「愚管抄」^[23,24]である。さらに、後述の日本古典文献25巻のデジタル化、Web上で英語と日本語(または、両言語)を利用した文献内検索と文献ページイメージ画像ファイルとの連携、文献間連携検索、閲覧と再利用を目標にしている。

以下に、本英日全文連携検索システムが直接対象にした文献のいくつかの概要を簡単に説明する。

(1) 古事記^[9,10,11]

「古事記」は、第40代天武天皇(在位西暦673-686年)の命により、諸家に伝わる天皇家の歴史(帝記)と神話や伝説(本辞)を収集し、虚偽を除き真実と思われるものを定め、それを稗田阿礼に覚えさせた。それをもとに第43代元明天皇が任命した太安万侶が和銅5年(712年)に撰録献上したとされる日本に現存する最古の歴史文学書である(「古事記」序文)。

上巻・中巻・下巻の3巻から構成され、上巻は神代の物語、つまり神話を語り、中巻から人代の物語、つまり天皇の代に入り、大和朝廷を創造したとされる初代神武天皇から第15代応神天皇まで、下巻は第16代仁徳天皇から第33代推古天皇までが記されている。

本英日全文連携検索システムが対象とした「古事記」の文献は、江戸時代中期の少壮気鋭の国学者であった本居宣長(1730-1801)が30余年の歳月をかけて完成した古事記研究からの「訂正古訓古事記 4刻」(永田文昌堂、1874年)である。

(2) 日本書紀^[14,15,16]

「日本書紀」は、官撰の国史という性格を持っており、漢文で編年体の体裁で書かれている。巻1と巻2が神代の物語、古事記上巻とほぼ同じく、天地開闢、男神の伊弉諾尊(いざなぎのみこと)と女神の伊弉冉尊(いざなみのみこと)の国生み・神生み、天照大神・素戔鳴尊(すさののみこと)の誕生などが続く。各段に「一書」と称する別伝を種々掲げているという特徴がある。巻3から初代神武天皇の東征から第40代持統天皇までの天皇の

代が記され、養老4年(720年)、舎人親王が合計30巻と系図1巻を撰進奏上した(「続日本紀」)。

本英日全文連携検索システムが対象とした「日本書紀」の文献は、江戸時代の儒学者で尾張藩士、河村秀根、益根父子が60年の月日を費やし刊行した「日本書紀」の注釈書である「書紀集解」30巻(1756年第一巻序)である。

(3) 続日本紀^[19,20]

「続日本紀」は、文武天皇元年(697)から桓武天皇の延暦十年(791)まで、九代95年間の歴史を記述した漢文の史書である。少なくとも8世紀中葉以後2度の編さん作業を経過した後、延暦16年(797年)に桓武天皇の命により、藤原朝臣継縄(ふじわらのあそんつぐただ)・菅野朝臣真道(すがののあそんまみち)らが、ほぼ現存する形に近い40巻を完成させ桓武天皇に進上した。「日本書紀」に次ぐ第2の正史として編纂され、古代の歴史・国文学を研究する上で重要な文献であり、奈良時代の資料として信頼できる史実性を備えており、また万葉集の歌の背景を理解するのに重要な資料ともなっている。

(4) 出雲国風土記^[25]

風土記は、第43代元明天皇の和銅6年(713年)における撰録の官命に基づいて、各国の国庁において順次、「解文」形式を主体とした風土記が編纂され言上された。当時の全国60余りの国すべてについて風土記がつくられたと思われるが、現在まで残っているのは「出雲国風土記」(島根県)、「常陸国風土記」(茨城県)、「播磨国風土記」(兵庫県)、「肥前国風土記」(佐賀県、長崎県)、「豊後国風土記」(大分県)の5か国風土記と諸書に引用された逸文数十か条がある。

(5) 神皇正統記^[27]

「神皇正統記」は、大日本者神國也。天祖ハジメテ基ヲヒラキ、日神ナガク統ヲ傳給フ。我國ノミ此事アリ。異朝ニハ其タグヒナシ。此故ニ神國ト云也。

で始まり、神話的神代から後村上天皇(南朝)までの出来事や皇位継承などが天皇を中心に書かれている。特に、鎌倉幕府を倒し建武の中興を開始した後醍醐天皇時代の歴史が詳しく、北条氏や足利氏は忠義がなかったことをとがめ、南朝(吉野朝廷)の正統性を強調している。その証拠としては、三種の神器(鏡、曲玉、剣)は南朝にあると伝えている。

「神皇正統記」は、後醍醐天皇の側近であった北畠親房(きたばたけちかふさ)が南朝の勢力を盛り返そうと次男顕信(あきのぶ)と吉野を出発し、南朝方の常陸の国に入り、北朝方との戦いを繰り返した筑波山麓にたつ小田城で完成された(延元4年:1339年)。

3. XML文書の設計概念

3-1. XMLによる文書標準化の特徴

格納文献数が少量でありデータ量が多くない場合、我々が既にインターネット上で公開している英日全文連携検索システムはかなり有効に作用すると思われる。しかし、格納文

献数が増大し、デジタル化された情報が膨大になると単純な文字列パターンマッチング技法のみでは、その検索効率が低下するという問題がある。また、歴史史料のようなデータに一定の形式をもっていない、文書構造(論理構造)を持つ文書から特定の項目だけを抽出することは困難であるなど、大きな制約と弱点が生じてくる。これらの問題を解決するため、文書の論理構造を定義可能なマークアップ言語(XML: Extensible Markup Language)を採用し全文検索システムを実現した。

XML形式はデータ交換規約として国際的に標準化された規格であり、その利用方法に関しても規格が整備されており、ツールや開発環境等の多くが公開されている。本テキストデータをXML形式とすることにより、厳密な構造チェックがツールにて容易に可能であり、またデータ利用においても、多くのツール、開発環境が利用でき、効率よくデータ整備が行える。

XMLの特徴として、(1)任意の新しいタグや属性名の定義が可能である。(2)階層型のタグ構造定義により、文書／データを構造化することが可能であり、章立てや段落など文書構造の情報を定義可能である。(3)文書構造を定義する文書型定義DTD(Document Type Definition)を設定することによって情報の操作、記述が容易である。(4)通常使用する文書(キャラクタベースの文書データ)により、プラットフォーム及びアプリケーションから独立している。

本設計では、各文献のテキストデータを本文和文、本文英文とノート英文に分割し保存した。さらに、各文献の共通事項を以下のように定義した。

- (1) データ形式：データ形式はXMLをベースとする。
- (3) 文字コード：UTF-8コードとする。
- (4) 外字定義：UTF-8コードにて定義されていない外字に関しては、%ufxxxx;形式にて定義し、外字を画像ファイルとして貼り付ける。
- (5) 構造定義：文書型定義DTDの設計で説明する。

