

研究の沿革

相田 満

1.研究目的

1.1.何をめざすか

哲学用語「存在論」に由来する「オントロジ」は、情報学においては「概念間の関係の明確な定義の集まり」として、それを実装する「トピック・マップ」とともに、情報リソースから独立した上位層に位置付けられ、情報を意味的に組織化、検索、ナビゲートするための新しいパラダイムとして注目を集めている。

しかしこれはまた、有史以来、日本や中国で幾度も編纂された、類概念(分類概念語彙)によってまとめられた古典的な辞書・辞典(=類書)が、きわめて継承性の強い、良質な「オントロジ(知識概念木)」の宝庫となっているように、伝統的な発想にのっとりしたものでもある。

古典的類書に使用される分類概念語彙の 1/4 は、現代日本でも使用され、自然景物・年中行事・人事関係の語彙に集中している。そこで、本研究では、我々研究グループが蓄積する 40 タイトル約 10 万件の、このような特性を持つ古典的な典籍に取材したオントロジを整理・分析するための「分析型データベースシステム」を構築することにより、次の3点の目標を設定し、古典と現代とを文化的・理念的に接合することに取り組みたい。

- 1)2000 年にも及ぶ叡智を注いで編まれ続けてきた古典的な類聚編纂物中の「部立て」に使用される「分類用語彙」の、語句の共通性だけでなく各典籍内における語句配列の一致を指標として、各典籍間の継承関係を考察することを支援する「分析型データベースシステム」を構築する。
- 2)オントロジの観点から、日本・中国の古典的な辞書・辞典(=類書)の継承関係を分析し、各々の典籍の特性や意義について考察する。
- 3)蓄積されたオントロジを利用して、和歌集、漢詩文集などの古典的類聚編纂物の編纂原理を分析するとともに、画像・映像などのコンテンツ分類への応用をはかる。

1.2.研究の特色

古来より、情報処理技術は、類書・辞書・事典などの形で、紙上で具現化されてきた。古典的な辞書・辞典(=類書)は、古典学研究において枢要な意義をもつものとしての認識は十分にあったが、その量の膨大さ故、これまで十分な評価もなされず、研究資料の提供さえも十分になされてこなかった。その意味で、本研究で扱う資料は、情報学というツールを得て初めて、その真価を発揮し得るものといえる。

また、本研究は古典的オントロジの集約と、配列の分析機能を付加したデータベースシステムの構築という2点で、知識ベースの構築、検索技術の高度化において情報学的に新見地を拓くものである。また、オントロジ分析の観点からは、諸典籍の構成を分析するために最適なレファレンスツールを提供するものとなっており、人文学研究分野において、漢字文化圏の知識の総体と基盤を具体的に分析・考察する形で提示される成果物は、人文情報科学研究の一つの典型を示すことにもなる。

1.3.国内外の関連する研究の中における本研究の位置づけ

21 世紀に入り、情報爆発の状況にいかに対処するかということが、緊急的課題として取りあげられ始めてい

る。その中で、熟成され、比較的傷を受けることの少なかった既存の知識大系すなわち分類概念の階層関係と語彙に着目し、それを集積・統合するという手法が、近年、知識工学・認知科学・図書館情報学など様々な分野で注目を浴びつつある。本研究もそれらの取り組みの一環として位置づけられる。

このことは、人文科学研究分野においても当然意識されるべき主題ではあるものの、残念ながら日本・中国における古典籍に取材する情報資源を整える当該分野における取り組みには類例がない。よって、本研究成果が今後の人文科学分野の研究進展に寄与するを願ってやまない。

2.研究組織(平成15年度)

2.1.研究代表者

相田 満(国文学研究資料館・研究情報部助手)

