様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年　4 月　1 日現在

機関番号：12601
研究種目：基盤研究（Ｂ）
研究期間：2009 年度　～　2011 年度
課題番号：21320121
研究課題名（和文）　歴史知識情報の正規化による古記録フルテキストデータベース高度化と記録語の解析研究

研究課題名（英文）　The Analytical Research on the Advanced Database of the Full Text of Old Diaries and Diary Terms by Normalizing Knowledge-based Historical Information

研究代表者
吉田　早苗（YOSHIDA, Sanae）
東京大学・史料編纂所・教授
研究者番号：00110693

研究成果の概要（和文）:『後龍翔院殿記』の構造化テキストにより、古記録に適合するタグ形式のあり方を研究し、史料編纂所情報処理システム全体との整合性を確認した。『後法成寺関白記』四を対象に TeX データを作成し、データ転換の方法論を追究するとともに、「大日本古記録」として公刊することにより、電子出版の有効性を検証した。

既刊古記録の索引データを電子化し、構造化テキスト中の語彙タグとの共通化の研究を進め、史料編纂所情報処理システム全体との有機的連携という新たな課題を発見した。索引データは、「中世記録人名索引データベース」を介して公開した。「編年史料カードデータベース」によって記録原本・写本情報を共有化し一般公開の準備を進めた。

研究成果の概要（英文）: This Project opened its research by structuralizing the text of *Goryushoindenki*, discussing its tug format and confirming the most suitable type for the whole Information Processing System of the Historiographical Institute. Next the project has completed the TeX data of the 4th Volume of *Gohojojikanpakuki* and sought the data converting method with examining the effectuality of the electrical publication by publishing its data in the style of *Dainihonkokiroku*.

At the same time, the project has digitizing the index of published old diaries and improved the standardization of term tugs in the structuralized text. As a result, the project has found the new subject which should be organized in cooperation with the whole Information Processing System of the Historiographical Institute. The digitized data of the index has been offered for the public use by “the Database of the Index of Medieval People Name in the Old Diaries”. Finally, the project has prepared for standardizing manuscripts and copies of diaries for the public use by “the Database of the chronological Card of Documents”.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	6,200,000	1,860,000	8,060,000
2010 年度	3,700,000	1,110,000	4,810,000
2011 年度	3,900,000	1,170,000	5,070,000
年度			
年度			
総　計	13,800,000	4,140,000	17,940,000

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：史料研究

１．研究開始当初の背景

史料編纂所が公開する「古記録フルテキストデータベース」は、史料原本の精査によって編まれた『大日本古記録』をもとに構築・公開されており、学界において最も信頼されているツールと言ってよい。また『大日本古記録』の索引データなどを搭載する「中世記録人名索引データベース」も、古記録研究において幅広く活用されている。従来、こうしたデータベース（以下 DB と記す）の構築にあたっては、活字化された史料集をもとにテキストをデジタル変換し、書誌に関するメタ情報を付加して登録していた。索引収載語彙も史料本文テキストとは別個にデジタル変換し、属性や語彙間の連関情報を与えて索引 DB に登録する方法をとってきた。すなわち紙媒体による編纂作業と DB 構築作業は、重複する部分が多いにもかかわらず、技術的な制約もあって、それぞれ別個になされてきたところである。かかる非効率を省きより合理的な作業スキームを確立するため、史料集編纂工程と DB 構築工程を一本化することが要請されていた。また同時に情報の共有化や高度利用の実現、DB の高度化についても永らく要望されてきた次第である。

２．研究の目的

史料編纂と DB 構築を一本化するためには、全ての工程を電算機で処理しうる形におきかえてゆく必要がある。その基礎的前提として、史料翻刻テキストならびにそこに付与された諸情報を、普遍的なルールに基づいて記述する体系を定義しなければならない。そこで本研究においては、テキストの体裁や付加情報に関して、電算機が識別しうる共通の記号―XML タグ―を定め、標準となる構造化テキストのスタイルを作り出すことを目的とした。あわせてこの構造化テキストから特定の語彙を機械的に抽出し、索引作成の基礎データとして蓄積する方法論を確立することや、構造化テキストを史料集版面データに移行させる方法論を案出することを目指した。これによって翻刻から出版・DB 公開に至る効率的なスキームの確立を目論んだ次第である。また古記録原本や諸写本を対象に、長年にわたり蓄積してきた調査情報についても、かかるスキームのうちで利活用できるようその環境整備を進めることも目標の一つに掲げた。

３．研究の方法

『大日本古記録』のテキストを素材として、いかに汎用的で機能的な XML タグを用いて構造化するかを検討し、その成果をふまえて現状の「古記録フルテキストベース」について機能改善と高度化を模索した。テキスト体裁注・人物注・地名注など明示する XML タグを設計し、既存 DB データとの間で比較検討を重ねた。また記録語彙に関する取組としては、タグを活用した語彙の機械的集積や、集積した語彙間の関係性・脈絡を記述するシステムの構築にむけて、実証実験を重ねたところである。平行して既刊史料集の索引をデジタルデータ化し、収載された語彙群の属性について、構造化テキストと同一のルールで標準化し、整合性を高める試みにも取り組みを進めた。構造化テキストから史料集版面電子データへの変換については、『大日本古記録』版面にもとづいた製版用電子データを Tex 形式で作成し、XML タグと TeX タグの比較検証を進め、問題点の析出に努めた。記録原本・写本に関する情報については、データの標準化を施し、既存のデータベース群に登録し、情報共有の条件整備に取り組むこととした。