3-2. 文書型定義DTD(Document Type Definition)の設計

日本古典史料のような文献は文献一次情報のままであり、内容的に複雑な意味を持つ文書構造を持っている。古事記、日本書紀や続日本紀は、巻番号、序文、要約、所見、本文などから構成され、その本文は本文を記述する基本構造であるパラグラフ、勅撰、割注、注釈(Note)、各項目などの物理構造から、さらに、編年体で記述されている続日本紀では重要な役割をする年構造、月日構造(西洋暦、和暦)などの論理構造から構成される。また、複数テキスト情報、テキスト画像ファイルと分割画像ファイルの連携検索を実現するために重要な物理構造としてのページ番号、パラグラフ番号、継続ページ数、ノート数、対応画像数(原本の文字画像)などで構成されている。これらのXMLでタグ付けされた文書の要素を利用した有効な検索を実現しなくてはならない。以下に、本文文書のツリー構造の例(図1)と文書型定義DTDの例を示す。

XMLを利用した日本古典史料の英日全文連携検索システムの構築

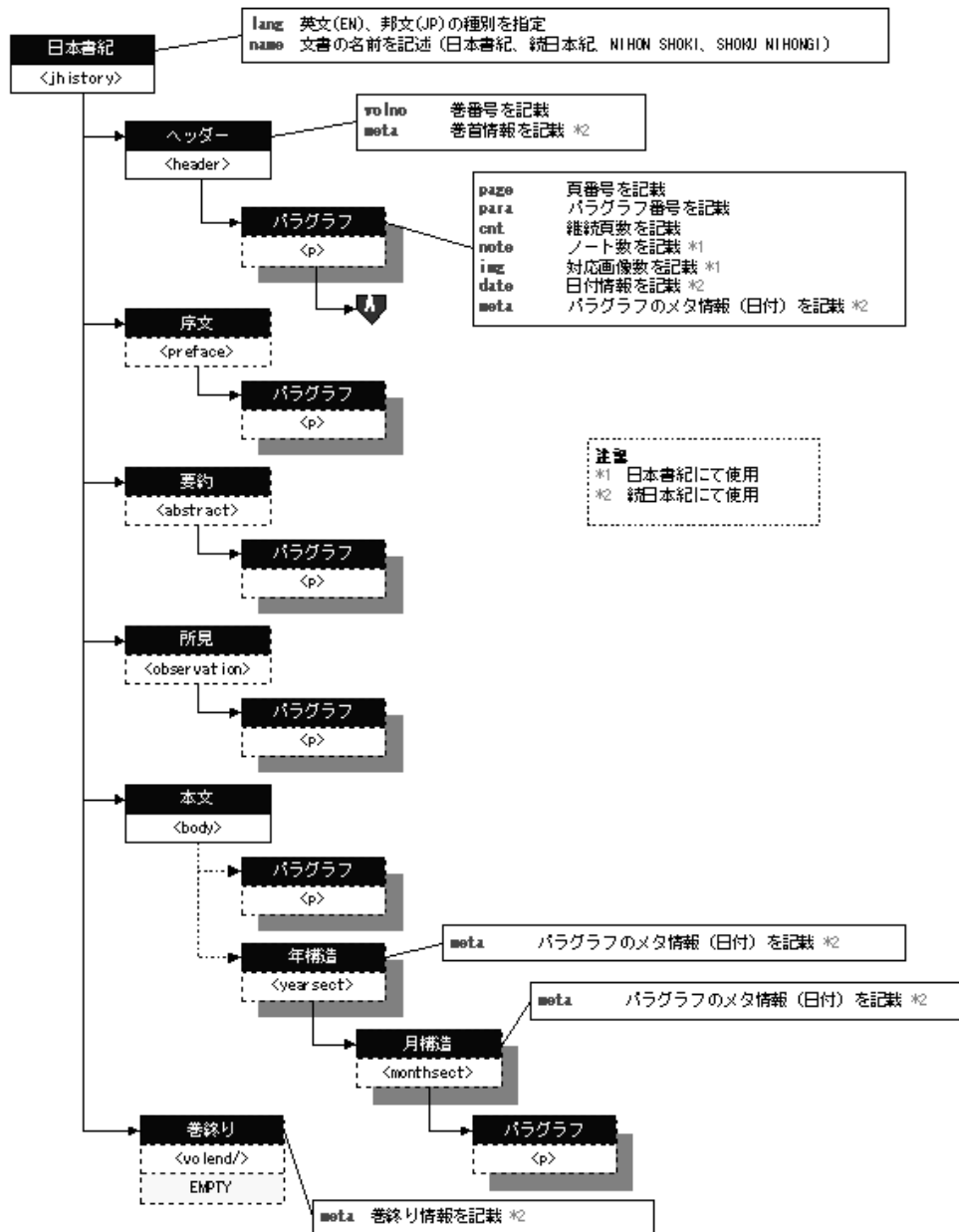


図 1. 日本書紀、続日本紀の本文文書のツリー構造の一部の例

[文書型定義D T Dの例]

```

<!--::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::-->
<!-- NIHONGI / SHOKU NIHONGI DTD Ver1.0 2005 -->
<!-- Osaka International University -->
<!--::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::-->
<!-- Parameter Entities for reference and float elements -->
<!--::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::-->
<ENTITY % refitem "imgref | noteref">
<ENTITY % floateitem "br | name | place | ritual | shrine |
other | annote">
<!--::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::-->
<!-- ELEMENT DECLARATIONS -->
<!--::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::-->
<ELEMENT jhistory (header, preface?, abstract?, observation?, body, volend? )>
<ATTLIST jhistory
lang
(JP|EN)
#REQUIRED
name
CDATA
#IMPLIED >
<ELEMENT header
(p*) >
volno
CDATA
#IMPLIED
meta
CDATA
#IMPLIED >
<ELEMENT body
(p* | yearsect* ) >
<ELEMENT yearsect
(monthsect*) >
<ATTLIST yearsect
meta
CDATA
#IMPLIED >
<ELEMENT monthsect
(p*) >
<ATTLIST monthsect
meta
CDATA
#IMPLIED >
<ELEMENT p
ateitem: ) * >
<ATTLIST p
page
CDATA
#IMPLIED

para
CDATA
#IMPLIED

cnt
CDATA
#IMPLIED

note
CDATA
#IMPLIED

```


XMLを利用した日本古典史料の英日全文連携検索システムの構築

```

img
CDATA
#IMPLIED

date
CDATA
#IMPLIED

meta
CDATA
#IMPLIED >
<!ELEMENT selector
<!ELEMENT volend
<!ATTLIST volend
meta
CDATA
#IMPLIED >
<!-- *****-->
<!-- Reference Elements -->
<!-- *****-->
<!ELEMENT imgref
<!ATTLIST imgref
rid
CDATA
#IMPLIED >
<!ELEMENT noteref
<!ATTLIST noteref
rid
CDATA
#IMPLIED >
<!-- *****-->
<!-- Float Elements -->
<!-- *****-->
<!ELEMENT br
<!ELEMENT annote
<!ELEMENT name
<!ELEMENT place
<!ELEMENT ritual
<!ELEMENT shrine
<!ELEMENT other
rid
CDATA
#IMPLIED >
<!-- *****-->
<!-- Optional Elements -->
<!-- *****-->
<!ELEMENT preface
<!ELEMENT abstract
<!ELEMENT observation
(p*) >
(p*) >
(p*) >

