

機関番号：62608

研究種目：基盤研究（B）

研究期間：2007～2010

課題番号：19300087

研究課題名（和文） 和漢古典学のオントロジモデルの応用

研究課題名（英文） The research to apply the model of ontology about the classical study of Japan and China to use.

研究代表者

相田 満 (AIDA MITSURU)

国文学研究資料館・研究部・准教授

研究者番号：00249921

研究成果の概要（和文）：

本研究においては、前近代の文献の分析を支援するために蓄積された概念語彙・暦日テーブル・歴史地名を利用して、(1) 和暦で記載される暦日情報を自動的に認識して太陽暦を自動的に付与するツールを開発し、(2) GIS情報の付与された歴史地名の形態素解析辞書を構築することに取り組んだ。そして、概念を蓄積した和漢オントロジデータベースと、和暦暦日の自動埋込ツールを利用した『古事類苑』の全文データベースの開発、歴史地名データベースを利用した研究モデルの提示などを通じて、日本文学・文化研究の高次化をはかった。

研究成果の概要（英文）：

To support the analysis of the document written in the former modern ages, we have straightened three kinds of dictionaries (Concept, Day of calendar of Japan, History name of a place.).

- (1) Developing the program which automatically recognizes the calendar day expression, and adds information converted into the solar calendar.
- (2) Making up the morphological analysis dictionary of the name of a place which is added GIS information.

We presented the research model by whom these were assumed to be a subject. And, it is the one to improve the research level of the study of the Japanese literature and the culture.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	3,100,000	930,000	4,030,000
2008年度	4,900,000	1,470,000	6,370,000
2009年度	4,200,000	1,260,000	5,460,000
2010年度	1,400,000	420,000	1,820,000
年度			
総計	13,600,000	4,080,000	17,680,000

研究分野：総合領域

科研費の分科・細目：情報学／図書館情報学・人文社会情報学

キーワード：文学情報システム、GIS、オントロジ、暦象、文書解析

1. 研究開始当初の背景

哲学用語「存在論」に由来する「オントロジ」は、情報学では「概念間の関係の明確な定義の集まり」として、情報リソースから独立した上位層(メタレイア)において、情報を意味的に組織化することによって、検索を支援するナビゲーションのために重要で、かつ不可欠な存在となっている。

しかし、その取り組みと発想は「オントロジ」という名称が象徴するように、有史以来の普遍的な命題への取り組みに通底するものといっても過言ではなからう。何となれば、和漢の古典においては、先人達が類概念(分類用概念語彙)によって、情報を統御する機能を持った類書・辞書・注釈書類に代表される類聚編纂物を数多く輩出してきているからである。そして、それらに内在する、宝庫とも形容すべき継承性の高いオントロジ群によって、長い間、情報の組織化と文化的価値の継承の役割を担ってきたのである。

研究申請者は、古典的類聚編纂物に見える、検索効率を向上させるための、そうした意匠の特性に着目して、当該書籍群の見出しに使用される分類用概念語彙を、概念構造もそのままに、情報資源として集積・整備する共同研究プロジェクトを企画し、取り組んできた。そして、和漢古典学のオントロジ[和漢オントロジと略称する]と称する情報資源群が、古典と現代とを文化的に結びつけるだけでなく、様々な分野の知的体系を包摂し得ることと、特に自然・人事を扱う絵画・映像と和漢オントロジとの親和性が高いことを情報学だけでなく、人文学研究も交えた複眼的視点によって明らかにしてきた。(平成15年度～17年度基盤研究(B)「和漢古典学のオントロジモデルの構築」[研究代表者:相田満])。

古典的世界観には、現代まで受け継がれる高い継承性がある一方で、地域・時代ごとの特色を示しながら変化していくものもある。和漢オントロジは、そうした不易と流行を体現するものでもあった。

そして、「応用」と名付けた今回のプロジ

ェクトで取り組んできたことは、これまでに集積された和漢オントロジをさらに充実させるだけでなく、集められた情報資源を高次かつ汎用的なものとするための基盤資源として、新たに日本の旧暦使用時代を扱う時間(日本旧暦オントロジ)・空間(日本歴史地名オントロジ)のオントロジ資源と組み合わせて利用することを構想するものである。