４．研究成果

構造化テキストについては、史料編纂所が所蔵する戦国期の記録『後龍翔院殿記』を素材として作成し、古記録に適合するタグ形式のあり方を探ることができた。このデータをもとに、既存のフルテキストデータベースへの適応方法を検証し、史料編纂所情報処理システム全体との整合性を確認したところである。また史料集刊行中の『後法成寺関白記』四を対象に Tex データを作成し、構造化テキストと比較対象を通じて、データ転換の方法論を追究した。なおこの Tex データについては、実際の編集・印刷に活用し、「大日本古記録」として既に公刊したところである。これにより電子出版にむけた有効性についてもあわせて検証することができた。古記録語彙をめぐる実践としては、既刊史料集のうち『御堂関白記』『後二条師通記』『殿暦』『猪隈関白記』『岡屋関白記』の索引データを電子化し、構造化テキスト中の語彙タグとの共通化にむけた取り組みを行った。今後、史料編纂所情報処理システム全体のなかで、いかに有機的に連携させてゆくかが課題として浮かびあがってきたところである。なお電子化した索引データについては、「中世記録人名索引データベース」を介して公開し、既に利用に供したところである。記録原本・写本

情報の共有化については、情報形式を「編年史料カードデータベース」に転換したうえで、登録を進めており、条件が整ったものから順次一般に公開することを予定している。

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）
本郷恵子「 遠藤基郎『中世王権と王朝儀礼』」『史学雑誌』118-11,2009 年、pp. 89-96、査読有
菊地大樹「上川通夫『日本中世仏教史料論』」『歴史学研究』855、2009年、pp.44-47、査読有
井上聡・馬場基「文字字形総合データベース作成の試み」『人間文化研究情報資源共有化研究会報告集』１、2010年、pp.99-112
菊地大樹「 東福寺円爾の印信と法流―台密印信試論―」『鎌倉遺文研究』26、2010年、pp.21-54、査読有
山田太造・井上　聡他「日本史史料における翻刻データの管理と編集支援」『ＦＩＴ2010』9、2010年、N-013
吉田早苗「宗忠と忠通―「中右記部類」に見える「法性寺関白記」『日本歴史』759、2011年、pp.89-94
本郷恵子「鎌倉幕府の特質について」『季刊　iichiko』110、2011年、pp.6-19、査読有
井上聡他「日本古文書ユニオンカタログ－古文書情報を網羅するための"古文書リンケージ"プラットフォーム－」『研究報告人文科学とコンピュータ』1、2012年、pp.1-8
井上聡他「日本史史料読解支援のための候補文字検索」『じんもんこん2011論文集』2011-8、2011年、pp.43-50、査読有
井上聡他「日本史史料における翻刻テキストの構造化支援手法」『情報処理学会研究報告（人文科学とコンピュータ研究会報告)』2011-5、2011年、pp.1-8
菊地大樹「イェール大学所蔵『応永三十二年具注暦』について」『東京大学史料編纂所研究紀要』22、2012年、pp.121-140

〔学会発表〕（計４件）
菊地大樹「東福寺円爾の印信と法流―台密印信試論―」鎌倉遺文研究会、2009年6月24日、早稲田大学
菊地大樹「中世金石文における「和習」への一視点―大陸の影響を視野に―」漢学ワークショップ、 2010年5月15日、プリンストン大学
本郷恵子「千葉氏にみる中世前期の再生産構造」鎌倉遺文研究会、2011年5月26日、早稲田大学
菊地大樹「中世における法華経持経者の系譜」興風談所研究会、2011年6月7日、興風談所

〔図書〕（計５件）
菊地大樹他『鎌倉の世界（史跡で読む日本の歴史６）』、吉川弘文館、2010年、pp.258
本郷恵子『将軍権力の発見』、 講談社、2010年、pp.242
本郷恵子『物語の舞台を歩く　古今著聞集』山川出版社、2010年、pp.158
井上　聡他『列島の鎌倉時代』 高志書院、2010年、pp.261
本郷恵子『蕩尽する中世』新潮社、2012年、pp.254

〔その他〕
ホームページ等
http://www.hi.u-tokyo.ac.jp/index-j.html

６．研究組織
(1)研究代表者
吉田早苗（YOSHIDA, Sanae）
東京大学・史料編纂所・教授
研究者番号：00110693

(2)研究分担者
本郷恵子（HONGO, Keiko）
東京大学・史料編纂所・教授
研究者番号：00195637
尾上陽介（ONOE, Yosuke）
東京大学・史料編纂所・准教授
研究者番号：00242157

(3)連携研究者
田中博美（TANAKA, Hiromi）
東京大学・史料編纂所・准教授
研究者番号：60111572
菊地大樹（KIKUCHI, Hiroki）
東京大学・史料編纂所・准教授
研究者番号：80272508
井上　聡（INOUE, Satoshi）
東京大学・史料編纂所・助教
研究者番号：20302656