```

4. 英日全文連携検索システムの概要と各種機能

XML化した各文献文書に対しては、利用者は所望する目的で、的確、高速、効率的に検索ができねばならない。また、我々は本英日全文連携検索システムを近年の有力な研究基盤となっているWeb上で設計し構築した。つまり、Web上で日本語と英語(または、両言語)を利用した文献内検索と文献間連携検索、閲覧、再利用を目標に設計し実現した。英語圏と日本語圏の研究者が、歴史学研究に有効な史料検索システムを利用し、研究を進めるには、日本語文書と英訳文書が連携して、検索可能にならなければならない。そのため、4種類の文献、つまり日本語文書、英訳文書、ローマ字読み文書、原文に近い底本のページイメージ画像ファイルに対して、文書構造が定義可能なXMLでタグ付けを行った。そのXMLタグ付けされた4種類の文書ファイルが連携して検索可能になる。

本英日全文連携検索システムは、XML化された文書を格納し、それに対して高速検索を実現するため、全文検索システムOpenText(カナダ、OpenText社)を採用した。近年のインターネットの急激な普及により、歴史学研究者も間便に利用できるWebからの検索を可能にするため、OpenTextに実装されている検索エンジンPAT70を使用し、Webブラウザからの融通性のある検索機能を実現した。以下に、このような視点で開発した本英日全文連携検索システムの概要と各種機能について述べる。

4-1. データベース管理システムOpenTextについて

本英日全文連携検索システムは、XML文書を格納するデータベース管理システムとして、研究機関等で実績のある全文検索システムOpenText(カナダ、OpenText社)を採用した。OpenTextは、検索アルゴリズムとしてパトリシア(Patricia: Practical Algorithm To Retrieval Information Code In Alphanumeric)ツリー方式を採用している。パトリシアツリー方式とは、検索キーの文字列がどの半無限部分文字列と一致するかを判定して検索する方式である。具体的には、文書の全ての文字から始まる半無限部分文字列(Semi-infinite strings)のインデックスを作成し、これを検索することで該当部分を探し出す。つまりインデックス作成時に与えられた文書に対してポインターを一文字ずつずらして置き、それぞれのポインターから文末までを半無限部分文字列とし、それをリスト化する。検索時は、検索パターンの文字列がどの半無限部分文字列と一致するかを調べる。実際の文字列はビットストリングで扱われ、0か1で分岐するツリー構造で表現されるため、単語の間をスペースで区切られる英語はもちろん、2バイトコードで膠着言語である日本語の文書に対しても高速に全文検索が可能である。

データベース管理システムOpenTextの特徴を下記に示す。

- (1) SGML/HTML/XML対応超高速全文検索システムである。
- (2) 「Oxford English Dictionaryの電子化プロジェクト」で開発され、最近国内の企業や政府機関、研究機関で様々な用途に利用されており、実績のある高速全文検索システムである。
- (3) 文書構造を利用した検索や文書の要素(フィールド)レベルでの検索が可能である。
- (4) 分散型のクライアントサーバ型での使用が可能である。

- (5) バイトストリームをインデックス付けするため、本英日全文連携検索システムが対象としているマルチバイトコードの日本語の検索にも対応している。

4-2. PAT70を利用した英日全文連携検索システムの概要

格納された文献の文字列や語彙に対する検索は、全文検索システムOpenTextが用意している検索機能(キーワード検索、範囲検索、複合検索)を利用して可能である^[5,35,36]。しかし、簡便なWebブラウザを利用し、XMLタグを有効に利用した、効率的で融通性のある検索を実行することが不可能である。これを実現するには、OpenTextに備わっている高速テキスト検索に有効なツールであり、コマンド・モードで実行可能な高速検索エンジンPAT70を利用する必要がある。

PAT70はテキスト検索に有効な高速検索エンジンであり、PAT70を使ってprefixes(プレフィックス)、words(単語)、phrases(フレーズ)を検索でき、長いフレーズに対しても単語と同じくらい高速に検索可能である。また、特定したテキストの近くに出現する文字列の検索(Proximity: 近接に基づく検索)、頻出するデータ(テキスト)の検索、頻出頻度(回数)などの取得も可能である。

さらに、PAT70はテキスト全体からの検索は当然として、検索エリアの限定が可能であり、テキストの一部を検索することも可能である。たとえば、ある文献中の所見、または要約だけに限定して検索を行うことができる。検索エリアの限定、つまり検索対象とするテキストのコンポーネントの選択には、各コンポーネントが何らかの始点と終点を示すタグが存在すればよい。この始点と終点を示すタグとして、例えば、XMLタグの<abstract>と</abstract>や<p page="66" para="1" note="7" cnt="3">と</p>などを利用する^[37,38]。

本英日全文連携検索システムは、インターネットのWeb上で、効率的で融通性のある検索機能を実現することを目標にした。そのため、高速検索エンジンPAT70をWebサーバのCGI(Common Gateway Interface)で動作させるアプリケーションプログラムを開発することが必要であった。このCGIプログラムは、Webクライアントからの動的な検索要求を受け付け、この要求を解析・処理し、高速検索エンジンPAT70にコマンド・モードで発信する。さらに、その検索結果を受信し、利用者が見易いようフォームを作成しブラウザに表示する必要がある。これらを実現するため、CGIプログラムとPAT70を別個のプロセスとして、動作させる必要があり、CGIプログラムはparentプロセスとして、PAT70はchildプロセスとして作動し、UNIXシステム固有のバッファであるパイプ(pipe)を利用して検索要求とその検索結果の相互やり取りが可能になるように設計・開発した。図2に、PIPEで接続されたアプリケーションプログラム(CGIプログラム)とPAT70の概念図を示す。