2. 研究の目的

本研究課題を実現するために、以下の4つの主題を設定し、それに基づく開発・データ整備、研究モデルの提示に取り組むものである。

(1) 暦象情報付加ツールの開発

電子化テキストの公開が、ますます活性化する現況にありながら、旧暦を含む文書进行处理可能とするツールは、相変わらず乏しい。そこで、旧暦・太陽暦を問わず時間情報が不規則に埋め込まれた文章から、自動的に時間情報を抽出し、それにユリウス通日、あるいは現在使用される太陽暦に変換した結果をタグとして埋め込む知識発見ツールを開発する。

(2) 暦日テーブルオントロジの充実

古暦に関する研究を深め、具注暦の収集などにより、暦日テーブルを充実させる。

(3) 歴史地名オントロジの構築(歴史地名形態素情報辞書の構築)

連携研究者(2007年度のみ研究分担者)・桶谷の作成した『大日本地名辞書』索引約52,000件にGIS情報が付されたデータベースデータに対して、地名・ヨミ情報の修訂、川・山・橋などの地名接続詞と地名本体との切り分け作業を行ない、日本の古典と親和性の高い、地名形態素辞書を作成する。また、作成された地名辞書を利用して、GIS情報と歴史地名を利用した研究を進めることにより、知識発見モデルを提示する。

(4) 和漢古典概念オントロジの整備

和漢古典概念オントロジを充実することにより、地名・暦注に使用される概念との接続させた検索を実現するための基盤を整える。

3. 研究の方法

上記研究目的の達成のために、以下の点に取り組んだ。

(1) 曆象情報付加ツールの開発

日本の前近代の文書における記載法は、太陰太陽曆に基づく。そのため曆日表現は、閏月を含み、現在使用される太陽曆と1-2ヶ月程度ずれ、しかも干支だけで表記されたり、同年同月などの省略法が使用されたりなど、不規則な形式によるものが多かった。そこで、曆日を補正した情報を補正した結果を、タグとして埋込み、以てテキスト分析をサポートするツールを開発するに際し、代表者がこれまでに蓄積していた尊卑分脈・公卿補任・地下家伝などの人物データベースをもとに、表記の不規則な出現率とパターンのおおよその傾向について予備調査を行い、仕様招請を行った上で開発に着手した。

開発ツールの基本機能は、前近代文献に不規則な表記で現れる和曆の曆日表現を自動的に認識し、それにユリウス通日や太陽曆に変換した結果を埋め込むものである。ただし、これまで多くの前近代の文書を扱ってきた経験から、処理対象の文字コードは UTF-16 に設定することにより、あらゆる文字種に対応することを念頭に置いた。その結果、ツールには MEGASOFT 社のテキストエディタ MIFES8.0 のマクロ言語である mil 言語を使用して UTF-16 文字コードに対応させるための開発コストの省力化・ツール運用の安定化を図った。

処理に際しては、100 万件を超える曆日テーブルを参照して、文字パターンマッチングによる逐次処理を行っている。また、曆日表記数字や干支の記載パターンや異体字辞書にも対応させ、干支のみの表記や同年同月同日などの記載にも対応させることにより、文書の自動解析を可能にした。

(2) 曆日テーブルオントロジの充実

日本で最も長期間使用された宣明曆を中心に、古曆に関する基本資料として、随心院蔵本の曆注資料の収集、および「江戸のものづくり」責任者の大谷光男氏より古曆の複写資料の提供を受けるなどして、曆注情報の分

析を進め、テーブルに情報を付加した。

また、開発ツールの認識精度を上げるために、100 万件超の曆日テーブルの和曆数字の数字異表記パターンを増加した。

(3) 歴史地名オントロジの構築(歴史地名形態素情報辞書の構築)

吉田東伍『大日本地名辞書』索引の約 52,000 件にそれぞれ GIS 情報が付されたデータの拡張を行った。

具体的な作業は、データの修訂・ヨミの追加、ローマ字化等である。

まず、データの修訂については、原資料に「伊豆沼(イズ)」としか書かれていないものを「伊豆(イズ) / 沼(ヌマ)」と、データの追加と切り分けを行い、それぞれの地名は地名接続詞と地名本体とに切り分けることで地名形態素辞書を作成し、さらに、ヘボン・訓令式を各 2 種類ずつ計 4 種類のローマ字を宛てた。