2.2.研究分担・協力者

坪 美奈子(日本大学・非常勤講師)
安保 博史(群馬県立女子大学・文学部助教授)
入口 敦志(国文学研究資料館・研究情報部助手)
江戸 英雄(国文学研究資料館・研究情報部助手)
蔵中しのぶ(大東文化大学・外国語学部教授)
後藤 幸良(相模女子大学・短期大学部)
佐伯 雅子(人間総合科学大学・人間科学部助教授)
塩野 友佳(筑波大学・大学院生)
谷本 玲大(相模女子大学・非常勤講師)
朽尾 武(成城大学・文学部・教授)
長崎 健(中央大学・文学部教授)
中島 和歌子(北海道教育大学・社会言語教育系助教授)
根木 優(成城大学・大学院生)[研究補助]
林 毅(バリアフリー・取締役社長)
原 正一郎(国文学研究資料館・研究情報部助教授)
三田 明弘(台湾:大葉大学助理教授)
山田 直子(国文学研究資料館・整理閲覧部助手)
渡辺 信和(同朋学園・仏教文化研究所研究室長)

3.研究課題

研究課題名:和漢古典学のオントロジモデルの構築

研究種目:基盤研究(B)

研究領域:

- ①分野:総合領域
- ②分科:情報学
- ③細目:情報図書館学・人文社会情報学

研究期間：2003年度(平成15年度)～2006年度(平成17年度)

課題番号：15300082

2003年度研究費：330万円

4.従来の研究経過および準備状況

本研究に密接に関連する研究課題で、研究代表者および分担者が受けた研究費は下記の通り。

①研究費名称：文部省科学研究費補助金・萌芽的研究／期間(年度)：(平成10-12年度)

研究課題名：「和漢古典分類語彙の階層化に関する基礎的研究」(No.10871058)

研究代表者氏名：相田満／研究経費：1,900千円

研究経過・研究成果：本申請の基礎研究にあたる。これまで必ずしも等質なものとして意識されたことのなかった類書・辞書・韻文・散文集、史書・儀礼書および注釈書等を類聚編纂物という概念によって統合して捉え直すための理論的裏付けと補強を行った。また、階層化データをデータベース化するために、和漢東西不問で枢要と思われる典籍の分類用語彙群を入力し、データベースマスタを作成した。今回申請のプロジェクトでは、この研究で蓄積されたデータを基礎データとして搭載するとともに、さらにデータの蓄積・研究内容の深化をはかりたい。

②研究費名称：文部省科学研究費補助金・基盤研究B(展開)／期間(年度)：(平成10-12年度)

研究課題名：「日本型歴史的人物系図情報データベースシステム構築の研究」(No.10551013)

研究代表者氏名：相田満／研究経費：12,800千円

研究経過・研究成果：『尊卑分脈』・『本朝皇胤紹運録』・『群書諸家系図』・『諸家知譜拙記』のデータを原本資料を使用してデータベース化を行った。XMLとVRMLを採用して検索でヒットした人物の関係情報を傍線でつなぎツリー型に、しかもインタラクティブに表示させるだけでなく、その人物に関する諸情報を分類し、原資料の記述を損なうことなく検索することができるデータベースシステムを開発した。Web上からの利用により約12万件のデータ検索が行える(現在館内試験公開中)。本研究により、多階層でかつ大量の情報を電子的に管理を行うための基本システムが完成し、今回申請を行う研究への道筋が拓けた。

③研究費名称：文部省科学研究費補助金・基盤研究A(一般)／期間(年度)：(平成8-11年度)

研究課題名：「知識支援型汎用検索デバイスエンジン・データの構築」(No.08401016)

研究代表者氏名：相田満／研究経費：20,800千円

研究経過・研究成果：現在通行のソーラス検索システムとは異なり、あるタームから異なる発想を引き出すための連想語・関連概念語を導き出せる可搬的な検索システムの開発とデータベースを構築した。具体的には、国文学研究に関連する現在通行の辞書・事典の見出し語と解説中の語句を、それぞれの関連語として集積し、それを利用した連想検索システムの開発を行ったものである。UNICODE日本語面約20,000字の文字セット内の異体字ソーラスを搭載し、異体字検索にも対応する。本研究で申請する、概念語の関係をあるがままに集積するという発想は、この研究が起点になる。

④研究費名称：文部科学省科学研究費補助金・萌芽研究／期間(年度)：(平成13-15年度)