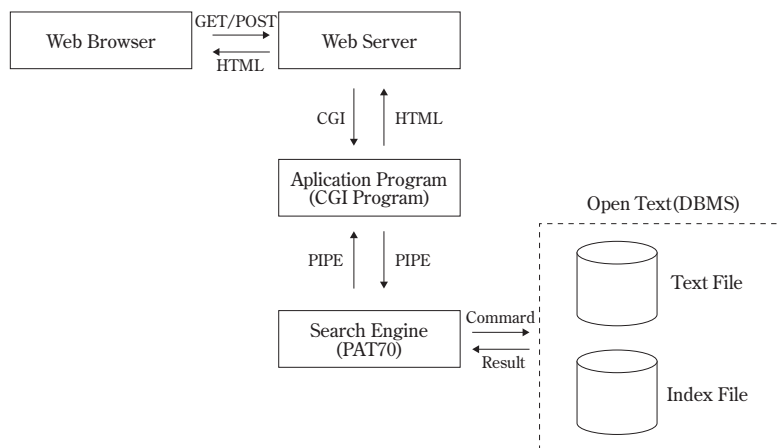


図2. PIPEで接続されたCGIプログラムとPAT50の概念図

4-3. 英日全文連携検索システムの各種検索機能

我々は、本英日全文連携検索システムを設計し、近年の有力な研究基盤となっているWeb上に開発・構築した。つまり、Web上で日本語と英語(または、両言語)を利用した文献内検索と文献間連携検索、閲覧、再利用を目標に実現した。英語圏と日本語圏の研究者が、歴史学研究に有効な史料検索システムを利用し、研究を進めるには、日本語文書と英訳文書が連携して、検索可能にならないといけない。そのため、各文献の日本語文書、英訳文書とノート文書に対して、論理構造(文書構造)と物理構造(各ファイル連携やページ構造)に基づいてXMLタグ付けを行った。本英日全文連携検索システムの各種機能について、最初に開発した江戸時代の儒学者で尾張藩士、河村秀根、益根父子が60年の月日を費やし刊行した「日本書紀」の注釈書である「書紀集解」と六国史に収録されている「続日本紀」を例に、各種機能を説明する。

(1) キーワード検索機能

「日本書紀」の巻第七(大足彦忍代別天皇 景行天皇)から具体的な検索例で説明する。巻第七は、景行天皇の代であり、有名な「日本武尊(やまとたけるのみこと)の東征と弟橘媛(おとたちばなひめ)の入水」の部分から具体的な例で説明する。文書の要約、英訳と本文(漢文)を下記に示す。

[文書の要約] 冬十月の壬子朔の癸丑(二日)に、日本武尊は出発された。戊午(七日)に、わざわざ寄り道をして、伊勢神宮を拝された。そうして叔母の倭姫命に暇乞いをされて、「今天皇のご命令をお受けして東方に赴き、全ての反逆者どもを誅伐しようとしております。それで、暇乞いに参りました」と申しあげられた。そこで倭姫命は、草薙剣を取って、日本武尊に授けて、「慎重にして決して油断なさるな」と言われた。

・・・・・・ 省略・・・・・・

それから相模に進まれ、上総に向かおうとして、海を望み、言霊の力を借りて言い立て、「これは小さな海だ。飛び越えてでも渡ることができよう」と言われた。ところが海の中ほどに至って、暴風がにわかに起り、御船は漂って進むことができなかった。時に、王に付き従って来た妾があった。弟橘媛(おとたちばなひめ)という。穂積氏忍山宿禰(ほづみのうじのおしやまのすくね)の娘である。王に謹んで、「今、風が起り、波浪が速く流れて、御船が沈もうとしております。これはきつと、海神の意によるのでありましょう。お許しを得て、卑しい私の身をもって、王のお命に代えて海に入りましょう」と申しあげた。言い終るや、すぐに波を押し分けて身を投じた。すると暴風はたちまちにして止み、御船は岸に着くことができた。そこで、時の人は、その海を名付けて馳水(はしりみず)といった^[17]。

[英訳本]

Winter, 10th month, 2nd day. Yamato-dake no Mikoto set out on his journey.

7th day. He turned aside from his way to worship at the shrine of Ise. Here he took leave of Yamato-hime no Mikoto, saying:—"By order of the Emperor, I am now proceeding on an expedition against the East to put to death the rebels, therefore I am taking leave of thee." Hereupon Yamato-hime no Mikoto took the sword Kusa-nagi and gave it to Yamato-dake no Mikoto, saying:—"Be cautious, and yet not remiss."

This year Yamato-dake no Mikoto first reached Suruga. The brigands of this place made a show of obedience, and said, deceivingly:—"On this moor there are large deer in very great plenty. Their breath is like the morning mist, their legs are like a dense wood. Do thou go and hunt them." Yamato-dake no Mikoto believed these words, and, going into the middle of the moor, sought for game. The brigands, desiring to kill the Prince, set fire to the moor. But the Prince, seeing that he had been deceived, produced fire by means of a fire-drill, and, kindling a counter-fire, succeeded in making his escape.

One version says:—"The sword Mura-gumo, which the Prince wore, wielded itself, and mowed away the herbage near the Prince, thus enabling him to escape. Therefore that sword was called Kusa-nagi."

The Prince said:—"I was almost betrayed." So he burnt all that robber-band and exterminated them. Therefore that place was called Yaketsu.

Next he marched on to Sagami, whence he desired to proceed to Kadzusa. Looking over the sea, he spake with a loud voice, and said:—"This is but a little sea: one might even jump over it." But when he came to the middle of the sea a storm suddenly arose, and the Prince's ship was tossed about, so that he could not cross over. At this time there was a concubine in the Prince's suite, named Oto-tachibana-hime. She was the daughter of Oshiyama no Sukune of the Hodzumi House. She addressed the Prince, saying:—"This present uprising of the winds and rushing of the waves, so that the Prince's ship is like to sink, must be due to the wishes of the God of the Sea. I pray thee let me go into the sea, and so let the person of thy

mean handmaiden be given to redeem the life of the Prince's Augustness." Having finished speaking, she plunged into the billows. The storm forthwith ceased, and the ship was enabled to reach the shore. Therefore the people of that time called that sea Hashiri-midzu.^[18]

本文は漢文であり、巻第三以降は編年体で記述されている。それを以下に示す。

冬十月壬子朔癸丑。日本武尊発路之。

戊午枉道拜伊勢神宮。仍辞于倭姫命曰。今被天皇之命。而東征将誅諸叛者。故辞之。於是倭姫命。取草薙劔。授日本武尊曰。慎之莫忘也。

是歳日本武尊。初至駿河。其処賊。陽從之欺曰。是野也。麋鹿甚多。気如朝霧。足如茂林。臨而應狩。信其言。入野中而覓獸。賊有殺王之情。放火焼其野。王知被欺。則以燧出火之。向焼而得免。

一云。王所佩劔叢雲。自抽之。薙攘王之傍草。因是得免。故号其劔。曰草薙也。

王曰殆被欺。則悉焚其賊衆而滅之。故號其處曰焼津。

亦進相模。欲往上総。望海高言曰。是小海耳。可立跳渡。乃至于海中。暴風忽起。王船漂蕩。而不可渡。時有從王之妾。曰弟橘媛。穗積氏忍山宿禰之女也。啓王曰。今風起浪泌。王船欲没。是必海神心也。願賤妾之身。贖王之命而入海。言訖乃披瀾入之。暴風即止。船得著岸。故時人號其海。曰馳水也。^[14,15,16] (注):「日本武尊」を下線で示す。

