

*Journal of Japan Society of Information and Knowledge*

# 情報知識学会誌

Vol.17 No. 2 (Feb. 2007)

~~~~~目 次~~~~~

## 特集 第15回(2007年度)年次大会(研究報告会&総会)

|                                          |                                         |
|------------------------------------------|-----------------------------------------|
| プログラム                                    | 59                                      |
| レファレンスデータに対する NDC の自動付与                  | 原田隆史, 江藤正己, 大西美奈子 … 61                  |
| 引用論文における引用箇所間の近さをとらえる尺度                  | 江藤正己 … 65                               |
| SRU/SRW を用いた教育図書館資料の書誌検索システムの構築          | 江草由佳, 高久雅生 … 69                         |
| 著者とキーワードの関連性に着目した研究領域ブラウジングシステムの試作       | レボウィッツ紀子, 松村敦, 宇陀則彦 … 75                |
| 営みの Lattice の構造化とアブダクション                 | 福永征夫 … 81                               |
| 音楽著作権についての歴史的研究 ~革命期から 19 世紀のフランスを中心に~   | 石井大輔 … 85                               |
| 科学技術文献の解析・可視化における辞書の活用について               | 甲田彰, 坂内悟 … 91                           |
| ユーザの Web 閲覧履歴を用いた検索支援システム                | 堀幸雄, 今井慈朗, 中山堯 … 95                     |
| 古典籍説明文からのドメイン知識の抽出                       | 長塚隆, 神門典子 … 101                         |
| 人文研究を支援するデータベースシステム - 聖教検索および系図表示 -      | 朴明哲, 森本雅史, 立花純児, 村川猛彦, 宇都宮啓吾, 中川優 … 105 |
| 科学研究費補助金による研究助成の効果に関する調査                 | 柿沼澄男, 西澤正己, 孫媛, 根岸正光 … 111              |
| キーワード分析による科研費におけるゲノムおよびナノテクノロジー関連研究の動向調査 | 西澤正己, 孫媛 … 117                          |
| 日本発行の科学技術雑誌数の分析                          | 時実象一 … 123                              |
| 文化資源オントロジの構築とその活用                        | 研谷紀夫, 馬場章 … 129                         |
| ターミノロジー論における階層関係の概念について                  | 山本昭 … 135                               |
| シンポジウム「Web2.0時代の情報システム」                  | 141                                     |
| 論文賞の投票方式の変更について                          | 142                                     |
| 情報知識学会誌 投稿規定                             | 143                                     |
| 事務局からのお知らせ                               | 146                                     |



情報知識学会

<http://www.jsik.jp/>

## 第15回(2007年度) 年次大会 (研究報告会 & 総会) プログラム

日程 : 2007年5月25日(金) ~ 26日(土)

会場 : 国文学研究資料館 (東京都品川区豊町 1-16-10)

### 【 2007年5月25日(金) 】

12:30 受付開始  
 13:00 ~ 13:10 開会挨拶 情報知識学会会長 細野公男  
 13:10 ~ 13:50 総会  
 13:50 ~ 14:00 休憩

#### 14:00~15:40 一般発表(1) 座長 : 中川 優(和歌山大学)

- 1 レファレンスデータに対する NDC の自動付与  
 ○原田隆史(慶應義塾大学), 江藤正己(慶應義塾大学大学院),  
 大西美奈子(NRI ネットワークコミュニケーションズ)
2. 引用論文における引用箇所間の近さをとらえる尺度  
 江藤正己(慶應義塾大学大学院)
- 3 SRU/SRW を用いた教育図書館資料の書誌検索システムの構築  
 ○江草由佳(国立教育政策研究所),  
 高久雅生(情報・システム研究機構 新領域融合研究センター)
4. 著者とキーワードの関連性に着目した研究領域ブラウジングシステムの試作  
 ○レボウィッツ紀子(筑波大学大学院), 松村敦, 宇陀則彦(筑波大学)

15:40 ~ 15:50 休憩

#### 15:50~17:30 シンポジウム 座長 : 宇陀則彦(筑波大学)

テーマ : Web2.0 時代の情報システム  
 パネリスト : 岡本 真(Academic Resource Guide)  
 兼宗 進(一橋大学情報処理センター)

18:30~ 懇親会(予定)

## 【 2007年5月26日(土) 】

**9:20~10:35 一般発表(2) 座長：細野公男(情報知識学会長)**

5. 営みの Lattice の構造化とアブダクション  
福永征夫(アブダクション研究会)
6. 音楽著作権についての歴史的研究 -革命期から 19 世紀のフランスを中心に-  
石井大輔(筑波大学大学院)
7. 科学技術文献の解析・可視化における辞書の活用について  
○甲田彰, 坂内悟(科学技術振興機構)

10:35 ~ 10:50 休憩

**10:50~12:05 一般発表(3) 座長：江草由佳(国立教育政策研究所)**

8. ユーザの Web 閲覧履歴を用いた検索支援システム  
○堀幸雄(香川大学), 今井慈朗(香川大学), 中山堯(神奈川大学)
9. 古典籍説明文からのドメイン知識の抽出  
○長塚隆(鶴見大学), 神門典子(国立情報学研究所)
10. 人文研究を支援するデータベースシステム -聖教検索および系図表示-  
○朴明哲(和歌山大学大学院), 森本雅史, 立花純児, 村川猛彦(和歌山大学), 宇都宮啓吾(大阪大谷大学), 中川優(和歌山大学)

12:05 ~ 13:30 昼休み

**13:30~14:10 論文賞表彰式 および 論文賞記念講演**

14:10 ~ 14:15 休憩

**14:15~15:05 一般発表(4) 座長：長塚 隆(鶴見大学)**

11. 科学研究費補助金による研究助成の効果に関する調査 -研究キャリアステージにおける科学研究費補助金取得パターンの分析-  
○柿沼澄男, 西澤正己, 孫媛, 根岸正光(国立情報学研究所)
12. キーワード分析による科研費におけるゲノムおよびナノテクノロジー関連研究の動向調査  
○西澤正己, 孫媛(国立情報学研究所)

15:05 ~ 15:20 休憩

**15:20~16:35 一般発表(5) 座長：芦野俊宏(東洋大学)**

13. 日本発行の科学技術雑誌数の分析  
時実象一(愛知大学)
14. 文化資源オントロジの構築とその活用  
研谷紀夫(東京大学大学院), 馬場章(東京大学)
15. ターミノロジー論における階層関係の概念について  
山本昭(愛知大学)

16:35 ~ 16:40 閉会挨拶 研究大会実行委員長 原田隆史

# レファレンスデータに対する NDC の自動付与 Automatically Classifying the Reference Records into NDC

○原田隆史<sup>1</sup>, 江藤正己<sup>2</sup>, 大西美奈子<sup>1,3</sup>

Takashi HARADA, Masaki ETO and Minako ONISHI

1: 慶應義塾大学文学部図書館・情報学専攻

School of Library and Information Science, Keio University

〒108-8345 東京都港区三田 2-15-45

Email: ushi@slis.keio.ac.jp

2: 慶應義塾大学大学院文学研究科図書館・情報学専攻

Graduate School of Library and Information Science, Keio University

Email: eto@slis.keio.ac.jp

3: 現在の所属は, NRI ネットワークコミュニケーションズ

本研究では, 国立国会図書館(NDL)のレファレンス協同データベースに収録されたレファレンス記録に対する NDC の自動付与実験を行った。手法としては決定木法, ナイーブベイズ法, SVM(Support Vector Machine)法の3つの機械学習法を用いた。5703 件のレファレンス記録中の質問文および回答文それぞれに含まれる単語を対象に学習させ, 634 件を対象に分類したところ, 質問文を対象に手法としては SVM 法を用いた場合にわずかに正しく付与される傾向が見られた。ただし, 再現率, 精度ともに対象や手法による差はそれほど大きくなかった。また, 分野カテゴリごとの再現率を調べたところ, 分野によって再現率に大きな差があった。分野によっては 60% を超える再現率で分類できる場合もあり, 機械学習法でレファレンス記録に NDC を付与することが有効であることが明らかになった。

In this research, we conducted the experiment of automatically classifying the reference records in the Collaborative Reference Database of the National Diet Library into Nippon Decimal Classification (NDC). Three machine learning methods: decision tree, Naïve Bayes, and Support Vector Machine (SVM) methods – were tested on the words included in the sentences of questions and answers in 634 reference records. The results indicate that the using the SMV method and answer sentences were classified successfully compared to others. As for the recall and precision ratio, there was small difference among the methods, and between question and answer sentences. The significant difference, however, was observed in the recall ratio by the NDC category; the result of some category obtained more than 60% recall ratio, which suggested that classifying reference records into NDC by using machine learning methods was effective.

## 1. レファレンス協同データベース

全国の公共図書館, 大学図書館, 専門図書館等における, レファレンス事例, 調べ方マニュアルなどのデータを蓄積し, インターネットを

通じて提供するレファレンス協同データベースが国立国会図書館によって平成 17 年 4 月 1 日より事業化された<sup>1)</sup>。レファレンス協同データベースには, 平成 19 年 3 月 31 日現在で公

共図書館，大学図書館，専門図書館など計 442 館が参加している。レファレンス事例の公開件数は平成 18 年 6 月には 1 万件を突破し，現在も毎日増加し続けている。

データベースに登録される項目は，第 1 表に示す 18 項目である。このうち，質問，公開レベル，管理番号，回答の 4 項目のみが必須入力項目であり，残り 14 項目は任意入力項目である。第 1 表に各項目に対する付与率を示す。

第 1 表に見られるように，寄与者，照会先，回答プロセスの登録率は極めて低く，それ以外にも全般に付与されていない項目が多い。NDC 項目は 12,062 件中 4,461 件が未登録である。また，NDC 項目が付与されているデータのうち約 17%については NDC の 1 桁目だけ(類項目のみ)が付与されている。

NDC はレファレンスデータを探す際にも有効であり，付与されていないデータにも後で自動的に付与することが期待される。しかし，これまで図書館に NDC 番号を自動付与する研究はいくつか行われてきたが<sup>2)-4)</sup>，レファレンスデータを対象に NDC を付与する研究はほとんど行われていない。そこで，本研究ではレファレンスデータに対して NDC を自動付与する実験を行った。

## 2. レファレンスデータに対する NDC の自動付与実験

### 2.1 実験データ

実験対象データとしては，2006 年 7 月 7 日に NDL レファレンス協同データベースで一般公開されていたレファレンスデータのうち，NDC が 2 桁以上(少なくとも綱項目まで)付与されていた 6,337 件を自動分類のための学習用集合および分類用集合として用いた。その際，レファレンスデータ中の質問文と回答文中に含まれる語それぞれを使用した場合効果の違いを検討するため，分類に用いる語は，質問項目と回答項目から抽出したものをそれぞれ用いて分類実験を行い比較した。

第 1 表 レファレンスデータベース入力項目と付与率

| 項目名    | 項目内容                         | 付与率    |
|--------|------------------------------|--------|
| 質問     | レファレンス質問の内容                  | 必須     |
| 公開レベル  | データ公開レベル                     | 必須     |
| 管理番号   | 各参加館が独自の事例管理番号               | 必須     |
| 回答     | 利用者に対して回答した内容                | 必須     |
| 事例作成日  | 事例データを作成した日付                 | 100.0% |
| 解決未解決  | 解決済か未解決か                     | 100.0% |
| キーワード  | 事例の中心的内容、主要概念の語              | 66.9%  |
| NDC の版 | 使用した『日本十進分類法』の版              | 66.3%  |
| NDC    | 日本十進分類法の分類番号                 | 63.0%  |
| 調査種別   | 調査の種類                        | 25.9%  |
| 内容種別   | レファレンス事例のジャンル                | 28.2%  |
| 参考資料   | 回答で参考にした資料                   | 28.8%  |
| 回答プロセス | 回答のために経た調査プロセス               | 3.7%   |
| 照会先    | 質問者に示した照会先機関情報               | 2.1%   |
| 事前調査事項 | 質問者が事前に調べていた情報               | 3.4%   |
| 備考     | 自由記入欄(補足事項等)                 | 8.0%   |
| 質問者区分  | 質問者の分類                       | 14.9%  |
| 寄与者    | 事例作成に際して情報提供等をした図書館外部の人および機関 | 0.6%   |

### 2.2 自動付与実験の流れ

本研究では機械学習による自動分類を行った。機械学習に用いるデータマイニングソフトとしては，Weka Ver. 3.4.8 を使用した<sup>6)</sup>。Weka はニュージーランドのワイカト(Waikato)大学を中心に開発されたフリーのデータマイニング用パッケージソフトである。機械学習アルゴリズムは，決定木，Support Vector Machine(以下 SVM と略す) ナイーブベイズという 3 種類を用い，それぞれの結果を比較した。

自動分類は，学習フェーズと分類フェーズの 2 つのフェーズに分けられる。学習フェーズでは，学習用集合から手がかりとなる語を切り出し，切り出した語を用いて分類ルールを作成する。また分類フェーズでは，学習用集合とは異

なる未分類のデータから切り出した語を元に分類を行う。

本研究では学習用の集合として、レファレンス協同データベースに収録されていたデータのうち、NDCが2桁以上付与されていた6,337件の9割にあたる5,703件から名詞または未知語を切り出して用いた。語の切り出しには「MeCab」を用いた<sup>5)</sup>。また、分類フェーズにおいては残り1割(634件)のデータを対象に分類を行い、その結果を人間が付与したNDC番号と比較した。なお、NDCの付与に際しては、NDCの1桁目(類項目)を自動分類する実験と、類項目が付与されたデータを元にして2桁目(綱項目)を付与する実験の2種類の実験を行った。

分類結果の評価に際しては、再現率(Recall)、精度(Precision)の両者を求めて分析を行った。アクセスポイントの補完という観点から考えた場合、特に再現率は重要であろう。

### 2.3 類項目の自動分類結果

質問項目と回答項目中の語を分類キーとして用いた分類の再現率と精度を第2表に示す。なお、回答項目を対象にSVMを用いた結果については計算時間が長すぎるために現時点では結果が得られていない。

第2表 類項目への分類結果

|        |     | バイズ   | 決定木   | SVM   |
|--------|-----|-------|-------|-------|
| 質<br>問 | 再現率 | 38.4% | 25.4% | 38.4% |
|        | 精度  | 38.5% | 34.9% | 46.5% |
| 回<br>答 | 再現率 | 35.6% | 34.0% | -     |
|        | 精度  | 39.9% | 40.5% | -     |

類項目への分類結果としては、第2表に見られるように、精度および再現率のいずれにおいてもSVMを用いた場合に高い値が得られた。

また、質問項目と回答項目の比較では、精度についてはナイーブバイズ、決定木ともに回答

項目を用いた時の方が高い値を示したが、再現率においてはナイーブバイズでは質問項目を用いた時の方が高かった。

さらに、質問文を対象としてSVMを用いた場合の各類ごとの再現率を第3表に示す。第3表に見られるように、33.8%から67.4%まで類項目によって再現率に大きな違いが見られた。

第3表 SVMを用いた時のNDC類項目ごとの再現率

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| 0類    | 1類    | 2類    | 3類    | 4類    |
| 35.4% | 33.8% | 43.5% | 42.7% | 53.1% |
| 5類    | 6類    | 7類    | 8類    | 9類    |
| 42.3% | 52.0% | 61.2% | 44.7% | 67.4% |

### 2.4 類項目が付与されたデータに対する綱項目の自動分類結果

類項目までNDCが付与されているデータに対し、綱項目の分類を行った実験の結果について各類ごとの精度を第4表に、また再現率を第5表に示す。ここで、最も値が高いものを網かけて示す。

第4表 分類結果の精度

| NDC | バイズ   |       | 決定木   |       | SVM   |       |
|-----|-------|-------|-------|-------|-------|-------|
|     | 質問    | 回答    | 質問    | 回答    | 質問    | 回答    |
| 全体  | 50.1% | 40.1% | 51.0% | 52.5% | 54.7% | 50.6% |
| 0類  | 52.1% | 35.2% | 48.7% | 44.1% | 55.1% | 47.9% |
| 1類  | 47.4% | 36.3% | 40.2% | 47.0% | 47.8% | 43.4% |
| 2類  | 53.6% | 32.6% | 54.1% | 48.3% | 60.6% | 51.9% |
| 3類  | 47.0% | 41.1% | 45.5% | 53.7% | 50.1% | 46.9% |
| 4類  | 45.6% | 36.7% | 50.0% | 53.2% | 49.3% | 49.3% |
| 5類  | 34.2% | 37.2% | 31.7% | 38.2% | 37.0% | 34.6% |
| 6類  | 51.1% | 51.5% | 48.7% | 51.5% | 55.4% | 55.6% |
| 7類  | 50.1% | 42.8% | 49.6% | 50.8% | 56.0% | 48.0% |
| 8類  | 56.3% | 46.2% | 71.9% | 66.9% | 66.3% | 60.0% |
| 9類  | 63.5% | 41.5% | 69.1% | 71.6% | 69.4% | 68.0% |

第4表に見られるように、最も値が高い対象や手法は異なるが、回答を対象に決定木を使用

する場合と質問中を対象に SVM を使用する場合に最も精度が高いものが多かった。

質問と回答全体の比較では、ナイーブベイズを用いた場合には質問中の語を用いた場合と回答中の語を用いた場合で大きな差がみられる類項目もあったが、それ以外の手法では対象の差による精度の大きな差は見られなかった。

第5表. 分類結果の再現率

| NDC | ベイズ   |       | 決定木   |       | SVM   |       |
|-----|-------|-------|-------|-------|-------|-------|
|     | 質問    | 回答    | 質問    | 回答    | 質問    | 回答    |
| 全体  | 26.9% | 26.3% | 25.6% | 27.1% | 27.6% | 26.9% |
| 0類  | 30.6% | 23.4% | 21.8% | 23.2% | 24.9% | 26.2% |
| 1類  | 24.4% | 25.0% | 20.8% | 24.4% | 24.2% | 21.2% |
| 2類  | 20.8% | 16.0% | 18.8% | 16.1% | 22.9% | 18.4% |
| 3類  | 37.5% | 35.8% | 36.5% | 44.3% | 39.7% | 36.9% |
| 4類  | 27.1% | 28.8% | 20.2% | 30.6% | 25.1% | 25.2% |
| 5類  | 24.1% | 33.0% | 27.8% | 32.7% | 29.3% | 28.7% |
| 6類  | 26.5% | 29.8% | 30.6% | 37.1% | 31.5% | 41.6% |
| 7類  | 37.0% | 34.1% | 41.6% | 30.4% | 40.5% | 37.0% |
| 8類  | 23.2% | 17.5% | 15.1% | 14.1% | 16.5% | 13.6% |
| 9類  | 17.5% | 19.8% | 22.8% | 18.2% | 21.2% | 20.6% |

再現率については、第5表に示すように、類によってどのアルゴリズムを用いた場合に再現率が高いかは異なり、また、質問項目と回答項目による再現率の差もあまり見られなかった。

### 3. レファレンスデータに対する NDC の自動付与

本研究では、ナイーブベイズ、決定木、SVM の3つの機械学習を用いてレファレンスデータという図書とは異なるデータに NDC を付与する実験を行った。

機械学習アルゴリズムの比較では、SVM を用いた場合が精度、再現率ともわずかに高い結果が得られた。ただし、他の手法との差はわずかであり、また SVM を用いた場合には他のアルゴ

リズムを用いた場合と比較して計算時間が長い。実用システムの開発を考える場合には計算時間の短縮も考慮する必要があるだろう。

また、対象中に出現する適切でない語が分類結果そのものに与える悪影響についても検討が必要である。たとえば、本実験において誤って分類されたレファレンスデータ中には「資料」や「意味」という単語が質問、回答項目どちらについても多く出現していた。これらの語は抽象的な内容を示す語であり、分類に寄与しない可能性が高い。今後は、誤って分類された結果の詳細な分析を行う必要があるだろう。

さらに、分類の手がかりとする語の選択について、質問文および回答文から抽出された語の適切な組み合わせ法を検討するなど、より効果的な分類法を検討していく予定である。

## 4. 引用文献

- 1) NDL レファレンス協同データベース.  
<<http://crd.ndl.go.jp/jp/public/use.html>>  
[2006-5-17]
- 2) 石田栄美. 宮田洋輔. 神門典子. 上田修一.  
目次と帯を用いた図書の自動分類.  
<<http://www.slis.keio.ac.jp/~ueda/paper/fi-82-12.pdf>>[2006-5-20]
- 3) 石田栄美. 図書を NDC カテゴリに自動分類する試み. Library and Information Science. No. 39, 1998, p. 32-45
- 4) 畑田稔. ニューラルネットワークによる図書の自動分類. 情報処理学会第 57 回全国大会講演論文集, No. 2, p. 360-361(1998)
- 5) MeCab: Yet Another Part-of-Speech and Morphological Analyzer MeCab.  
<<http://mecab.sourceforge.jp/>>  
[2006-06-29]
- 6) The University of Waikato. Weka.  
<<http://www.cs.waikato.ac.nz/ml/weka/index.html>>[2006-06-29]

# 引用論文における引用箇所間の近さをとらえる尺度 Measuring Distance between Cited Papers in a Citing Paper

江藤正己  
Masaki ETO

慶應義塾大学大学院文学研究科

Graduate School of Library and Information Science, Keio University

E-mail:eto@slis.keio.ac.jp

従来の共引用では、一つの引用論文における被引用論文間の類似度は全て同じであることを前提としている。しかし、引用論文における引用箇所間の意味的な近さをとらえることで被引用論文間の類似性の強弱を推測できると考えられる。そこで、近さをとらえる尺度として、「物理的距離（文字数）」、「（引用箇所の）周辺語間の類似度」、「論文構成からみた距離」を設定し、尺度としての適切さを検討する。

Co-citation, which is a typical similarity indicator among papers, has a premise that the degrees of similarity among cited papers are equal in one citing paper. But the similarity strength can be guessed from the distance between the places where cited papers are shown in a citing paper. In this paper, three distance measures (“physical distance”, “co-occurrence of citing words”, and “structural distance”) are introduced for the purpose and these measures are evaluated.

