

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 山尾正之*³ 大久保祐子*⁴

**A Collaborative Research between Japan and the U.S. on
Constructing the Full Text Coordinated Retrieval System
of Japanese Historical Resources**

Ikuo Oketani*¹ Delmer Brown*² Masayuki Yamao*³ Yuko Okubo*⁴

Abstract

This project develops interactive retrieval system for English and Japanese texts, with a focus on language structures and historically descriptive methods. Furthermore, it promotes international collaboration by making these interactive retrieval system available to the public through the Internet, and by giving assistance to historical research.

The literature we are dealing with is as follows: 1) Japanese ancient chronologies such as *Kojiki*, *Nihon Shoki*, and *Shoku Nihongi*, and *Jinno Shotoki*, 2) *Engishiki*, a collection of laws and regulations on shrines, 3) *Izumo Fudoki*, a local document of a certain area, and 4) *Gukansho*, a seven-volume representative history text of the Middle Ages, and so on. Moreover, we intend to digitize the twenty-five volumes of Japanese classic texts.

The goals of the development of this system and this research are two-fold: 1) to contribute to the research of Japanese history and literature conducted by English-speaking researchers and students, and 2) to pursue the synergistic effect of research by promoting collaboration between English-speaking researchers and Japanese researchers. Such collaboration can be significant for researchers on Japanese history in both English-speaking countries and Japan. The areas of the collaborative research range from Japanese ancient history and research on the history and foundation of the ancient Japanese state to ethnological research. We

* 1 : おけたに いくお : Professor, 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 : やまお まさゆき : Assistant Professor, Osaka International College

* 4 : 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

<2004. 6. 30受理>

expect our project to pioneer the retrieval of and access to historical information from Japan.

With these goals, we constructed 1) a search program within texts using English and/or Japanese, 2) a search program connecting different texts, and 3) a browsing function. We also created 4) an English translation-assisting program on the Internet which makes the best use of the characteristics of *Shoku Nihongi* as a chronicle, the years, seasons, months, and dates are automatically displayed on the screen. We, furthermore, plan to connect the current system to geographic and historical information that is important in dealing with the information on literature and ancient historical materials.

Key Words

Full Text Retrieval System, Japanese Historical Resources, Internet,
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

近年のインターネットの急速な普及によって、それを利用した電子情報の公開が一般的になってきている。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。つまり、研究者（ユーザ）が必要とする情報を高速・的確・効率的に見つけ出すための情報検索システムをWWW（World Wide Web, 以下、Webという）上で提供することで、歴史学研究を援用し創造的な活動の活性化を推進することが期待される。

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。

本稿では、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（簡易型タグ付け）を作成した。その簡易型タグ付けをされた文書に対して、検索機

能を設計し実現した。まず、対象とした日本古典史料の概要、本英日全文連携検索システムの目的と概要、各種検索機能とその問題点について述べる。また、続日本紀の英訳（翻訳）支援システムの特徴について述べる。さらに、カリフォルニア大学パークレー校と国文学研究資料館との共同研究について述べる。

2. 日本古典史料の英日全文連携検索システムの目的と概要

2-1. 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」、特定の地方誌的文書である「出雲国風土記」、歌集「万葉集」、天台座主の慈円の子で承久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」^[22]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^[8 9 10]

「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。^[8 9 10]

(2) 日本書紀^[13 14 15]

「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(すさのをのみこと)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。^[13 14 15]

(3) 出雲国風土記^[24]

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け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。

年代および内容が官命により規定されており、それは次の5項目であり、

- (1) 畿内七道の国名、郡名、郷名には好字(漢字2字の嘉き字)を付ける。
- (2) 郡内に産する産物(鉱物・植物・動物など)について色目(物産品目)を筆録する。
- (3) 土地の肥沃状態。
- (4) 山川原野(自然地)の名称の由来。
- (5) 古老の相伝する旧聞異事(伝承)。

について筆録文書として言上報告させた。

「出雲国風土記」は、巻首の総記と各郡記と巻末記の3部を共に有する唯一の巻本である。内容は、第1部:国の総括的な地勢、編集方針、国名由来、神社や郡・郷・里などの統計など、第2部:出雲九郡に関する統計、郡名由来、伝承、寺院、神社、山・川・池・島、産物など、第3部:主要道路の順路と路程、国境、各駅間の距離、軍制などが記述されており、巻末に、