具体的な検索例として、上記本文(漢文)の2、4、5行目に出現する「日本武尊」を検索文字列(キーワード)として指定し、検索した結果である。まず、JHTI (Japanese Historical Text Initiative) プロジェクトのホームページ(<http://pnc-ecai.oiu.ac.jp/jhti/>)のプルダウンメニューから検索する文献を選択し、本英日全文連携検索システムを実行する。図3に示す入力フォーム画面(Interactive Searching of Nihon Shoki)の Search words: ボックスに、検索文字列(キーワード)として「日本武尊」を入力し、検索した例を示す。

なお、各文献で研究者に有効な検索方法の採用が必要である。例えば、「続日本紀」は、編年体の史書の特徴として、巻番号の後に年・季・月・日が記述されている。そのため検索機能として文字列(キーワード)のみでなく年月日や指定期間内の検索も可能にした。年を西暦710から712年で指定し、検索文字列(キーワード)として「平城」を入力した例を図4に示す。

図5は検索結果表示画面であり、検索文字列「日本武尊」でヒットしたパラグラフ一覧が表示される。画面の上部から、検索文字列(日本武尊)とヒットしたパラグラフ数(:19)が表示され、それ以降に10個単位で該当するパラグラフ一覧が巻番号、ページ番号とパラグラフ番号と共に表示される。また、指定したキーワードは、見やすくするため赤色太字で表示される。詳細表示(英日対応文書)を希望するとき、該当パラグラフのDetailボックスをクリックすると日本語と英語の対応パラグラフが表示される。また、図6に示すように、[Show Original Paragraph] ボックスをクリックすると表示画面の右側に、「書紀集解」の画像ファイルから切り出された該当文字列が画像で表示される。さらに、画面上部の[Show page image] をクリックすると、「書紀集解」の該当ページ画像が表示される。