研究課題名：「和漢古典籍における「標題文芸」の基礎的研究」(No.14651078)

研究代表者氏名：相田満／研究経費：3,000千円

研究経過・研究成果：書名・目次における「標題」をめぐるさまざまな意匠と継承関係を分析し、その

意義を文学的に評価することを試みるもの。広義には、オントロジをめぐる諸現象の一端を深化させた研究に位置づけられる。

⑤研究費名称:文部省科学研究費補助金・基盤研究C/期間(年度):(平成12-13年度)

研究課題名:「国文学データを統合利用するためのモデル論的研究」(No.1251006)

研究代表者氏名:原正一郎/研究経費:3,300千円

古典籍・論文目録・マイクロ画像を横断的に検索するための書誌情報の統合、および検索支援データの構築についての基礎的研究。あわせてダブリンコアによる統合検索システムの開発を試みた。

⑥研究費名称/期間(年度):文部省科学研究費一般研究D(昭和51年度)

研究課題名:「類書の史的 연구」

研究代表者氏名:朽尾武

研究経過・研究成果:先秦～元代までの中国に編纂された類書史についての研究で、その成果は、「類書の研究序説(1)～(3)」(「成城国文学論集」10,11,12/1978.3,1978.10,1979.3)にて報告した。本報告以後、日本の古典学との関係において重要な位置をしめる中国類書について通観した考察は著されていない。その意味で、本研究にて構想するデータベースを利用しての研究の深化によって、格段の成果が期待される。

以上、これまでの研究で蓄積された基礎的データをもとに、システムの運用をはかりたい。また、和漢の古典学におけるオントロジの意義と研究手法の提示については、人文学系研究者によるシステム評価と、データ蓄積、各データセットの考察を通じて、各分担者が研究を深めるだけでなく、広くプロジェクト協力者データセットの提供を求め、コラボレーションによるデータ蓄積をはかることにより、当該研究者の必須レファレンスツールの存在となるデータベースを立ち上げたい。

5.研究第1年度の研究計画

平成15年度の研究内容は、これまでのプロジェクトと研究資源を活用して、今回のプロジェクトに適用可能な範囲を洗い出す。今回の研究プロジェクトにおける検討項目は、以下の通りである。

①構成・配列情報の比較結果を出力・明示する方法の検討

各データセット(典籍)におけるオントロジの構成・配列を比較して、その差異を明示することにより、それぞれの類縁関係を分析するための方法について検討する。

②データ実装(加工)方法とデータフォーマットの策定

データの実装は、事前に検索用と表示用のメタデータを生成し、その双方を実装・公開する手法の採用を予定する。その際の、汎用的な利用に供するためのフォーマットなどについて検討する。

③文字セットの問題

ほぼ100%で漢字文字列が使用されるデータの性質上、文字セットは大規模な文字コードセットが必要となる。そのため、運用に際しては、UCS(ユニコード)に対応したシステムを採用する。異体字の扱いは、相田の作成したUCS2.0日本語面約20,000字の文字セット内の漢字シソーラスを使用する。ただし、UCSはすでにバージョンが3.1に上がり、収録文字セットも約27,000字に拡張された状況に対応するため、シソーラスの拡張を行う。

④各オントロジ取材源となった典籍の評価

各オントロジ取材源となる典籍の意義について考察し、各資料の解題を作成するとともに、新規データの

採集作業を行う。

各担当と分野は、以下の通り。

漢籍類書(朽尾)・和製類書(相田・中島)・和製辞書(入口)・仏典(蔵中)・漢詩集(佐伯)

和歌集(江戸)・類書型源氏物語注釈書(後藤)・書籍目録類(山田)

⑤研究支援システムの開発

研究初年度は、データ構築を支援するためのテキスト処理プログラムおよび分析支援ツールを開発する。

6.研究活動の記録

6.1.共同研究会

◎第1回 共同研究会

【開催日】平成15年7月2日(水)