キーワード: 共引用, 論文検索, 類似検索

co-citation, document retrieval, similarity search

## 1 はじめに

### 1.1 共引用とは

論文間の類似度を算出する代表的な指標として共引用<sup>[1]</sup>がある。この指標では対象となる二つの論文間の類似度は、一般的に算出対象の論文のペアが同一の論文から共に引用される回数によって求められる。

ただし従来の共引用では、本文における引用のされ方の違いが考慮されない。つまり一つの引用論文においては、どの被引用論文間の類似度も全て同じことが前提になっている。しかし、そのような違いを考慮し被引用論文間の類似性の強弱を推測できれ

ば、共引用の類似度が精密になり、類似論文検索の性能を向上させることができる。

被引用論文間の類似性の強弱を推測するために引用論文本文を解析する方法として、引用箇所間の意味的な近さをとらえることが挙げられる。引用論文の本文において引用箇所間が意味的に近ければ被引用論文間の類似性が強く、引用箇所間が意味的に遠ければ被引用論文間の類似性が弱いと判断できるためである。実際筆者のこれまでの研究により、引用箇所間の意味的な近さを利用することで、被引用論文間の類似性の強弱を推測できる示唆を得ている<sup>[2]</sup>。

## 1.2 目的

引用箇所間の意味的な近さをとらえる場合、その尺度としては様々なものが考えられる。尺度が引用箇所間の意味的な近さを適切にとらえていれば、尺度の値と被引用論文間の類似度の間には相関がみられるはずである。そこで、本稿では引用箇所間の意味的な近さをとらえる尺度を複数設定し、相関を調べることによってそれぞれの尺度の適切さを検討する。

## 2 近さをとらえるための尺度

引用箇所間の意味的な近さをとらえるために、以下の三つの尺度を設定した。

### 2.1 物理的距離

二つの引用箇所が物理的に短い距離内であれば、両者は意味的に近いといえよう。そこで、引用箇所間の文字数を距離とする尺度を設定する。その際、論文の長さに対する補正をおこなうため、「文字数」を「論文の総文字数」で除算する正規化処理をおこなった。

### 2.2 周辺語間の類似度

引用論文中の引用箇所の前後  $N$  語を周辺語とする。測定対象の二つの引用箇所の周辺語に同じ語が多く含まれていれば、その引用箇所間は意味的に近いと考えられる。そこで、周辺語間の類似度を語の共出現数によって求め、それを距離とする尺度を設定する。具体的には、出現頻度で重み付けをおこない、コサイン係数によって類似度を求める。

### 2.3 論文構成からみた距離

論文には段落や文などの構造の単位がある。二つの引用箇所が同一の構造単位内であれば、引用箇所間は意味的に近いと考え

られる（たとえば、二つの引用箇所が同一段落内にある）。そして、同一段落よりも、同一文内にある方が意味的に近い（小さな構造単位内にあるほど意味的に近い）とみなすことができる。そこで論文の構造の単位を用いた尺度を設定する。本稿では、構造単位として「段落」「文」「列挙」の三つを用いる。「列挙」とは、[AA99,BB98,CC95]のような一つの引用箇所において複数の論文を並列列挙する形式の引用を指す。「論文構成からみた距離」の尺度は「列挙」「同一文」「同一段落」「非同一段落」の4種類から成る。

## 3 実験

引用論文を解析して2節で述べた尺度で引用箇所間の近さを測り、「各尺度の測定値」と「被引用論文間の類似度」との間の相関を調べる。両者の間に相関があれば、尺度として適切と判断できる。

### 3.1 実験データ

実験には、科学技術分野の主要なデータベースである *CiteSeer* が公開しているデータセット *CiteSeerMetadata* <sup>[3]</sup> を利用した。このデータセットには、論文の書誌情報、全文を入手するためのURL、引用関係の情報が含まれている。

データセットに含まれる論文のうち、タイトルかディスクリプタに「database」が含まれる論文を選び全文のダウンロードを試みた。ダウンロード出来た論文のうち、引用記号が機械処理しやすい（アルファベットと数字から構成される）ものを選び引用論文集合とした。引用論文集合は1,468件となった。

### 3.2 被引用論文間の類似度の算出指標

被引用論文間の類似度は、論文間の類似度に算出によく用いられる「tf\*idf/cosine」「正規化書誌結合」「正規化共引用」の三つの指標を使用した。

### 3.3 集計方法

集計は次の二通りの方法でおこなった。(集計方法 1) 測定した「引用箇所間の近さ」とその「被引用論文間の類似度」との間の相関を求める。(集計方法 2) 最初に「引用箇所間の近さ」と「被引用論文間の類似度」のそれぞれを引用論文毎に平均し、その平均値同士の相関を求める。

### 3.4 算出・集計結果

算出・集計に用いたデータ数は、表 1 となった。(被引用論文の全文を入手できないなどの理由により) 類似度の算出ができなかったものや二つ以上の引用箇所を引用論文から特定できなかったものは、対象から除外している。

表 1 算出・集計に用いたデータ数

|        | tf*idf/cosine | 正規化<br>書誌結合 | 正規化<br>共引用 |
|--------|---------------|-------------|------------|
| 集計方法 1 | 35,064        | 33,626      | 44,198     |
| 集計方法 2 | 1,014         | 973         | 1,055      |

#### 3.4.1 物理的距離

「物理的距離」と「被引用論文間の類似度」の相関係数を求めた。相関係数が  $-1$  に近づけば、引用箇所間の近さを適切にとらえていると判断される。結果を表 2 に示す。

表 2 物理的距離と被引用論文間の類似度との相関

|        | tf*idf/cosine | 正規化<br>書誌結合 | 正規化<br>共引用 |
|--------|---------------|-------------|------------|
| 集計方法 1 | -0.075        | -0.107      | -0.067     |
| 集計方法 2 | -0.124        | -0.127      | -0.105     |

#### 3.4.2 周辺語間の類似度

引用箇所の周辺語の語数  $N$  として、 $15 \cdot 25 \cdot 50 \cdot 100 \cdot 150 \cdot 200$  の 6 種類を設定して、各  $N$  において「周辺語間の類似度」を求めた。「周辺語間の類似度」と「被引用論文間の類似度」の相関係数を求めた結果が表 3 である。この尺度では、相関係数が 1 に近いほど適切と判断される。

#### 3.4.3 論文構成からみた距離

「論文構成からみた距離」と「被引用論文間の類似度」の相関をみた結果が表 4 である。この尺度は順序尺度であるため、それぞれの種類毎の類似度の平均値を示す。

## 4 考察

### 4.1 各尺度の適切さに関する検討

実験した尺度のうち、文字数からみた「物理的距離」に関してはほとんど相関がないといえる。したがって、この尺度は引用箇所間の意味的な近さを適切にとらえていないと判断される。次に、「周辺語間の類似度」では相関がみられた。このため、「周辺語間の類似度」は、引用箇所間の意味的な近さをとらえていると判断できる。三つ目の尺度である「論文構成からみた距離」では、構造単位が小さいものほど類似度が高くなっていることがわかり、引用箇所間の近さをとらえていることが分かった。

### 4.2 共引用ペアの分布による尺度の比較

相関がみられた「周辺語間の類似度」と「論文構成からみた距離」に関して、尺度の値毎の共引用ペアの分布状況を調べて比較した。なお、「周辺語間の類似度」では、「tf\*idf/cosine」の集計方法 1 で最も相関係数の高かった  $N = 100$  の場合の分布状況を調べた。その結果が表 5、表 6 である。

表3 周辺語間の類似度と被引用論文間の類似度との相関

|       |               | 15語   | 25語   | 50語   | 100語  | 150語  | 200語  |
|-------|---------------|-------|-------|-------|-------|-------|-------|
| 集計方法1 | tf*idf/cosine | 0.259 | 0.290 | 0.321 | 0.352 | 0.347 | 0.337 |
|       | 正規化書誌結合       | 0.228 | 0.258 | 0.267 | 0.278 | 0.290 | 0.292 |
|       | 正規化共引用        | 0.214 | 0.256 | 0.278 | 0.307 | 0.298 | 0.295 |
| 集計方法2 | tf*idf/cosine | 0.352 | 0.384 | 0.382 | 0.364 | 0.381 | 0.382 |
|       | 正規化書誌結合       | 0.187 | 0.217 | 0.238 | 0.232 | 0.204 | 0.206 |
|       | 正規化共引用        | 0.209 | 0.241 | 0.245 | 0.230 | 0.258 | 0.264 |

表4 「論文構成からみた距離」の尺度値毎の被引用論文間の類似度の平均値

|       |               | 列挙    | 同一文   | 同一段落  | 非同一段落 |
|-------|---------------|-------|-------|-------|-------|
| 集計方法1 | tf*idf/cosine | 0.292 | 0.234 | 0.185 | 0.153 |
|       | 正規化書誌結合       | 0.200 | 0.156 | 0.109 | 0.075 |
|       | 正規化共引用        | 0.248 | 0.204 | 0.155 | 0.119 |
| 集計方法2 | tf*idf/cosine | 0.349 | 0.257 | 0.214 | 0.178 |
|       | 正規化書誌結合       | 0.234 | 0.163 | 0.129 | 0.093 |
|       | 正規化共引用        | 0.274 | 0.222 | 0.179 | 0.142 |

表5 「周辺語間の類似度」の尺度値における共引用ペアの分布

| 周辺語間の類似度 | ~0.1  | ~0.2   | ~0.3  | ~0.4  | ~0.5  | ~0.6 | ~0.7 | ~0.8 | ~0.9  | ~1.0  |
|----------|-------|--------|-------|-------|-------|------|------|------|-------|-------|
| 共引用ペアの数  | 8,854 | 14,622 | 9,850 | 4,624 | 1,991 | 832  | 474  | 447  | 1,079 | 1,425 |

表6 「論文構成からみた距離」の尺度値における共引用ペアの分布

| 種類      | 非同一段落  | 同一段落  | 同一文   | 列挙    |
|---------|--------|-------|-------|-------|
| 共引用ペアの数 | 34,596 | 6,013 | 1,719 | 1,870 |

## 5 まとめ

実験結果から、「周辺語間の類似度」と「論文構成からみた距離」が尺度として適切さを持つことが分かった。また、両者を共引用ペアの分布状況からみた場合、前者の方が共引用のペアを詳細に分類していることがわかった。

ただし、それぞれの尺度が類似論文検索の性能向上に寄与することができるか、あるいはどちらがより適切な尺度であるのか判断は、実際の検索システム上でさらに検証する必要がある。なお、必ずしも一方の尺度のみに限定する必要はなく、両者を組み合わせることも有用であろう。

## 参考文献

- [1] Small, H. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*. vol. 24, no. 4, 1973, p. 265-269.
- [2] 江藤正己. 引用箇所の間隔に基づいた共引用の検討. 電子情報通信学会第18回データ工学ワークショップ/第5回日本データベース学会年次大会 (DEWS2007), 広島, 2007, L1-1.
- [3] CiteSeer.PSU OAI, <http://citeseer.ist.psu.edu/oai.html>

## SRU/SRW を用いた教育図書館資料の 書誌検索システムの構築

### A SRU/SRW-based Information Retrieval System for Bibliographic Data at Library of Education

江草由佳<sup>\*1</sup>, 高久雅生<sup>2</sup>

Yuka EGUSA, Masao TAKAKU

\*1 国立教育政策研究所教育研究情報センター

Educational Resources Research Center, National Institute for Educational Policy Research

〒153-8681 東京都目黒区下目黒 6-5-22

E-mail: yuka@nier.go.jp

2 情報・システム研究機構新領域融合研究センター

Transdisciplinary Research Integration Center, Research Organization of Information and Systems

〒101-8430 東京都千代田区一ツ橋 2-1-2 国立情報学研究所内

E-mail: masao@nii.ac.jp

日本語書誌データを対象として SRU/SRW プロトコルに対応した書誌検索システムを開発した。国立教育政策研究所教育図書館が提供している「教育研究論文索引」の日本語論文書誌データ約 12 万件を対象とし、さらに同図書館 OPAC ログデータを元にしたテストクエリセットを作成し、これらの検索が SRU/SRW を通じて行えるかを確認するとともに、実運用規模における検索システムの可能性について考察した。

This paper reports an information retrieval system for Japanese bibliographic data based on SRU/SRW protocol. The system consists of a client which can access to our SRU/SRW server, and of a server which provides bibliographic data for educational papers at Library of Education. The server has about 120 thousands bibliographic records of “*Kyoiku-Kenkyu-Ronbun-Sakuin*”, and are capable for basic retrieval functions in SRU/SRW, such as Boolean queries, queries for specific search indexes, etc. Its capabilities are confirmed by using two kinds of query datasets, functional test-set and OPAC log-based one, in a batch search experiment.

キーワード: SRU/SRW, 情報検索システム, 検索プロトコル,  
SRU/SRW, information retrieval system, information retrieval protocol

## 1 はじめに

1980 年代末から 1990 年代を通じて標準化された情報検索プロトコル Z39.50[1] は、欧米の図書館業界を中心に普及が進んだ。しかし、その一方で、Z39.50 プロトコル自体は、現在のインターネット普及以前にその開発が始まり、OSI ベースの通信プロトコルとして標準化が進められ、かつ、その標準化作業が WWW 普及以前にほぼ終わっていたという 2 点から、プロトコル自体が他のオンラインシステムの開発において使われ

ている環境から次第に外れた存在となっていく。また、多くの機能を盛り込んだ標準仕様であったため、その全ての機能を網羅することは難しく、また簡易な検索機能を実装するだけで良い場合にも、仕様全体のどの部分を実装すれば良いかを確認するのに苦労するなど、新たに実装を行なうことが難しい点もあった。とりわけ、WWW サーブエンジンが普及し、簡易な検索条件のみ事足りるとする一般の利用者が増えたため、これらの要求に応える場合に Z39.50 を採用することはそのコストに見合わない

といったケースも出てきている。

これらの問題点を克服するために Z39.50 標準管理機構において次世代プロトコルの検討が行われ、HTTP を通信プロトコルとし、XML を要求/応答メッセージの記述に用いるような、より現代的な方向に開発が進められた。この結果として開発されたプロトコルが SRU/SRW (Search/Retrieve via URL, Search/Retrieve Web Service) [2] の 2 つの次世代プロトコルである。SRU は REST (Representational State Transfer) フレームワークに基づくプロトコルとして設計され、SRW は SOAP 仕様に基づく検索要求を行うプロトコルである。これらのプロトコルは Z39.50 から離れ、完全な別仕様として開発され、互換性を犠牲にすることにより、より現代的なプロトコルとしてシステムを実装できるようになった。SRU/SRW は現在 1.1 版の仕様が公開されており、1.2 版のリリースも近々予定されている。日本国内でも、SRU/SRW の紹介記事 [3] が出ており、国立国会図書館 [4]、筑波大学 [5] など SRU/SRW を用いた検索システムが提供されつつある。

一方で、現在のところ、SRU/SRW プロトコルの実装は実験段階にとどまっており、これらが Z39.50 プロトコルと比して、どのようなインパクトをもたらすかは明らかではない。そこで、筆者らはこれまで Z39.50 日本語検索システムを開発してきた経験 [6][7][8][9] を踏まえて、本稿では SRU/SRW のクライアントおよびサーバの実装を通じて、その特性を考察する。

## 2 SRU/SRW

SRU/SRW の検索要求は CQL (Common Query Language<sup>\*1</sup>) を使用する。CQL は SRU/SRW の次世代プロトコルとともに新たに開発された。CQL はシステム間のやりとりのためだけでなく、人による可読性も考慮に入れて設計されているために、利用

\*1 SRU 1.2 版からは Contextual Query Language という名称となる予定である。

者が直接 CQL を入力しての検索もできるような検索式である。

SRU における要求を図 1 に示す。ベース URL は検索サーバを表わす。クエリパートは、プロトコルのバージョンや検索結果の返戻要求、検索式などプロトコル上のパラメータを表現する。この例では「読書」で検索したヒット件数を要求している。

サーバからのレスポンスは XML で返される。図 2 は、本研究で開発した SRU サーバから 1 レコードを返した際のレスポンス例である。

## 3 書誌データ：教育研究論文索引

本研究で使用した書誌データは、国立教育政策研究所教育図書館が作成し、Web 上で提供している「教育研究論文索引」[10] である。「教育研究論文索引」は、教育図書館が受け入れた逐次刊行物の中から教育に関する記事を採録したものである。本研究では、1968 年および 1983~2005 年の 24 年分 122,510 レコードを使用した。

論題、著者名、著者名よみがな、掲載誌名、巻号、掲載ページ数、ISSN、請求記号、発行年月、文献番号、キーワード、登録日、冊子体の書誌項目がある。

## 4 開発システム

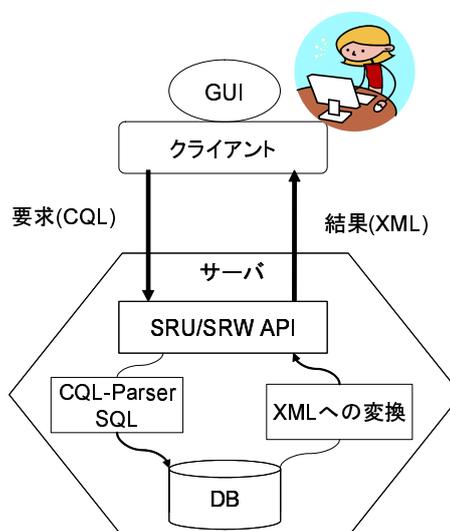


図3 システムの構成図

URL 例：  
 http://kaede.nier.go.jp/epi?version=1.1&operation=searchRetrieve&query=%E8%AA%AD%E6%9B%B8  
 ベース URL：http://kaede.nier.go.jp/epi  
 クエリパート：?version=1.1&operation=searchRetrieve&query=%E8%AA%AD%E6%9B%B8

図 1 SRU 要求: 検索式は「読書」を指定している。なお、URL 中での日本語文字列は UTF-8 として表現した上で URI エンコードするため、数値参照となっている。

```
<?xml version="1.0" encoding="UTF-8"?>
<zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
  <zs:version>1.1</zs:version>
  <zs:numberOfRecords>1</zs:numberOfRecords>
  <zs:records>
    <zs:record>
      <zs:recordPacking>xml</zs:recordPacking>
      <zs:recordData>
        <xml>
          <pubdate>1996.3</pubdate>
          <journal_id>370.5-59-29</journal_id>
          <page>p.1 ~ 175</page>
          <regdate>1997.9.30</regdate>
          <author>林部一二, 堀井啓幸, 杉本真理子, 佐藤晴雄</author>
          <keywords>教育内容・方法 図書館教育 社会教育・生涯学習 家庭教育</keywords>
          <volnum>29</volnum>
          <book_mapping>教育研究論文索引 1996 年版</book_mapping>
          <title>子どもの読書の実態と家庭における指導に関する調査研究</title>
          <journal>日本教材文化研究財団 調査研究シリーズ</journal>
          <paper_id>9603611</paper_id>
        </xml>
      </zs:recordData>
      <zs:recordPosition>1</zs:recordPosition>
    </zs:record>
  </zs:records>
</zs:searchRetrieveResponse>
```

図 2 SRU/SRW からの返戻レコードの一例: 教育研究論文索引独自のスキーマにもとづく XML 形式

開発システムの構成を図 3 に示す。システムは SRU/SRW クライアントと SRU/SRW サーバに分かれる。

#### 4.1 SRU/SRW サーバ

サーバは基本的な検索要求およびレコード返戻要求に応える機能を有している。

クライアントからの検索要求があったら、まず検索要求に含まれる CQL 検索式を解析し、内部のデータベースで用いている SQL に変換する。内部データベースにおいて SQL を用いた検索が行われたら、その結果を XML 形式に変換してクライアントに返す。この間、検索式にエラーがある場合や、対応していない検索式を渡された場合、対応していない返戻レコード形式を要求された場合には、SRU/SRW 仕様に沿ったエラーメッセージを返す。

キーワード検索、フレーズ検索、and, or, not 演算子を使用した論理検索、検索項目を指定した検索、Dublin Core (DC) の検索項目を指定した検索などの基本的な CQL 検索式に対応した。

検索結果として返されるレコードは教育研究論文索引用の独自 XML スキーマに基づくものを採用した。図 2 に返戻レコードの例を示す。

#### 4.2 SRU/SRW クライアント

図 4 に本研究で開発した SRU/SRW クライアントのスクリーンショットを示す。

クライアントはシンプルな構成として構築しており、Web ブラウザ上のフォームで利用者に CQL を入力してもらい、検索をサーバ側に渡す構成となっている。

サーバからの応答として返ってきた XML

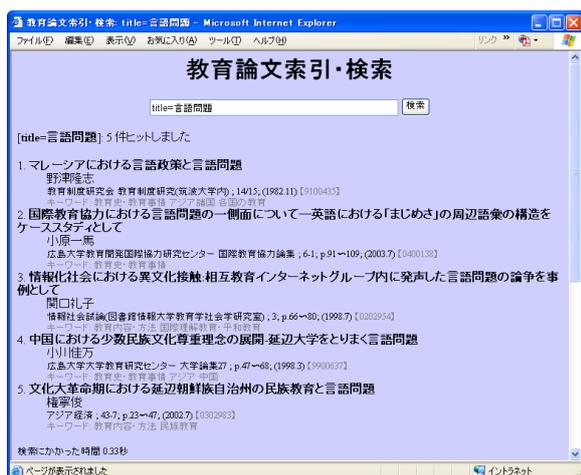


図4 SRU/SRW クライアント

メッセージは内部でパースして、100 件毎に表示する。

### 5 テスト用クエリ集合の開発

開発したサーバに対する検索実験として、以下の2種類の観点から、テスト用クエリ集合を作成し、検索が行えるか確認した。

1. 機能テストセット: CQL の様々な機能に対してどの程度対応しているかどうかチェックする目的のもの
2. ログベーステストセット: 現実の検索要求に対してどれだけ検索可能かをチェックする目的のもの

機能用テストセットについては、既存のCQL 関連ソフトウェア、仕様書で提供されているテストセット<sup>\*2\*</sup>を元に作成した。ログベーステストセットについては、教育図書館蔵書検索(OPAC)<sup>\*4</sup>の検索ログを使用して作成した。このOPACでは主に書籍を対象としており、今回対象としたデータは論文を対象としている点は異なるものの、教育分野に関する情報ニーズである点と教育図書館が所蔵している資料の検索という観点から類似しているため、擬似的な検索

<sup>\*2</sup> SRWLucene <http://wiki.osuosl.org/display/OCKPub/SRWLucene>  
<sup>\*3</sup> CQL::Parser <http://www.textualize.com/cql-parser>  
<sup>\*4</sup> <http://opac.nier.go.jp/>

表1 テストセットのヒット件数の分布

| ヒット件数      | テストセットクエリ数 |      |
|------------|------------|------|
|            | 機能         | ログ   |
| 0          | 49         | 7231 |
| 1 ~ 100    | 24         | 885  |
| 101 ~ 1000 | 12         | 53   |
| 1001 ~     | 27         | 47   |
| 計          | 97         | 8216 |

要求とみなすことができると判断し、この検索ログを使用することとした。OPACのログデータから収集した検索式はCQL形式に変換した。これらテストセットの検索式は、1行にCQLで表現された1つの検索要求を記述するという形式で作成しておき、それをバッチで検索実行する形で使用した。

機能用テストセットとしては128の検索式を生成し、その全てについて検証を行った。これら128件の検索式のうち、6件は意図的にエラーとなるように作成した検索式であり、これらがエラーを返すことを確認した。また、10件については、未対応のためエラーを返すことも確認している。なお、この他15件の近接演算を用いた検索式には未対応であるが、正しくエラーが返せていない。これら以外の97件の検索式においては検索要求を受け付けられることを確認した。ログベーステストセットとしては、8216件の検索式を生成し、その全てについて検索を実行した。それぞれのテストセットのヒット件数の分布は表1に示す。

### 6 考察

日本語書誌データを使用したSRU/SRW検索システムにおける留意点を挙げ、それぞれについて考察する。

#### 6.1 日本語検索

Z39.50とSRU/SRWは、欧米を中心にプロトコル仕様の開発が進んだために、日本語のような膠着語で分かち書きの必要な言語、複数の文字コードが使用されている言語についてはあまり考慮されていない。日

本語検索式への対応という面においては、これらの指定をどのようにして日本語環境において実装していくか、その手法についての議論は十分に行われたとはいいがたい。この点は、既に Z39.50 にも同様の問題が指摘されていた [11] が、これらの問題は SRU/SRW にも引き継がれている [12][13]。

現在、比較的一般に行われているのは形態素解析器等を用いた辞書、接続ベースの分かち書きであるが、これらは書誌レコードや検索式などの短い文字列に対しては、必ずしも高精度な分かち書きを行える訳ではないという欠点があり、検索精度の低下、検索漏れの発生といった懸念もある。ただ、SRU/SRW および Z39.50 では、指定がなされない限り、データベース内部でどのようなマッチング手法を使用しているかは大部分データベース側で自由に設定できるよう、抽象化されている。そのためプロトコル仕様（この場合は CQL）自体が欧米語を想定していたとしても、ある程度自由な設定が効くはずである。一方で、CQL の `=` や `all`、`any` といったリレーションと呼ばれる文字列マッチングもしくは項目マッチングの指定項目を、どのようにデータベースにおけるマッチングに適用させていくかは、日本語の場合は一定の方式とされるほどの方式は存在せず、また、効率性の面からも様々な実装方式に分かれることと思われる。この点については、実装方式のガイドライン制定や、仕様に対するフィードバックを行っていく必要があるように思われる。

なお、本報告において開発したシステムにおける実装では、検索要求で送られたキーワードが中間一致するレコードをヒットするように開発したため、そもそも「ワード」概念を意識せずに検索が可能となっており、利用者が入力した文字列が含まれるレコードは必ずヒットする。そのため、対象としている言語が日本語だけの場合はこのような実装も一手法といえる。

## 6.2 検索要求における文字コード

SRU/SRW では、XML の結果返戻を前提にしている点などから、文字コードは XML

における Unicode 指定を基本として設計されている。さらに、SRU では簡易な URI 中における検索要求の引き渡しのために検索要求中の文字コードは UTF-8 に限定されている。仕様として文字コードが指定されたことから、日本語のように様々な文字コードを使用している環境でもシステムは XML 処理および URI 中の UTF-8 指定にのみ対応すればよく、システム間での文字コードにおけるネゴシエーションなどは必要なくなった。この点は Z39.50 と異なる。

とりわけ SRU においては使用文字コードを UTF-8 に限定しているため、日本語文字等を使用して検索したい場合には、必ず UTF-8 への変換を行う必要がある。既存の環境が UTF-8 環境ではなく Shift\_JIS や EUC 環境である場合、サーバ側であれば、レコードデータもしくは検索サーバにおける対応が必要になり、また、クライアント側であれば表示および入力時の UTF-8 対応が必要となってくる。

今回開発したクライアントでは、エンドユーザの入力からサーバへの受け渡しまで全てを UTF-8 で行うシステムとして開発した。一方、サーバ側では、クライアントからの要求は UTF-8 を始めとする要求に応えるものの、内部処理は元々の日本語データとの関係から EUC-JP を用いた。このため、要求が UTF-8 で来た場合、内部で一旦文字コード変換を行ってからデータベース検索を行い、検索結果を再度 UTF-8 に戻すなどの処理を行った。

## 6.3 サーバ側における選択の明示

SRU/SRW においては、様々な機能や CQL において、指定が何も無い場合はサーバ側における選択 (serverChoice) にゆだねるという仕様になっている。これには、サーバを開発する際に、トレードオフを考慮して実装の手間を減らして開発工数を下げたり、サーバ側での工夫で多様なマッチング手法等を採用できるなどの点で貢献があると考えられる。特に、検索のアクセスポイントに関しては、DC などの既存の標準スキーマとのマッピング等は必須ではないため、現

在運用しているデータベースを SRU/SRW 対応にして提供していく際にそのまま XML 形式における提供を検討するだけで良いなど、便利な点もある。しかし、横断検索等のニーズを考えた場合、DC などのよく使われている標準とのマッピングを行なって、検索可能にすることを検討するべきである。

さらに、どのようなサーバ実装が選択されているか分からない場合は、利用の際に不都合が起こるケースが想定されるため、サーバについての説明を利用者に対して提供する必要が出てくる。これへの対策としては、SRU/SRW のサーバ情報を提供するための Explain 機能を用いて提供することが可能なのではないか。その際にはどのような情報が利用者にとって必要なのかについて検討する必要がある。

## 7 おわりに

本研究では、SRU/SRW のクライアントおよびサーバの開発を行った。実装にあたっては国立教育政策研究所教育図書館が提供する教育研究論文索引の主に日本語論文書誌データ約 12 万件を対象とし、さらに同図書館 OPAC ログデータを元に、擬似的な検索要求としてテストセットを作成し、これらの検索が SRU/SRW を通じて行えるかを確認し、実運用規模における検索システムの可能性を示した。また、特に SRU/SRW の日本語環境における対応状況の確認も行ない、あわせて考察した。

## 謝辞

本研究は科学研究費補助金（課題番号：18800080）の助成を受けた。

## 参考文献

- [1] Z39.50 Maintenance Agency. ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50) : Application Service Definition and Protocol Specification. 1995, 156p. (online), available from <http://www.loc.gov/z3950/agency/1995doce.html>.
