様式 C-19

科学研究費補助金研究成果報告書

平成２１年６月１０日現在

研究種目：基盤研究(B)
研究期間：2006～2008
課題番号：１８３２０１１４
研究課題名（和文）　XML を利用した日本古典史料の英日連携検索システムの設計と構築に関する研究
研究課題名（英文）　A study of the Design and the Construction of the Full Text Coordinated Retrieval System of Japanese Historical Resources using XML (Extensible Markup Language)

研究代表者
桶谷　猪久夫
大阪国際大学・国際コミュニケーション学部・教授
研究者番号：９０１６９２６９

研究成果の概要：
日本古典歴史史料を対象に、文書構造や歴史的記述方法に着目し設計され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研究である。出雲風土記、愚管抄、読史余論、古語拾遺、栄華物語、大和物語等の英日連携検索システムを構築した。各文献のデジタル化、日本古典文献の文書を実体と共にその論理構造や属性を定義可能な XML でマークアップし、それら XML 化された各種文書ファイルをデータベース管理システムに格納し、インターネットの Web ブラウザから高速で有効な検索を実現するためのインターフェースを開発した。また、地理情報との連携化の基礎資料であるデジタル地名辞書を構築し、文献情報と古典史料を取り扱うとき重要な地理情報（地理的・歴史的データの連携)との連携化を開発した。

交付額

（金額単位：円）

	直接経費	間接経費	合　計
２００６年度	2,800,000	840,000	3,640,000
２００７年度	2,300,000	690,000	2,990,000
２００８年度	1,800,000	540,000	2,340,000
年度			
年度			
総　計	6,900,000	2,070,000	8,970,000

研究分野：人文学
科研費の分科・細目：史学・日本史
キーワード：全文検索システム、日本古典文献、XML、インターネット、外字属性データベース、外字処理機能、デジタル地名辞書

１．研究開始当初の背景

日本古典史料を対象に、英日両言語でインターネットを利用したフルテキスト形式の検索システムは、バージニア大学の Electronic Text Center から Japanese Text Initiative としてサービスされている。これは、古典から現代までの文学を中心に提供されている。本研究は、特に日本神道・日本の古典史料を中

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、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の開発と歴史的変遷を考慮した履歴データベースの開発、文献情報と古典史料を取り扱うとき重要な地理情報（地理的・歴史的データの連携)との連携化を開発することは、歴史学研究に新たな視点を与え、新しい研究課題・方法を生み出す契機となり、コンピュータ応用技術やインターネット利用技術を新たな段階へ進展させる意義を持つ。

２．研究の目的

歴史史料を対象に、文書構造や歴史的記述方法に着目し設計され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研究である。本研究の目的は、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 31 巻のデジタル化とデータベース化を目標にしていた。

本システムの開発と研究の目的は、英語を話す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、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、日本の古代史研究、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。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。しかし、歴史学研究で利用される古典史料のデータベース化や情報検索においては、歴史的に関連ある史料の効果的な横断的(統合的)検索機能の実現、外字や異体字の問題、原テキストの入力方法、出力方法など解決すべき種々の問題が存在し、いまだに有効な手法がないのが現状である。

しかし、世界的規模での情報検索と情報発信が可能になったインターネット上の Web を利用した研究は、歴史学研究分野においても、例外なく急速に普及している。さらに、インターネット上でデータベースを利用したり流通させたりするとき、情報交換用漢字の不足による深刻な問題がある。これらの問題を解決するため、いくつかの基礎的実験と研究を行ってきた。

本研究の目的は、日本古典史料を対象にこれまでの基礎的実験と研究をさらに推進し、
(1)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。
(2)関連する日本古典資料の横断的（統合的）検索機能の開発と歴史的変遷を考慮した履歴データベースを開発する。
(3)日本古典史料の定量的解析の試みと歴史的事例に基づく各種検索機能プログラムを開発する。
(4)外字を対象とした漢字データベース（漢字属性ファイル）の拡張とインターネット上での利用・転送技術を開発する。
(5)文献情報と古典史料を取り扱うとき重要な地理情報（地理的・歴史的データの連携)との連携化を開発する。

本研究は、歴史学研究分野における膨大な日本語･漢文文献を対象とする。直接対象とする文献は、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」、日本初の解釈歴史書である「愚管抄」、「読史余論」、「太平記」、「古語拾遺」、「栄華物語」、「太平記」、「大鏡」、「万葉集」等であり、さらに、日本古典文献 31 巻のディジタル化、日本古典文献の文書を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 XML でマークアップし、それら XML 化された各種文書ファイルをデータベース管理システムに格納し、インターネットの Web ブラウザから高速で有効な検索を実現するためのインターフェースを開発する。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、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た。