天平五年二月卅日 勘造 秋鹿郡人 神宅臣全太理(みやけのおみかなたり)

國造帶意宇郡大領外正六位上勳十二等 出雲臣廣嶋(いずもおみひろしま)

と編述年月日、勘造者、同責任者(編集責任者)の署名がある。

また、「出雲国風土記」の最初、「意宇郡(おうぐん)」の冒頭に、有名な出雲国の成り立ち、いわゆる、「国引き神話」がある。八束水臣津野命(やつかみずおみつぬのみこと)が国をつくるのに、出雲国は小さすぎるので各地から引いてきて継ぎ合わせた、と記されている。継ぎ足されたところは島根半島の部分である。

東端の「三穂(みほ)の埼」は北陸の能登から、西端の「支豆支の御埼(きづきのみさき)」は朝鮮半島の新羅から、その間の「闍見(くらみ)の国」と「狭田(さだ)の国」はそれぞれ「北門(きたど)の良波(よなみ)国」、「北門の佐伎(さき)国」から引いてきたと記述されている。

(4) 神皇正統記^[26]

「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、神話的の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、劍)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）。

2-2. 日本古典史料の英日全文連携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と文献間連携検索、閲覧、再利用を目標にしている。これら対象とする文献の一部は、既に研究者によりフルテキスト・画像ファイルとして入力済みで研究整備がされている。

3. 日本古典史料の英日全文連携検索システムの設計と各種機能の概要

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、統計的処理やデータベース処理が可能であることなどがあげられる。つまり、網羅的な研究（全ての単語や全ての文字単位まで）が追求できることである。

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、原文に近い底本の画像ファイルに対して、文書構造が定義可能な簡易型のタグ付けを行った。そのタグ付けされた4種類の文書ファイルが連携して検索可能になる。

簡易タグ付けの基本は、(1)既にデジタル情報として入力された英訳版を元にタグ付けする、(2)検索と表示の単位は、パラグラフ単位とする。そのため、漢文や英文は複数文になる場合があるので、それら複数文を1つのタグで囲む。タグ付けは、結果としてタグ個数が少なく、また簡易になり作業日程を短縮できた、(3)検索効率を向上させるため、注釈（Note）を別ファイルとして保存し、本文の位置情報を示すタグを付加した。

本英日全文連携検索システムは、現在Web上のCGI（Common Gateway Interface）^[31]機能を利用しインタプリタ言語Perl（Practical Extraction and Report Language）^[32,33]で各種検索機能を実現している。

以下に、簡易型タグ付けされた各文献（日本語、英訳文、注釈文）と簡易型タグの簡単な説明を示す。

* 日本文献のタグ例：日本書紀の注釈書である「書紀集解」^[13,14,15]

253#1#0#日本書紀 卷第三

253#2#0#神日本磐余彦天皇 神武天皇

254M#1#0#神日本磐余彦天皇。諱彦火火出見。彦波瀲武^{%u9e15;%u9dc0};草葺不合尊。第四子也。母曰玉依姬。海童之少女也。天皇生而明達。意確如也。年十五。立爲太子。長而娶日向國吾田邑吾平津媛。爲妃。生手研耳命。研耳命。

255T#1#0#及年四十五歲。謂諸兄及子等曰。昔我天神。高皇產靈尊。大日^{%u5b41};尊。拳此豐葦原瑞穗國。而授我天祖彦火瓊瓊杵尊。於是彦火瓊瓊杵尊。闢天闕披雲路。驅仙蹕以戻止。是時運屬鴻荒。時鍾草昧。故蒙以養正。治此西偏。皇祖皇考。乃神乃聖。積慶重暉。多歷年所。自天祖降跡。以逮于今。一百七十九萬二千四百七十餘歲。而遼^{%u9088};之地。猶未霑於王澤。遂使邑有君。村有長。各自分疆。用相凌轢。

* 英訳本のタグ例：W. G. Aston "NIHONGI"^[17]

253#1#0#NIHON SHOKI BOOK III.

253#2#0#THE EMPEROR KAMI-YAMATO IHARE-BIKO.1 (JIMMU TENNO.)