- [2] The Library of Congress. SRU: Search and

- Retrieve via URL (Standards, Library of Congress). available from <http://www.loc.gov/standards/sru/>.
- [3] 平山 亮. 情報検索および図書館相互貸借の標準規格. 情報の科学と技術. Vol.56, No.7, 2006, p.307-311.
- [4] 国立国会図書館. 国立国会図書館デジタルアーカイブポータル (ndl-dap) - SRW とは？-SRW-システム紹介 <http://www.dap.ndl.go.jp/home/modules/pukiwiki/?SRW> (参照 2007-04-19).
- [5] 筑波大学. 知的コミュニティ情報システム仕様書. <http://www.kc.tsukuba.ac.jp/spec060621.pdf> (参照 2007-04-18).
- [6] 宇陀 則彦, 江草 由佳, 高久 雅生, 石塚 英弘. Z39.50 による日本語書誌データ検索システム. 情報知識学会誌. Vol.9, No.2, 1999, p.1-15.
- [7] 江草 由佳, 真野 泰久, 宇陀 則彦, 石塚 英弘. Z39.50 プロトコルによる日本語書誌データ情報検索システム. 第 6 回研究報告会講演論文集. 情報知識学会. 東京, 1998-05. 情報知識学会, 1998, p.29-36.
- [8] 高久 雅生, 江草 由佳, 宇陀 則彦, 石塚 英弘. “Z39.50 による書誌データ検索システムの構築: Dublin Core を共通スキーマとして”. デジタル図書館. No.16, 1999, p.97-106. (オンライン), 入手先 [http://www.dl.slis.tsukuba.ac.jp/DLjournal/No\\_16/12-masao/12-masao.html](http://www.dl.slis.tsukuba.ac.jp/DLjournal/No_16/12-masao/12-masao.html).
- [9] 江草 由佳, 高久 雅生, 宇陀 則彦, 石塚 英弘. Z39.50 データベース選択支援 環境. 情報知識学会誌. Vol.11, No.2, 2001, p.1-10.
- [10] 江草 由佳. 教育図書館における複数コレクションの提供. デジタル図書館. No.32, 2007, p.23-33.
- [11] 安齋 宏幸. Z39.50 の技術解説. 情報の科学と技術. Vol.48, No.3, 1998, p.134-139.
- [12] 宮澤 彰. CQL/ZING-SRW. <http://www.jsa.or.jp/stdz/instac/committe/H15report/report-contents/aidos/wg1-23/23-10.ppt> (参照 2007-04-18).
- [13] 宮澤 彰. CQL と非ヨーロッパ書法の検討. <http://www.jsa.or.jp/stdz/instac/committe/H15report/report-contents/aidos/wg1-26/26-05.ppt> (参照 2007-04-18).

# 著者とキーワードの関連性に着目した 研究領域ブラウジングシステムの試作 A Prototype for a Research Area Browsing System focusing on Author-Keyword Relationship

レボウィッツ 紀子 松村 敦 宇陀 則彦

Noriko LEBOWITZ, Atsushi MATSUMURA, Norihiko UDA

筑波大学大学院 図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

E-mail: {nlebo,matsumur,uda}@slis.tsukuba.ac.jp

新しい研究領域に取り組もうとしている研究者は、文献調査により対象領域の研究内容について把握する。しかし、研究領域を把握していない段階では情報要求に曖昧性が生じ、効率的な文献調査を行なうことができない。この問題を解決するには、文献調査の前段階において対象とする研究領域の全体像を把握し、さらに個々の研究テーマを概観できる機能が必要であると考えられる。本稿では、著者と著述文献キーワードの関連性を利用して、各著者の研究内容の傾向を読み取り、それを研究テーマとして解釈し、研究領域をブラウジングする手法を提案する。また、提案手法に基づいて試作したシステムを紹介する。

A researcher studies a new target domain through a search of papers. However, an efficient search is difficult if the information demands are ambiguous. To solve this problem, it is necessary prior to a search of the papers to have an understanding of the entirety of the research field and an overview of the research topics. In this paper, we propose a technique to extract the research topics by using author-keyword relationship in a specified research area. And we introduce the prototype system based on our proposal technique.

キーワード：研究領域，ブラウジングシステム，ISPモデル，PubMed，MeSH，Research Area，Browsing System，Information Seeking Process Model

## 1 はじめに

これから新しい領域で研究に取り組もうとしている研究者の多くは、文献調査により対象とする研究領域の全体像を把握した

上で研究内容を具体化していく。この際、検索システムを用いるのであれば、ユーザは情報要求を的確なキーワードとして表現する必要がある。しかし、ユーザが研究領域について把握していない段階では情報要求

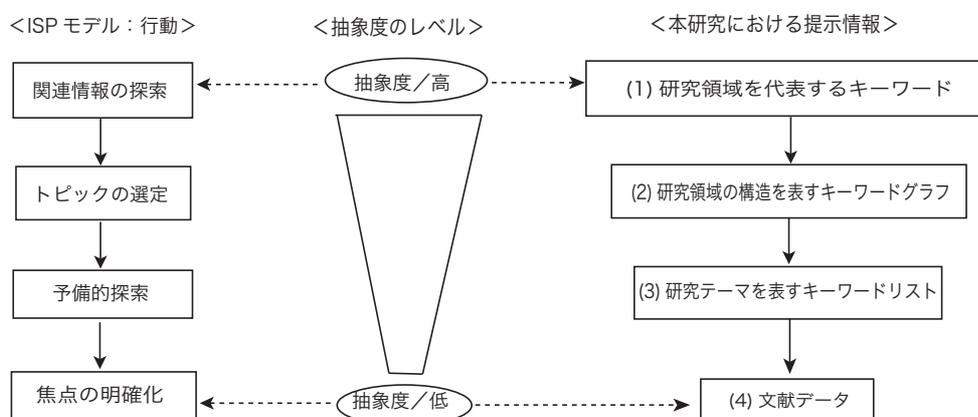


図 1: 本研究における提示情報の抽象度についての定義

に曖昧性が生じるため、的確なキーワードの選択が難しく、抽象度の高いキーワードを用いた検索となる。その結果、提示される文献データの数が膨大になり、適合文献の判定が困難になるという問題が生じる。

本研究では、この問題の原因をユーザの情報要求とシステムにより提示される情報との間にある抽象度の不一致と捉え、クールソのISPモデル[1][2]に沿って、ユーザの情報要求に応じた抽象度の情報を提示する仕組みを提案する。具体的には、研究領域におけるキーワードを関連づけて提示することで全体像の把握を促し、その上で、著者と著述文献のキーワードを関連づけて研究テーマとして提示することで、曖昧なユーザの情報要求を段階的に特定化させることを試みる。以下では、提案手法の詳細と、それに基づいて試作した研究領域ブラウジングシステムについて説明する。

## 2 提案手法

クールソは情報探索過程について探索開始、選択、探求、焦点形成、収集、表現という6つの段階を設定し、それぞれの段階における感情・思考・行動の変化を Information Seeking Process (ISP) モデルとして表現している。本研究の提案手法は、このISPモデルのうち探索開始から焦点形成までを対象とする。文献選択を行うブラウジングの過程において提示する情報の抽象度をISPモデルの行動に沿った形で段階的に下げることにより、ユーザの情報要求の曖昧性を解消することを試みる。提示情報の抽象度として以下の4段階を設定した。

- (1) 研究領域を代表するキーワード
- (2) 研究領域の構造を表すキーワードグラフ
- (3) 研究テーマを表すキーワードリスト
- (4) 文献データ

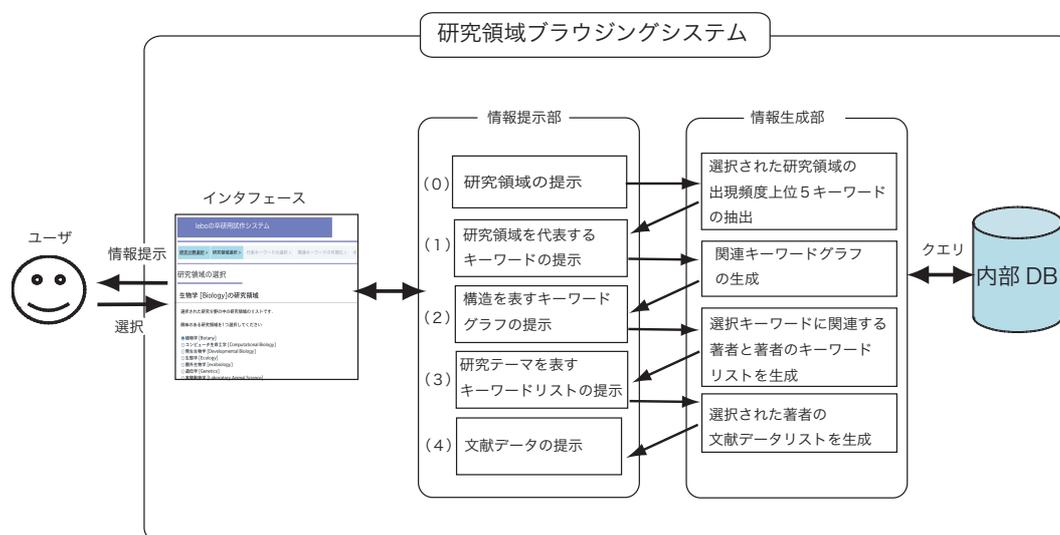


図 2: 試作システムの処理の流れ

齊藤 [3] により具体化されている ISP モデルでの探索開始から焦点形成おける「行動」と、本研究で提示する情報の抽象度を対応付けたものを図 1 に示す。まず (1) では、ブラウジングを開始するユーザの最初の手がかりとなる研究領域を代表するキーワードを提示する。その上で、(2) では研究領域を代表するキーワードとそれに関連するキーワードをグラフ構造により可視化して提示する。ユーザは可視化されたキーワード間の関連を繰り返しブラウズすることで、研究領域の構造を知ることができる。これは ISP モデルにおける「関連情報の探索」から「トピックの選定」の段階に対応する。次に (3) では、著者と著述文献のキーワードを関連づけ、研究領域における研究テーマとして著述の多い著者の順に提示する。これは ISP モデルにおける「予備的探索」の段階となるものである。著者と著述文献のキーワードを関連づけて提示することで、ユー

ザは著者ごとに研究領域の傾向がわかる。研究領域の傾向を把握するためには、通常多数の文献を読む必要があるが、実際には関係のない文献も多く含まれており、非効率的である。研究領域に強い影響を与えている文献から読むことができれば、効率よく傾向を把握できる。そこで本研究では、研究領域の主要なキーワードを含み、かつ著述数の多い著者の文献を影響の強い文献と仮定した。実際に植物学に関する文献データから著者とキーワードを抽出して分析した結果、著述の多い上位約 20% の著者の文献には研究領域の約 80% のキーワードが含まれていることがわかった。このことは、著述の多い著者の文献が研究領域の傾向に強い影響を及ぼしていることを示している。

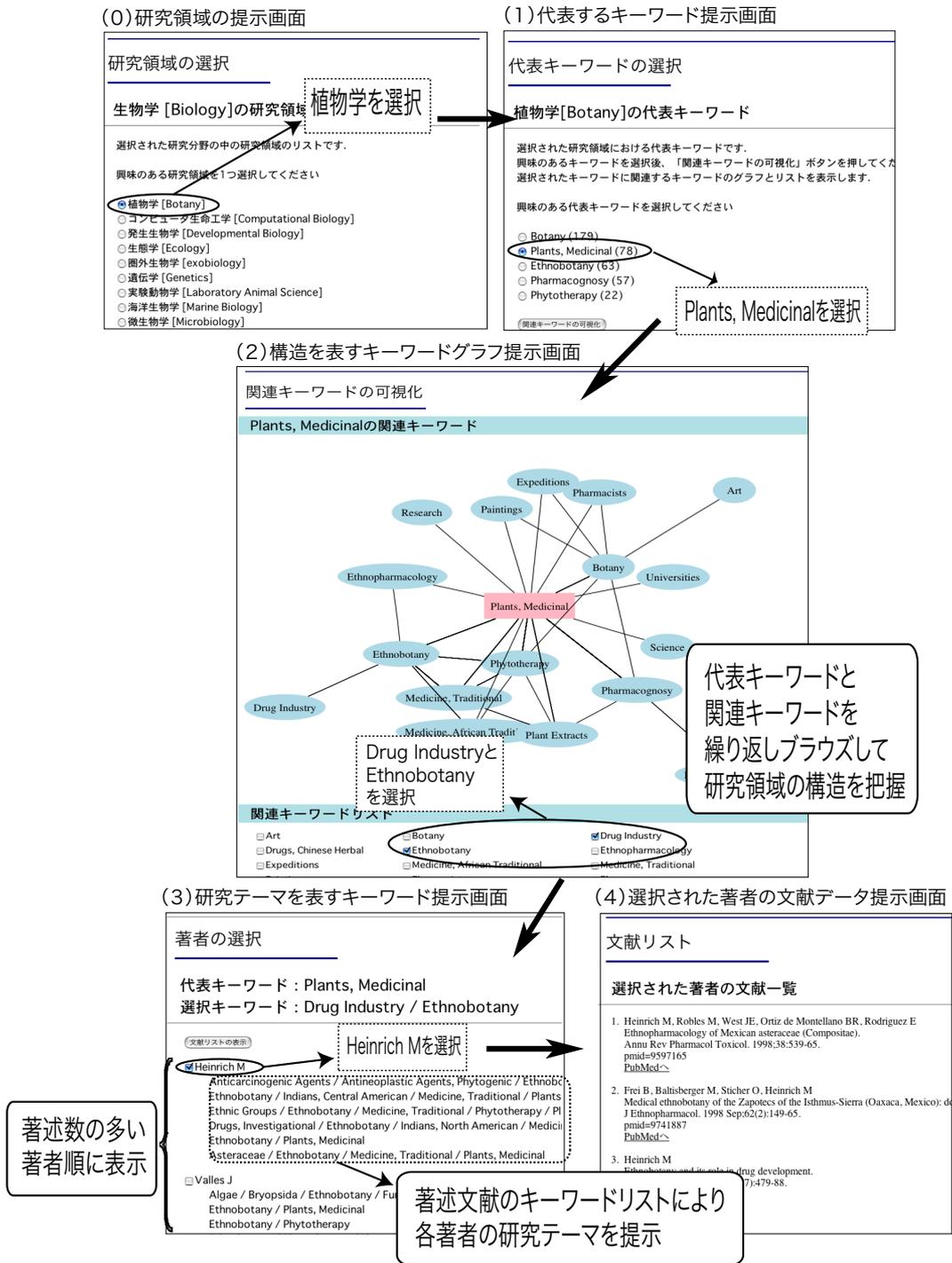


図 3: 試作システム上での情報探索の流れ

### 3 研究領域ブラウジングシステム

提案手法に基づいて、対象とする研究領域の把握と適合文献選択の際のブラウジングを支援するシステムを試作した。

試作システムの構築には、PubMed[4]の文献データを用いた。PubMedの文献データには階層構造を持つMeSH(Medical Subject Headings)[5]タームと呼ばれる索引語が付与されているため、これをキーワードとして扱った。MeSHタームの階層構造において、Biology(生物学)の下位概念であるBotany(植物学)を研究領域として設定した。次にシステムにおける処理の流れを図2に示す。また図3には試作システムのインタフェースを用いて、システム上での情報探索の流れを示す。

次に、各段階における処理の流れについて説明する。

#### 研究領域の提示

研究領域の提示はシステム利用の出発点である。本システムでは、MeSHの階層構造とMeSHタームを利用して研究領域の設定を行なった。ユーザにより研究領域が選択されると、システム側では検索対象をそのキーワードを持つ文献データとする。

#### 代表キーワードの提示

決定された検索対象の範囲において、出現頻度の高い順に5つのキーワードを研究領域を代表するキーワードとして抽出し提示する[図3-(1)]。ユーザはこれらのキーワードから興味のあるものを一つ選択する。

キーワード間の関連を表すグラフの提示  
前段階でユーザにより選択されたキーワードと同一文献内に共起するキーワードとを関連づけて、グラフとして可視化する[図3-(2)]。ユーザは、研究領域を代表するキーワードと構造を表すキーワードのグラフを繰り返しブラウズし、これらのキーワードの中から、情報要求に合致するものを選択する。

#### 研究テーマを表すキーワードリストの提示

ユーザにより選択されたキーワードを著述文献に持つ著者のリストを生成する[図3-(3)]。このリストは著者名と著述文献のキーワードリストで構成されており、これを研究テーマとして著述数の多い順に提示する。ユーザは提示された研究テーマのリストをブラウズして自分の興味に近い研究を行っている著者を選択する。

#### 文献データの提示

ユーザにより選択された著者の著述文献リストを提示する[図3-(4)]。さらに、テストデータの入手先であるPubMed上のデータへのリンクも提示する。ユーザはこのリンク部分をクリックすることで、PubMed上の該当データに移ることができる。

以上が試作システムによる一連の処理の流れである。

### 4 おわりに

本研究では、学術情報アクセスを対象として、情報要求に適した抽象度の情報を提

示する仕組みを提案し，それに基づいてシステムを試作した．研究領域を把握するという観点から，著者と著述文献のキーワードの関連性を利用して，研究テーマの提示を行なった．

今後の課題は，システムが各段階で提示する情報の妥当性の評価である．そのためには，まず，提示する情報に抽象度を設定したことによる効果について調査を行なう必要がある．また，特に，研究テーマの抽出に関して検討を行い，研究者の文献選択の支援を行える，より実用的な手法を得たいと考えている．

## 参考文献

- [1] Carol C. Kuhlthau. “Inside the Search Process: Information Seeking from the User’s Perspective”. *Journal of the American Society for Information Science*. vol. 42, no. 5, 1991, p. 361-371.
- [2] 渡辺智山. “利用者研究史と情報探索過程モデル”. *同志社図書館情報学*. 第7号, 1996, p. 38-74.
- [3] 斎藤泰則. “3 情報探索の論理”. *情報探索と情報利用*. 東京. 勁草書房. 2001, p. 153-188. (図書館・情報学シリーズ 2) (ISBN 4-326-04801-8).
- [4] U.S. National Library of Medicine. PubMed. 更新 2007-04-25. (オンラインデータベース), 入手先 <<http://www.pubmed.gov>>, (参照 2007-04-25).
- [5] U.S. National Library of Medicine. Medical Subject Headings. 1999, 更新 2006-07-25. (オンラインデータベース), 入手先 <<http://www.nlm.nih.gov/mesh/>>, (参照 2007-04-27).

## 営みの Lattice の構造化とアブダクション

### Realizing Abduction, and Making up the structure of Lattice of behaviors

福 永 征 夫

Masao Fukunaga

アブダクション研究会

Email:jrfd117@ybb.ne.jp

要旨:地球規模の問題群に真に対処するためには、自己や人間などの<この部分域>を最適にすることと、他者や環境などの<あの部分域>をも含む、<全体域>を最適にすることを、矛盾なく両立させるような広域的な知を発見して、より高次の領域的な知を確立し、営みと営みの間に Lattice という融合の構造を作ることが必要である。そのために、人間の経験や学習の深さと広さを拡充して、アブダクション Abduction を実現する能力を進化させなければならない。

Abstract:To act for the solution of the global problems, as the damages of environment and ecosystem of the earth , the shortage of energy and resources, we should not depend on using skill of trade-off which enables compromise conflicting values of different behaviors of life, but we should make up the structure of Lattice which means fusion of different values of different behaviors of life. And to attain the purpose , we have to evolve our abilities of realizing Abduction.