XMLを利用した日本古典史料の英日全文連携検索システムの構築

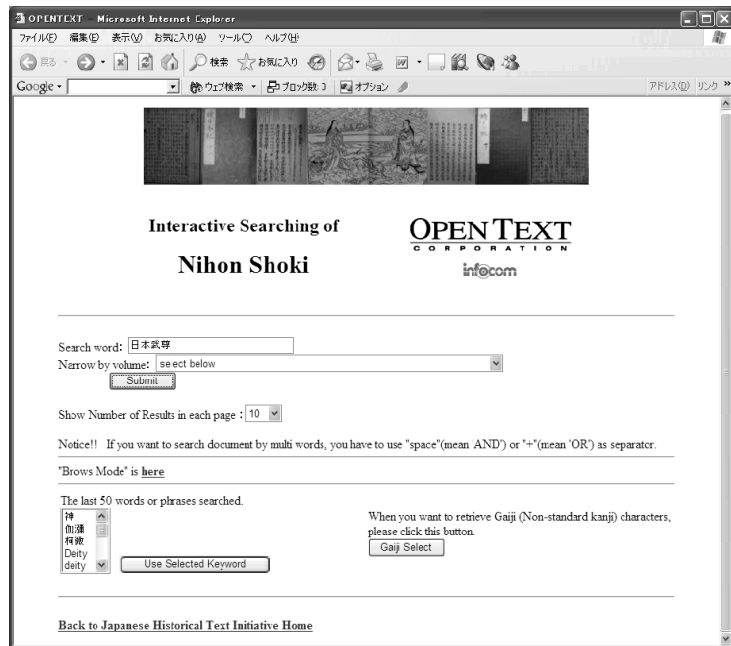


図3. 検索画面で検索文字列（キーワード）「日本武尊」を指定した例

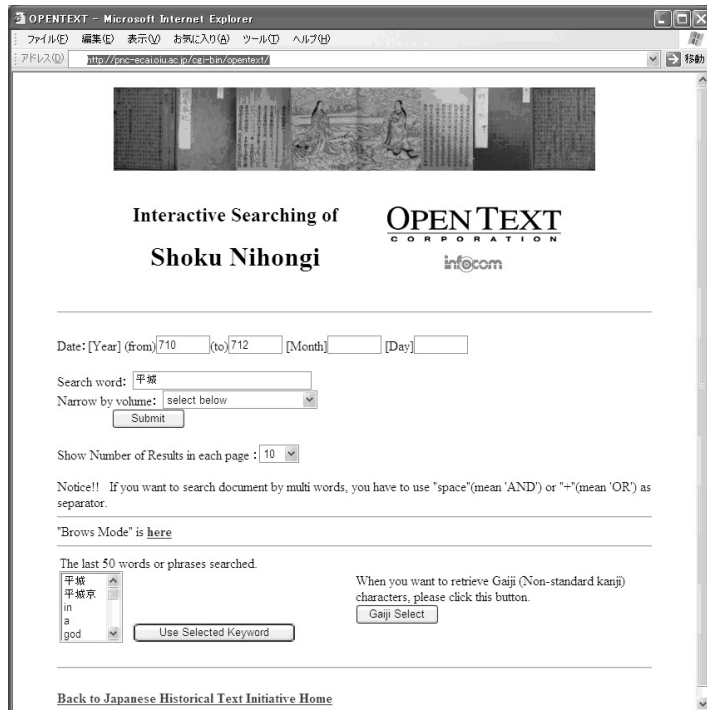


図4. 「続日本紀」の検索画面で指定期間(西暦710-712年)と「平城」を指定した例

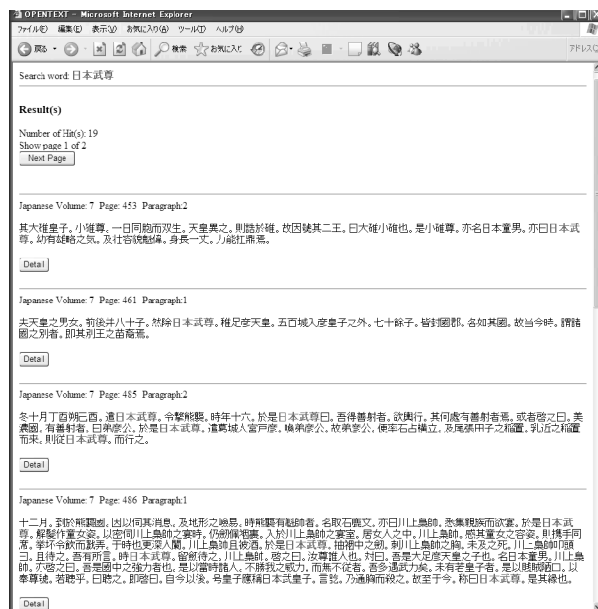


図5. 英日全文連携検索システムの検索結果画面

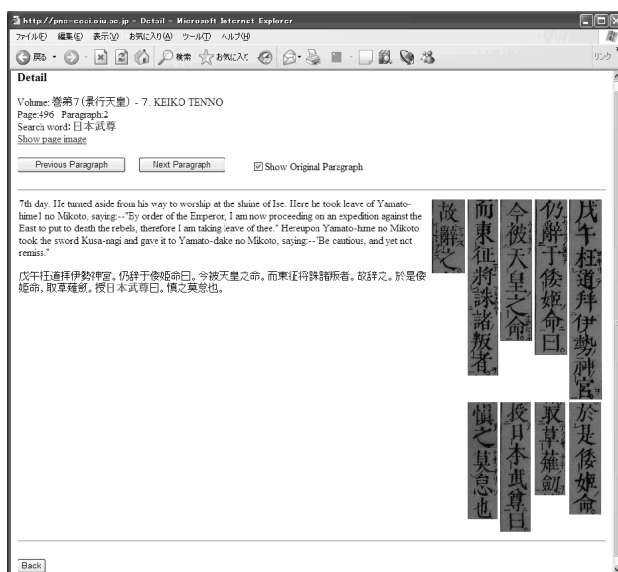


図6. 英日対応パラグラフの表示例

(2) 閲覧(ブラウジング)機能

閲覧(ブラウジング)機能は、言語(英語、日本語、両言語対応)を選択し、文献の巻番号を指定して先頭から、またページ指定やパラグラフ番号指定で、連続して閲覧することが可能である。さらに、閲覧するパラグラフ数を指定可能である(5,10,20,30 デフォルト値:10)。閲覧(ブラウジング)機能は、日本史・国文学を学習する初心者にとって、また、それらの教育用に有効であると思われる。

(3) 項目検索機能

項目検索機能は、神の名前、神社名、神社の場所名、儀式などから文献を効率的に検索することが可能である。そのため各項目のテーブルを作成し、タグ付けをプログラムで自動的に行うと共に、関連する画像(人物像や神社など)や絵画を格納している。

4-4. 史料の定量的解析の実験

電子化情報の特徴として、検索、加工、複写、転送が容易であり、また、統計的処理や計量言語学などの手法を利用した研究の展開の可能性を持っている。その一環として、史料の定量的解析の実験として、巻単位、文献単位や複数文献単位に語彙検索や文字頻度分析ツールなども実装している。

今回、日本語文書で漢字「神」の出現する頻度の調査と英訳本で「神」に該当する単語の調査を行った。漢字「神」は、神代の物語である巻1に、290個(27.8%)、巻2に、210個(20.1%)出現し、巻3の初代神武天皇から巻30の第40代持統天皇までに、544個(52.1%)出現し、その総数は、1,044個であった。

調査した英語の単語は、日本語「神」に該当する“god”, “divine”, “deity”である。当然、その単語の英大文字、英小文字、単数形と複数形も含んでいる。これら単語の各文献毎の出現頻度を表1に示す。その出現した単語総数が804個で、「神」の出現した1,044個と個数が異なるのは、「書紀集解」が漢文で記述されており、天照大神(47個):Ama-terasu no Oho kami、大己貴神(6個):Oho-na-muchi no Kami、甕速日神(4個):Haya-hi no Kamiなど人名に「神」の字が使用されていることや地名(神部:the Kambe of Kibi)等にも使用されているためである。

研究社、新英和大辞典 第6版によると^[39]、

[god]

1. (各種の信仰に基づく)神、神霊(deity, divinity); 男神(←→goddess)
2. 神像、偶像(idol); 神に祭られた人; 崇拝の対象、神
3. [G-] a 《キリスト教》神、創造の神、造物主、万有の神、天主
b 《クリスチャンサイエンス》無限の神霊、神
c 《ユダヤ教》=YHWH、 d 《イスラム教》=Allah
4. [G-] 「感嘆・のろい・祈願などの成句に用いるが、不敬を避けるためにしばしば Godを省略したり他の語に代えたりすることがある]

[deity]

1. [しばしばD-] 神位、神格、神性
2. 神
3. [the D-] (特に、自然宗教用語として)宇宙創造の神、天帝(the Supreme Begin)
4. この上なく善[強力、偉大]であるもの、神としてたたえられる[崇拝される]人
[もの]；(ある分野で)強力な力[影響力]をもつ人[もの]

[divine]

形容詞

1. 神の、神に関する、神による(⇒holy SYN)；神性の
2. 神授の、天与の、天来の
3. a 神にささげた；神をたたえる；神聖な(holy)；宗教的な(religious)
b 《廃》神学(上)の、神学に関する
4. 神のような(godlike)；神々しい、この世のものならぬ；超人的にすぐれた、非凡な
5. 《口語》たまらなくいい、すばらしい、すてきな ★主に女性の用いる強意語

名詞

1. 神学者；聖職者、牧師
2. [しばしばthe D-] 神；(人間にみられる)神性、神性を有するものと訳されている。

「日本書紀」では、大文字・小文字、単数形・複数形を含めて、godで477個、deityで237個、divineで90個の英訳単語が使用されている。

god, deity, divineが使用されている例は以下の通りである。

大神	: the Great God	神宮	: the Shrine of the God
河神	: the River-God	國神	: a God of the country
陽神	: the male Deity	陰神	: the female Deity
天神	: a Heavenly Deity	悪神	: a malignant Deity
神意	: the Divine behest	神劔	: a divine sword

英語のGodは日本語の「神」とはその宗教観からも基本的に異なると思われる。ここで出てくる「神」は、Godのような全知全能の神、創造の神、万有の神のみでなく、deityの3.のように、もっと神性な、または、自然宗教用語としての宇宙創造の神、神としてたたえられる(崇拝される)人・もの、ある分野で強力な力影響力をもつ人・もの、とたとえられ訳されるものだと思われる。

表1. 各文献に出現する「神」関係の英訳単語の出現数

巻 名	god	gods	God	Gods	GODS	goddess	Goddess	小 計
巻第1(神代上)	2	3	28	39	1	0	42	115
巻第2(神代下)	0	1	47	26	1	0	2	77
巻第3～巻第30	6	30	97	135	0	8	9	285
合 計	8	34	172	200	2	8	53	477

巻 名	divine	Divine	deity	deities	Deity	Deities	小 計	合計
巻第1(神代上)	8	4	8	8	41	32	101	216
巻第2(神代下)	5	0	0	0	29	14	48	125
巻第3～巻第30	68	5	16	6	48	35	178	463
合 計	81	9	24	14	118	81	327	804