254M#1#1#THE Emperor Kami Yamato Ihare-biko's personal name was Hiko-hohodemi. He was the fourth child¹ of Hiko-nagisa-take-u-gaya-fuki-ahezu no Mikoto. His mother's name was Tama-yori-hime, daughter of the Sea-god. From his birth, this Emperor was of clear intelligence and resolute will. At the age of fifteen he was made heir to the throne. When he grew up, he married Ahira-tsu-hime, of the district of Ata in the province of Hiuga, and made her his consort. By her he had Tagishi-mimi no Mikoto and Kisu-mimi no Mikoto.

255T#1#0#When he reached the age of forty-five, he addressed his elder brothers and his children, saying: "Of old, our Heavenly deities Taka-mi-musubi no Mikoto, and Oho-hirume no Mikoto, pointing to this land of fair rice-ears of the fertile reed-plain, gave it to our Heavenly ancestor, Hiko-ho no ninigi no Mikoto. Thereupon Hiko-ho no ninigi no Mikoto, throwing open the barrier of Heaven and clearing a cloud-path, urged on his superhuman course until he came to rest. At this time the world was given over to widespread desolation.

* 英訳本の注釈文 (Note) のタグ例：W. G. Aston "NIHONGI"^[17]

254M#1#1# Emperor is as near an equivalent as possible of the Chinese 天皇. Both are foreign words. The Japanese interlinear gloss is Sumera Mikoto "supreme majesty," sumera having the same root as suberu, "to unite as a whole"; hence, "to have general control of" See Satow, "Rituals," T.A.S.J., VII., ii., p. 113. Yamato, see above, note to p. 13. Ihare is the name of a district of Yamato; Hiko means prince. Jimmu (divine valour) is a posthumous name.

(注) : タグの説明

記載形式 : 254M#1, ¥IMG:2#1

日本文献 : ページ番号, 複数ページタグ#パラグラフ番号, ¥IMG:分割画像の数, #ノート数#本文 : Mは2ページにまたがる。T: 3、Q:4、…

英訳本 : ページ番号#パラグラフ番号#ノート数#本文

注釈文 (Note) : ページ番号#パラグラフ番号#ノート番号#ノート本文

(注) : その他に、項目検索用タグや画像 (絵画など) リンク用タグが存在する。

(注) : %u9e15;, %u9dc0;, %u5b41;, %u9088; は外字、ただしUnicodeにあり。

3-1. 日米共同研究の推進

本英日全文連携検索システムは、Web上で日本語、英語 (または、両言語) と原典のページ単位の画像ファイルを利用した文献内検索と文献間連携検索、閲覧、再利用を可能にする。そのため、後述の25文献の日本古典史料 (漢文、日本語) と英訳本のデジタル情報が必須になる。このため、国文学研究資料館の科学研究費補助金 (基盤研究S) 「国際コラボレーションによる日本文学研究資料情報の組織化と発信」と「JHTI (カリフォルニア大学バークレー校) プロジェクト」との共同研究を推進するため協定書を締結している。これにより、JHTIプロジェクトが保有していない日本古典文学本文データベース