キーワード keywords:部分と全体の最適化 Optimization of each part and the whole, ラティス Lattice, アブダクション Abduction

#### 1. 地球規模の難題の主たる要因

地球環境問題,資源・エネルギーの枯渇,貧富の差の拡大,人口の爆発,難病の発生,災害や事故の巨大化,民族・宗教・文化・政治・経済をめぐる対立と紛争の激化,凶悪な犯罪やいじめ・虐待行為の多発 などの地球規模の問題群の根底に共通する,それらの主たる要因は,人や集団が特定の領域的な知による人間中心・自己中心の特定の部分最適を,一様な営みとして追求し,非連続な生産・消費・廃棄の一様な営みの量的規模を拡大して,資源やエネルギーの大量使用と不経済な費消を招き,自然や生存環境への負荷を増大させたことにある。

#### 2. 部分と全体の最適化

人間が,省資源・省エネルギーの実を挙げて,地球のシステムや生態系の偏芯を回復するためには,人間の生産・消費・廃棄の活動を,多様な領域的な知による多様な省資源・省エ

ネルギーの営みとして行くと共に,営みの<この部分域>の最適化と,営みの<あの部分域>をも含む,<全体域>の最適化を,矛盾なく両立させるような広域的な知を発見し,より高次の領域的な知を創造して,実践に移し,人間の生存の営みを,より高次の領域的な知に則したところの,互いに矛盾がなく,連続性のあるものとして実現して,自然や生存環境への負荷を軽減することが重要である

#### 3.トレード・オフからラティスの構造へ

省資源・省エネルギーの営みの実を挙げ,環境問題などの地球規模の難題に対処するためには,営みと営みの間に矛盾を来たさないように,広域的な知を発見し,より高次の知を再構成して,実践に移さないことには,問題の根源的な解決にならないのである.ある立場を貫く一つの領域的な知と,他の立場を貫く一つの領域的な知とが自己主張を衝突させると,資源やエネルギーを浪

費させて環境に負荷を与える。そこで、それらの領域的な知を連ねる広域的な知を探索して発見し、より高次の領域的な知を再構成して、実践することが極めて重要になる。従来のように、省資源・省エネルギーの営みと営みの間の矛盾をトレード・オフ (trade off) の妥協で対応しては、問題の根本的な解決にはならないばかりか、かえって、問題を複雑にってしまう。そうしないで、営みの<この部分域>の最適化と、営みの<あの部分域>をも含む、<全体域>の最適化を、矛盾なく両立させるような広域的な知を発見して、より高次の領域的な知を再構成するためには、現実性と可能性の両方の視点から個人や集団の経験や学習の深さと広さを拡充して、人間のアブダクション (abduction) を実現する能力を進化させなければならない。そのような広域的な知を発見して、より高次の領域的な知を再構成することは、営みと営みの間にラティスという融合の構造を作ることである。ラティス (lattice : 束) とは、2つの部分と、その間の論理積の関係と論理和の関係からなる数学的な構造のことである。

#### 4. アブダクションの学術的な定義

アブダクションとは、C・S パース (PEIRCE, 1839~1914) の用語であるが、その概念や過程は、必ずしも明確になっていない。一般的には、『不可解な事実Yが観察された場合、これを、その結論として説明しようとする仮説Xを構想し、提起する推理であり、一般に「ある事象Yは不可解である。しかし、もしXならばYなりとすれば、Yは不可解ではない。したがって、Xは真であろう」という形式をとるもの』(哲学事典・平凡社) とされている。

#### 5. アブダクションの新たな定義

人間のアブダクションの能力を進化の可能なものにするために、次のような形で、『ア

ブダクション』の新たな定義を試みることにする。『アブダクションとは、既存の領域的な知では説明できない不可思議な事物・事象に当面する一方で、別の事物・事象との遭遇をトリガーとして、新たな領域的な知を探索し積み重ね、既存の領域的な知との間に広域的な知を発見し、より高次の領域的な知を創造することによって、不可思議な事物・事象を説明可能なものにするところの蓋然的な推論である。』『それは、不可思議な事物・事象とは異なる別の事物・事象を仲介者として、新たな広域的な知を発見すると共に、その広域的な知という、パターンの表象を高次の領域的な知という、シンボルの表象に展開して、新たな認知と行動の基準を確立しようとする、人間の環境への適応の行為である。』

〔既存の領域的な知では説明できない不可思議な事物・事象に当面する〕場合には、次の二つのケースが想定される。

第一は、現実性の視点に基づいて、既存の領域的な知のもとで、それを深く掘り下げながら思考や行動をするときに、既存の領域的な知とは整合しない、ある種の異質な事物・事象を見出すに至って、これが不可思議な事物・事象となる場合である。第二は、現実性の視点に基づいて、既存の領域的な知のもとで、それを深く掘り下げながら思考や行動をするときに、既存の領域的な知とは整合しない、異質な新たな観念を見出すに至る。そして、これまで説明できていた事物・事象が、新たな観念では説明できなくなって、不可思議な事物・事象となる場合である。ニュートンが行った万有引力の法則の定立というアブダクションの例は、第二のケースに当たる。ニュートンの仕事をアブダクションの新たな定義に重ね合わせて見ると次のようになる。①ニュートンは、現実性の視点に立って、既存の領域的な知であるケプラーの法則とガリレオの慣性の法則について深く思案していたが、『物体に

力が加わらなければ、静止するか、等速円運動をする』というガリレオの考えとは異なり、『物体に力が加わらなければ、静止するか、等速直線運動をする』と考えるに至った。つまり、既存の領域的な知という現実性の視点に立ちながら、それとは整合しない、異質な新たな観念を見出すに至った。既存の領域的な知を深く掘り下げることで見出した異質な新たな観念をもってすると、ニュートンには、月が等速直線運動をすれば、軌道から外れて無限のかなたに行ってしまう筈なのに、地球の周りを回り続けているのが『不可思議な事象』に見え、思い悩むところとなった。②あるとき、農園で、リンゴの木から実が落ちるといふ、別の事象に遭遇したのがトリガーとなり、可能性の視点に立って、思考や行動を拡大し、『物体が地球の重力に加速され引き寄せられて落下する』という物体と物体の間の重力的な相互作用に関する新たな領域的な知を探索して蓄積していった。③あるとき、探索し蓄積してきた新たな領域的な知が、以前に掘り下げていた既存の領域的な知と結びつき、新たな広域的な知がパターンの表象として成立した。そして、それが、より高次の領域的な知として、シンボルの表象に展開され、万有引力の法則となって、すべての説明が可能になったのである。

## 6. 『3軸認知場』のモデル

### (1) 領域的な知と広域的な知

領域的な知とは、生存する上での特定の目的を達成するための理解と働きかけをするのに要した、または、要すると思われる、ひとまとまりの事実の情報または価値の情報または目的の情報、および、それらを構成する要素、要素の関係や、それらの集合のことをいう。

広域的な知とは、『部分と全体の択一的最適化』の作用によって、異なる領域的な知の間に広域的に創発する知のパターンのこと

をいう。

### (2) 領域的な知を蓄積

人間は、未来を想像して予期し今ここに対処する営みを通じ、現在から未来に向けて新たな記憶の部分域を形成し、＜事物・事象＞の事実の情報、＜自己の思考・自己の行動＞の目的の情報、＜事物・事象＞や＜自己の思考・自己の行動＞の価値の情報を、脳という[時間][事実・目的の空間][評価(感情)の空間]の3つの軸からなる認知の場に、時間の知と空間の知を積み重ねて、領域的な知として蓄積する。

### (3) 広域的な知を創発

人間は、過去を想起する営みを通じて、広域的な知を発見する。脳は、現在から過去に向けて、新たな記憶の部分域から過去の記憶の全ての部分域へエネルギーを循環させ、新たな記憶の部分域と過去の記憶の全ての部分域を『融合』という関係で結びつけて、記憶の過去の全てのネットワークと多重させながら、記憶の最新のネットワークを構築する。このプロセスで、1つの新たな記憶の部分域と2つの過去の記憶の部分域からなる、3つの記憶の部分域の間の関係が、小域で隣接するものから、中域で近接するものや、大域で離隔するものに組み換えられて、より脱領域的で、より広域的な知が創発される。

### (4) より高次の領域的な知への発展

領域的な知として蓄積された事実の情報は、広域化を経て、＜事物・事象＞に関するより高次の領域的な知、すなわち、環境世界に関するより高次の理解の基準として発展していき、主体という自我の理解の能力が、＜事物・事象＞を短期的・小域的に理解する能力から、中期的・中域的に理解する能力へ、さらには、長期的・大域的に理解する能力へと、発展を遂げていく。

領域的な知として蓄積された＜事物・事象＞や＜自己の思考・自己の行動＞に対する価値の情報は、広域化を経て、＜自己＞が保有する、より高次の領域的な知、すなわち、＜

自己>が保有する、より高次の価値づけの基準として発展していき、主体という自我の価値観が、生理的・物質的な価値観から、社会的・技術的な価値観へ、さらには、精神的・文化的な価値観へと、発展を遂げていく。領域的な知として蓄積された目的の情報、広域化を経て、<自己の思考・自己の行動>に関するより高次の領域的な知、すなわち、<事物・事象>や評価(感情)という<自己>に対するより高次の働きかけの基準として発展していき、自己や人間、などの<この部分>を最適にすることと、他者や環境などの<あの部分>をも含めた<全体>を最適にすることを矛盾なく達成しようとする主体という自我の目的行動が、短期的・小域的な目的行動から、中期的・中域的な目的行動へ、さらには、長期的・大域的な目的行動へと、発展を遂げていく。

## 7. 『部分域と全体域の誘導合致』

### (1) 論理和と論理積

認知における『部分と全体の択一的最適化』とは、論理和の機能と論理積の機能のいずれをも実現し、かつ、いずれかを非同期的・択一的に実現することである。

『部分域と全体域の誘導合致』のモデルは認知における『部分と全体の択一的最適化』の機能を実現することができる。

そして、『部分域と全体域の誘導合致』のモデルは、ラティスの構造を表わす数的な論理モデルである。

### (2) 自然の系における『融合』

脳を含む自然のシステムがエネルギーのレベルの高い部分域  $P_2$  から低い部分域  $P_1$  へエネルギーを循環させるプロセスにおいて、 $P_2$  と  $P_1$  は、ある側面でプラスに関係し、エネルギーの移動を促進し合うと共に、他の側面ではマイナスに関係して、エネルギーの移動を抑制し合い、トータルでは、両方の関係が平均されて、プラスでもマイナスでもない中立的な関係の結合を実現する。

この中立的な関係での結合を『融合』といい、『融合』をもたらすプロセスを『部分域と全体域の誘導合致』という。

自然のシステムでは、1つの全体域になろうとする作用の力と、2つの部分域のままで留まろうとする作用の力が、互いを誘導しあって拮抗し、合致して、『融合』をもたらすのである。

### (3) モデル式

結合前の2つの記憶の部分域に関して、より新しい記憶の部分域  $P_2$ 、より以前の記憶の部分域  $P_1$ 、のそれぞれが保持するエネルギーの量の相対的な比率を  $\theta P_2$ 、 $\theta P_1$  とし、 $\theta P_2 = 1$ 、 $1 > \theta P_1 > 0$ 、とすると、モデルは、次のように表わすことができる。

<1>  $P_2$  と  $P_1$  の間のプラスの関係は、  
 $\theta P_2 / \theta P_1 > \theta P_2 + \theta P_1$  で、解は、  
 $\theta P_1 < (\sqrt{5} - 1) / 2$  となる。

<2>  $P_2$  と  $P_1$  の間のマイナスの関係は、  
 $\theta P_2 / \theta P_1 < \theta P_2 + \theta P_1$  で、解は、  
 $\theta P_1 > (\sqrt{5} - 1) / 2$  となる。

<3>  $P_2$  と  $P_1$  の間のトータルとしての中立的な『融合』の関係は、  
 $\theta P_2 / \theta P_1 = \theta P_2 + \theta P_1$  で、解は、 $\theta P_1 = (\sqrt{5} - 1) / 2$  となる。

<4>  $[(\sqrt{5} - 1) / 2 \doteq 0.61803398]$  を『融合定数』F S [FUSIONAL CONSTANT] と名づける。

<5>  $P_2$  から  $P_1$  へ移動するエネルギーのレベルを、『循環定数』C R [CIRCULATIVE CONSTANT] と名づける。FSとCRの間には、  
 $(FS + CR) (FS + CR) = FS$  という関係が存在し、  
 $CR = \sqrt{\{(\sqrt{5} - 1) / 2\} - (\sqrt{5} - 1) / 2} \doteq 0.168117389$  となる。

そして、脳の記憶の部分域  $P_2$  と  $P_1$  がプラスに関係するときは、それらの担う事実または価値または目的の情報が論理積の関係で結びつき、マイナスに関係するときには、情報が論理和の関係で結びつくことになる。

<2

$\theta P_1$

# 音楽著作権についての歴史的研究—革命期から 19 世紀のフランスを中心に— The History of the Music Copyright in France between from the Revolution to the 19th Century.

石井 大輔\*  
Daisuke ISHII

\* 筑波大学大学院 図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

〒305-8550 茨城県つくば市春日 1-2

E-mail: di141@slis.tsukuba.ac.jp

音楽著作物は歴史の中で次第に社会性を帯び大量消費されていくことにより、著作権制度自体に大きな影響を及ぼし、著作権を集中管理するためのシステムがつくられていく。本稿はフランスにおける、アンシャン・レジームから革命期を経て 19 世紀に SACEM の設立に至る法制度的、社会的側面から音楽著作物の流通・利用と音楽著作権の生成、発展の過程を追う。そこでは楽譜出版の歴史、フランス著作権制度、音楽著作権の管理団体の勃興が取り扱われる。

This article dealt with a music copyright. Mass consumption of the musical works gave big influence to the copyright system, and a system to control a copyright intensively was made. I chase a process of the circulation of the musical works, the use and generation and the development of the music copyright from the legal system and the social side. This article passes through the Revolution period from *Ancien régime* in France, and to reach the establishment of SACEM in the 19th century. Therefore it is considered the history of the score publication, a French copyright system, the sudden rise of the management system of the music copyright.

キーワード: フランス著作権法, 音楽著作権, 楽譜出版, フランス革命, SACEM

French Copyright Law, Music Copyright, Score Publication, The French Revolution, SACEM

## 1. はじめに

本稿は知的財産制度の中の著作権制度、特には音楽著作権についての研究である。具体的には 19 世紀を中心としたフランスにおける音楽著作権成立の歴史について検討を行った。研究の対象としてフランスを取りあげたのは、フランスが法的に音楽著作物の権を世界で最も早い時期に保護した国であるからである。近世から近代にかけて、ヨーロッパ社会に

おいて音楽著作物に対するニーズが増大し、次第に社会性が増していくにつれ法制度的、社会的な仕組みがフランスを中心として立ち現れてきたということをふまえて、近代的な音楽著作権制度の発祥の地としてのフランスの事例をみていくことが重要である。

フランス著作権制度発展の背景には、ヨーロッパ社会における近代市民社会の成立があった。1789 年にはじまるフランス革命が大きな役割を果たした。19 世紀のフランスは、さまざま

な社会的な仕組みが大きく変動した時代であった。これは音楽の著作物についても同じである。音楽著作物が大量消費されていく中で、音楽著作権を集中的に管理する仕組みが誕生する。フランスは音楽著作権を集中管理する組織が世界で初めて誕生した国である。

本研究の特徴のひとつは、音楽著作物利用の社会的な側面の強調にある。これまでの著作権の研究は法制度の解説のみにとどまっていた、その社会的な背景についての研究が十分に行われていないと思われる。

## 2. 印刷楽譜の発達と音楽著作権

### 2.1 音楽の著作物としての楽譜

まず、「音楽の著作物」とは具体的に何を指すのかということを確認する。「著作物」というからには、その形態は表現されたモノでなければならないが、そのモノに当たるのは音楽の場合は楽譜である。現在のようにレコードやCDのような録音技術がまだ発明されていなかった時代において、音楽という無形のメロディーは楽譜というモノを媒介にして広まる。ただ、当初の楽譜は手書きであり、複製の方法も基本的に写筆であった。写筆楽譜は出版部数も限られ高価であり、楽譜が社会に影響を及ぼすほどに大量に出回るためには印刷技術の発達を待たなければならなかった。言語の著作物である書籍の大量印刷といった出版事業の隆盛は15世紀のゲーテンベルクによる活版印刷術の発明に端を発しているが、これは楽譜の出版についても同じである。ひとつの楽譜を大量に複製できる楽譜印刷術の創始は、音楽の歴史だけでなくヨーロッパ社会の文化に非常に大きな影響を与える。すなわち、書籍出版において印刷術の発展が比較的安価な書物の普及を後押しする中で、文字を読めるということが知的活動を営む上での必要条件となるのと同様に、教養ある人々にとって楽譜を読めるということが芸術的なステータスとな

るとともに、音楽的趣味の向上につながったのである。

このように出発地点において書籍の印刷術と歴史を共有している楽譜出版であるが、その発展は製造方法において独自の展開をみせている。これは楽譜の機能的な特性に起因する。楽譜というものは水平な五線と垂直の音符という構造になっているために、活版印刷の場合はその組み合わせの活字をすべて用意しなくてはならないということになる。これは、文字を主とした書籍にくらべてはるかに複雑で、技術的にも経済的にも非常に負担が大きかった。この状況は18世紀に彫版印刷というコストの低い技法が確立されるまで続くことになる。

### 2.2 楽譜出版業の成立と保護

つぎに楽譜出版業の状況である。アンシャン・レジーム、すなわち革命以前における楽譜出版は、国王や自治体から“印刷許可”を得て、その保護のもとで独占的に事業を展開していた。先に事業として成立していた書籍出版者の間では、国王から得た印刷許可による専売権の他に、さらに組合などによって互いの利益を保護していたが、これは独占的な利益を得るといった目的の他に、当時まだリスクの大きかった出版事業を同業者同士の競争や、いわゆる海賊版から保護するという目的があった。このような事情は楽譜出版業にとってはさらに深刻であった。技術的にも経済的にも負担の大きい楽譜出版事業におけるリスクは多大であったが、そのリスクに対して当時の楽譜の需要はまだ限られていたからである。楽譜を購入する人々というのはオーケストラを抱えるような裕福な人々に限られていたばかりか、楽譜を読める層も識字層に比べて非常に少なかったのである。このような状況の中で、印刷楽譜は長い間事業として成立することが難しく、印刷許可を得た出版者もごく少数に限られていた。例えば、パリのアドリアン・ル・ロワとその従兄弟にあたるロベール・バラールは、1553年にアンリ2世から王室印刷家として楽譜印

刷の独占権を受けてから、彼らの子孫たちはその後 2 世紀に渡って他の出版者の新規参入を許さなかったという<sup>[1]</sup>。

このように技術的にも社会的にも、まだ莫大な利益を生み出せるような事業として成立していなかった楽譜出版であるが、17 世紀から 18 世紀にかけてその状況は一変する。大崎<sup>[1]</sup>は楽譜出版の歴史の中で、18 世紀は印刷楽譜の社会的需要が増え、楽譜出版事業の態勢が整っていく過渡期であるとし、これは 18 世紀が「音楽の大量消費が始まった時代」ということに他ならないと指摘する。この頃になると、一般市民の楽譜の需要が増え、さらに印刷コストの低い彫版印刷術が発明された。彫版印刷術は簡単な道具で印刷が可能なることから印刷許可を得ていない出版者たちが暗躍することになるが、皮肉なことに結果的に楽譜出版は繁栄期を迎えるのである。正規の出版楽譜に対して非常に安価な海賊版といったものが出回るようになると、正規の楽譜出版者の利益を保護するための制度が必要になった。これは言語の著作物である書籍でも同じである。たとえばイギリスでは、出版業者たちは権力者に働きかけたことによって、著作者の権利を認めた「アン女王法(1710年制定)」が成立する。白田<sup>[2]</sup>は、アン女王法は、学問の振興のために創作者の権利を保護する名目で制定されたが、結局著作者がそこから利益を生み出すためには、出版者が本を出版しなくてはならないので、出版業者の利益は結果的に守られるというカラクリであったと指摘する。このような書籍出版者の行動は、後に楽譜出版者もまったく同じような行動をとる。先にも述べたが、出版事業として成功するまでに時間を要した音楽著作物は、法的な保護が書籍に比べて遅れることになる。フランスにおいて音楽著作物の著作者の権利が公式に認められるようになるのは、革命期の「1793 年法」であった。しかしながら、実際はそれまで出版者の権利であった複製権を著作者にいったん返上しながらも、後から著

作物に対するすべての権利を出版者に譲り渡すような方法をとる。音楽家は出版者との契約で楽譜を売り渡した際にその場で対価を受け取るのみで、その後のすべての権利を出版者に譲渡する場合がほとんどであった。

### 3 フランス著作権法の誕生と音楽著作権

#### 3.1 フランス著作権法の誕生と特徴

次に、フランス著作権法成立の歴史について述べる。フランス革命の起こる 1789 年は政治的に重要な政策が多く打ち出された。特にアンシャン・レジームにおける数々の特権の廃止(8 月 4 日)、人および市民の権利宣言(Déclaration des Droits de l'homme et du Citoyen de 1789)(8 月 26 日)、そしてさまざまな制度的枠組みが作られ、国民憲法会議で可決された。そのうちの重要な議題のひとつが「著作者の権利」の扱いであった。この時期にフランス著作権法は「1791 年法」と「1793 年法」という 2 つの法律が成立する。1957 年に著作権法が総合法として大々的に改正されるまでの 150 年以上の間、この 2 つの法律に基づいて積み重ねられた判例と、10 以上の個別の法改正のみがフランスの著作権制度の基本を形成することになる。もちろん、この間に著作権法を改正しようという動きがまったくなかったわけではない。フランス政府は法改正に向けて度々立法委員会を設置したが、その議論のほとんどは著作権の存続期間の延長についてであった。フランスでは 1793 年法によって著作権の存続期間は著作者の死後 10 年間に限定されることになったが、アンシャン・レジームで認められていた特権認可は、相続によって半永久的に継承することができた。そういった背景から宮澤<sup>[3]</sup>は、著作者の間には著作権の存続期間を、土地や家屋の所有権と同じように永久にしてもらいたいという願望が根強く残っていたため、その後の法改正運動は常に保護期間についての議論を中心に行われていくことにな

つたと指摘する。

### 3.2 音楽著作権

先に述べた通り、フランスにおいて音楽著作物の著作権の権利が法律として成文化されたのは「1793 年法」である。1793 年法は著作者の財産的利益に関する法律であり、7 月 19-24 日の法律、9 月 1 日の法律が含まれる。7 月 19-24 日の法律は著作者の権利と保護する著作物の種類の拡大を目的としているが、この法律によって文学的著作者、作曲家、画家、図案家(デザイナー)の著作者の権利が保障された(1 条)<sup>[4]</sup>。先行した「1791 年法」においても著作者の権利が認められたが、これは劇場で上演・演奏される著作物についてであった。1791 年法は興行に関する法律であり、劇場の自由と劇場で上演される著作物のあり方を定めたものであった。1791 年 1 月 13 日法では著作者の権利として、存命中は作品が劇場で無断に上演されるのを禁じることができるとした(3 条)。しかし、これはオペラをはじめとした劇場の著作物以外の文学的著作物、音楽の著作物などは保護の対象外であった。音楽の著作物については、劇場の著作物の一部として保護されることはあったが、ここでいう保護されなかったものは音楽単体の著作物、例えば音楽家によって作られる協奏曲などの作品や、シャンソンといった、単独の音楽著作物である。

## 4. 近代市民社会の発展と音楽著作物

### 4.1 フランス革命と音楽著作物

フランス革命が勃発すると、国民議会はアンシャン・レジームにおける数々の封建的特権を廃止する。近代市民社会の幕開けである。

音楽著作物である楽譜を利用するコンサートなど演奏会は、アンシャン・レジームでは王立アカデミーが与える特権として、許可されたもの以外の開催が禁じられていた。また、許可されたコンサートも、演奏できる曲目は制限され、

開催期間や時間までも細かく制限されていた。革命によってそれらの特権認可が廃止されると、ようやく私的なコンサートも比較的自由に開催できるようになり、かつての高貴で裕福な層を対象としていたものから、次第に一般の市民を対象とした世俗的なものとなり、パリでは古典的な格式張ったコンサートの開催頻度が減少する。アディライード・ド・プラーズ<sup>[5]</sup>は、その原因はコンサートの構成の質ではなく、革命の混乱からコンサート自体が開ける状況になかったことや、コンサート協会の主催者や裕福な聴衆はまさに革命の当事者として財産や身の危険を抱えていたことを指摘する。さらには、それまでのコンサートの代わりに街中のカフェやレストランなどで音楽を楽しむといったスタイルが主流となる。音楽が市民の娯楽として次第に浸透していくのである。一般市民の社会的な力が強まるにつれて、音楽の消費形態はカフェや大道芸などに、市民の興味が向いていくことになるが、その流れを後押ししたのが、近代市民社会の誕生とともに 18 世紀から 19 世紀かけて本格化していた音楽ジャーナリズムの台頭であった。

### 4.2 音楽ジャーナリズムの発展

近代市民社会の誕生とともに開花したジャーナリズムは、イギリスのコーヒー・ハウスが起源であると言われているが<sup>[6]</sup>、カフェは世の中で起きた出来事の情報交換場所であったということで、まさに都市の流行発信の拠点であった。このような中で、新聞や雑誌では市民の身近な娯楽としての音楽が次第に取り上げられるようになる。新作の告知、演奏会や舞踏会の広告を載せ、作品に対する批評記事や有名音楽家に焦点をあてた記事なども多く書かれるようになる。一般の新聞が音楽ジャーナリズムにおいて果たした役割も見逃せない。専門の音楽雑誌は情報の質と量で優れていたが、あまり教養のない人々にとっては難しいものであった。街で開催される音楽会や舞踏会の情報を手軽に知るには、街の新聞が手っ取り早か

った。そのような新聞を読むところがカフェであった。カフェには、何種類かの新聞や雑誌が備えつけられ、日常のさまざまな情報交換の場所となっていた。このようにして近代市民社会の発展とともに、人々が音楽に親しむ選択肢が広がり、ジャーナリズムはそのような動きを促すのに欠かせない存在となっていた。

また、そういったカフェ自体で簡単なコンサートが開かれることも多かった。そのようなコンサートは「カフェ・コンセール」と呼ばれている。カフェ・コンセールでは当時流行のシャンソンなどが演奏されたが、たいへんな人気となりパリでは最盛期に 100 軒あまりの店が営業されていたという<sup>[3]</sup>。

#### 4.3 著作権意識の高まりと音楽著作権の集中管理

カフェやレストランなどで著作者以外の演奏家や歌手が楽曲を利用することに対して、著作者に演奏利用料といったものが支払われることはなかった。著作者は楽譜の印税を受けたことはあったとしても、演奏利用料を受け取ることはなかったのである。革命以前には、作曲家の地位は大雑把に 1784 年 3 月 13 日の議会で、王立音楽アカデミーの劇場で演奏された楽曲の著作者の権利について定められていた<sup>[4]</sup>。また、劇場における上演権、演奏権は 1791 年法によって認められていたが、保護の対象となるのはあくまで劇場において上演される劇やその伴奏音楽、バレエやオペラといった劇場の著作物に限定されると解釈されていて、音楽の著作物は対象外であった。1822 年のパリ裁判所の判決によると、劇場以外の場所で著作者の許諾なく音楽著作物を演奏することは自由であり、シンフォニー、室内音楽などといった単独の音楽の著作物は保護の対象から除外されるということであった<sup>[3]</sup>。

そうした状況の中で、音楽著作物の利用を巡ってカフェの経営者と一部の著作者の間で騒動が起こる。有名な「カフェ・アンバサドル事件」である<sup>[7]</sup>。この事件は、作詞家ブルージェ (Ernest Bourget)、作曲家のアンリヨン (Paul

Henrion) とパリゾ (Victor Parizot) が、アンバサドルというカフェで飲食をした際に、彼らの楽曲が無断で演奏されていたことに対して抗議し、飲食代金の支払いを拒否したことに始まる。彼らの主張は、カフェの経営者たちは莫大な収益を上げて事業として成功していたのにもかかわらず、そこで演奏された音楽の演奏利用料が支払われることがなかったが、カフェの集客力は音楽を演奏したことによるから、著作者に対して報酬が支払われるべきであるということであった。この事件は訴訟になり、音楽出版社のコロンビエ (Jules Colombier) もブルージェらを支援、最終的に 1849 年 4 月 28 日パリ控訴院の判決において、作曲家や作詞家といった著作者たちの演奏権が保障されることによって一応の解決をみることとなる。

ところが、彼ら著作者側の主張が全面的に認められたとしても、それまで普通のこととして街中で演奏されていた音楽に対して、急に演奏利用料を徴収するとしたところでそれを著作者個人で行うことは現実的ではなかった。そこで著作者たちが共同して音楽著作権を集中的に管理する仕組みを誕生させる。1850 年 3 月 18 日、ついに作詞家、作曲家、音楽出版社らは連名で音楽作詞者作曲者出版者組合 (Syndicat des Auteurs, Compositeurs et Éditeurs de Musique) を創設した。そしてまもなく 1851 年 2 月 28 日に“組合 (Syndicat)” から“協会 (Société)” になる。音楽作詞者作曲者出版者協会 (Société des Auteurs, Compositeurs et Éditeurs de Musique)、すなわち今日まで続く SACEM の誕生である。

こうして、1791 年法の成立から半世紀以上経った後に、ようやく音楽著作物の演奏利用料を徴収するための組織として、SACEM は世界初の音楽著作権の集中管理団体となったが、その活動は、当初は激しい批判にさらされていた。音楽著作物の著作権を管理して、さらに演奏利用料を徴収するというのが世界で初めての試みであったこともあって、これまで自

由に音楽を利用してきた人々は、SACEMの活動に対して協力的でなく、先の控訴院の判決に対しても“常軌を逸した判決”と考え、依然無断使用し続けていたという<sup>[7]</sup>。これに対して、最終的には1852年6月24日の破毀院の判決において作詞者と作曲者の上演・演奏権に関する判例法規が確立され、さらには1860年9月11日の破毀院の判決ではSACEM自体の利用料の徴収と徴収率の裁量権について認められた<sup>[8]</sup>。これによりSACEMの活動が保障されることになる。

SACEMは国内でその存在が認められたことにより、まもなく徴収事業の範囲を海外に広げる。1869年のベルギーを皮切りに利用料徴収の代理店を次々と海外に設立していく。これは国内の楽曲が海外で利用される場合においても演奏利用料が徴収されて、著作者に分配されるようにするためということである。SACEMは20世紀の半ばまでの間に最大の徴収額を誇る著作権管理団体になった。北アフリカを含む27の地域に300の代理店と約1000人の駐在員を常駐させている。1901年以降はフランス植民地帝国を中心に、マダガスカルやインドシナ半島にまで範囲を拡げ、数年後には仏領西アフリカ(A.O.F.)、仏領赤道アフリカ(A.E.F.)まで代理店を置いている<sup>[7]</sup>。

SACEMは海外においても著作権利用料徴収の主導権を握り、徴収のネットワークを拡げることでより多くの使用料を徴収し、当時としては著作者の多くがこの制度から効率的に利益を享受できるようになった。当時はSACEMの他にこのようなシステムを持つ団体が存在しなかったことから、SACEMは著作権を集中的に管理するシステムを生み出した先駆者として、後に世界中で創設される著作権管理団体のモデルとなったのである。

## 5. おわりに

以上のように、革命期から19世紀末までのフ

ランスにおける音楽著作権の歴史を概観してきた。また、音楽著作物の歴史として15世紀に始まる楽譜出版業の成立についても取り扱った。そして、最後に音楽著作物を集中的に管理することで効率よく著作者に最大限に経済的な利益をもたらすための仕組みとして、SACEMという組織の設立に至った経緯を明らかにした。

こうした一連の歴史からわかることは、著作物の需要が増すことで、そこから初めて経済的な利益を得ようとする行動が起こるということである。逆に言うなら、需要が少なく経済的な利益が得られそうもないうちには、著作権者たちはそこからあえて経済的な利益を引き出そうとは思わないのである。その点、フランスにおいてはやはりフランス革命という市民革命が著作物の需要を大きく喚起する要因となったことは間違いない。

## 参考文献

- [1] 大崎滋生: 楽譜の文化史, 音楽之友社, 222p., 1993.
