

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・アーカイブズ研究系・助教

研究者番号：00249921

研究代表者の専門分野：人文情報学・和漢比較文学

科研費の分科・細目：情報学／情報図書館学・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. 研究計画の概要

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、情報元から独立した上位層で、情報を組織化し、検索を支援するために重要な存在であるが、その実は、その名称の来歴が示す通り、有史以来の普遍的な営みでもあった。和漢の古典においては、類概念（分類用概念語彙）で統御される辞書や辞典が、継承性の高いオントロジの宝庫として、情報の組織化と文化的価値の継承の役割を担い続けてきたからである。

研究代表者は、かかる古典学の継承性に着目して、類聚編纂物中の階層的に表現される分類用概念語彙を集積し、情報資源として現代生活や社会に応用し得ることを明らかにしてきた（2003～05 年度基盤(B)「和漢古典学のオントロジモデルの構築」[代表:相田]）。

本研究は、情報処理機器でサポートされる機会の稀な、日本の前近代文献を対象に、古典的語彙オントロジ・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 3 者のオントロジを組み合わせた利用を構想し、原資料から、時間情報を切り分け、知識発見を促すためのツールを開発、その利用と開発研究を通じて、研究モデルの提示を進める実践的応用研究に取り組むものである。

具体的には、和漢オントロジ・日本旧暦オントロジ・日本歴史地名オントロジ 3 種の階層的オントロジを、日本の古典籍本文へ埋め込むことにより、情報ナビゲート機能を有するフルテキストコンテンツを構築することを構想するものである。そして、そのためのツールの開発のために、不規則な記述にも対しても暦日情報が自動付加されるツールを

開発し、あわせて、知識発見をサポートするためのオントロジの整備を進めるのみならず、関連するオントロジ資源の共有を促すものである。そして、そのことを通じて、前近代文献世界の知識発見モデルの提示と研究の進展と活性化をめざす。

2. 研究の進捗状況

以下の 4 点を中心に研究を進めてきた。

(1) 暦象情報付加ツールの開発

『古事類苑』記載本文を試験データとして、旧暦による時間情報が不規則に埋め込まれた文章から、時間情報を抽出、それにユリウス通日・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知識発見ツールを開発、95%以上の正答率を実現した。

(2) 暦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暦を中心に、古暦に関する基本資料および、原典資料の収集を行い、暦注情報を分析・付加した。

(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）

研究分担者（2007年度、以後連携研究者）・桶谷の作成した『大日本地名辞書』索引データ約 52,000 件のデータについて、地名表記・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、日本の古典籍と親和性の高い、地名形態素辞書を作成した。また、構築されたオントロジと GIS 情報を利用した知識発見モデルの成果を国内外の学会で発表した。

さらに、千島・樺太・台湾についても GIS 情報だけでなく、新たな地名やローマ字読み（訓令・ヘボン式の 6 通りの読み）を付加することにより、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・暦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を充実させた。

また、和漢古典概念語 DB を人間文化研究機構資源共有集中システム (NIHU One) より公開し、各典拠に時間情報を付加した。

3. 現在までの達成度

①当初の計画以上に進展している。

(理由) 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、さらに今年度からは東京大学史料編纂所の基盤データとして採択され、時空間検索システムのコアデータとして利用された。

また、暦日情報の自動抽出・情報付加ツールの開発の機能向上の方策については、システム的にはほぼ完成している。今後は、エラーデータの分析により、切り出し用のテーブルを個別に修正するために、さまざまなデータを試料として検証を重ねる段階にある。

『大日本地名辞書』のデータについては、補訂分に相当する樺太・千島・台湾のデータに大幅な増補を加えた。これにより、切り出された地名形態素を使用して、さまざまな文化的語彙の検証を行うことが可能になった。

たとえば、その検証の一環として、日本人の姓と地名の相関関係を、平成元年生まれの 911,817 件の人名から抽出した異なり姓数 37,862 件を使用して、姓の頻度と地名との一致についての分析を試みた。その結果、出現頻度が上位 1,000 位までの姓は 90%近くが地名と一致するが、以後は一致率が漸減し、5,000 位以下では 50%を切る一致率となり、23,000 位以後は 10%前後の一致率のまま推移していることが判明した。その結果、地名と姓が一致するという事については、あくまで耳慣れた姓と地名が一致することが多いということであって、必ずしも日本人の姓全般に及ぶものではないことが判明した。

なお、『大日本地名辞書』DB は、連携研究者桶谷・原との協力により、京都大学地域情報統合センターから公開を行っている。

4. 今後の研究の推進方策

これまでは、データ形成とシステム形成に重点を置いた研究の進め方が主であった。データ形成、システム開発は、研究期間終了後も、より充実したものを構築するために引き続き継続されることになるだろうが、最終年度となる研究第 4 年度においては、それまでの研究資源を活用した、オントロジの利用モデルの提示を重点的に進める予定である。特に、暦日・地理情報切り出しツールおよび、それらのデータを搭載したデータ分析ツールを活用することにより、文学研究に有用な知識発見モデルの提示を進めたい。その際、地名オントロジの構築・分析と、和漢古典籍より抽出された概念語彙との統合、暦日情報と組み合わせた知識発見モデルなどは、これ

までに無かった視点からの研究成果となるだろう。そこで、本研究手法のマニュアル化を行いたい。

また、開発ツールは、日本の暦日情報を切り出す機能に特化したものであるが、ツールの仕様としては、日本以外の暦日情報、さらには地名を抽出して位置情報を付加する機能にも対応はしている。ただし、その実現のためには、切り出し用辞書の整備が不可欠なため、辞書整備の基本的指針を策定する。

5. 代表的な研究成果

[雑誌論文] (計 3 件)

- ① 相田満、地名オントロジ—大日本地名辞書からの出発—、情報処理学会研究報告 人文科学とコンピュータ、2009-CH83-2、PP1-8(8)、2009 年、査読無
- ② 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
- ③ 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, pp pp55-62(8)、2007 年、査読有

[学会発表] (計 2 件)

- ① 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「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、2009 年 12 月 19 日、立命館大学 びわこ・くさつキャンパス
- ② Mitsuru Aida, 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 2 件)

- ① 相田満、秋道智彌、小松和彦、中村康夫「水」と「瑞」、『人と水 3 水と文化』(勉誠出版)、PP79-106、2010 年、査読有
- ② 相田満、柴山守、原正一郎、貴志俊彦、『大日本地名辞書』から広がる地名オントロジの可能性、『アジア遊学 113 地域情報学の創出』(勉誠出版)、pp44-51、2008 年、査読有

[その他]

H20 年度情報処理学会山下研究論文賞受賞。

ホームページ

<http://nihuone.nihu.jp/nihu/SelectLocal>

e.do(NihuOne:和漢オントロジ検索 DB)

<http://www.h-gis.org/> (京大地域情報統合センターHGIS)