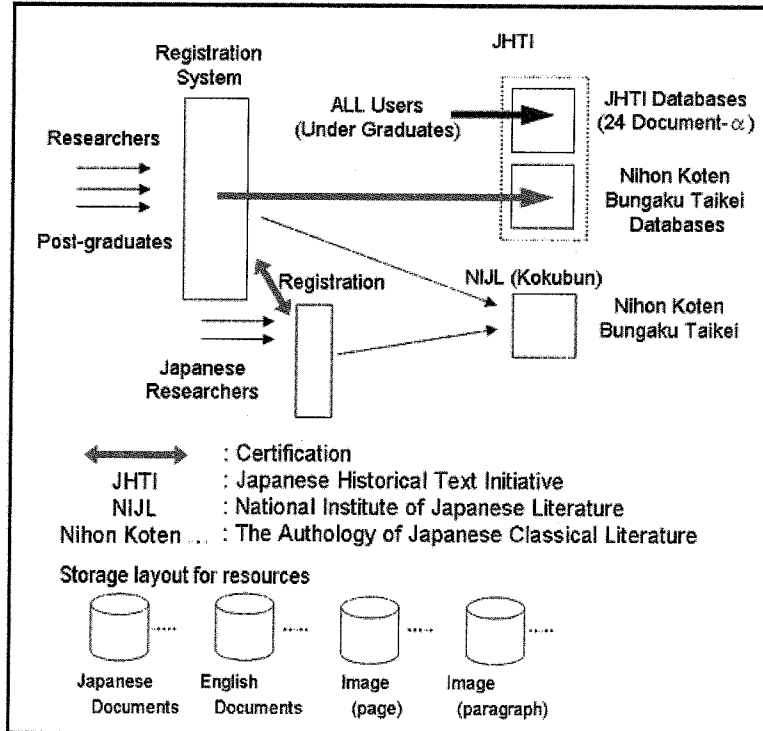


図1. JHTI認証システムの概念図と各文献の格納図

(日本古典文学大系)のデジタル情報を使用し研究が遂行可能である。当然、この部分の使用は国文学研究資料館プロジェクトとの協定書に許可された範囲内であり、また国文学研究資料館に登録された日本文学・史学研究者やそれらを専攻した大学院生に限定される。例えば、大学生の場合は卒業研究に限定され登録申請後指導教員の承認書が必要になる。現在、この協定書を前提に、「神皇正統記」、「出雲国風土記」や「大鏡」等の構築と開発を進めている。そのため国文学研究資料館の本文データベース検索システムの申請が承認されているかどうかの認証システムを必要とする。そのJHTI認証システムの概念図と各文献の格納図を図1に示す。当然、研究者・大学院生や英語圏の学生も含めて、他の文献はJHTIプロジェクトで構築した英日全文データベースが使用可能である。

3-2. 日本古典史料の英日全文連携検索システムの各種検索機能

本英日全文連携検索システムは、現在Web上のCGI (Common Gateway Interface)機能を利用し、適切な文書の部分をパターンマッチングして抽出するのに最適化されたインタープリタ言語Perl (Practical Extraction and Report Language)で各種検索機能を実現している。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、(1)古事記、(2)日本書紀、(3)続日本紀、(4)出雲風土記、(5)延喜式、(6)愚管抄、(7)神皇正統記、(8)明治以降神社関係法令史料、である。

以下に、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」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」の巻第三(神日本磐余彦天皇 神武天皇)から具体的な検索例で説明する。巻第三は、神武天皇の東征から書き始まり、文章の要約は、「神日本磐余彦天皇は諱(いみな)を彦火火出見(ひこほほでみ)といい、…(省略)…天孫が降臨されてから今日まで、百七十九万二千四百七十余年が過ぎた。しかしながら遼遠の地は、今なお王化の恩恵に浴していない。大きな村には君がおり、小さな村には首長がいて各々がそれぞれ境を設け、互いに抗争し鎬を削っている。…(省略)…その年の冬十月の丁巳朔辛酉(5日)に、天皇は自ら諸皇子・舟軍を率いて、東征の途に就かれた。速吸之門に着かれた時に、一人の獵師がいて、小舟に乗って近づいて来た。天皇はこれをお召しになり、そして、問うて、お前は誰かと仰せられた。…」^[16]

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。

自天祖降跡。以逮于今。一百七十九萬二千四百七十餘歲。而遼%u9088;之地。猶未霑於王澤。遂使邑有君。村有長。各自分疆。用相凌轢。抑又聞於塩土老翁曰。東有美地。青山四周。其中亦有乘天磐船而飛降者。余謂彼地。必當足以恢弘天業。光宅天下。蓋六合之中心乎。厥飛降者。謂是饒速日歟。何不就而都之乎。諸皇子對曰。理實灼然。我亦恆以爲念。宜早行之。是年也。太歲甲寅。

其年冬十月丁巳朔辛酉。天皇。親帥諸皇子。舟師東征。至速吸之門。時有一漁人。乘艇而

至。天皇招之。因問曰。汝誰也。……

(注)：「東征」を下線で示す。

具体的な検索例として、上記本文の6行目の「東征」を検索文字列(キーワード)で検索した結果である。まず、図2で示すカリフォルニア大学バークレー校のJHTI (Japanese Historical Text Initiative) プロジェクトのホームページ^[29]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)から「日本書紀」の検索対象巻番号の巻第3(神武天皇)－3. JIMMU TENNO)を選択する。次に、MODE:ボックスでRetrievalを選択し、Find word or phrase:ボックスに、神武天皇が東方に美しい国ある。四方を青山が囲んでいる。天下に君臨する地であり、我が国の中心の地であると言われ、諸皇子と舟軍を率いて東征の途に就かれたと書かれている箇所から、検索文字列(キーワード)として「東征」を入力し、Word(s)retrieval: Which version?ボックスでJapaneseを指定し検索した例を示す。もし、外字を含んだ検索文字列が必要ならば、図3の下部のWhen you want to retrieve Gaiji (Non-standard Kanji) characters:で、(a)部首の画数選択、(b)直接部首名の入力、(c)総画数の入力、(d)音読みの入力、を利用して候補外字を画面表示し視覚的に指定可能である。また、直接、大漢和(諸橋)コードやUnicode入力で外字を直接指定可能である。