- [2] 白田秀彰: コピーライトの史的展開, 信山社, 431p., 1998.
- [3] 宮澤溥明: 著作権の誕生 フランス著作権史, 日本ユニ著作権センター, 309p., 1998.
- [4] Bertrand, André R.: Le droit d'auteur et les droits voisins, Masson, 796p., 1991.
- [5] アディライード・ド・プラーヌ(長谷川博史訳): 革命下のパリに音楽は流れる, 春秋社, 419p., 2002.
- [6] 西原稔: 音楽家の社会史 19世紀ヨーロッパの音楽生活, 音楽之友社, 212p., 1987.
- [7] SACEM: Centenaire de la SACEM 1850-1950, 80p., 1950.
- [8] ルモアール, J. J.(宮澤溥明訳): 「著作権とSACEM-SACEMの歴史-(2)」, 日本音楽著作権協会会報, 37号, pp.3-4, 1960.

# 科学技術文献の解析・可視化における辞書の活用について

## Usage of a Scientific Dictionary for Analysis and Visualization for Scientific and Technological Document

甲田彰<sup>1\*</sup>, 坂内悟<sup>2</sup>

Akira KOUDA, Satoru BANNAI

\*1 独立行政法人 科学技術振興機構 情報提供部 営業推進課

Sales Promotion Division, Department of Service, Japan Science and Technology Agency

E-mail:kouda@jst.go.jp

2 独立行政法人 科学技術振興機構 情報提供部 普及企画課

Marketing Division, Department of Service, Japan Science and Technology Agency

E-mail:sbannai@jst.go.jp

独立行政法人科学技術振興機構(JST)では、科学技術文献検索サービス「JDream II」の検索結果を出現頻度の経年変化や類似度の分析を行い、利用者に提供する「解析・可視化機能」の開発を進めている。JDream IIには出典がはっきりした文献データ約 4,000 万件を搭載しており、そのうちの 2,300 万件については、JST 自らが抄録、索引語を付与している。今回は、現在の JDream II に搭載されている分析機能である「頻度分析機能」を説明した上で、解析・可視化機能の概要や機関名辞書を用いた解析可視化機能の分析精度向上の工夫や今後の課題について報告する。

At Japan Science and Technology Agency (JST), we are developing a new function "Analysis and Visualization" for our document retrieval system "JDream II". In this report, I describe the frequency analysis function on the current JDream, and Analysis and Visualization using our organization dictionary. Future developments are also described.

キーワード: データベース, 検索システム, 解析, 可視化, 辞書

### 1.はじめに

科学技術振興機構(以下「JST」)は 2006 年 4 月から、新たな科学技術文献

検索サービス「JDream II」の提供を開始した。これは、1976 年にサービスを開始した JOIS と、2003 年に主に大学や病院等の利用者向けに開発しサービスを提

供してきた JDream を統合し、機能を飛躍的に向上させた新しいシステムである。

本稿では、まず、JDream II の概要と、現在 JDream II に搭載している「頻度分析機能」を説明する。また、頻度分析機能に基づき開発を進めている解析・可視化機能について、機関名辞書の活用による精度向上と今後の課題を説明する。

## 2 JDream II の概要

### 2.1 JDream II の搭載データ

JST では、国内外の科学技術分野の資料として、科学技術系のジャーナルを初め、学会誌、協会誌、企業・大学・独立行政法人・公設試験場等の技術報告、業界誌、臨床報告等の科学技術文献を収集し、科学技術文献検索サービス「JDream II」で提供している。JDream II に搭載されている文献データは約 4,000 万件で、そのうち、JST が収集した約 2,300 万件については、各文献データに、主題分析に基づく抄録

と統制された索引語を付与している。

### 2.2 検索における辞書の活用

科学技術文献中には様々な異表記語や同義語が存在しているため、利用者が思いついた単語のみで検索した場合、必ずしも網羅的に検索することは困難である。JDream II では、網羅的な検索を可能とするため、同義語辞書、異表記語辞書を検索支援機能として搭載している[1]。また、19年度中には、機関名辞書も搭載する予定である。

## 3 JDream II の頻度分析機能

### 3.1 頻度分析機能の概要

JDream II では検索結果の文献データ中の索引語や著者名、機関名等をカウントする「頻度分析機能」を提供している。図 1 に検索結果を著者所属機関で頻度分析した結果の画面例を記載する。



図 1 著者所属機関による頻度分析結果画面の例

### 3.2 頻度分析機能の活用方法

頻度分析機能を用いることにより、検索結果の文献データを一つ一つ参照することなく、全体の傾向を把握することが可能である。また、頻度分析の結果に対して、さらに頻度分析を実行することで、項目間の関連性を把握することが可能で

ある。図2に、著者所属機関で頻度分析した結果に対して、再度著者所属機関で頻度分析を行った場合の例を記載する。

こうした分析を行うことで、ある概念と別の概念の文献における関連性や、機関と科学技術分野との関連性などを把握することが可能である。

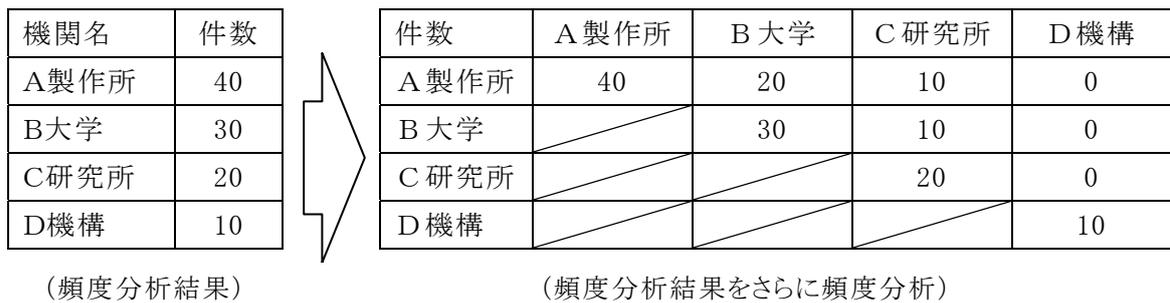


図2 頻度分析結果に対する再分析の例

## 4 解析・可視化機能の開発

### 4.1 解析・可視化機能開発の背景

#### 4.1.1 文献検索に対する環境の変化

これまで、JSTの文献検索サービスでは、「漏れなく検索する」ことを重視し、サービスを提供してきた。しかしながら、研究開発のスピードアップ、ボーダレス化が進む一方で、研究開発の現場で扱う情報の量は過去に比べ格段に増加している。こうした状況の変化により、研究開発の現場では、文献をひとつひとつ参照するのに加え、文献の全体的な傾向を迅速に把握する必要性が生じている。

#### 4.1.1 JSTの文献データの特徴を活かした取り組み

JDream II に搭載した文献データは前記 2.1 の通り、

- a. 出典が明らか
- b. 幅広い範囲を収集

c. 抄録と統制された索引語を人手で付与という特徴があり、全体的な傾向を正確に把握するのに適している。JDream II の利用者からは「頻度分析機能を拡張し、解析し可視化する機能」を求める声が寄せられている。

### 4.2 解析・可視化機能の概要

以上の背景のもと、JSTでは、検索結果の文献データ中の索引語や機関名の出現頻度や共出現回数を分析し、グラフやマップの形式で提供する「解析・可視化機能」の開発を進めることとした。

#### a. 統計グラフ

機関や索引語の出現頻度を経年グラフの形で表現したもの。

#### b. 概念マップ

索引語の共出現や索引語と機関名の共出現をスケルトンマップで表現したもの。図3に索引語の共出現を表現したスケルトンマップの例を記載する。

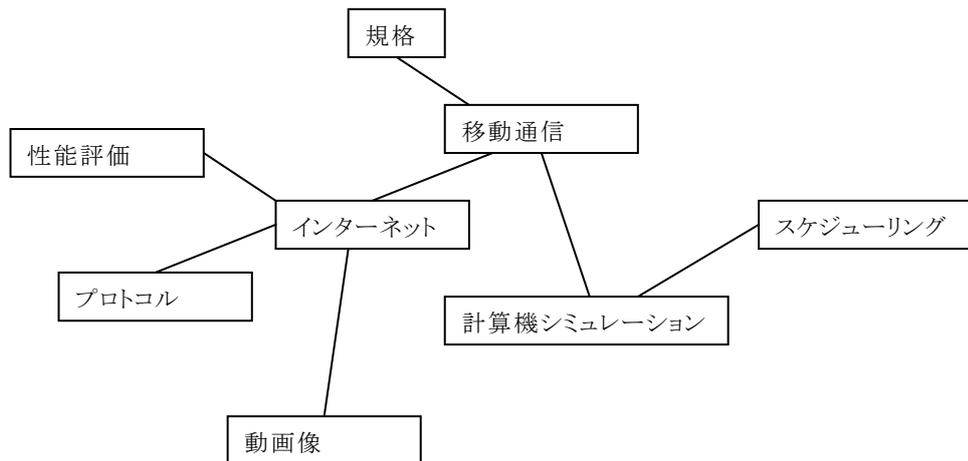


図3 スケルトンマップ(索引語)の例

#### 4.3 精度向上のための方策

##### 4.3.1 異表記による分析精度の低下

前記 2.1 の通り JDream II の文献データには統制された索引語が付加されているので、これを用いることで、科学技術用語については精度の高い頻度分析が可能である。

一方、著者所属機関の場合は、同一機関であっても様々な記述が存在するため、別の機関としてカウントしてしまい、出現頻度が正確に算出できない、という問題がある。

##### 4.3.2 機関名辞書の利用による分析精度の向上

このため、19年度に JDream II に検索支援機能として搭載する予定の機関名辞書を用いて、機関名の変換したうえで解析・可視化を行うことを予定している。JST の機関名辞書には、組織名(第1階層)に加え、部署名(第2階層)、部門名(第3階層)やそれらの履歴として、約10万個の機関名データが登録されている。

なお、日本語機関名の場合は大学名・学部名・学科名等のように、先頭に第1階層の機

関名の記載があるが、英語表記の場合、必ずしも先頭に第1階層の組織名が記載されていないため、機関名の表記を、(カンマ)や.(ピリオド)等で分解してから辞書を適用し分析する予定である。

## 5 今後の課題

企業の場合には資本提携のような名称からは判断できないような関係があり、辞書で分析精度の向上に対応することには限界がある。こうした情報も解析・可視化に反映するためには、将来的には利用者が機関間の関連性情報を登録し解析・可視化に反映できるような機能が必要と考えられる。

### 参考文献

[1]甲田彰,森田歌子. 科学技術文献検索システムにおける大規模用語辞書の活用について, 情報知識学会誌, vol. 16, no.2 P.45-48

## ユーザの Web 閲覧履歴を用いた検索支援システム

Improving Web Retrieval Precision  
base on User Web browsing behavior

堀 幸雄<sup>†</sup>                      今井 慈朗<sup>¶</sup>                      中山 堯<sup>‡</sup>  
Yukio HORI                      Yoshiro IMAI                      Takashi NAKAYAMA

The information on the Web is increasing explosively, and the Internet/Web has become the information infrastructure of our own time. Web users gather and find information through search engines using keywords related to their interests. However, it is not always the case that users conceive suitable keywords relevant to the problem in question. In this paper, we propose a novel method for query expansion based on the user profile generated from browsing histories of a user. This enables a user the context-sensitive keyword expansion using profiles specific to the user. Here, we assume that a user's web browsing behavior corresponds to the user's intended query. We illustrate the usefulness of our proposed method for query expansion with some examples.

## 1 はじめに

近年あらゆる情報が Web 上で提供されるようになり, Web 上で提供される情報の効率的な検索に対する要求はますます高まっている. 多くのユーザは自分の欲しい情報を探すために検索エンジンを利用している. 検索エンジンのうち主流となっているのは, キーワードに基づく検索システムである. ユーザは自分が調べたい事柄を検索語として入力することで, 様々な情報を得ることができる. しかし従来から指摘されているように, ユーザがいつも自分の知りたい情報に関連する適切なキーワードを知っているとは限らない<sup>[1]</sup>.

本稿はユーザの Web 閲覧履歴に出現する単語をクラスタリングし, その結果を検索語の拡張に利用する手法を提案する. 閲覧履歴を元に検索語の拡張を行なうのは, ユーザのページ閲覧行動と検索意図には何らかの関係があると考えられるからである. また提案手法を実際のユーザの Web 検索セッションに適用し, 検索語の拡張がユーザの所望のページの検索結果順位をどの程度向上させるかの評価実験を行なった.

## 2 関連研究

Google や Yahoo! などの検索エンジンがユーザに対して検索語に関連する関連語を提案する「Google サジェスト」<sup>1</sup>、「関連検索ワード」<sup>2</sup>と呼ばれるサービスがある. それぞれユーザが入力した検索語に対し, 想定される検索語や関連するキーワードを提案する機能である. 関連するキーワードを得る方法について詳細な情報は公開されていないが, 不特定多数のユーザの検索語ログや Web ページ中に出現するキーワードの様相などの要因を元に関連語を得ていると思われる. また大塚らはパネルログと呼ばれる大規模アクセスログをもとに, 検索語の関連語群を得るシステムを提案している<sup>[4]</sup>. この研究はアクセスログからサイト名, ウェブコミュニティ, ページに出現する単語などの情報を用いて関連語を抽出する. これらの手法は不特定多数の検索語ログや閲覧履歴を用いているが, これに対して本研究はユーザ個人の閲覧履歴を元にしている点が異なる.

河重らはユーザ個人の閲覧履歴の内容を用いることで意図を考慮した Web 検索手法を提案している<sup>[5]</sup>. この研究では検索結果, 閲覧履歴における検索語の周辺テキストから検索語の拡張を行なっている. Lev Finkelstein らはユーザが Web 閲覧中に選択したリンク周辺から単語を抽出し, その単語群を用いて検索語の拡張を行なう手法を提案している<sup>[2]</sup>. これらの研究はユーザ個人の閲覧履歴という情報をベースにし

<sup>†</sup>香川大学情報基盤センター Information Technology Center, Kagawa University horiyuki@itc.kagawa-u.ac.jp

<sup>¶</sup>香川大学工学部 Faculty of Engineering, Kagawa University imai@eng.kagawa-u.ac.jp

<sup>‡</sup>神奈川大学理学部 Faculty of Science, Kanagawa University, nakayama@info.kanagawa-u.ac.jp

<sup>1</sup><http://www.google.co.jp/webhp?complete=1&hl=ja>

<sup>2</sup><http://developer.yahoo.co.jp/search/webunit/V1/webunitSearch.html>

ている点は同じであるが、単語群の抽出の仕方が本研究と異なる。また検索語の拡張により検索を行なった場合の評価が行なわれていない。

### 3 提案手法

ここではユーザの閲覧履歴を元に検索語に関連する拡張語を得る手法について述べる。一般にユーザは様々な目的で Web ページの閲覧を行っており、このような Web 閲覧履歴をそのまま検索語の拡張に用いても異なる目的で入力されたキーワードが混在するため良い結果は得られない。そこでユーザの Web 閲覧履歴に出現する単語を自己組織化マップ<sup>[8]</sup>を用いてクラスタリングを行ない、形成されたクラスタに着目した検索語拡張を行なう手法を提案する。ユーザの閲覧履歴から検索語拡張を行なうまでの流れを図 1 に示す。

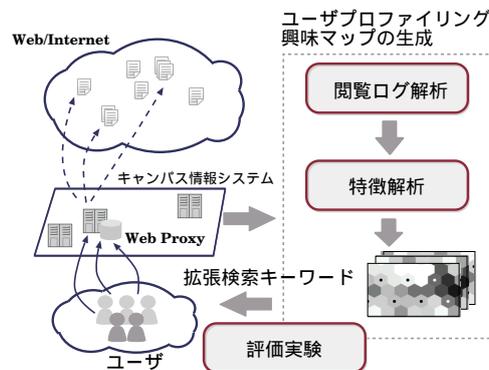


図 1: 検索語の拡張システム 概要

#### 3.1 閲覧履歴における特徴空間の定義

まず各ユーザの Web 閲覧履歴を Web プロキシサーバのログ情報から獲得する。このログ情報には全ての対外 Web アクセスに関するクライアント IP アドレス、HTTP コマンド (GET|POST)、日時、URL、ステータスコードが記載されている。クライアント IP アドレスからその時刻にその IP アドレスを利用していたユーザ ID を得る。この情報を元にユーザ  $u$  がある期間  $span$  において閲覧した Web ページ集合  $B_u$  を得る。

$B_u$  に対し、形態素解析を行ない、名詞だけを抽出する<sup>3</sup>。Web サイト  $j^4$  で得られた単語  $t_i (i = 1, \dots, N)$  の重みを  $w_{ij}$  とし、閲覧した Web サイト ( $j = 1, \dots, M$ ) により、 $M$  次元ベクトル  $w_i = (w_{i1}, w_{i2}, \dots, w_{iM})^T$  として表現する<sup>[9]</sup>。単語の重み付けは tf-idf 法を用いた。

$$w_{ij} = \text{tf}(i, j) \cdot \text{idf}(i)$$

$$\text{idf}(i) = \log(M/\text{df}(i))$$

ただし、 $\text{tf}(i, j)$  はサイト  $j$  における単語  $t_i$  の出現回数であり、 $M$  は総閲覧 Web サイト数、 $\text{df}(i)$  は単語  $t_i$  が 1 回以上出現する Web サイト数である。

またストップワード処理として下記のもの除外した。

1. ひらがな、カタカナ 1 文字の単語: 形態素解析に失敗している可能性が高く、検索語の拡張語候補に使う可能性が少ないために除外する。
2. 低頻度語を除外: 上位 70% までの単語を採用した。

<sup>3</sup>具体的には名詞・一般、名詞・固有名詞、名詞・副詞可能、名詞・形容動詞語幹、名詞・サ変接続を対象とした。

<sup>4</sup>ここではドメイン名ごとを個別のサイトとした。

### 3.2 ユーザプロファイリング作成

ユーザの閲覧履歴中から得られる特徴ベクトル  $w_i$  は多次元ベクトルで構成されるため、このままでは検索語に関連するキーワードを求めるのが困難である。そこで類似する単語同士をクラスターとして抽出する。自己組織化マップを用いてユーザの閲覧した Web ページの特徴相互の距離関係を可能な限り保持した状態でこの特徴ベクトルを 2 次元平面上に写像する。

自己組織化マップは 2 層のニューラルネットワークで構成され、入力層に入力される特徴ベクトルを  $w_i$ 、出力層のノードに連結される参照ベクトルを、

$$m_k = \{\mu_{k1}, \mu_{k2}, \dots, \mu_{kM}\}$$

とする。自己組織化マップ生成の手順は次の通りである。

1. 参照ベクトル  $m_k$  を初期化する。
2. 次式により入力  $w_i$  に対する最近傍ノード  $c$  を決定する。

$$c = \arg \min_k \|m_k - w_i\|$$

3. 最近傍ノード  $c$  とその近傍ノードを更新する。

$$m_k(t+1) = h_{ck}(m_k(t) - w_i(t))$$

$$h_{ck} = \alpha \exp(-(\|r_k - r_c\|^2)/2\sigma^2)$$

ここで  $\alpha$  は学習係数であり、 $h_{ck}$  は近傍関数である。 $r_k, r_c$  はそれぞれノード  $k, c$  の位置ベクトルを現わす。

以上の手順により 2 次元平面上に射影されたユーザの閲覧履歴中に出現する単語をまとめたマップをユーザプロファイルとする。得られたユーザプロファイルマップの例を図 2 に示す。図 2 の例ではマップ右上に就職活動関連の単語、左下にショッピング関連の単語といったように閲覧履歴中で類似した単語が表出している。

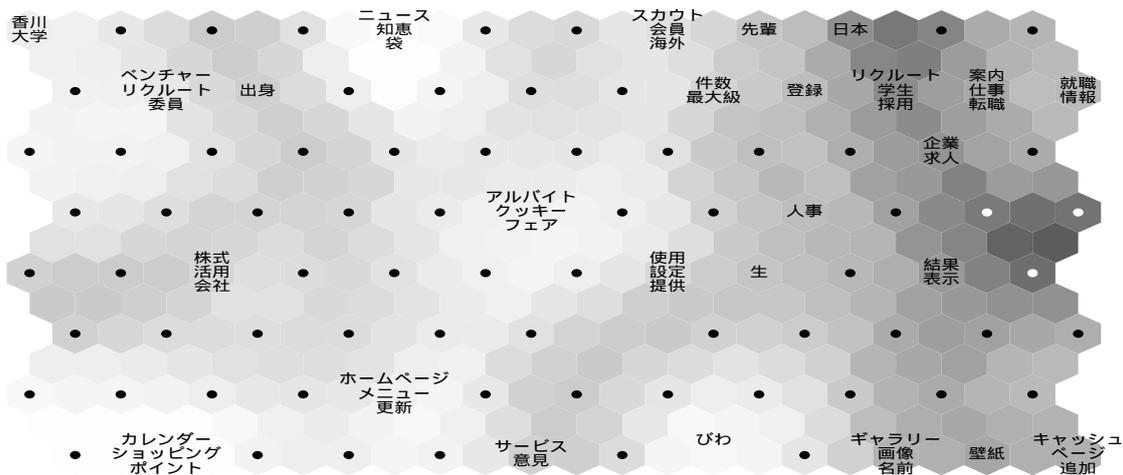


図 2: ユーザプロファイルマップ

### 3.3 検索語から関連語の獲得

ユーザが検索エンジンに投入した検索語  $K$  をそのユーザのマップに当てはめることで  $K$  に関連する拡張語  $K'$  を得る。今回ユーザごとのマップを生成するための閲覧期間  $span$  を一日、ユーザのプロファイルマップのサイズを  $12 \times 8$  とし、マップ上の 1 つの要素に最大 5 つのキーワードを割当てることとした。したがって閲覧履歴から最大 480 のキーワードがマップに表出することになる。

## 4 評価実験

### 4.1 実験方法

各ユーザのプロファイルマップを用いて、Web 検索時の検索語を拡張し、その有効性を確かめる実験を行なった。拡張語を得るプロファイルマップは、検索時において最新のマップを使うこととした。

検索語をユーザプロファイルマップに当てはめて得られる拡張語の有効性を考える上で、実際のユーザの検索セッションに着目する。通常ユーザは次の手順で検索エンジンを利用する。

1. 検索語  $K$  を検索エンジンに投入する。
2. 検索結果から所望のページ  $LP$ (Landing Page) をクリックする。このとき  $LP$  の検索結果順位を  $R$  とする。

この検索セッションに着目し、検索順位  $5 \leq R \leq 50$  へ遷移した検索セッションについて検索語  $K$  をユーザプロファイルマップに当てはめ関連語  $K'$  を得る。 $R$  を  $5 \sim 50$  と設定したのは、ユーザは所望のページを検索語に絞り込めていないと思われるからである。

学内の検索セッションにおける検索順位とその頻度の関係を図 3 に示す。調査期間は 2006 年 7 月 1 日～31 日の 1 ヶ月間である。

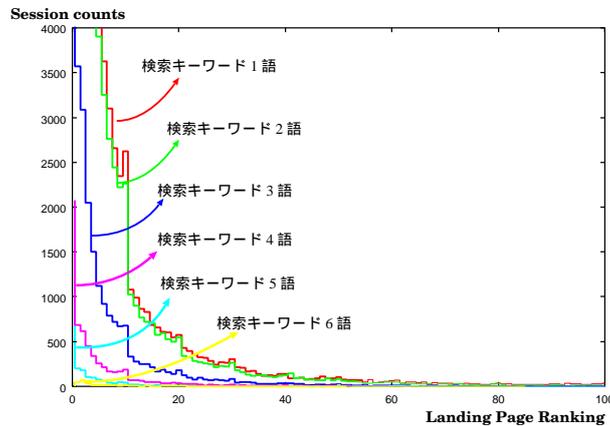


図 3: Web 検索セッションにおける検索順位-度数分布

検索エンジンにキーワード投入から他ページへの遷移したセッション数は 193212 あり、このうち検索結果の上位 5 位以内のページに遷移したセッション数は全体の 59% であり、検索結果の 5 位から 50 位までのページに遷移したセッション数は全体の 39% であった。すなわち半数弱のセッションにおいてユーザは自分の探したい情報に容易に辿りつけていないと思われる。

次に  $K + K'$  で検索した場合の  $LP$  の検索順位  $R'$  を調べる。このとき改善度  $A = R - R'$  とし、 $A > 0$  ならば  $K'$  により所望のページ  $LP$  が  $K$  で検索したときに比べ、検索結果の上位に表示されたことになる。また  $5 \leq R \leq 50$  の検索セッションにおいて  $A > 0$  の割合である改善率  $AP$  を求め、 $K'$  が有効に働く割合を調べた。

実験対象とした期間、検索セッション数を表 1 に示す。

表 1: 実験対象データ

|                               |                |
|-------------------------------|----------------|
| 実験期間:                         | 2006-7-02 ~ 31 |
| 全検索セッション数:                    | 259,758        |
| $5 \leq R \leq 50$ の検索セッション数: | 94,675         |
| ユニークユーザ数:                     | 3,290          |
| 検索語の種類:                       | 40,811         |

## 4.2 実験結果

実験結果を表 2 に示す．検索セッションについて 3 割程度の場合において本手法による検索語の拡張が有効であることがわかった．

表 2: 実験結果

|          |       |
|----------|-------|
| 改善度 $A$  | 10.04 |
| 改善率 $AP$ | 0.29  |

表 3 に  $K'$  が有効に働いたいくつかの例を示す．

表 3: キーワード拡張例

| $K$       | $K'$   | $R$ | $R'$ |
|-----------|--------|-----|------|
| イタリア 人口問題 | 医療     | 17  | 1    |
| ロボット開発 歴史 | コンピュータ | 8   | 1    |
| ベルリン オペラ  | ウィーン   | 14  | 7    |
| 直島 香川     | 岡山     | 5   | 3    |
| 個性化教育     | 興味     | 40  | 7    |

「イタリア 人口問題」を検索語にしていたケースでは，ユーザは社会問題に関連するページを閲覧していたため「医療」といった人口問題と関連するキーワードが拡張語に選ばれている．同様に「ロボット開発 歴史」を検索語にしたケースでは，「ロボット」，「開発」という単語から「コンピュータ」という拡張語を得ることでユーザの目的のページを絞り込んでいる．また「ベルリン オペラ」という検索語を用いたケースでは，ユーザは世界各地の地名が出現するページを見ていたため，意味的にあまり関連性のない「ウィーン」という単語が拡張語に選ばれている．このように本手法は閲覧履歴から検索語に類似する単語を拡張語とするため，意味的には関連が強いとは限らない．

## 4.3 考察

ここで本手法で得られた拡張語と Google 検索で得られる関連語と Overture が提供するキーワードアドバンスツール<sup>5</sup> から得られる関連語を比較した．表 3 の検索語に対して，提案手法，Google，Overture によって得られる関連語のリストを表 4 に示す．

表 4: キーワード拡張の比較

| $K$       | 本手法                | Google           | Overture         |
|-----------|--------------------|------------------|------------------|
| イタリア 人口問題 | 医療 家族 婚姻           | 食料 対策 先進国        | 研究所 社会保証         |
| ロボット開発 歴史 | コンピュータ ユーザ ラジオ アニメ | 大企業 発表 会社        | 歴史 年表            |
| ベルリン オペラ  | ウィーン ドイツ語 エリア      | 服装 ペトレンコ 鑑賞      | ドイツ 国立 劇場 スケジュール |
| 直島 香川     | 岡山 イベント まつり 花火     | フェリー シーサイドパーク 宿泊 | 旅行 アート ホテル       |
| 個性化教育     | 興味 活用 喜び 技術        | 生活科 小学校 PDF      | 全国 連盟            |

Google や Overture からは「人口問題」に対して「食料，対策，社会保証」，「ロボット開発」に対して「発表，歴史，年表」といった一般的に関連の高いと思われる単語を得ることができる．また本手法から得られる拡張語のリストは既存のサービスと比較して同等またはそれ以上の関連語を示していることがわかる．

また「個性化教育」というマイナーな検索語に対して，Google や Overture はあまり関連のある単語を出してはいないと思われる．それに対し提案手法では「興味 喜び 活用」といった教育に関連する単語を得ている．提案手法ではマイナーな単語からも閲覧履歴に類似する単語を選ぶことで関連語を正しく得ていると思われる．

<sup>5</sup><http://inventory.jp.overture.com/>

## 5 おわりに

本稿ではユーザのウェブ閲覧履歴を用いて、検索語の拡張を行なう手法の提案を行なった。そして提案手法が検索セッションにおいてどの程度有効に働くかどうかを評価した。また不特定多数のユーザのクエリログより関連語を得るシステムとの比較について考察を行なった。本システムは検索語がマイナーな場合においても個人の文脈に対応した関連語を得られることがわかった。

今後の課題としてユーザプロファイルを作成する期間の調整、閲覧セッション単位での閲覧履歴のモデル化、検索語の拡張だけでなくキーワード間の上位下位関係についての発見などが挙げられる。

## 謝辞

本研究の一部は平成 19 年度香川大学若手研究 (萌芽研究) 経費によるものです。ここに記して謝意を表します。

## 参考文献

- [1] P. Ingwersen: *Information Retrieval Interaction*, 邦訳: 細野公男他, 情報検索研究 -認知的アプローチ-, トッパン, 1995.  