３．研究の方法

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。しかし、格納

される日本古典史料文献数が増大し、デジタル化された情報が膨大になると、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。また、日本古典史料のようなデータに一定の形式をもっていなく、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、それら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、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。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。さらに、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。

これらの実現のため、XML の階層構造、リンク構造やポインターを用い、文書本体への高速全文検索との統合を実現する文書構造検索方式を開発する。

まず、文書の論理構造を定義可能な XML を採用し、Web を利用した英日連携検索システムを構築することを計画している。また、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、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全文検索システム OpenText（OpenText 社、カナダ）を採用する。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開発する。

各文献は文献の文書構造の解析から XML 化（タグ付け）され、インターネットでの表示を可能にするため、CGI プログラムでユーザーインターフェースを開発する。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・開発する。

さらに、文献情報と古典史料を取り扱うとき重要な地理情報（地理的・歴史的データの連携)との連携化を開発する。レイヤー構造を持つ古地図と現在のデジタル化された標高地図を用い、現存する何百もの地理的、歴史的データを挿入していく計画である。これにより、研究者は、あらゆる文献に記されたすべての事象、たとえば神社の空間的、時間的な設定を、瞬時に同じ画面上に参照可能になる。この地理情報システム（GIS :Geographic Information System)は、デジタル地図情報のマッピングのみでなく、歴史研究に重要な時間軸（年代）の設定や該当地点の注釈の閲覧や付加が可能である。この地理情報・資源共有化システム(GIS)は、研究代表者が参加している HGIS (Humanities Geographical Information Science)研究会で設計・構築し利用可能になっている。その地理情報・資源共有化システムと日本古典史料の英日連携検索システムとの連携化を実現する。そのための人文科学分野のデジタル地名辞書を構築していく。

４．研究成果

本研究の目的は、日本古典歴史史料を対象に、文書構造や歴史的記述方法に着目し設計され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研究である。具体的には、日本古典文献 31 巻のデジタル化とデータベース化を目標にしている。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、続日本紀や神皇正統記、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、日本初の解釈歴史書である愚管抄、歴史物語の大鏡、太平記、栄花物語、吾妻鏡、歌物語である大和物語等であり、さらに、日本古典文献 31 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、再利

用を目標にして構築した。日本古典文献の一部を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 XML でマークアップし、それら XML 化された各種文書ファイルをデータベース管理システム（OpenText）に格納し、インターネットの Web ブラウザから高速で有効な検索を実現するためのインターフェースを開発した。さらに、文献情報と古典史料を取り扱うとき重要な地理情報（地理的・歴史的データの連携)との連携化を開発した。

以下に具体的な研究成果を記述する。

(1) 歴史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た。

(2) 日本古典文献の対象文献を OCR を利用し入力した（文字認識でデジタル情報化：「太平記」、「読史余論」、「愚管抄」、「大鏡」、「読史余論」、「古語拾遺」、「大和物語」など)。

(3) デジタル化された文献（日本文と英訳文）に対して、文書構造や相互関連から有効で効率的な検索を可能にする XML のタグ付けを定義し、全文検索と組み合わせた文書構造検索方式を開発した。また、一部文献に対しては、文献ページ画像ファイルや検索部分の日本語画像ファイルを連携化させた。

(4) 文献情報と古典史料を取り扱うとき重要な地理情報（地理的・歴史的データの連携)との連携化を開発した。延喜式巻 9 と巻 10（延喜式神名帳）に記載された神社（式内社：2,861 社、3,132 座)の神社名、地域名、地名の変遷、位置情報（緯度、経度）などを作成し式内社データベースを構築し、延喜式検索システムにリンクした。

(5) 人文科学分野のデジタル地名辞書データベースを構築した。その地名辞書は、吉田東吾（1864 年：元治元年～1918 年：大正 7 年）が 32 歳から 44 歳にわたる 13 年間で書き上げ、1907 年（明治 40 年）に完成した全国の地誌である「大日本地名辞書」（全 11 巻)、5,580 頁以上、1,200 万字の地名辞書から構築した。国土の国名、郡名、町村名、郷名、神社、寺院、山川、湖沼、港湾などの地名に関する考証・変遷などが詳しく解説されている。索引には 53,676 に及ぶ地名が「漢字」と「仮名」により記述されている。これらに漢字、仮名表示、ローマ字表記、緯度、経度、地名属性などを付加し構築した。また、寺院名鑑の 78,588 寺院、式内社の 2,842 神社、日本で最初に測量された地図である仮製図・迅速図から抽出した地名 19,356 も付加し構築した。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１１件）