4-5. 外字処理について

今回はXML化と同時に文字コードのUTF-8化を行った。これにより、「日本書紀」で、文献内に出現する外字総数：969字数、外字種類数：305字が、それぞれ外字総数：131字数、外字種類数75字に、また、「続日本紀」で、外字総数：145字数、外字種類数：54字が、それぞれ外字総数：11字数、外字種類数7字に大きく減少した。表2に、各文献に出現した外字総数と外字種類数を示す。もし、利用者がUTF-8コードに定義されていない外字を含んだ文字列の検索が必要ならば、図3に示した検索画面の下部分の[Gaiji Select] ボックスをクリックすると各文献に出現するすべての外字が別ウィンドウに表示される^[40]。この画面から該当する外字を視覚的に指定可能である。外字に関しては、%ufxxx;形式で定義することで検索可能になり、また外字の表示はgif画像ファイルとして貼り付ける。「日本書紀」と「続日本紀」に出現する外字一覧のウィンドウを図7と図8に示す。このことにより利用者から本英日全文連携検索システムの使用が簡便になることを期待している。

表2. 各文献に出現する外字総数とその種類

	日本書紀	古事記	延喜式	続日本紀	神皇正統記	愚管抄	出雲風土記	総 数
外字総数	969	339	1357(1328)	145(97)	71	96	147	1,622
ユニコード内	838	194	1243(1214)	134(90)	71	96	135	1,334
作成外字	131	145	114	11(7)	0	0	12	402
内大漢和内	83	139	28	7(3)	0	0	11	261
内大漢和外	48	6	86	4	0	0	1	145
外字種類	305	83	109	54	24	12	57	644
ユニコード内	230	54	94	47	24	12	49	510
作成外字	75	29	15	7	0	0	8	134
内大漢和内	44	23	10	5	0	0	7	89
内大漢和外	31	6	5	2	0	0	1	45

(注) 括弧内は、日本語文献のみ



図7. 「日本書紀」に出現する外字一覧表

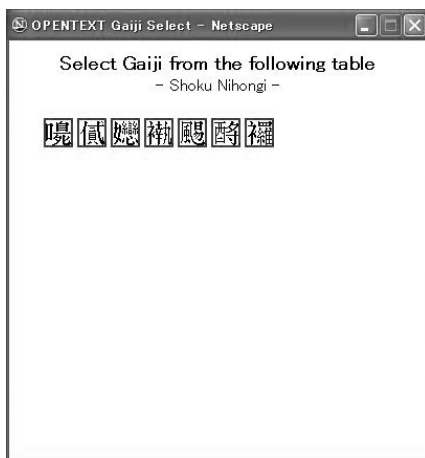


図8. 「続日本紀」に出現する外字一覧表

5. 今後の課題

本稿では、日本古典史料を題材に、インターネット上のWebを利用した英日全文連携検索システムを実現し、複数文献ファイル、つまり日本語ファイル、英訳ファイルと画像ファイルの連携と各種検索機能について述べた。本英日全文連携検索システムはインターネット環境下で、複数文献のテキスト連携表示機能、ページ画像連携表示機能、外字の混在した文字列の検索機能と外字表示機能／転送機能に対して、有効に作用し、URL: pnc-ecai.oiu.ac.jp/jhti/とURL: sunsite.berkeley.edu/JHTI/で運用されている。

今後、表3に示す25文献の格納を計画している。我々は、これまでに英日全文連携検索システムで開発・構築した文献は、以下の12文献である。

- | | |
|---|---------------------------|
| (1) Kojiki (古事記) | (2) Nihon Shoki (日本書紀) |
| (3) Shoku Nihongi (続日本紀) | (4) Izumo Fudoki (出雲風土記) |
| (5) Engi Shiki (延喜式) | (6) Okagami (大鏡) |
| (7) Gukansho (愚管抄) | (8) Jinno Shotoki (神皇正統記) |
| (9) Taiheiki (太平記) | |
| (10) Meiji igo Shukyo kankei Horei (明治以降神社関係法令史料) | |
| (11) Kokutai no Hongi (国体の本義) | (12) Dokushi Yoron (読史余論) |

さらに、神社名と神社の場所(位置情報)などを利用したGIS(Geographic Information System)との結合を図りたい。

私たちが開発した検索システムは、日米の研究者や学生が使うための基本的なTOOL作りであり、今後、他の文献の翻訳・原本を入力し、解釈書を格納し、利用できるようにしていくことを目標にしている。そのためにも、古典文献に対するコンテンツの充実、翻訳、原本の提供やシステムの改良に向けて、国際的なコラボレーションを一層推進したい。

本英日全文連携検索システムが日本の史料の新たな解釈・解析など歴史学研究を促進し、研究支援へのコンピュータの有効性を示し、新しい視点を与え、新しい研究課題と研究方法を生み出す契機になっていくことを期待したい。

最後に、本検索システム構築の機会を与えてくれたカリフォルニア大学バークレー校のECAIコーディネータLewis Lancaster教授、歴史史料に対するご教示やご討論を頂いた東アジア図書館(East Asian Library)図書館長Peter Zhou氏、ライブラリアン石松久幸氏、文献の英訳文書と日本語文書の校正・編集をやっていただいた新谷廣一氏ほか関係各位に謝意を表する。

なお、本研究は科学研究費基盤研究(B)(2)「Webを利用した歴史史料英日全文連携検索システムの設計と開発に関する研究」(平成15～17年度、研究代表者 桶谷猪久夫)の下で行った。

表3. デジタル化対象文献 (*: 検索可能文献、**: 現在構築中文献)

- Text 1: Kojiki (古事記)*
- Text 2: Nihon Shoki (日本書紀)*
- Text 3: Shoku Nihongi (続日本紀)*
- Text 4: Izumo Fudoki (出雲風土記)*
- Text 5: Kogoshui (古語拾遺)**
- Text 6: Engi Shiki (延喜式)*
- Text 7: Eiga Monogatari (栄華物語)**
- Text 8: Okagami (大鏡)*
- Text 9: Azuma Kagami (吾妻鏡)
- Text 10: Gukansho (愚管抄)*
- Text 11: Jinno Shotoki (神皇正統記)*
- Text 12: Taiheiki (太平記)*

- Text 13: Daijingu Jin'iki (大神宮神威記)
Text 14: Dokushi Yoron (読史余論)*
Text 15: Meiji igo Shukyo kankei Horei (明治以降神社関係法令史料)*
Text 16: Kokutai no Hongi (国体の本義)*
Text 17: Tenri-kyo (天理教)
Text 18: Kurozumi-kyo (黒住教)
Text 19: Konko-kyo (金光教)
Text 20: Omoto-kyo (大本教)
Text 21: Itto-en (一燈園)**
Text 22: Tensho Kotai Jingu-kyo (天照皇太神宮教)
Text 23: Risscho Kosei-kai (立正佼成会)
Text 24: Tsubaki Ookami Yashiro (椿大神社)
Text 25: Manyousyu (万葉集)**