.33-40, 2006.
- [2] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman and E. Ruppim.: Placing search in context: The concept revisited, In Proceedings of the 10th WWW Conference, 2001.
- [3] 原田昌紀, 清水奨: WWW 検索システムにおける不特定多数の操作履歴の活用, 情報処理学会研究報告, 97-OS-74/97-DPS-81, pp.61-66, 1997.
- [4] 大塚真吾, 喜連川 優: 大規模アクセスログを用いた検索支援システムの提案, 日本データベース学会 Letters Vol.5, No.1, pp.13-16, 2006.
- [5] 河重 貴洋, 大島 裕明, 小山 聡, 田中 克己: 質問修正と再ランキングを用いた文脈依存 Web 検索, 電子情報通信学会, 第 17 回データ工学ワークショップ (DEWS 2006), 3C-i14, 2006.
- [6] 山口雅史, 大島裕明, 小山 聡, 田中克己: サーチエンジンのクエリログを利用した同位語の発見, 日本データベース学会 Letters Vol.5, No.2, pp .17-20, 2006
- [7] 安川美智子, 山田 篤: Web 検索エンジンを用いた用語検索履歴からのシソーラス自動構築, 日本データベース学会 Letters Vol.3, No.1, pp.105-108, 2004.
- [8] Kohonen, T.: *Self-Organizing Maps*, 3rd Edition, Springer Verlag, 2001.
- [9] Salton, G. and McGill, M. J.: *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.

# 古典籍説明文からのドメイン知識の抽出

## Extraction of Domain Knowledge from the Explanations of Japanese Classics

長塚 隆<sup>\*1</sup>, 神門典子<sup>2</sup>

Takashi Nagatsuka, Noriko Kando

\*1 鶴見大学

Dept. of Library, Archival and Information Studies, Tsurumi University, Tsurumi 2-1-3, Tsurumi-ku, Yokohama, Japan  
230-8501 E-mail:nagatsuka-t@tsurumi-u.ac.jp

2 国立情報学研究所

National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan 101-8430 E-mail:kando@nii.ac.jp

鶴見大学図書館で公開している「古典籍」の説明文を対象にメタデータの分析を行い、抽出したメタデータをグループ化した。さらに、基本的な古典籍の形態である「写本」、「版本」、「切れ」、および「その他の資料」に分け、メタデータの違いについて比較したところそれぞれの古典籍の形態が反映されていることが明らかとなった。古典籍説明文について、ドメイン知識の視点から整理を試みた。

The metadata elements were extracted from the explanations of Japanese Classics on the web of Tsurumi University Library. They were divided into several groups and compared among four types of Japanese Classics, the manuscript, the printed book, the detached segment, and the other. It presents a clear difference of the metadata elements among four types of Japanese Classics. We discussed the difference of them from an aspect of domain knowledge.

キーワード: 古典籍, 説明文, ドメイン知識, オントロジ, メタデータ, Japanese Classics, explanation, domain knowledge, ontology, metadata

## 1 はじめに

文化遺産をデジタル化し、インターネットで公開するさいに、多様な文化遺産に関する情報により効率的にアクセスできるための試みがなされている。例えば、博物館情報を用いて複合的メタデータスキーマを構築する手法についての検討<sup>1</sup>、文化遺産のメタデータセットに従ってデータを分類する手法について研究し意味論的な相互運用の管理に関するデータ分類管理についての検討<sup>2</sup>、あるいは、木下らによる民俗学研究資料の電子化を行い民俗

学分野に特化したシソーラスとOntologyの関連性についての検討<sup>3</sup>など、様々な試みがなされている。

また、最近では、文化遺産をデジタル化しインターネットで公開に際して、専門家により該当の文化遺産を紹介する説明文を付与するケースが増加している。しかし、多くの説明文はその分野についての専門的な知識を前提としているものが多い。

本報告では、江戸時代末までに日本で著された古い書物、いわゆる「古典籍」の説明文についてどのようなメタデータを付与すべきかを

検討し、自動的にメタデータを付与するための要件について考察した。

## 2 古典籍の説明文情報

古典籍については、様々な角度から多くの研究がなされている。古典籍についての書誌情報データベースも作成され、メタデータもそれぞれのデータベースの特徴にあわせて付与されている。

一方で、古典籍の一般への理解や普及などのために、大学、研究所や博物館などで多くの展示会が開催されている。これらの展示会では展示品を広く理解してもらうために、作品を紹介する説明文が作成されることが多くなっている(表1)。

表1 Webで提供される古典籍データベースと古典籍説明文

| 利用の用途 | 区分       | 提供名称              | 提供機関              |
|-------|----------|-------------------|-------------------|
| 専門家向け | 書誌・目録    | NDL-OPAC          | 国立国会図書館           |
|       |          | 日本古典籍総合目録         | 国文学研究資料館          |
|       | 書誌・目録・画像 | 日本古典資料調査データベース    | 早稲田大学図書館          |
|       |          | 古典籍総合データベース       | 名古屋大学図書館          |
|       | 書誌・目録・内容 | 古典籍内容記述的データベース    | 名古屋大学図書館          |
| 一般向け  | 書誌・画像    | 館蔵品一覧             | 東京国立博物館           |
|       |          | 収蔵品データベース         | 京都国立博物館           |
|       | 説明文      | 奈良国立博物館所蔵写真検索システム | 奈良国立博物館           |
|       |          | 貴重書展              | 鶴見大学図書館           |
| 子供向け  | 説明文・画像   | 文化遺産オンライン         | 文化庁               |
|       |          | ヴァーチャル展示          | 国文学研究資料館          |
|       |          | 貴重書展              | 鶴見大学図書館           |
|       |          | 鶴見大学コレクション        | Global Memory Net |
|       |          | 館蔵品百選             | 名古屋市立博物館          |
| 子供向け  | 説明文・画像   | 名品紹介              | 奈良国立博物館           |
|       |          | お子様向けDictionary   | 京都国立博物館           |

「古典籍」についての書誌情報は、表1にあるように古典籍を所蔵している国会図書館や大学・研究機関・博物館・美術館などで、図書目録が作成され、インターネットで公開されている。

これらの中には国文学研究資料館の「日本古典資料調査データベース」や早稲田大学図書館の「古典籍総合データベース」のように、目録情報だけでなく書物の画像情報とあわせて提供しているところもある。

また、名古屋大学図書館では個々の古典籍の内容も記述する「古典籍内容記述的データベース」を提供している。この他に博物館や美術館では、東京国立博物館の「館蔵品一覧」、京都国立博物館「収蔵品データベース」、奈良国立博物館「所蔵写真検索システ

ム」などが所蔵品の書誌と画像を提供している。

最近では、これらの専門的なメタデータ情報の提示だけでなく、古典籍の一般への理解や普及などのために、大学、研究所や博物館などで多くの展示会が開催されている。これらの展示会では展示品を広く理解してもらうために、作品を紹介する説明文が作成され、インターネットで公開されることも多くなっている。

インターネットで公開されている古典籍の説明文は、まだそれほど多くないが、表1で示したように広がりつつある。

しかし、多くの展示会では古典籍の紹介を印刷体のみで行っている。これは、博物館などで、他の所有者が所蔵するものを展示することが多いことも影響しているようである。



図1 鶴見大学コレクション (Global Memory Net)

これらの説明文は画像と説明文とが組み合わされているケースが多いが、説明文のみのケースもある(表1)。

図1のように、筆者の一人は鶴見大学図書館が所蔵する古典籍や古地図などをデジタル化し、インターネットで提供するに際して、画像とともにメタデータを付与して説明文だけより、内容の理解がし易いようにした<sup>4, 5</sup>。

### 3 メタデータ項目の比較

表1に示した提供データベースのメタデータ項目を比較した。

これらのデータベースは基本的なメタデータ項目のみからなる文化遺産オンラインや早稲田大学図書館の「古典籍総合データベース」などから古典籍に特化した詳細なメタデータが付与された国文学研究資料館の「古典籍総合目録データベース」・「日本古典資料調査データベース」や名古屋大学図書館の「古典籍内容記述的データベース」まで多様である。多くのデータベースで共通するメタデータとしては①「書名」、②「著者／作者」、③「成立年」、④「形態」などであった。

国文学研究資料館の「日本古典資料調査データベース」や名古屋大学図書館の「古典籍内容記述的データベース」ではこの他に書誌学的な事項からの多くのメタデータが付与されていた。

一方で、文化遺産オンラインやGlobal Memory Netでの鶴見大学コレクションなどでは、他では余り付与されていない「材料/材質」や地理的区分なども付与されていた。これらとダブリンコアの基本メタデータセットとの比較も行った。

### 4 メタデータの抽出

比較的多くの古典籍説明文がネット上に掲載されている鶴見大学図書館の古典籍説明文からメタデータをマニュアルで抽出した。

本報告で使用した鶴見大学古典籍説明文は、最近開催された古典籍についての展示会の際に、来場者への説明のために主として鶴見大学日本文学科の教員により作成されたものである。また、一部は出版物として、参加者に配布されたものもある。

本報告では、「第110回鶴見大学貴重書展－歌書の版本」、「第109回鶴見大学貴重書展－曾我物語とその周辺」、「第108回鶴見大

表2 説明文からの抽出メタデータの階層化 (全38レコード)

| 大項目番号 | レベル-1     | レベル-2       | 出現レコード数 | 全レコード中の出現率 | 共通メタデータ |
|-------|-----------|-------------|---------|------------|---------|
| 1     | 書名(タイトル)  |             | 38      | 100%       | ○       |
|       |           | 副書名(サブタイトル) | 8       | 21%        |         |
| 2     | 著者/作者/筆者  |             | 18      | 47%        | ○       |
|       |           | 著者/作者/筆者の説明 | 7       | 18%        |         |
|       |           | 系統・出版社の説明   | 12      | 32%        |         |
| 3     | 所蔵者       |             | 6       | 16%        | ○       |
| 4     | 写本/版本     |             | 36      | 95%        |         |
|       |           | 写本の系統       | 4       | 11%        |         |
|       |           | 版本の種類       | 11      | 29%        |         |
|       |           | 版面          | 2       | 5%         |         |
| 5     | 成立年       |             | 35      | 92%        | ○       |
|       |           | 成立年推定の説明    | 3       | 8%         |         |
| 6     | 寸法(表紙縦横)  |             | 35      | 92%        | ○       |
|       |           | 形態          | 6       | 16%        |         |
| 7     | 書型        |             | 5       | 13%        |         |
| 8     | 数量(巻数)    |             | 33      | 87%        | ○       |
| 9     | 装訂        |             | 28      | 74%        |         |
|       |           | 装訂の説明       | 19      | 50%        |         |
| 10    | 由来        |             | 24      | 63%        | ○       |
| 11    | 料紙        |             | 22      | 58%        |         |
|       |           | 料紙の説明       | 5       | 13%        |         |
| 12    | 本の各部      |             | 16      | 42%        |         |
|       |           | 蔵書印         | 1       | 3%         |         |
|       |           | 釈文          | 5       | 13%        |         |
|       |           | 極め札:極め:折紙   | 7       | 18%        |         |
| 13    | 内容の説明     |             | 14      | 37%        | ○       |
|       |           | 原本の紹介       | 5       | 13%        |         |
| 14    | 絵の説明      |             | 4       | 11%        |         |
| 15    | 切れ/断簡     |             | 6       | 16%        |         |
|       |           | 切れ/断簡の説     | 6       | 16%        |         |
|       |           | 切れの系統       | 1       | 3%         |         |
| 16    | 塗箱(保存用の箱) |             | 2       | 5%         |         |
| 17    | 参照ページ     |             | 12      | 32%        |         |
| その他   | 展示会内の小区分  |             | 38      | 100%       |         |
|       |           | 書風          | 3       | 8%         |         |
|       |           | 参考          | 2       | 5%         |         |
|       |           | 利用目的        | 1       | 3%         |         |
|       |           | 残存状態(残欠)    |         |            |         |
|       |           | 叢書合冊目録分類    |         |            |         |

学貴重書展－物語の古筆切」、「第105回鶴見大学貴重書展示 古今集1100年」、「第99回鶴見大学 特別展示 古典籍の魅力」の各展示会からの説明文122件とGlobal Memory Netに収録の109件、合計231件の説明文を使用した。

古典籍の説明文231レコードから、メタデータの解析をするために、各展示会の資料区分ごとに最低1レコードとなるように、説明文38レ

コードを抽出した。これは、全レコード231件の15%強に相当する。

表1の各種データベースのメタデータ項目から、基本メタデータ項目および古典籍の説明に必要と考えられる拡張メタデータ項目とを選択した。表2に示したように、この他に説明文中より新たなメタデータ項目に適切と思われる語句も抽出し、新たなメタデータ項目として追加付与した。

表2に示したように、抽出した全38レコード中で、各メタデータの付与率が50%以上となる項目は全41のメタデータ中で、「書名(タイトル)」、「展示会ごとの区分」、「写本/版本」、「成立年」、「寸法(表紙 縦 横)」、「数量(巻数)」、「装訂」、「由来」、「料紙」、「装訂の説明」の合計10項目であった。

これらのメタデータのうち、古典籍の説明に必要とされる拡張メタデータ項目は、「写本/版本」、「装訂」、「料紙」、「装訂の説明」の4つである。

そのほかのメタデータで、基本メタデータと考えられるものは「書名(タイトル)」、「成立年」、「寸法(表紙 縦 横)」、「数量(巻数)」、「由来」などであると判断した。しかし、このうちの「寸法」や「数量」については、古典籍に即した

サイズの計測、あるいは数量で巻数など固有の単位が必要となる項目である。また、「由来」も古典籍に即した内容となり、単一の語句であらわすには難しく、多くの場合に説明文中に文章で由来についての説明がなされている。

## 5 まとめ

今回の検討サンプルである鶴見大学古典籍説明文から、「展示会ごとの区分」、「本の各部」、「内容の説明」、「参照ページ」、「系統・出版社の説明」、「版本の種類」、「料紙の説明」、「書型」、「写本の系統」、「絵の説明」、「成立年推定の説明」、「活字(版本)」、「塗箱(保存用の箱)」、「切れの系統」の14項目を本説明文の解析から新たなメタデータとして抽出できた。

説明文からのメタデータをグループ化し、メタデータ間の階層関係について検討したところ、基本的な古典籍の形態である「写本」、「版本」、「切れ」および「その他資料」の間で、付与されたメタデータが異なり、それぞれの古典籍の形態が反映されていることが明らかとなった。

## 参考文献

- [1] 秋元 良仁; 亀山 渉: 「博物館情報を用いた複合的メタデータスキーマ構築手法「ファジィ・スキーマ」の言語設計」, 情報処理学会研究報告, Vol.128, pp.25-27, 2006.
- [2] 藤沢 仁子; Frederic Andres: "Metadata Integration based on Multi-facet Thesaurus", 「シソーラスによる複数のメタデータの統合」, 情報処理学会研究報告, Vol. 105, pp. 33-40, 2005.
- [3] 木下 慶子; 村上 敦志; 稲積 泰宏; 森住 哲也; 木下 宏: 「デジタルアーカイブにおけるOntologyの活用」 電子情報通信学会技術研究報告. Vol.105, No.466, pp. 1-6, 2005.
- [4] Nagatsuka, Takashi; Chen, Ching-chih: "Global Memory Net Offers New Innovative Access to Tsurumi's Old Japanese Waka Poems and Tales, and Maps", In the Proceedings of the eighth International Conference on Asian Digital Libraries (ICADL2005), Springer, LNCS3815, pp.149-157, 2005.
- [5] Egusa, Yuka; Nagatsuka, Takashi: "New innovative access to educational and cultural multimedia contents", World Library and Information Congress: 72nd IFLA General Conference and Council(Seoul) [http://www.ifla.org/IV/ifla72/papers/097-Egusa\\_Nagatsuka-en.pdf](http://www.ifla.org/IV/ifla72/papers/097-Egusa_Nagatsuka-en.pdf), 2006.

# 人文研究を支援するデータベースシステム —聖教検索および系図表示— Database Systems for Humanities —Scripture Retrieval and Genealogical Visualization—

朴明哲<sup>\*1</sup>, 森本雅史<sup>2</sup>, 立花純児<sup>2</sup>, 村川猛彦<sup>2</sup>, 宇都宮啓吾<sup>3</sup>, 中川優<sup>2</sup>  
Ming Zhe Piao<sup>\*1</sup>, Masashi Morimoto<sup>2</sup>, Junji Tachibana<sup>3</sup>,  
Takehiko Murakawa<sup>2</sup>, Keigo Utsunomiya<sup>3</sup>, Masaru Nakagawa<sup>2</sup>

<sup>1</sup>和歌山大学大学院システム工学研究科 (〒640-8510 和歌山市栄谷930)

<sup>2</sup>和歌山大学システム工学部 (〒640-8510 和歌山市栄谷930)

<sup>3</sup>大阪大谷大学文学部 (〒584-8540 大阪府富田林市錦織北3-11-1)

<sup>1</sup>Graduate School of Systems Engineering, Wakayama University

<sup>2</sup>Faculty of Systems Engineering, Wakayama University

<sup>3</sup>Faculty of Literature, Osaka Ohtani University

寺院に所蔵されている様々な文書や典籍(聖教)において, その奥書には書写・伝授などの由来が記録されており, 聖教の形成や当時の人間関係, およびその教学的な環境を知る手掛かりとなる. このような人文研究を支援するため, 全文検索エンジンHyper Estraierを用いた聖教検索システムと, 既発表の僧侶人間関係検索システムの発展版として, 複数の師匠に対応した系図表示システムを構築した. 両システムの連携についても述べる.

The documents and scriptures in the possession of temples are a key to making clear the transcription of the writings, the human relationships and the religious circumstances behind them, by means of the colophons that state the background of themselves. We have been developing the applications for supporting the activities of humane studies. In this paper, we introduce a scripture retrieval system powered by the full-text search engine Hyper Estraier, and an improved version of genealogical visualization system which allows any priest to have more than one master in a genealogical table. In addition the prospect of the cooperation between these two applications is discussed.

キーワード: デジタルアーカイブ, 古典籍, 全文検索, 系図

digital archive, ancient document, full-text search, genealogical table

## 1. はじめに

近年, コンピュータは急速に発展し, 研究に欠かせない重要な構成要素になりつつある. これは, 人文系・芸術系分野にお

いても例外ではない[1]. その中でも, 古代・中世の文献を対象とする分野の研究動向の一つとして, 古代・中世の仏教並びに仏教界を巡る種々の問題に注目が集まっている. この問題の例として, 仏教活動の成

果である聖教自体の形成・継承の問題や聖教を巡る僧侶の活動などが挙げられる[2][3][4]. 近年, これらの問題の重要性が認識され, 古代・中世においてどんな僧侶が潜在し, どのような活動を行っていたのかを知ることは非常に有益であると言える.

その問題の解決手法の一つとして, 寺院に所蔵されている聖教の全体把握と個別の聖教の精査が行われている. 『聖教』とは, 「寺院社会内で教義・行法に関して記録したもので僧尼の修学や宗教活動の実践に際して活用され, かつ師弟間における原本授受または書写伝授によって法脈継承を根拠づける文献」[5]であり, 経典も含まれる. また, 聖教には「奥書」という書誌情報があり, 「いつ誰が命じて誰が書写をしたか」「経典がどこの寺院に所蔵されたか」などといった情報が記載されている. 国文学や歴史学において奥書は, 聖教自体の形成・継承の問題や聖教を巡る僧侶の活動を把握するのに重要な情報になっている.

本稿では, 上述した人文研究を支援するために構築した, 全文検索エンジン **Hyper Estraier** を用いた聖教検索システムと, 既発表の僧侶人間関係検索システムの発展版として, 複数の師匠に対応した系図表示システムを紹介する. さらに展望として, 両システムの連携についても述べる.

## 2. 聖教検索システム

人文系研究において, 史料などをテキスト化してコンピュータに格納し, 後に検索することはよく行われている. 総合的な検索に関しては, NDL デジタルアーカイブポータル (<http://www.dap.ndl.go.jp/home/>) が知られている. また漢訳仏典では, 大正新脩大

蔵経のテキスト化が多くの人の手によってなされており, 全文検索システムも公開されている (CBETA : <http://www.cbeta.org/>, 広済寺 仏教典籍検索 : <http://www.kosaiji.org/~kyoten/> など). しかし, 本研究のデータベースでは, 同一の経典名でも, 所蔵や奥書が異なっていれば別の文献とみなしている. このような聖教に関するデータを管理・検索するようなシステムはまだ少なく, 現状は, 各研究者が **Microsoft Access** などを用いて管理し, 検索している状態である. そのため, ある条件で複数の目録から検索したい場合, 目録ごとに操作をする必要があり, 手間を要するほか, 見落としが発生する可能性もある.

そこで, 上記の問題を解決するために, 数万に及ぶ聖教書誌情報の中から, 目的の情報を簡単に, かつ短時間で獲得できるような検索システムの構築を行った[6]. すでに筆者らの一人が聖教データベースを構築し, 検索できるようにしている (<http://www.orcaland.gr.jp/~utsunomiya/>) が, 本システムは質・量の両面においてこれを凌駕することを目指している.

この目的を実現するため, 本システムでは **Hyper Estraier** と呼ばれる全文検索エンジンを用いた. 内部で転置インデックスを用いており, 前処理に時間を要するが, インデックスが出来上がれば検索は極めて高速である. さらに, 文書には任意の属性を付加することができる. この特徴を用いて本システムでは「属性検索」も実装している.