図4は検索結果表示画面であり、検索文字列「東征」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、検索対象文献名(Nihon-shoki)、検索文字列(東征)、文献の巻数名(“巻第3(神武天皇)－3. JIMMU TENNO”)とマッチしたパラグラフ数(Found: 2 matches)と該当するパラグラフ一覧がページ数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More Details...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。パラグラフ数1の場合は、図5に示すように、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。画面上部の[This Page's Image]をクリックすると、Original Image of Documentの該当ページ画像が表示される。また、検索文字列の出現した文書の前後パラグラフへのナビゲーション機能や表示パラグラフ数の変更(3、または5パラグラフ単位)が可能である。さらに、英訳文書においては、注釈文(Note)が多く出現する。そのため、[Show Notes]をクリックすることにより、本文に続き注釈行(Note)が該当箇所に展開表示可能である。

(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。

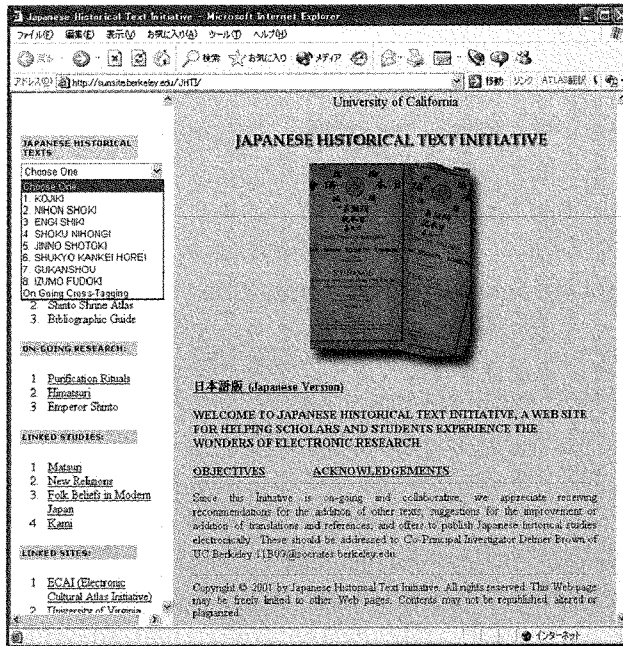


図 2. JHTIプロジェクトのホームページ

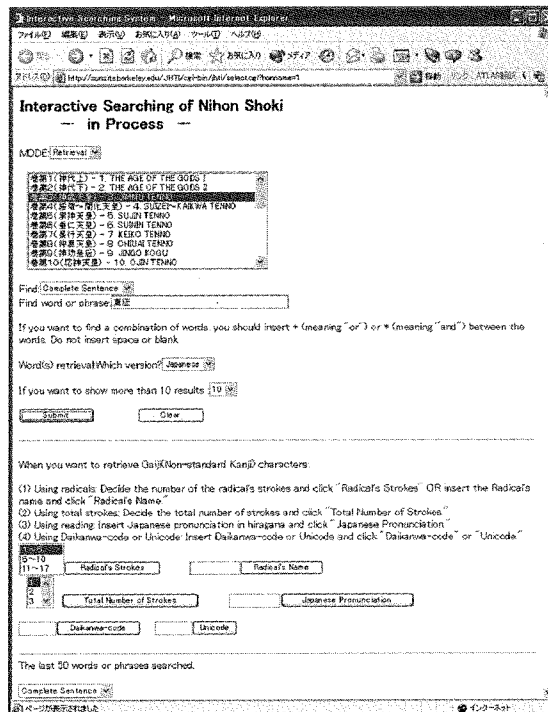


図 3. 検索画面で検索文字列 (キーワード)「東征」を指定した例

日本古典史料の英日全文連携検索システムの構築

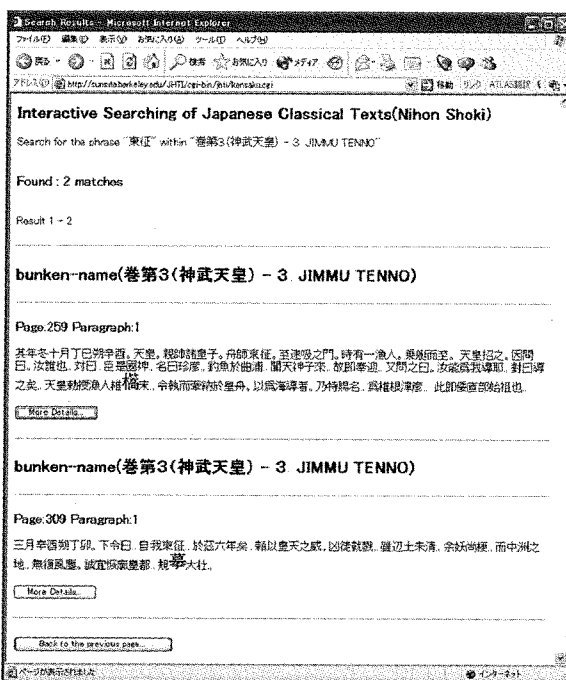


図 4. 日本古典史料英日全文連携検索システムの検索結果画面

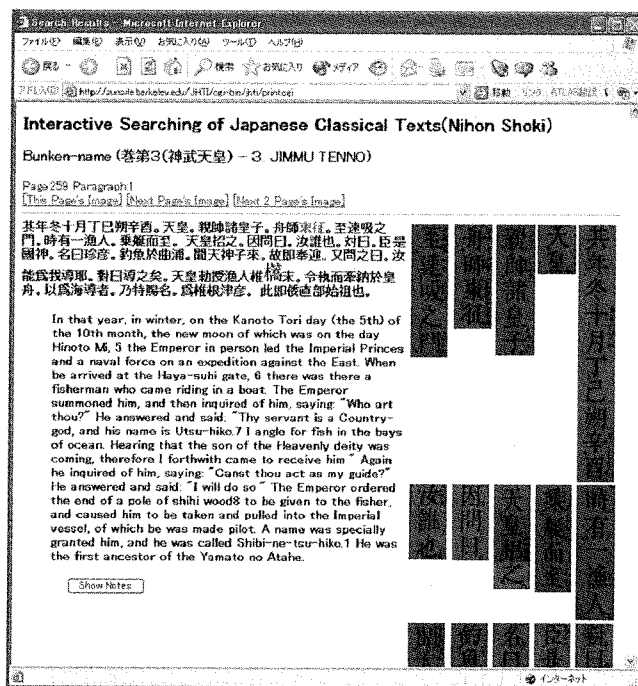


図 5. 英日対応パラグラフの表示例

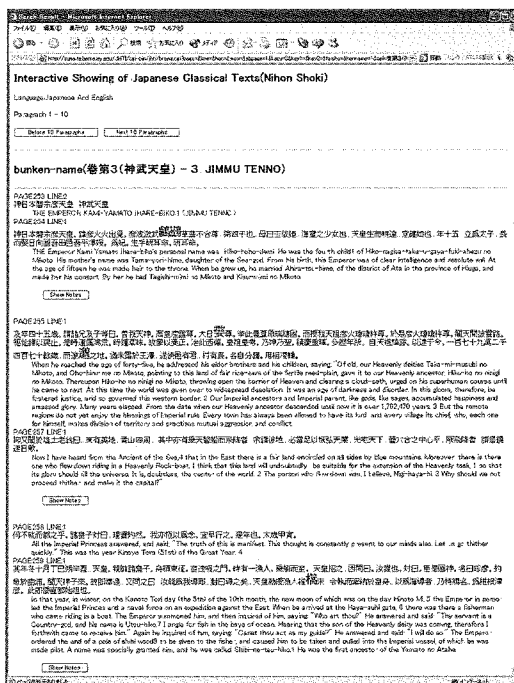


図 6. ブラウジング機能の表示例 (巻第 3 神武天皇)