この地名形態素辞書は日本の古典概念語と親和性の高い、歴史地名を分析するのに重要な情報資源となっており、本データベース化の作業によって、それらがいずれも GIS 情報と組み合わせられて利用可能となる。この取り組みによって、これまでに類例のない特色を持つデータが出来上がったといえよう。

本研究では、そうした情報資源を利用した知識発見モデルを提示する研究成果を国内外の学会で発表することを通じて、利用の可能性を模索するとともに、データベースの可能性や、取材典拠の価値などについて評価を問うた。

また、千島・樺太・台湾についても、GIS 情報に加えて、同様の情報付加を施し、今後さらに高次の応用研究を可能とする準備を整えた。

(4) 和漢古典概念オントロジの整備

地名・曆注に使用される概念との接続のために、和漢古典概念オントロジデータベースに搭載する典拠辞書を新たに追加し、『倭名類聚抄』、俳諧歳時記類など、バリエーションの追加につとめた。

4. 研究成果

(1)前近代文献の構造化支援と暦象情報付加ツールの開発

本研究で作成した暦日テーブルは、人間文化研究機構資源共有化システムと京都大学地域情報統合センターにて採用され、Hu-Timeに代表される時空間検索システムのコアデータとして利用されるに至っている。

また、暦象情報付加ツールの基礎辞書として使用し、開発ツールを使用して、これまでに入力・公開を行った『古事類苑』データベースのXMLデータに暦日情報を付加した。ツールは旧暦による時間情報が不規則な記述形式で出現する文章の時間情報を認識してユリウス通日・太陽暦を埋め込むことにより、情報ナビゲートを実現するもので、95%以上の正答率を実現した。そこで、これまでに構築した『古事類苑』データ(天部・歳時部・地部・称量部)を対象に、旧暦を太陽暦に変換した時間情報を埋め込んだHTMLデータ約6000頁の構築を行った。[現在、当該データのデータベース評価版を公開中]

さらに、開発ツールと同じプログラムを使用して地名抽出して位置情報を付加する機能の実現の可能性評価試験も行った。この試みについては、切り出し用辞書の更なる充実・整備が必要なことが判明した。ただし、その実現への取り組みは今後の課題のために、人名との切り分け実験用に、人名データを現代人・歴史人名と併せて20万人分の収集を行い、今後の検証用に資した。

なお、2007年度に発表した「暦象オントロジの構築：日本旧暦時代の文献分析支援のために」により、情報処理学会山下研究論文賞を受賞した。

古事類苑 全文検索 (試験版)

ホーム Powered by research

検索 長閑

対象言語 全ての言語 更新日 指定なし

表示件数 10件ずつ

インデックス 最終更新日時 2011年04月04日 18時16分03秒

Unicode版msearch Version 1.52(U2)を使った検索方法

例	意味
AAA	キーワード「AAA」を含むページを検索。
AAA BBB	キーワード「AAA」と「BBB」を両方とも含むページを検索。
-AAA	キーワード「AAA」を含まないページを検索。
AAA BBB	キーワード「AAA」か「BBB」のどちらか一方もしくは両方を含むページを検索。
{AAA BBB} {CCC}	{AAA BBB}のようにつづつ以上のキーワードでもOKです。ただし、ネストはできません。
{AAA}	キーワード「AAA」がページのタイトルに含まれるページを検索。{は大文字HJでもOKです。
{...}	キーワード「AAA」をページのURLにも含むページを検索。uは大文字UでもOK

全言語のページ検索。34件ヒットしました。1件目から16件目を表示します。検索にかかった時間0:00:16

[1] [2] [3] [4] [進む]