【参考文献】

- [1] 桶谷猪久夫、Delmer Brown、山尾正之、大久保祐子、『日本古典史料の英日全文連携検索システムの構築 - 日米共同研究について -』、大阪国際大学紀要「大阪国際論叢」第18巻第1号、pp.1-17, 2004.9.30
- [2] Ikuo Oketani, Delmer Brown, Shigeo Kikuchi, Chizuko Saito “A Collaborative Research between Japan and the US on Designing and Constructing the Full Text Coordinated Retrieval System of Japanese Historical Documents using the Internet” The 2003 PNC (Pacific Neighborhood Consortium) Annual Conference and Joint Meetings, Bangkok, Thailand, Proceedings (CD-ROM), pp.1-18, November 7-9, 2003
- [3] 桶谷猪久夫、Delmer Brown、藤本雅彦、大久保祐子、『Webを利用した歴史史料英日全文連携検索システムの開発 - 日米共同研究について -』、情報処理学会研究報告、Vol.2003, No.107, 2003.10.24
- [4] 桶谷猪久夫、Delmer Brown、才藤千津子、新谷廣一、『Webを利用した歴史史料の英日全文連携検索システムの開発』、大阪国際大学紀要「大阪国際論叢」第16巻、pp.1-20, 2002.11.30
- [5] 桶谷猪久夫、新谷廣一、『SGMLを利用した琉球王国評定所文書と琉球家譜の全文連携検索システムの設計と実現』、大阪国際女子大学紀要、第27-2, pp.1 - 18, 2002.3.31
- [6] 桶谷猪久夫、才藤千津子、Delmer Brown、『簡易型タグを利用した歴史史料の英日全文連携検索システムの設計と開発 - 日本書紀、古事記における事例 -』、情報処理学会人文科学とコンピュータシンポジウム論文集、pp.65 - 72, 2001.12.14 - 15
- [7] Ikuo Oketani, Chizuko Saito, Delmer Brown “The Construction and the Future Development of the Full Text Coordinated Retrieval System of Historical Documents using the Internet” PNC (Pacific Neighborhood Consortium) Interim Conference, Guadalajara (Mexico), pp. 11-11, December 2, 2001
- [8] Ikuo Oketani, Chizuko Saito, Delmer Brown, “Shinto Project: A Design and Construction of the Full Text Retrieval System using Simple-tagged Nihon-shoki Texts (the Imperial Chronicle of Japan)”, Ninth ECAI (Electronic Cultural Atlas Initiative) Conference, Sydney, 12-16 June, 2001, p. 12 (abstract)
- [9] 尾崎暢映編、『訂正古訓古事記 / [本居宣長訓]、上』、新典社, 1971

- [10] 尾崎暢弼編、『訂正古訓古事記 / [本居宣長訓]、中』、新典社、1978
- [11] 尾崎暢弼編、『訂正古訓古事記 / [本居宣長訓]、下』、新典社、1978
- [12] Norinaga Motoori “Kokun Kojiki teisei 4-koku” Nagata Bunshodo, 1874, stored in UCB East Asian Library
- [13] Donald L. Philippi “Kojiki /Translated with an Introduction and Notes by Donald L. Philippi” University of Tokyo Press, 1968, pp. 1-655
- [14] 河村秀根・益根、『書紀集解(二)』、臨川書店、1969、pp.1 - 656, stored in UCB East Asian Library
- [15] 河村秀根・益根、『書紀集解(三)』、臨川書店、1969、pp.657 - 1256, stored in UCB East Asian Library
- [16] 河村秀根・益根、『書紀集解(四)』、臨川書店、1969、pp.1257 - 1916, stored in UCB East Asian Library
- [17] 小島憲之他校注・訳、『日本書紀1』、新編日本古典文学全集2、小学館、2002、pp.340-397(巻第七、景行天皇)
- [18] W. G. Aston “NIHONGI: Chronicles of Japan from the Earliest times to A.D. 697” Printed by the Japan Society, 1896
- [19] 補増 六国史巻三、『続日本紀上巻』、朝日新聞社、1940、pp.1 - 455
- [20] 補増 六国史巻四、『続日本紀下巻』、朝日新聞社、1940、pp.1 - 516
- [21] Translated and annotated by J.B. Snellen “Shoku Nihongi : Chronicles of Japan, continued, from 697-791 A.D.” (Vol.1-3) In The Transactions of the Asiatic Society of Japan. Second Series. Vol.11. 1934. Asiatic Society of Japan; 151-239
- [22] Translated and annotated by J.B. Snellen “Shoku Nihongi : Chronicles of Japan, continued, from 697-791 A.D.” (Vol.4-6) In The Transactions of the Asiatic Society of Japan. Second Series. Vol.14. 1937. Asiatic Society of Japan; 209-278
- [23] 岡見正雄、赤松俊秀校注、『愚管抄』、日本古典文学大系86、岩波書店、1981、pp.3 - 547
- [24] Delmer M. Brown and Ichiro Ishida : The Future and the Past : a translation and study of the Gukansho, an interpretative history of Japan written in 1219, University of California Press, 1979. p.1 - 479
- [25] 秋本吉郎校注、『風土記(出雲風土記)』、日本古典文学大系2、岩波書店、1980、pp.93 - 256
- [26] Translated with an introduction by Michiko Yamaguchi Aoki : Izumo Fudoki, Sophia University, Tokyo, 1971, pp.1 - 157
- [27] 岩佐正他校注、『神皇正統記』、日本古典文学大系87、岩波書店、1981、pp.5 - 542
- [28] Translated by H. Paul Varley : A Chronicle of Gods and Sovereigns : Jinno Shotoki of Kitabatake Chikafusa, Columbia University Press, 1980, pp.1 - 300
- [29] Electronic Cultural Atlas Initiative : <http://ecai.org/>
- [30] Japanese Historical Text Initiative : <http://sunsite.berkeley.edu/JTHI/>
- [31] <http://pnc-ecai.oiu.ac.jp/>, 「Japan PNC, ECAI」 ホームページ
- [32] Shishir Gundavaram “CGI Programming on the World Wide Web” O'Reilly & Associates, Inc., 1996.11
- [33] Larry Wall and Randal L.Schwartz “Programming Perl” O'Reilly & Associates, Inc., 1992.3
- [34] Randal L.Schwartz “Learning Perl” O'Reilly & Associates, Inc., 1994.4
- [35] Infocom Corporation, OpenText7 データベース管理者ガイド
- [36] Infocom Corporation, OpenText7 データベース管理リファレンス・ガイド
- [37] Infocom Corporation, OpenText7 PAT Tutorial、マニュアル
- [38] Infocom Corporation, OpenText7 PATリファレンス、マニュアル

[39] 竹林滋編、「研究社、新英和大辞典 第6版」、(株)研究社、2002.3, p.1048, p.647, p.712

[40] 国際符号化文字集合(UCS)－第1部 体系及び基本多言語面。

(注) 漢字フォント(20,902セット)、島根県立大学メディアセンター勝村哲也教授提供