(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を想定している。現在、プログラム言語Perlを使用し、Webからの利用者の要求(文字列やその論理結合質問)を解釈し、格納されたデータに対して、適切な文字列部分をパターンマッチングして抽出している。そのため、文字列検索と同様な機能しか有していないが、今後の拡張でXML(Extensible Markup Language)タグなどを付加したときに有効に作用すると思われる。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。

その他の機能として、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。

本英日全文連携検索システムは、研究・教育に試験的に公開され使用されている。^[29] 開発における問題点として、①単語の表記上の違いがあった。たとえば「古事記」の英訳本では、男神伊弉諾尊は、“IZANAGI”となり、「日本書紀」の英訳本では、“Izanagi”となっている。これらの問題は、検索用プログラムで対処(解決)した。②既にデジタル化された英訳本の注釈文(Note)が冊子体のイメージに忠実に入力されていた。この問題に対しては、検索プログラムのバッファリングで対処しても検索効率の低下をもたらすため、注釈文(Note)を別ファイルとして保存し、本文の位置情報を示すタグで対処した。

本文と注釈文 (Note) の分割管理と連携タグの追加により、検索速度が5倍以上にあがることが確認された。③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」^[18,19] 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。

3-3. 日本古典史料の英訳 (翻訳) 支援システムの構築

「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。^[18,19] 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、藤原朝臣繼縄 (ふじわらのあそんつぐただ) ・菅野朝臣真道 (すがのあそんまみち) 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。

本英日全文連携検索システムが対象にした「続日本紀」の英訳本 (J.B. Snellen)^[20,21] は、巻一から巻六までしか完成していない、そのためそれ以降の英訳の完成が強く求められている。それらの要望に答えて、Webベースの英訳 (翻訳) 支援システムを実現した。「続日本紀」の編年体の史書の特徴を生かし、年・季・月・日を自動的に挿入している。また、既に検索可能になっている文献の古事記、日本書紀、神皇正統記などや既に翻訳済みの第六巻までの続日本紀の参照を可能にしている。

現在、國學院大学で文部科学省21世紀COEプログラム「神道と日本文化の国学的研究発信の拠点形成」の一環として、神道・日本文化研究国際シンポジウム「<神道>はどう翻訳されているか」が開催されている。歴史書、神道思想が展開されている書物、その他、神道に関わる書籍や文書を外国語に翻訳する際には、神道の基本用語をどう理解し、どのようなコンテキストに位置付けるか、原文をそのまま直訳できるのか、原文を解釈するだけで終わるのか、などが討論されている。また、國學院大学日本文化研究所が編集した「神道事典」の英訳作業が進行している。これらが早急に利用できることがこの翻訳作業にとって重要な意味を持つと思われる。

その英訳 (翻訳) 支援システムの入力画面を図7に示す。当然、翻訳済みの英日対応文書閲覧画面・印刷機能や校正もWebから可能になっている。現在、この日本古典史料の英訳 (翻訳) 支援システムを利用し第七巻以降の翻訳が遂行されている。

4. 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、各種検索機能と英訳 (翻訳) 支援システムについて述べた。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、

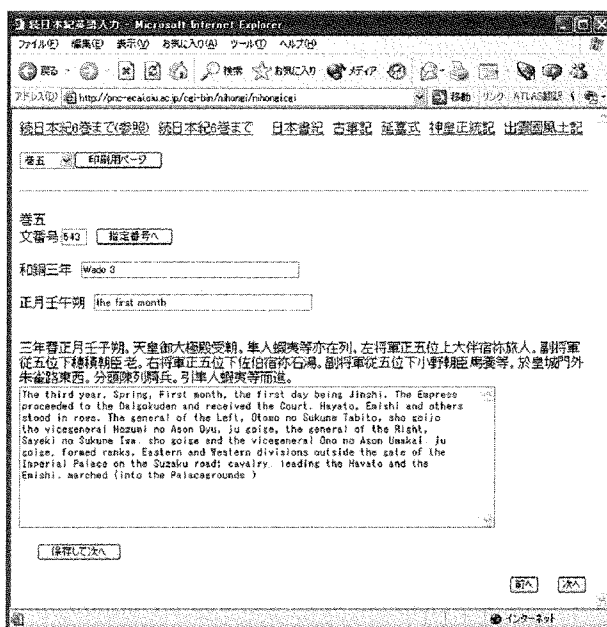


図7. 英訳(翻訳)支援システムの入力画面

有効に作用し、URL: <http://sunsite.berkeley.edu/JHTI/>で試験運用されている。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。今後、表1に示す25文献の格納を計画している。