1. <http://base1.nijl.ac.jp/~kojien/saijibu/frame/000183.html> 晴天徳也。徳年有登壇。又有内裏徳也。長保〔日本紀略 十一傳〕0997年3月25日長徳三年二月十四日己酉。百匡前。即可譯連年號之由。0999年2月1日長保元年正月十三日丁卯。詔改0899年1月20日...
2. <http://base1.nijl.ac.jp/~kojien/syoryoubu/frame/000075.html> 二十五年。立積六十二年。以慶雲大尺量。長保新製官寸法立積全同。此并慶長新製之京弁以前。天下諸國通用。〔日本年代記 五〕世宗川には定難のはたらき方の貴族する共。其人と次第に念出にならぬやうに兼任の心...
3. <http://base1.nijl.ac.jp/~kojien/saijibu/frame/001342.html> 上。即召少外記系助問詳事。○去年閏十二月。一條母后藤原詮子前。圖陣日結御酒造事。申云。正...
4. <http://base1.nijl.ac.jp/~kojien/saijibu/frame/001373.html> 詳略年如所行之。但1001年1月27日長保三年停此處。依セウ日内也。○此年閏十二月。一條母后藤原詮子前。專任御所御定行耳。登議一人行事0863年1月28日元和三年0865年2月4日應保二年...
5. <http://base1.nijl.ac.jp/~kojien/saijibu/frame/001137.html> 非遺使。實樂人等登壇。1000年2月8日長保二年家主進云。望傳員。金剛大寺一休。在多種向蓮花二本。向花瓶。二口云。裏書云。御書傳傳作法。東寺傳。先法用。次表日。持分并供養淨真言了。次下座口智調議。次講師...

[図1] 暦日タグを挿入した古事類苑HTML検索(試行検索版)

(2)大日本地名辞書におけるGIS情報の整備とGIS情報を利用した知識発見モデルの提示

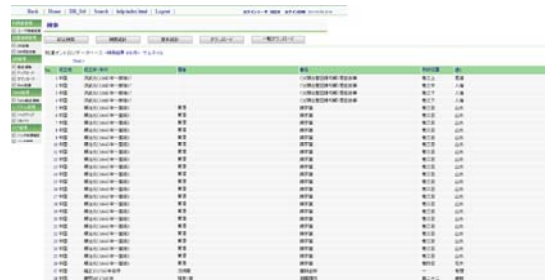
千島・樺太・台湾を含む吉田東伍『大日本地名辞書』中の地名を地名形態素に分別する作業と、ローマ字読み(訓令・ヘボン式の6通りの読み)の付加作業を行った。また、その成果に基づいて、各地名に付されたGIS情報を利用した知識発見モデルの提示や、国文学研究における『大日本地名辞書』辞書の意義付けを論じた成果の発表を行った。さらに、台湾編については、新たな地名を付加するとともに、現地調査を行って、日本地名と台湾地名とが重層的に重なる様相を具体的に分析し、発展研究の糧とした。

(3)和漢古典概念オントロジの整備

平成21年度から人間文化研究機構資源共有集中システム(NIHU One)にて、それまで蓄積してきたオントロジ資源の一部を和漢古

典概念語データベースとして公開した。

また、『古事類苑』の部立だけでなく、引用各書に含まれる概念語(見出し)について分析し、『古事類苑』公開データベースの見出しナビゲーションの改良の可能性について分析を行い、成果発表を行った。



【図2】NIHU-One データベース

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 20 件)

- ①相田満,『枝葉抄』から見た『胡曾詩抄』, 説話 11, 説話研究会, pp117-130(14), 2011 年, 査読無
- ②相田満, データを支えるデータ人間文化研究における知識ベース構築の試みについて, 人間文化研究情報資源共有化研究会報告集 2, pp75-111(37), 2011 年, 査読無
- ③山田奨治・相田満, 古事類苑全文データベースの構築について, 人間文化研究情報資源共有化研究会報告集 1, 人間文化研究機構, pp151-159(8), 2010 年, 査読無
- ④相田満, テキストコンテンツのポテンシャル—『古事類苑』データベースの取り組みから—, 情報処理学会論文集「人文科学とコンピュータシンポジウム(2009)」デジタル・ヒューマニティーズの可能性, Vol.2009.No16, pp253-260(8), 2009 年, 査読有
- ⑤相田満, 地震と六国史—キャラクター生成装置としての六国史—, 東洋研究 171, pp111-133(23), 2009 年, 査読有
- ⑥相田満, 地名オントロジの可能性, 情報処

理学会論文集「人文科学とコンピュータシンポジウム 2007」, Vol.2007, No.15, pp55-62(8), 2007 年, 査読有

⑦原 正一郎・柴山 守, 地域情報学の構築と時空間情報解析ツール, 情報処理学会論文集「人文科学とコンピュータシンポジウム 2007」, Vol.2007, No.15, pp71-78(8), 2007 年, 査読有

⑧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, 情報処理学会論文集「人文科学とコンピュータシンポジウム 2007」, Vol.2007, No.15, pp79-86(8), 2007 年, 査読有