検索に先立って, 聖教情報の登録が必要である. ここでその聖教データの流れについて説明する. 目録などを単位として, 表形式のファイル (Excel 形式または Access 形式) で管理されている聖教目録ファイル

に対して、これを CSV (Comma Separated Values)形式に変換しておく。この CSV ファイルに対して、登録に適した形に項目(列)を編集する。この作業は、Excel などで行ってもよいが、記述されたルールをもとに新しい CSV ファイルを出力するプログラムも作成している。編集された CSV ファイルを、登録ページを介してアップロードすると、サーバ側ではそれを行ごとに分割し、属性を付加して文書ファイルを構成して(例を図 1 に示す)、Hyper Estraier のサーバに登録する。

```
@bikou=(奥)保延五年九月二日巳時奉書
念仏宗僧運覚 願以書写功 必為往生因 法界衆生 生西方淨刹
@digest=a1d7adfe7c3d027d4d0593df0b46e0a0
@id=2022
@kansuu=一卷
@kunntenn=>
@line=2022
@mokuroku=基本聖教目録
@okugaki=>
@soutei=>
@syomei=無量清浄平等覚経巻下
@syosyajiki=保延五年 (1139)
@syozou=河内金剛寺
@uri=./syogyo_data/12943_20070330-150804_ShiftJIS.csv_2022
```

無量清浄平等覚経巻下,一卷,河内金剛寺,保延五年(1139),->,>,>,>,(奥)保延五年九月二日巳時奉書 念仏宗僧運覚 願以書写功 必為往生因 普法界衆生 生西方淨刹

図 1 文書ファイルの例

検索画面の例を図 2 に、また文献[4]で着目した僧侶「運覚」を検索語として、得られる結果画面の例を図 3 に示す。C 言語を中核とする Web アプリケーションとして構築している。検索可能な項目には、書名、出典目録、書写時期、装幀、訓点、奥書、備考がある。また検索対象の目録を限定するこ

ともできる。結果では、指定可能な各項目のほかにも巻数も表示される。またいくつかの聖教情報では、装幀、訓点、奥書の分割が容易ではなく、備考にまとめて記載している。



図 2 検索画面



図 3 検索結果画面

これまでにも本検索システムに登録した聖教情報は、7 種類の目録から合わせて 44,135 件である。

### 3. 系図表示システム

ここでは、電子化した僧侶の系図書に対して、人物の検索およびその系図表示ができる新たなシステム[7]について報告する。系図書として、師資相承を記した血脈や、姓氏家系を記した尊卑分脈を対象としているが、このような系図の電子化については、中世ヨーロッパにおける修道院僧の活動の分析[8]などのように、日本に留まらない国際的なニーズがある。

筆者らは過去にいくつかのシステムを構築してきた。最初は、人間関係を格納する関係データベースを定義し、Java アプリケーションを用いて系図を表示していた[9]。次に、SVG (Scalable Vector Graphics)を用いて縦書き表示や外字表示を可能にした[10][11]。しかし、それらのシステムでは、系図書は木構造をなしていること、すなわちどの僧侶も師匠は高々一人であることを前提として、描画するようにしていた。そのため、ある僧侶に複数の師匠がいる場合、表示できない。複数の師匠がいる系図の具体例を図4に示す。この図で徳円の師匠は、伝教大師、広円、円澄、広智の4人である。

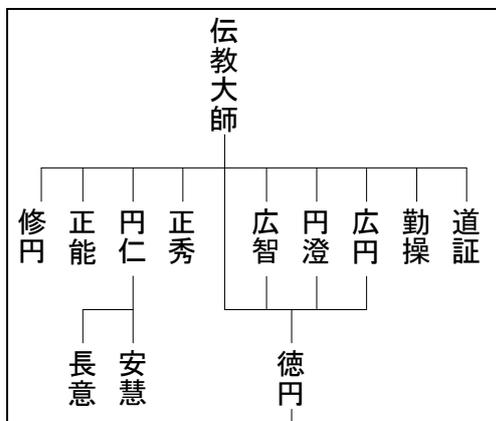


図4 複数の師匠を持つ僧侶を含む系図[2]

そこで、複数の師匠に対応するためのデータ構造を検討するとともに、その系図表示を試みた。複数の師匠に対応させる際、師弟関係に注目した。すべての師弟関係の抽出を行い、そこから1人の人物の師匠・弟子を求め、系図を表示させる。

系図表示の画面例を図5に示す。操作の効率化を考え、検索、表示の操作が1画面で行えるよう、表示ウィンドウをフレームで分割している。検索の方法として、「人物名からの検索」と「系図の種類からの検索」の2種類を提供している。前者は、人物名を入力し、該当する人物が所属する系図を選択して系図を表示するものであり、ウィンドウを左上、右上、右下の順に見ることになる。後者は、系図の種類を選択し、その系図に所属する人物の一覧から人物を選択して系図を表示させるもので、ウィンドウを左上、左下、右下の順に見ることになる。

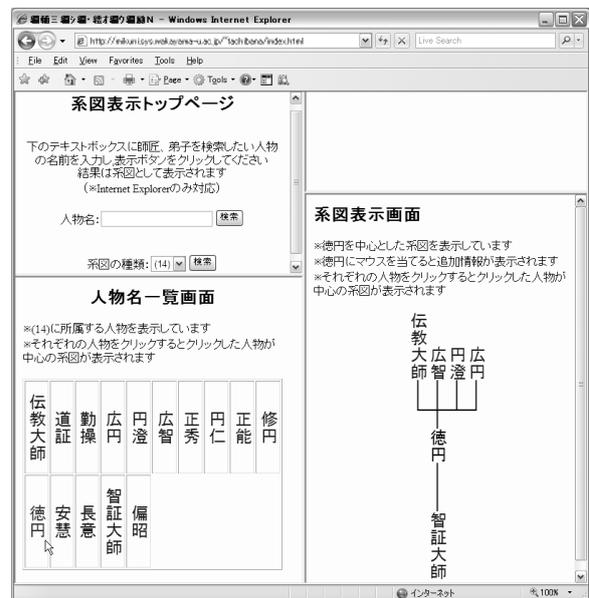


図5 系図表示画面

図5の右下には、徳円を中心として、その師匠と弟子で構成される部分系図が表

示されている。中心以外で出ている人物をクリックすると、その人物が中心となる系図に変わる。中心人物にマウスカースルを当てると、その人物の情報が(もしあれば)ポップアップ表示される。

#### 4. 連携した検索システム

ここで、2 節および 3 節で独立して紹介した二つのシステムについて、その連携の必要性・有用性と、実現方法を述べることにする。手掛かりとなるのは、東大寺図書館蔵『弘賛法華伝』にある図 6 の奥書である[3]。

大日本国保安元年七月八日於太宰府勸  
俊源法師書写畢、宋人蘇景自高麗国奉渡  
聖教之中、有此法花伝、仍為留多本令書  
写也、羊僧覚樹記之

図 6 『弘賛法華伝』の奥書

ここで「俊源」と「覚樹」はいずれも僧侶の名前であることが、系図書から確認できる。

もし、聖教検索における奥書・備考や、系図表示における人物情報の中に、他の検索で利用できる語句がリンクとなっていれば、初学者に参考になるほか、多面的な情報探索が期待できる。このようなリンクのない現状では、利用者が毎回検索語を与えなければならず、多大な不便を強いている。見落としや、打ち間違いによるタイムロスも起こりやすい。

他への検索に利用できる語句として、人物名の他に何があるか、検討する。まず年代が挙げられる。図 6 でも、「保安元年」に見られる。他の奥書を見たところ、地名および寺院名も抽出できると望ましい。

奥書や人物情報のテキストデータが与えられたとき、計算機で上述の情報が求めら

れる(機械抽出できる)か否かについて、筆者らは以下の見通しを立てている。まず、年代についてはルールのみでほぼ抽出できる。次に地名および寺院名については、いくつかのケース(「於」の直後など)で機械的に求められるが、地名や寺院名のデータベース(以下「DB」という)を別途用意しておき、照合するほうがより精度が高まる。最後に、人物名については、人物 DB との照合なしには困難である。ただし年代についても、専用の DB を設置することで、年代(西暦, 和暦, 元号, 西暦による区間指定)を検索して条件に合う聖教や人物名が瞬時に求められるようにすべきであろう。

これまでに挙げた地名 DB, 寺院 DB, 人物 DB, 年代 DB は、聖教検索や系図表示において、システム内部で検索することを念頭に置いて説明してきたが、利用者が直接検索するのも差し支えなく、そのニーズも少なくないと考えられる。

以上の構想に基づいて得られる、連携システムの利用イメージを、図 7 に示す。

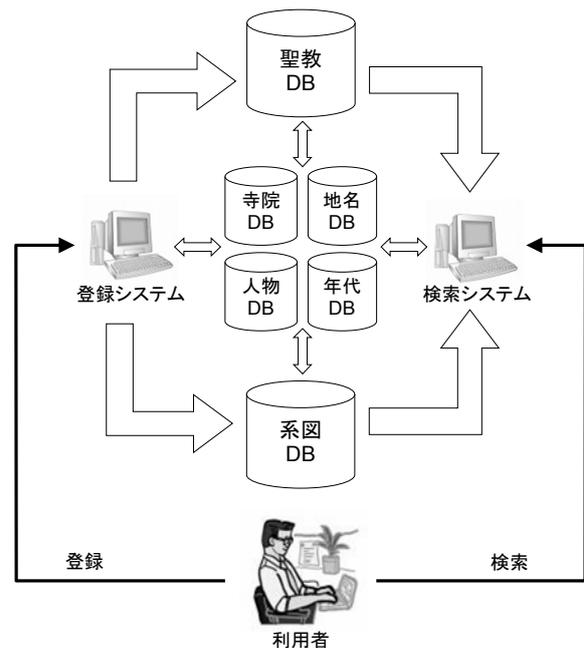


図 7 聖教検索と系図表示の連携利用

## 5. おわりに

本稿では、人文研究を支援するために構築した、全文検索エンジンを用いた聖教検索システムと、複数の師匠に対応した系図表示システムを紹介した。さらに展望として、両システムの連携についても述べた。

最後に、利用方法の将来像について触れておく。今のところ、聖教の奥書の電子的検索を調査の場で行うことができず、インターネット接続可能なところでサーバと通信する必要がある。しかしパソコンの高性能化とコストダウン、仮想化技術の普及などにより、1台のノートPCに本稿で述べてきたシステムを安価に搭載することが可能となっている。それにより、原資料を目の前にして即時的にデータベースにアクセスし、掘り下げて検討することもできる。これは文化財情報に関するユビキタス化とも言え、人文研究に大きく貢献するであろう。さらにその場で情報を登録できるようにすれば、電子化作業の障壁がより小さくなる。著作権などの問題に対しては、登録情報ごとに、登録者や公開の可否を指定できるよう、データベースに属性を追加すればよい。

## 参考文献

- [1] 八村広三郎:「人文科学とデータベース」, 情報処理学会誌, Vol.38, No.5, pp.377-382, 1997.
- [2] 宇都宮啓吾:「東寺観智院金剛藏『天台血脈』について」, 大谷女子大学紀要, Vol.86, pp.14-17, 2004.
- [3] 宇都宮啓吾:「血脈資料『諸嗣宗脈紀』について—龍谷大学本を手懸かりとしたその成立とデータ公開を巡る問題—」, 大阪大谷国文, Vol.1, pp.45-60, 2007.
- [4] 宇都宮啓吾:「聖教調査におけるデータの整備を巡る問題—『念仏宗僧運覚』・書誌情報を中心として—」, 基盤研究(A)(1)「金剛寺一切経の総合的研究と金剛寺聖教の基礎的研究」報告書, 2007.
- [5] 上川通夫:「中世聖教史料論の試み」, 史林, No.79, Vol.3, pp.108-129, 1996.
- [6] 森本雅史:「全文検索を用いた聖教検索システムの構築」, 和歌山大学システム工学部卒業論文, 23p., 2007.
- [7] 立花純児:「複数の師匠に対応した僧侶の系図表示システム」, 和歌山大学システム工学部卒業論文, 27p., 2007.
- [8] Jean-Loup Leutre (加納修訳):「修道院の日常生活における聖人伝の朗読と書物(9世紀~13世紀西欧)」, 21世紀COEプログラム「統合テキスト科学の構築」第1回国際研究集会報告書, 名古屋大学大学院文学研究科, pp.179-193, 2003.
- [9] 田中猛彦; 富金原賢次; 宇都宮啓吾; 中川優:「平安・鎌倉時代を対象とした僧侶データベースシステム」, 情報知識学会誌, Vol.13, No.2, pp.18-31, 2003.
- [10] 田中猛彦; 富金原賢次; 宇都宮啓吾; 中川優:「平安・鎌倉時代を対象とした僧侶データベースシステム:概要とその後の展開」, 情報知識学会誌, Vol.14, No.3, pp.44-49, 2004.
- [11] 金平晃尚:「SVGを用いた系図表示インタフェースの開発」, 和歌山大学システム工学部卒業論文, 26p., 2005.

## 科学研究費補助金による研究助成の効果に関する調査

—研究キャリアステージにおける科学研究費補助金取得パターンの分析—

### The Effect of Grants-in-Aid for Scientific Research: its analysis through researchers' career stages

柿沼 澄男\*, 西澤 正己, 孫 媛, 根岸 正光

Sumio KAKINUMA, Masaki NISHIZAWA, Yuan SUN, Masamitsu NEGISHI

国立情報学研究所 情報社会相関研究系

Information and Society Research Division, National Institute of Informatics

〒101-8430 千代田区一ツ橋 2-1-2

Email: {kakinuma, nisizawa, yuan, negishi}@nii.ac.jp

科学研究費補助金採択課題データベースを利用し、著名な研究者の研究キャリアステージにおける科学研究費補助金の取得状況を分析した。30代前半から60代前半に亘る研究キャリアステージにおいて、各研究分野に特徴的な研究費の取得パターンと各分野に共通する取得パターンを見出した。

The system of Grants-in-Aid for Scientific Research from Ministry of Education, Culture, Sports, Science and Technology of Japan is one of the oldest funding systems, which has been supporting researchers belonging to universities and research institutes in Japan. This paper examines the researchers' need of funding chronologically through their career stages by the amount of grants awarded.

キーワード：科学研究費補助金, キャリアステージ, パターン分析

#### はじめに

文部科学省科学研究費補助金(以下、科研費という)は古くから学術研究の下支え、底上げの役目を果たしてきたと言われている。現在では、科研費以外にも多くの競争的研究資金が配分されているものの、研究費総額では科研費が約半分を占めている。この他、学術研究に対する研究資金としては、各大学・研究所等から配分される基盤的研究資金があるが、基盤的研究資金に

ついては近年減少の傾向にあり、学術研究における科研費の果たす役割はますます増大している。

本研究においては、研究者のキャリアステージにおける効果的な研究費配分を解明することを目的として、著名な研究者の各キャリアステージにおける科研費の取得パターンを分析した。

#### 1. 科学研究費補助金データベース

国立情報学研究所では、科研費の採択課題の速報を収録した「科学研究費補助金採択課題データベース(KAKENK)」およびその成果概要を収録したデータベースである「科学研究費補助金成果概要データベース(KAKEN)」を作成している。KAKENKにおいては、1965年度から2006年度までの採択課題、KAKENにおいては1985年度から2003年度までの成果概要が収録されている。KAKENKデータベースには研究代表者、採択課題名、研究種目、分科細目、新規・継続の別、配分額等が収録されている。KAKENデータベースには、配分額を除くKAKENKの情報に加え、研究分担者、成果概要、キーワード、発表文献リスト等が収録されている。本報告の分析には1965年度から2005年度までのKAKENKデータベースを使用した[1]。

## 2. データの分析

分析に当たって、著名な研究者として1995年から2004年までの文化功労者、日本学士院賞(エジンバラ公賞を含む)、紫綬褒章の受賞・受章者のうち70歳以下(2005年現在)の者168名を対象とした。これらの研究者について、科研費研究者番号が付

与されている1996年以降は研究者番号、それ以前は氏名および所属によりKAKENKデータベースから採択課題(研究代表者のみ)を抽出した。これによって、採択課題を抽出できた153名が分析の対象者である。この対象者に対して、科研費の採択課題ごとに採択時の年齢を計算し付与した。対象者の学問分野ごとに各年齢の平均科研費獲得金額を算出し、各年齢段階における科研費の獲得パターンを分析した。

表1は、対象となった153名の分野別構成人数、一課題当たり科研費配分額の平均、および採択総件数である。

本報告に先立ち、筆者らは科研費の分析を行った[2, 3]。そこにおいては研究者個人に着目して科研費獲得の経年分析を行い、研究者の特性に応じた科研費の獲得パターンや分野別の特色を解明した。しかしながら、これらは事例研究的なものであり統計的な分析は行っていなかった。今回の分析はサンプル数は少ないものの統計的分析を試みたものである。

## 3. 研究キャリアステージ

研究者は、博士課程・ポスドク等を経て教

表1: 対象者の分野構成人数, 一課題当り研究費配分額および採択総件数

| 分野      | 人数 | 一課題当り配分額(千円) | 配分額の標準偏差 | 採択総件数 |
|---------|----|--------------|----------|-------|
| 社会      | 11 | 5967.7       | 11933.4  | 127   |
| 人文      | 17 | 4110.3       | 7167.8   | 139   |
| 理工-化学   | 19 | 14746.2      | 46432.6  | 632   |
| 理工-工学   | 17 | 19945.8      | 60194.3  | 478   |
| 理工-数物   | 31 | 9713.9       | 29148.7  | 665   |
| 生物-医歯薬学 | 27 | 19699.3      | 48862.1  | 1070  |
| 生物-生物学  | 16 | 26931.2      | 60320.6  | 509   |
| 生物-農学   | 15 | 5849.0       | 8065.7   | 417   |

育研究者に採用され、助教、准教授、教授等とキャリアを積んでいく。学問分野によりキャリアパスは異なるであろうが、年齢を目安として区切ると次のように考えることができる[4].

1st ステージ:学部または修士課程修了後 30 歳位まで

2nd ステージ:30 歳位から 30 歳代後半ないし 40 歳位まで

3rd ステージ:30 歳代後半ないし 40 歳位から 55 歳位まで

4th ステージ:55 歳位より上

一般的には、1st ステージは研究者にとっては修業の時代であり、2nd ステージで研究費を取得して独り立ちする段階になる。3rd ステージは研究者にとって一番脂の乗った時期であり、分野によっては大型の研究費を獲得し研究を発展させる時期である。4th ステージは、管理職の道を歩む者と研究者の道をさらに発展させる者とに分かれる時

期である。

#### 4. 研究キャリアステージにおける科研費獲得金額の推移

図1, 2は、学問分野別に上記の研究キャリアステージに対応した科研費の獲得状況を示したものである。初期ステージと、後期ステージ、人文系と生物系の乖離が大きいので、見やすいように、リニアスケール(図1)と対数スケール(図2)の異なるスケールで示している。

##### 4.1 人文系

2nd ステージにおいては年間平均 100 万円以下の科研費を獲得する。3rd ステージに入りようやく 100 万円を超えるようになり、40 歳代後半にピークとなり 1000 万円～1500 万円の額となる。50 歳代になると獲得額は減少し 200 万円～400 万円となる。4th

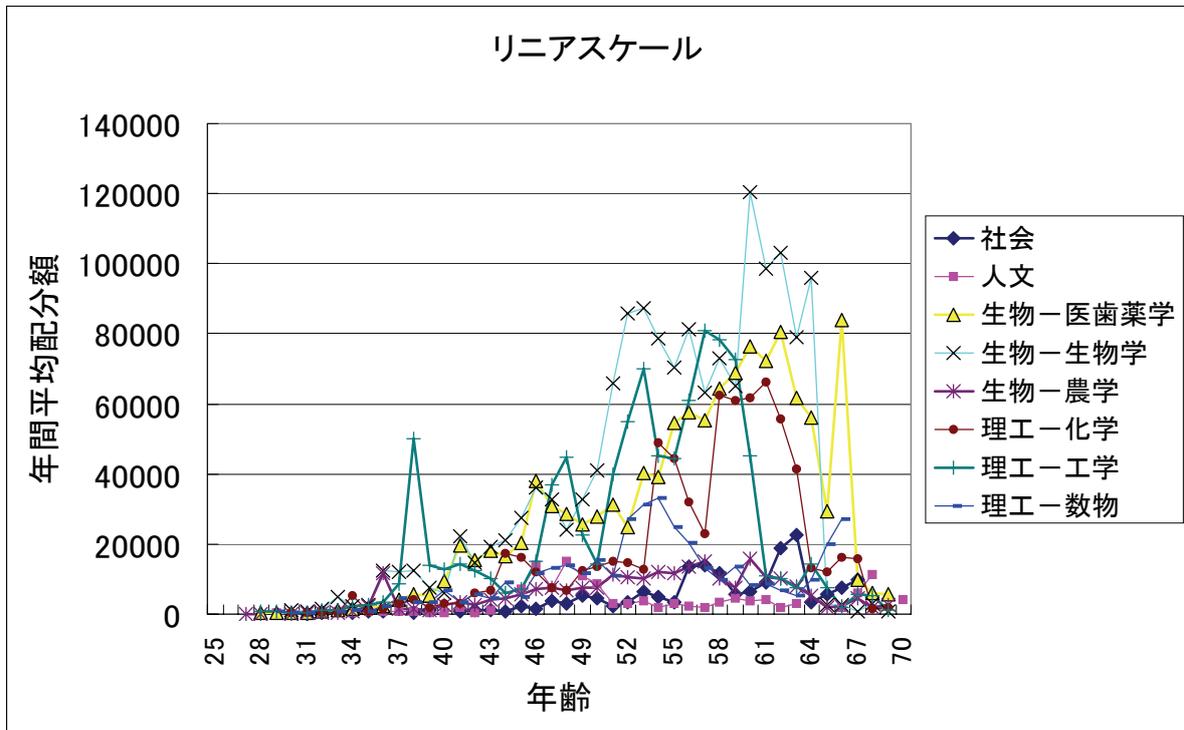


図1: 分野別の年齢ごとの科研費の獲得状況 (リニアスケール)

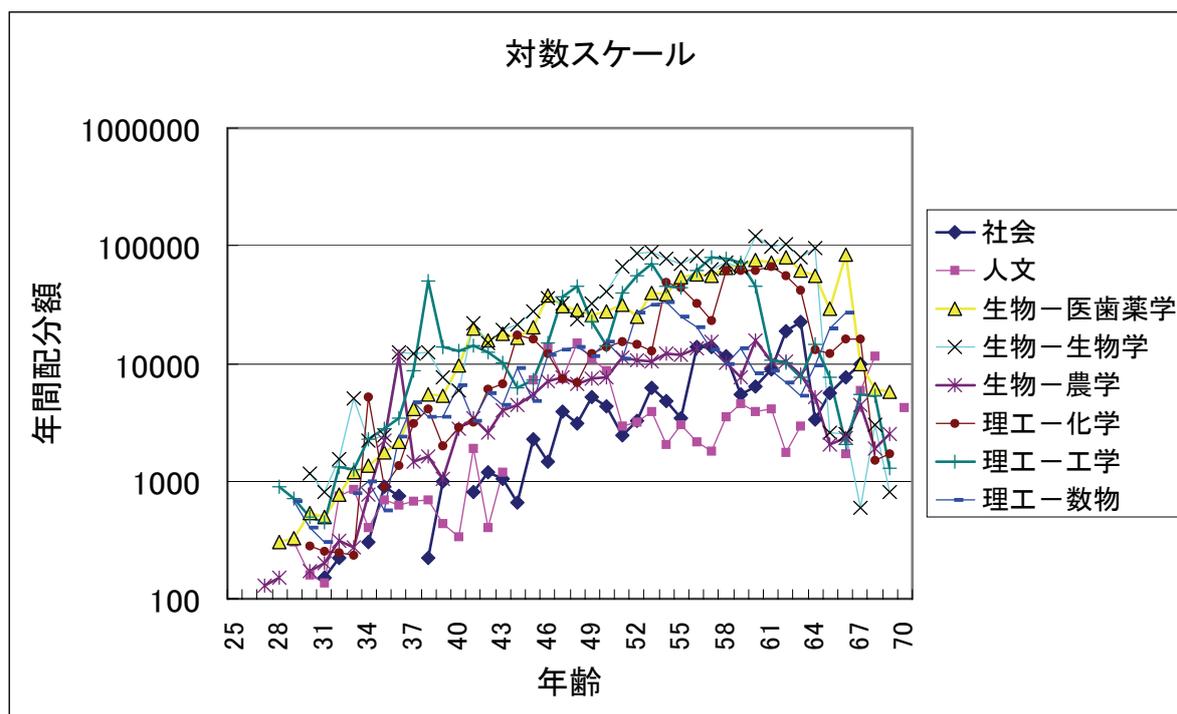


図2： 分野別の年齢ごとの科研費の獲得状況（対数スケール）

ステージになってもその状況が続く。

## 4.2 社会系

2ndステージにおいては年間平均100万円以下の科研費を獲得する。3rdステージで年間100万円を超えるようになり、100万円～600万円の額となる。4thステージになると獲得額のピークとなり、1000万円を超える額を獲得するようになる。

## 4.3 理工系-化学

2ndステージの前半は年間平均100万円以下の科研費獲得額であるが、後半には300万円前後の額を獲得するようになる。3rdステージにおいては年間1000万円台の額を獲得し、4thステージになると年間5000万円を超えるようになる。この状況は60歳代前半まで維持される。

## 4.4 理工系-工学

2ndステージは年間平均100万円以下の科研費獲得額から始まり、徐々に金額が増え後半には1000万円前後の額を獲得するようになる。38歳の額が大きくなっているが、これは一名に対して、この年から始まった特別推進研究の1年目という特異値があるため大きくなっている。3rdステージにおいては、初期には年間1000万円程度であるが、40歳代後半には2000万円～4000万円の額を獲得するようになる。50歳代は5000万円を超え、4thステージの50歳代後半でピークとなり年間平均8000万円を獲得する。60歳を超えると科研費獲得額は急激に減少する。

## 4.5 理工系-数物

2ndステージの前半は年間平均100万円以下の科研費獲得額であるが、後半には

300 万円前後の額を獲得するようになる。3rd ステージにおいては年間 1000 万円台の額を獲得し、55 歳前後の 3rd ステージから 4th ステージにかけて獲得金額はピークとなり 2000 万円～3000 万円の金額を獲得する。50 歳代の終わりからは 1000 万円前後の金額となる。65 歳と 66 歳で金額が高いのは大型研究費を獲得した者が一人いるためである。

#### 4.6 生物系—医歯薬学

2nd ステージの前半は年間平均 100 万円以下の科研費獲得額であるが、後半には 500 万円～1000 万円の額を獲得するようになる。3rd ステージには 1500 万円～4000 万円の額を獲得している。4th ステージには 5000 万円を超える額を獲得するようになり、60 歳代半ばまでこの状況が維持される。

#### 4.7 生物系—生物学

2nd ステージの前半は年間平均 100 万円～300 万円の科研費獲得額であるが、後半には 1000 万円程度獲得するようになる。3rd ステージの前半は 1000 万円程度であるが、後半には 2000 万円～5000 万円の獲得額となっている。4th ステージ初期の 55 歳～60 歳は獲得額のピークとなり、5000 万円から 8000 万円の獲得額となっている。60 歳を超えると獲得額が急激に減少する。

#### 4.8 生物系—農学

2nd ステージは概ね年間平均 100 万円以下の科研費獲得額である。3rd ステージになると着実に獲得額は増大するが、40 歳代は 1000 万円以下にとどまっている。50 歳代になるとようやく 1000 万円を超えるようになる。4th ステージになると獲得金額のピーク

を迎えるが、それでも 1000 万円～1500 万円の範囲に収まっている。60 歳を超えると獲得額は減少に向かい 1000 万円以下になる。

### 5. 2005 年度科研費採択者と主要学術賞受賞者に対するアンケート調査

文部科学省科研費に対する研究者側からの意見、要望を直接調査するために、2005 年度科研費採択者と主要学術賞受賞者に対するアンケート調査を、(財)松尾学術振興財団と共同して 2006 年 3 月に実施し、873 名から回答を得た。結果の概要は既に公表したが[5]、2006 年度も引き続き結果の分析検討を進め、科研費に対する政策提言をまとめつつある。総じて科研費による基礎研究に対する支援の充実、デュアルサポート・システムの維持継続の要望が強いこと等が明らかになっている。

### 6. まとめ

今回の分析により、著名な研究者においては、学問分野により研究費の金額に差異はあるものの、研究キャリアの 2nd ステージにおいて小額の研究費を獲得して独り立ちし、3rd ステージにおいては更に高額の研究費を獲得して研究を発展させ、4th ステージにおいては 60 歳位までは研究活力が維持されることが分かった。また、学問分野間に、特に人文社会科学と自然科学の間に、使用する研究費の額の隔たりが大きく存在することが分かった。

#### 参考文献

[1] 「科学研究費補助金採択課題・成果概要

データベース」 <http://seika.nii.ac.jp/>  
(KAKENとKAKENKの統合検索サービス)

[2] 西澤 正己, 根岸 正光, 柴山 盛生, 孫  
媛, 野村 浩康, 光田 好孝, 前田 正史, 「科  
学研究費データベースによる採択パターン分  
析」, 情報知識学会誌, Vol.15, No.2, pp.85-88,  
2005

[3] 西澤 正己, 根岸 正光, 柴山 森生, 渡  
邊 恵子, 野村 浩康, 光田 好孝, 前田政史,  
「科学研究費データベースによる採択と研究成果  
のパターン分析」, 情報知識学会誌, Vol.16,

No.2, pp.1-6, 2006

[4] 吉武 博通, 「大学教員の能力・資質の持  
続的向上に向けて-大学教員の任用・育成と評  
価のあり方-」, カレッジマネジメント, vol.143,  
Mar.-Apr. 2007

[5] 「科学研究費補助金の学術貢献に関する  
アンケート調査-調査結果の概要」, 松尾研究  
会報, vol.14 (2005), 松尾学術振興財団  
(2006.5)

[http://www.matsuo-acad.or.jp/research/14-ka  
kenhi.pdf](http://www.matsuo-acad.or.jp/research/14-ka<br/>kenhi.pdf)

# キーワード分析による科研費におけるゲノムおよびナノテクノロジー 関連研究の動向調査

## Investigation into Genome-related and Nanotechnology Research at Grants-in-Aid in JAPAN

西澤 正己\*、孫 媛

Masaki NISHIZAWA and Yuan SUN

国立情報学研究所 情報社会相関研究系

Information and Society Research Division, National Institute of Informatics

〒101-8430 千代田区一ツ橋 2-1-2

Email: { nisizawa, yuan }@nii.ac.jp

科学研究費補助金において、科学技術基本計画の重点分野に対してどのように研究費が配分され、どのように変化してきたかは興味を持たれるところである。我々は、これらの重点領域に対するキーワードを抽出したうえで、科研費のすべての採択課題の課題名からこれら重点領域に関連する課題を選び出し、配分額の変遷を調べた。また、科研費の分科細目表に対しても同様に特徴的なキーワードを抽出し、それらの一致度を見ることにより、重点領域との関連分野を対応分析により示した。最終的にはすべての重点領域について行う予定であるが、ここでは試験的に(1)ライフサイエンス分野のゲノム関連分野と(2)ナノテクノロジー・材料分野の2分野について調査した結果を述べる。

The system of *Grants-in-Aid for Scientific Research* from MEXT is one of the oldest ones, which is the funding system for researchers belonging to universities and institutes in Japan. In this *Grants-in-Aid*, it is very interesting to see how the adoption situation of the research of these prioritized areas on the 2nd Science and Technology Basic Plan has changed. We measure strength of a related level each other by comparing "Field feature keyword" that has been developed so far and the research subjects and the applied detail research fields in the *Grants-in-Aid*. Especially in this analysis, the relation between a field related to the genome, a nanotechnology in the prioritized area and the applied research subjects and their applied detail research fields are examined by using this method, and the secular distortion of the adoption situation and the mutual relation are examined by using Correspondence Analysis etc.