本英日全文連携検索システムは、大量で特別に加工されていない、つまりデータに一定の形式をもっていない日本古典史料を対象に、歴史研究を支援することを目的に構築・実現した。そのためには、何らかのプログラムを介した検索を必要とするため、ブラウザの要求に対してWebサーバから呼び出されコンテンツに対して動的で対話的な環境を提供するCGIと呼ばれる機能を利用した。しかし、格納歴史文献数が増大し、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、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、日本古典史料の文書構造と歴史的記述方法に着目した検索手法を開発しシステムを設計・開発することが重要である。

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、WWW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。文書本体とは、別に管理された注釈や相互参照などを文書本体と同様に検索可能にすることにより、Web閲覧上での相互参照を実現する。これらを実現するため、文書の論理構造を定義可能なXMLを採用し全文連携検索システムを構築することを計画している。

本英日全文連携検索システムが扱うデジタル情報（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を行えると考える。また、文字コードのUnicode（UTF-8コード）化をも計画している。UTF-8コード化することにより、外字数は大幅に減少する。例えば、今までに構築した8文献の外字総数1,622字、外字種類数644字は、UTF-8コードに多くの漢字フォントが存在するため、外字総数402字、外字種類数134字に減少する。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。

最後に、本英日全文連携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）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バークレー博士課程の大久保裕子氏ほか関係各位に謝意を表す。

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1. 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）**

国際研究論叢

- 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 : Rissho 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] 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.
- [2] 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
- [3] 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, 採録決定
- [4] 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
- [5] 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,
- [6] 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.
- [7] 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)
- [8] 尾崎暢殃編, 『訂正古訓古事記 / [本居宣長訓]、上』, 新典社, 1971.
- [9] 尾崎暢殃編, 『訂正古訓古事記 / [本居宣長訓]、中』, 新典社, 1978.
- [10] 尾崎暢殃編, 『訂正古訓古事記 / [本居宣長訓]、下』, 新典社, 1978.
- [11] Norinaga Motoori "Kokun Kojiki teisei 4-koku" Nagata Bunshodo, 1874, stored in UCB East Asian Library.
- [12] Donald L. Philippi "Kojiki /Translated with an Introduction and Notes by Donald L. Philippi" University of Tokyo Press, 1968, pp. 1-655.
- [13] 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.

- [14] 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.
- [15] 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.
- [16] 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.192-237
- [17] W. G. Aston “NIHONGI : Chronicles of Japan from the Earliest times to A.D. 697” Printed by the Japan Society, 1896.
- [18] 補増 六国史巻三、『続日本紀上巻』、朝日新聞社、1940、pp.1-455
- [19] 補増 六国史巻四、『続日本紀下巻』、朝日新聞社、1940、pp.1-516
- [20] Translated and annotated by J.B. Snellen “Shoku Nihongi : Chronicles of Japan, continued, from 697-791 A.D.” (Vol.1c3) In The Transactions of the Asiatic Society of Japan. Second Series. Vol.11. 1934. Asiatic Society of Japan; 151-239
- [21] 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
- [22] 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
- [23] 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
- [24] 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
- [25] Translated with an introduction by Michiko Yamaguchi Aoki : Izumo Fudoki, Sophia University, Tokyo, 1971, pp.1-157
- [26] 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
- [27] Translated by H. Paul Varley : A Chronicle of Gods and Sovereigns : Jinno Shotoki of Kitabatake Chikafusa, Columbia University Press, 1980, pp.1-300
- [28] Electronic Cultural Atlas Initiative :<http://ecai.org/>.
- [29] Japanese Historical Text Initiative :<http://sunsite.berkeley.edu/ITHI/>.
- [30] <http://pnc-ecai.oiu.ac.jp/>, 「Japan PNC, ECAI」ホームページ。
- [31] Shishir Gundavaram “CGI Programming on the World Wide Web” O'Reilly & Associates, Inc., 1996.11.
- [32] Larry Wall and Randal L.Schwartz “Programming Perl” O'Reilly & Associates, Inc.,1992.3.
- [33] Randal L.Schwartz “ Learning Perl” O'Reilly & Associates,Inc., 1994.4.
- [34] 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。
- [35] 今昔文字鏡（単漢字10万字TTF版）、文字鏡研究会。