⑨相田満, 暦象オントロジの構築: 日本旧暦時代の文献分析支援のために, 情報処理学会 人文科学とコンピュータ研究会報告, 2007-CH-76, pp25-32(8), 2007 年, 査読無

〔学会発表〕(計 27 件)

- ①相田満, 『古事類苑』と『古事類苑』データベースのこと, 第 1 回国文研フォーラム, 2011 年 4 月 20 日, 国文学研究資料館
- ②相田満, デジタル地名辞書活用の可能性, 第 6 回インド日本文学会, 20/2/2011, Tagore Hall, Japan Foundation, Delhi,
- ③相田満, 歴史地名のオントロジと GIS—『大日本地名辞書』を腑分けして見えてくるもの—, シンポジウム「日本の歴史的時空間情報の現在」, 2010 年 9 月 11 日, 国際日本文化研究センター
- ④ AIDA MITSURU, The Origins and Current State of Digitization of Humanities in Japan--About National Institute of Japanese Literature--, Digital Humanities(DH)2010, 22/7/2010, King's College London
- ⑤ Mitsuru Aida, Ontology of the Japanese place name-The presentation of the knowledge discovery model using Japan place name dictionary database and GIS information-, 2009PNC Annual Conference and Joint Meeting, 6/10/2009, Academia

Sinica in TAIPEI,TAIWAN

⑥相田満, 地名オントロジー大日本地名辞書からの出発―「Historical GISの地平」シンポジウム(人文地理学会歴史地理研究部会・情報処理学会人文科学とコンピュータ研究会共催),2009年7月25日, 帝塚山大学東生駒キャンパス

⑦AIDA MITSURU, OKETANI IKUO. The Construction of the Digital Gazetteer based on Humanities GIS and the Analysis of Characteristic, GIS-IDEAS, PNC and ECAI 2008 Joint International Conferences,8/12/2008,Hanoi University of Technology

⑧相田満,地震と六国史―キャラクター生成装置としての六国史―,2008年中國文化大學日本語文學系所主催日本研究國際學術研討會―教育・社會・語言・文學―,2008年5月24日, 中國文化大學[台湾台北]

⑨相田満, 前近代の時空間(パネル・ディスカッション「時間空間テーマで探る・解く・語る人間文化」), 人間文化研究機構研究資源共有化シンポジウム「研究資源共有化―その展開と可能性―」,2008年3月14日, 情報・システム研究機構会議室

⑩AIDA MITSURU, Possibility of Name of a Place Ontology: That Centers on Old Calendar Document of Japan, ECAI / PNC Joint Meetings: Area Studies, 20/10/2007, U.C..Berkeley, USA

〔図書〕(計3件)

①相田満,「水」と「瑞」,『人と水3 水と文化』(勉誠出版),PP79-106,2010年

②相田満,『大日本地名辞書』から広がる地名オントロジーの可能性,『アジア遊学 113 地域情報学の創出』(勉誠出版),pp44-51,2008年

③渡辺信和,天保年間巡拝記録の意義と略縁起,『近世略縁起論考』(和泉書院),pp97-124,2007年

〔その他〕

ホームページ等

http://mgr.nihu.jp/ndp_top.htm

(人間文化研究機構 nDP[nihu Data Provider]和漢オントロジー検索DB)

<http://basel.nijl.ac.jp/~kojiruien/index.html>

(国文学研究資料館「古事類苑データベース」)

<http://f-lab.jp/kokubu/msearch/msearch.cgi>

(古事類苑全文検索〔暦日タグ付データベース 試行検索〕)

6. 研究組織

(1)研究代表者

相田満 (AIDA MITSURU)

国文学研究資料館・研究部・准教授

研究者番号: 00249921

(2)研究分担者

連携研究者に同じ(平成19年度のみ)

(3)連携研究者(平成20年度～平成22年度)

原正一郎 (HARA SHOICHIRO)

京都大学・地域研究統合情報センター・教授

研究者番号: 50218616

山田 奨治 (YAMADA SHOJI)

国際日本文化研究センター・准教授

研究者番号: 20248751

桶谷 猪久夫 (OKETANI IKUO)

大阪国際大学・人間科学部・教授

研究者番号: 90169269

渡辺 信和 (WATANABE NOBUKAZU)

同朋大学・文学部・主任研究員

研究者番号: 40350996[平成22年12月逝去]