【場所】国文学研究資料館大会議室

【参加者】相田・江戸・蔵中・後藤・佐伯・塩野・朽尾・長崎・中島・中村・根木・林・原・山田・渡辺

【協議事項】

- ①研究プロジェクトの趣旨説明・メンバー紹介・経過報告
- ②研究発表・討議
 - ネットワーク統合型データベースによる東アジア資料の共有化に関する研究(報告[原])
 - 「標題学」方法序説(報告[相田])
 - 和漢古典学におけるオントロジ・モデルの構想(報告[相田])

◎第2回 共同研究会

【開催日】平成15年11月16日(日)

【場所】同朋大学知成館

【参加者】相田・坏・安保・江戸・蔵中・佐伯・谷本・朽尾・長崎・中島・根木・原・山田・渡辺

- ①開会の挨拶(相田)
- ②事務事項確認
- ③参加者自己紹介
- ④研究発表・討議
 - ネットワーク統合型データベースによる東アジア資料の共有化について(原)
 - 『山海経』と猿について(朽尾)
 - 国立中央図書館台湾分館の和書調査報告―典拠の参照を中心に―(山田)
 - 仏典における和漢古典学のオントロジモデルの構築―唐・長安西明寺の学僧の類聚的編纂物のもつ有機的関係から―(蔵中)
 - 『田舎』冠称型標題の一視点(安保)
 - 近世初期の笑話本と標題文芸(入口)
 - 『本朝麗藻』の標題文芸(佐伯)
 - 標題文芸―『枕草子』をめぐって―(坏)
 - 同朋大学仏教文化研究所寺院襲蔵物データベース―相田・渡辺による共同作成―デモンストレーション(渡辺)

◎第3回 共同研究会

【開催日】平成16年2月16日(月)

【場所】国文学研究資料館

【参加者】相田・坪・安保・入口・蔵中・後藤・佐伯・谷本・朽尾・中島・根木・三田・山田・渡辺

①開会の挨拶(相田)

②事務事項確認

③参加者自己紹介

④研究発表・討議

○データ形成支援プログラム(仮称)ーデモンストレーション(相田)

○改題の事例報告〔①『西鶴諸国はなし』 ②『本朝水滸伝』〕(安保)

○長安西明寺の類聚編纂書のオントロジとその受容ー『諸経要集』述意部・『法苑珠林』述意部と『三宝絵』上巻の構成ー(蔵中)

○類聚性のある源氏物語の注釈書(後藤)

○『本朝麗藻』の分類意識(佐伯)

○『枕草子』初段と和漢の類書的首巻部類標題の関係について(中島)

6.2.成果発表(口頭発表)

○相田 満「「標題」学方法序説」

第11回(2003 三年度)情報知識学会研究報告会[2003 年5月24日、於:国立情報学研究所]

○相田 満「和漢古典学のオントロジの可能性について」

特定領域研究『東アジアの出版文化』G班研究集会[2003 年11月2日、於:学士会館]

○相田 満「古典的オントロジ資源の可能性ー和漢古典学のオントロジー」

人文科学とコンピュータシンポジウム 2003 『デジタルアーカイブー情報資源の活用と共有の深化をめざしてー』[情報処理学会・人文科学とコンピュータ研究会主催、2003 年12月18日発表、於:国立歴史民俗博物館]

○相田 満「古事類苑のデータベース化」

総合研究大学院大学共同研究プロジェクト「文化科学研究分野における情報資源共有化のためのコラボレーション研究」第1回研究集会[2004 年1月19日、於:総合研究大学院大学講義室]

○相田 満「目録分類における「文学」項目とその周辺」

国文学研究資料館共同研究会[(2004 年1月30日)、於:国文学研究資料館中会議室]

○相田 満:「古典的オントロジ資源の形成」

東洋学へのコンピュータ利用第15回研究セミナー(京都大学学術情報メディアセンター第75回研究セミナーと併催)[主催:京都大学人文科学研究所附属漢字情報研究センター・京都大学学術情報メディアセンター、2003 年3月28日、於:京都大学学術情報メディアセンター北館3F 講習室]