①Ikuo Oketani、Mitsuru Aida、The Construction of the Digital Gazetteer Based on Humanities GIS and the Analysis Of Characteristic、PNC2008 Annual Conference Joint Meeting with ECAI and JVGC, Hanoi University of Technology, Hanoi, Vietnam、Program and Abstract, 2008, PP. 95-95, PP. 1-27 (CD-ROM), 査読無

②柴山守、原正一郎、貴志俊彦編、桶谷猪久夫他、デジタル地名辞書構築とその利用、アジア遊学 No.113, 特集「地域情報学の創出」、勉誠出版、2008、PP. 182-187、査読有

③桶谷猪久夫、延喜式と吉田東伍「大日本地名辞典」からのデジタル地名辞典の構築、地域情報学ニューズレターNo.3、京都大学東南アジア研究所、2008、PP. 13–14、査読無

④桶谷猪久夫、前川武、日本地名辞書の開発と地名属性からの特徴分析、大阪国際大学紀要「大阪国際論叢」第 21 巻第 2 号、2008、PP. 19-39、査読無

⑤桶谷猪久夫、人文分野における日本地名辞書の構築と地名属性の特徴分析、情報処理学会、「人文科学とコンピュータシンポジウム(IPSJ Symposium Series)」、Vol.2006、No.17、2007、PP. 79-86、査読無

⑥Ikuo Oketani, Mitsuru Aida, The Construction of the Gazetteer of Japanese Place Names based on Humanities GIS, and the Analysis of Characteristic on Distribution of Place Names Attribute, PNC (Pacific Neighborhood Consortium) and ECAI 2007 Annual Conference and Joint Meetings, University of California, Berkeley, U.S.A、2007、PP. 1-49(CD-ROM), 査読無

⑦桶谷猪久夫、JHTI (Japanese Historical Text Initiative) Project: － 日本古典史料の英日全文連携検索システムの構築 －、国際コミュニケーション学科編集委員会編、異文化コミュニケーション研究 —探求・発見・教育—、2007、PP. 99–125、査読無

⑧桶谷猪久夫、前川武、式内社データベースの構築と分布の調査、大阪国際大学紀要「大阪国際論叢」第 20 巻第 2 号、2007、PP. 49-62、査読無

⑨原正一郎、桶谷猪久夫、景観の計量的解析～GISを利用した聖なる場所の統計的分析～、情報処理学会、「人文科学とコンピュータシンポジウム(IPSJ Symposium Series)」、Vol.2006,No.17、2006、PP. 235–240、査読無

⑩Shoichiro Hara, Ikuo Oketani,

Reconstruction of Historical Landscape - In the context of Ancient Shrines and their Surrounding Features -, PNC (Pacific Neighborhood Consortium) 2006 Annual Conference in Conjunction with PRDLA & ECAI, Seoul National University, Korea, Program and Abstract, 2006、PP. 52-52, PP. 1-36 (CD-ROM) 、査読無

⑪Ikuo Oketani、Mitsuru Aida、Creation and Application of Japanese Historical Gazetteer – Ontological Approach to Geographical Name and Place -、PNC (Pacific Neighborhood Consortium) 2006 Annual Conference in Conjunction with PRDLA & ECAI, Seoul National University, Korea, Program and Abstract、2006、PP. 49-50, PP. 1-24 (CD-ROM), 査読無

〔学会発表〕（計３件）

①Ikuo Oketani, Spatiotemporal Tools and Metadata for Area Studies, ECAI Meeting in conjunction with CAA、2009.03.24, Williamsburg, VA, USA

②桶谷猪久夫、デジタル地名辞書、平成 20 年度第 4 回PASIS・HJIS合同研究会、2009 年 2 月 22 日、京都大学地域研究統合情報センター

③桶谷猪久夫、デジタル地名辞書の開発とデモ、京都大学地域研究統合情報センター平成 19 年度全国共同利用研究報告会、2008 年 4 月 27 日、京都大学地域研究統合情報センター

〔図書〕（計 0.件）

〔産業財産権〕

○出願状況（計０件）

○取得状況（計０件）

〔その他〕

http://pnc-ecai.oiu.ac.jp/
http://www.berkeley.edu/jhti

６．研究組織

(1)研究代表者
桶谷　猪久夫
大阪国際大学・国際コミュニケーション学部・教授
研究者番号：９０１６９２６９

(2)研究分担者
藤本　雅彦
大阪国際大学・人間科学部・教授
研究者番号：３０１７３４７０

(3)連携研究者
柴山　守
京都大学・東南アジア研究所・教授
研究者番号：１０１６２６４５