### 1. はじめに

政府および総合科学技術会議は科学技術基本法のもと、1996年から第1期、第2期科学技術基本計画[1]を進めてきた。第2

期科学技術基本計画(平成13~17年度)では重点分野として「ライフサイエンス分野」、「情報通信分野」、「環境分野」、「ナノテクノロジー・材料分野」が特に重点を置かれ、資源配分が重点的に行われてきた。こ

これらの分野は第3期計画(平成18~22年度)においても、引き続き重点分野とされている。

文部科学省科学研究費補助金(科研費)[2]は大学等の研究者に対しての個人申請・ピアレビュー審査による研究費配分制度である。この制度は1939年に始まり、古くから基礎研究の下支え、底上げの役目を果たしてきたと言われている。現在では、科研費以外にも多くの競争的資金が配分されているものの、研究費総額では、科研費が約半額を占めている。この科研費において、科学技術基本計画の重点分野に対してどのように研究費が配分され、どのように変化してきたかは興味を持たれるところである。

我々は、科学研究費補助金データベース(KAKEN、KAKENK)[3]からこれらの重点領域に対するキーワード(トピック関連キーワード)[4][5]を抽出したうえで、科研費のすべての採択課題の課題名から切り出したキーワードに対して、これら重点領域に関連する課題を選び出し、これらの関連課題の配分額の変遷を調べた。また、科研費の分科細目表に対応して、それぞれの細目に採択された課題からその細目に特徴的なキーワードを抽出し(分野特徴キーワード)、先の「トピック関連キーワード」との一致度を見ることにより、重点領域との関連分野を対応分析により示した。最終的にはすべての重点領域について行う予定であるが、ここでは試験的に(1)ライフサイエンス分野のゲノム関連分野と(2)ナノテクノロジー・材料分野の2分野について調査した結果を

述べる。

## 2. 科学研究費補助金データベース

国立情報学研究所では、科学研究費補助金の採択課題の速報を収録した「科学研究費補助金採択課題データベース(KAKENK)」[3]およびその成果概要を収録したデータベースである「科学研究費補助金成果概要データベース(KAKEN)」[3]が作成されている。KAKENKにおいては1996年度から最新の2005年度までの採択課題(現在では遡及入力が進み、一部の項目を除いては1965年からのデータが収められている)、KAKENにおいては1985年度から2003年度までの成果概要が収録されている。KAKENKデータベースには採択課題名、研究種目、分科細目(大型研究費、奨励研究を除く)、新規・継続の別、配分額等が収録されており、KAKENについてはKAKENKの情報に加え、成果概要、キーワード、発表文献リスト等が収録される。なお、KAKENデータベースは成果概要の報告書を収録するため、採択後少なくとも約1年半後の課題が収録されていることになる。今回の分析には1996-2005年度のKAKENKデータベース、および2000-2003年度のKAKENデータベースを分析に使用した。

表1に1985年度から2005年度までの科学研究費補助金の予算額の推移を示す[2]。1991年に大学院重点化政策が始まり、1995年には科学技術基本法が施行されている。この表からわかるように、1992年から

表1: 1985年度から2005年度までの科学研究費補助金の予算額と増加率の変遷[2]

| 年度        | 1985 | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996  | 1997  | 1998  | 1999  | 2000  | 2001  | 2002  | 2003  | 2004  | 2005  |     |
|-----------|------|------|------|------|------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 予算額(10億円) | 42.0 | 43.5 | 45.1 | 48.9 | 52.6 | 55.8 | 58.9 | 64.6 | 73.6 | 82.4 | 92.4 | 101.8 | 112.2 | 117.9 | 131.4 | 141.9 | 158.0 | 170.3 | 176.5 | 183.0 | 188.0 |     |
| 増加率(%)    |      | 3.7  | 3.6  | 3.6  | 8.4  | 7.6  | 6.1  | 5.6  | 9.7  | 13.9 | 12.0 | 12.1  | 10.2  | 10.2  | 5.1   | 11.5  | 8.0   | 11.3  | 7.8   | 3.6   | 3.7   | 2.7 |

2003 年度の伸びが大きく、10%を超える成長がほとんどこの時期に集中している。2003 年度以降も伸びは続いているが、成長率は3%程度に留まっている。

### 3. 分析手法および結果

#### 3.1 トピック関連キーワード

上記のKAKENデータベースには採択課題に対して申請者が付与したキーワードが収められている。(1)ゲノム関連研究に関しては、ゲノム4領域(統合ゲノム: comp、ゲノム医科学: med、ゲノム生物学: bio、ゲノム情報科学: info)の特定領域研究(2000-2004 年度)から KAKEN データベースに含まれる 2000-2003 年度の課題を抽出し、この課題に付与されたキーワードを「ゲノム関連キーワード(comp, med, bio, info)」とした。

(2)ナノテクノロジー関連研究に関しては適当にまとめた特定領域研究がないため、文部科学省ナノテクノロジー総合プロジェクトセンター[6]が公表する科学研究費のナノテクノロジー関連研究課題を 2000-2003 年度の KAKEN データベースから抽出した。また、振興調整費、21 世紀 COE に関しては公表された関連研究課題の代表者が科研費に採択された課題を 2000-2003 年度の KAKEN データベースから抽出し、さらに課題の分野に応じて、バイオ関連(bio)、化学関連(chem.)、デバイス関連(device)、素材関連(matel)、ナノ構造関連(struct)の5つに分類した。それらの課題に付与されたキーワードをあわせて「ナノテクノロジー関連キーワード(bio, chem., device, matel, struct)」とした。

#### 3.2 関連研究費配分額

科研費の研究費配分額については

KAKEN データベースにしか収められていない。よって、1996 年度から 2005 年度までの KAKEN データベースに対して、研究課題からキーワードを抽出し、(1)「ゲノム関連キーワード(comp, med, bio, info)」および(2)「ナノテクノロジー関連キーワード(bio, chem., device, matel, struct)」との一致を調べた。それぞれのキーワード群と一致した課題を(1)ゲノム関連研究課題(comp, med, bio, info)、(2)ナノテクノロジー関連課題(bio, chem., device, matel, struct)として年度別にそれぞれ課題数を示したのが図 1-1, 1-2 である。

課題数に関しては、どちらも 1998 年の伸びが大きい。またゲノム関係は単調な伸びが続いている。ナノテック関係では 200 年まではほとんど伸びていないが、2002 年度以降の伸び、特に素材(matel)やデバイス(device)関係の伸びが大きい。

これらの図を、各年度の科研費全体の総採択課題数に対する比率で示したのが図 2-1、図 2-2 である。これからわかるように、全体の伸びに対して、1998 年度の伸びはほぼ同様の伸びであり、これらの分野特有でないことがわかる。ゲノム関連に関しては 2001 年度まで緩やかに比率が上昇し、それ以降は全体の比率からはやや減少していることがわかる。ナノテック関連では、分野により分かれるが、全体的に課題数の比率はやや減少かほぼ横ばいの状態である。

同様に、それぞれの課題の研究費配分額を図 3-1、3-2 に示す。配分額に関してはどちらも伸びは著しい。特にどちらも 2001 年度から 2002 年度の伸びが急激である。これは第 2 期科学技術基本計画(2001-2005 年度)の影響が大きく現れているものと思われる。

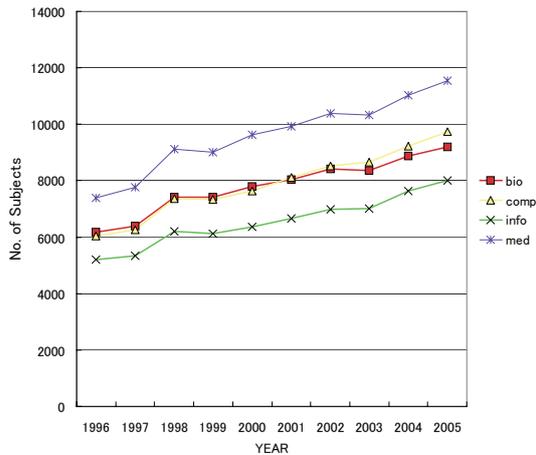


図 1-1 ゲノム関連課題数の変遷

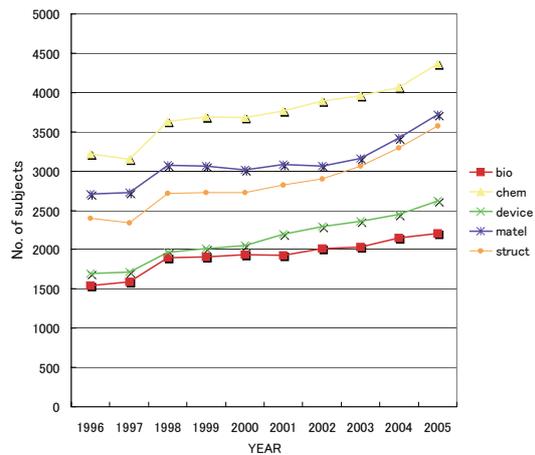


図 1-2 ナノテク関連課題数の変遷

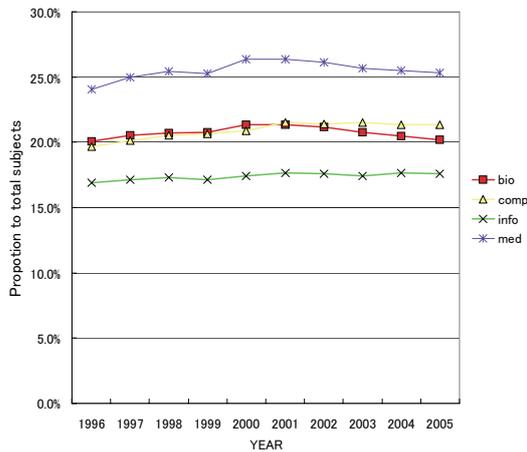


図 2-1 ゲノム関連課題数比率の変遷

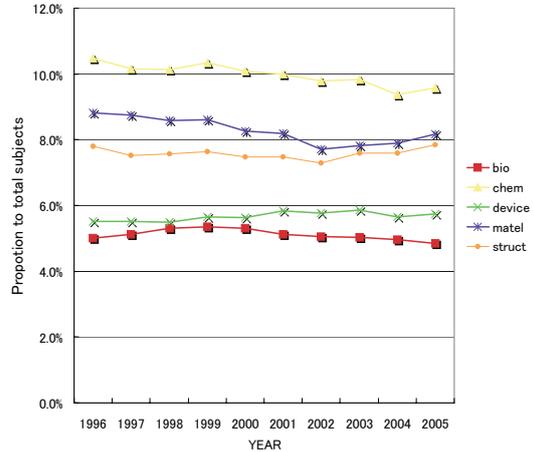


図 2-2 ナノテク関連課題数比率の変遷

配分額に関しても、科研費全体の配分額に対しての比率として表したのが図 4-1、図 4-2 である。この図では明らかに全体の伸びに対して 2001 年度の伸びがゲノム、ナノテクどちらの分野においても大きいことがわかる。また、ゲノム関連が 2001 年度の伸びに比べ他の年度が全体のほぼ平均の伸びにとどまっているのに対して、ナノテクノロジー関連は、化学関連領域(chem.)等の例外を除いては 2003 年度近くまで全体の平均を上回る伸びを示している。

これらのキーワード分析では、全体に対

する比率でもわかるように、かなり課題数、配分額ともに過大見積りしている。「トピック関連キーワード」に関してはさほど問題がないと思っているが、対象が課題名のみであるため、情報不足であると言えるのであろう。最新の課題は分析できないが、成果概要が使える KAKEN(成果概要データベース)を使用してさらに精度を上げるべきであろうが、現在の方法でも、これらの結果からは大まかな傾向はつかめるのではないと思われる。

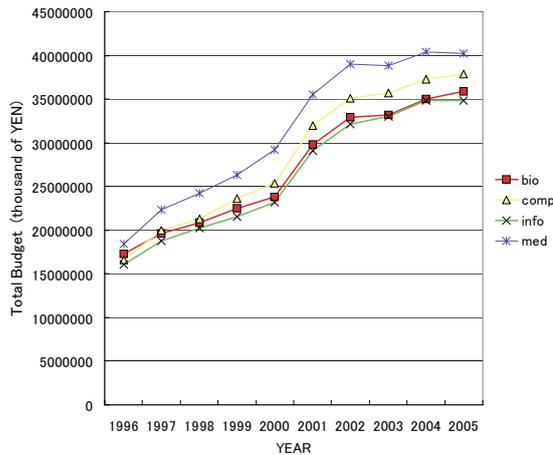


図 3-1 ゲノム関連配分額の変遷

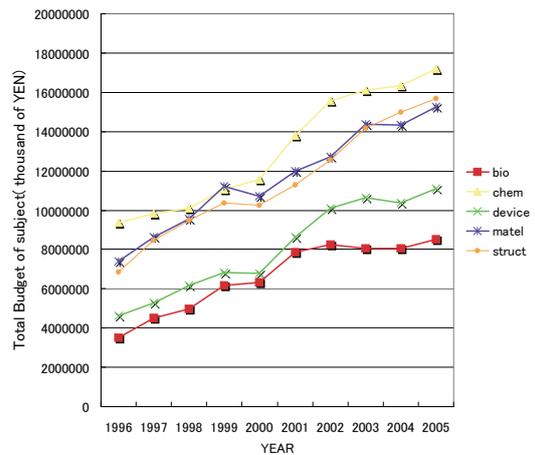


図 3-2 ナノテク関連配分額の変遷

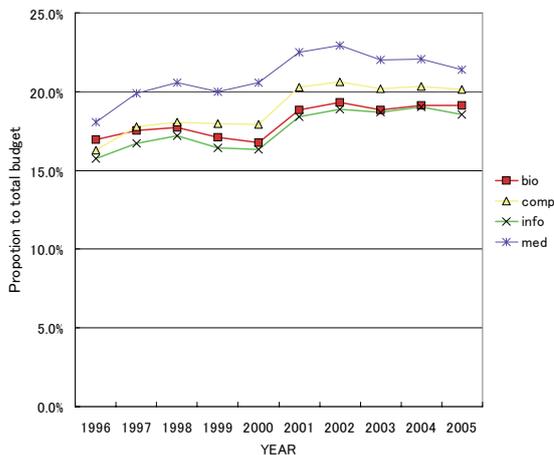


図 4-1 ゲノム関連配分額比率の変遷

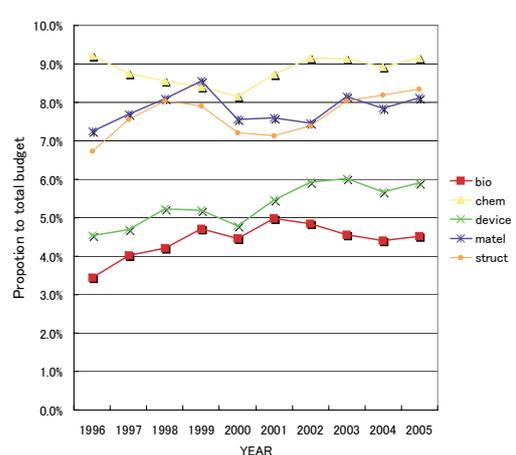


図 4-2 ナノテク関連配分額比率の変遷

#### 4. ゲノム関連、ナノテク関連研究と他分野との関係

3.1 章で述べた(1)「ゲノム関連キーワード(comp, med, bio, info)」および(2)「ナノテクノロジー関連キーワード(bio, chem., device, matel, struct)」と科学研究費補助金の分科細目表の細目分野との関連を調べるため、改定後の 2002 年度以降(2002-2005 年度)のそれぞれの細目分野に属する研究課題からキーワードを抽出し、それぞれの細目分野の「分野特徴キーワード」を作成した。この「分野特徴キーワード」と(1)および(2)の「トピック関連キーワード」

の一致度を見ることにより、トピックと科研費の分科細目との関係を調べた。一致度の測定に関しては[4][5]に詳しく述べている。

図 5-1、図 5-2 には(1)および(2)と「分野特徴キーワード」の一致度(距離)計算の結果上位 30 細目分野に対する対応分析の結果を示す。図 5-1 では(1)ゲノム関連 4 領域と関連の強い細目分野コードとの位置関係、図 5-2 では(2)ナノテクノロジー 5 領域と関連の強い分科細目コードの位置関係を示している。この図により、それぞれの関連の位置関係が示されるが、それぞれの相対距離で位置を示しているので関

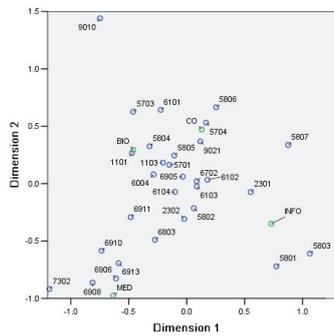


図5-1 ゲノム関連4領域と関連細目の関係

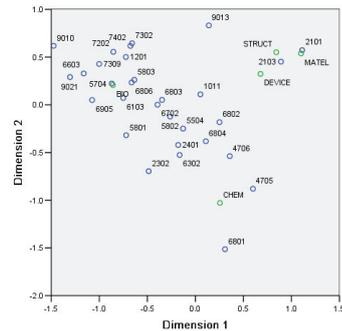


図5-2 ナノテク関連5領域と関連細目の関係

連の強さが直接示されているわけではないことに注意されたい。

(1)ゲノム関連に関しては、ゲノム科学(2301、2302)生物学、農学、医科学関係の細目が目立つ。(2)ナノテクノロジー関連に関しては、ナノ・マイクロ科学、工学、生物学、医学関連が目立つが、ゲノム関連に比べて、広い分野に分布している傾向が見受けられる。

現在はまだ分析の途中段階なので詳しい分析はできていない。今後さらに詳しく、また、科学技術基本計画の他の重点分野についても分析をおこない、その効果を検証したいと思っている。ここで用いた分科細目表とそのコード表は紙面の都合上割愛させていただいたが、参考文献[3]の「サービス概要・研究分野について」に示されているので、参照してもらいたい。

## まとめ

我々は、トピック関連キーワードと科研費の採択課題の課題名から切り出したキーワードを比較することによりゲノム・ナノテクノロジー関連課題の課題数・配分額の変遷を調べた。また、科研費の分科細目表に対応

した分野特徴キーワードとの一致度を見ることにより、重点領域との関連分野を対応分析により示した。過大見積りや相互関連関係の可視化等の問題もあり研究途上であるが、最終的にはすべての重点領域について分析を行う予定である。

## 参考文献

- [1] 総合科学技術会議、科学技術基本計画, <http://www8.cao.go.jp/cstp/kihonkeikaku/index3.html>
- [2] 科学研究費補助金, <http://www.jsps.go.jp/j-grantsinaid/index.html>
- [3] 2005年4月からKAKENK, KAKENデータベースを統合して科学研究費補助金データベース(KAKEN)としてサービスされている。: <http://seika.nii.ac.jp/>
- [4] 西澤 正己、孫 媛、「科研費分野変更に伴う情報学分野の変遷マッピング」, 情報知識学会 第12回研究報告会講演論文集, 2004年5月, pp.9-12
- [5] Masaki Nishizawa, Yuan Sun, “Research Fields related to Information Science under a new classification in Japan”, Proceedings on 4th International Conference on University Evaluation and Research Evaluation, pp67-73, Whan University, China, 27 September (2004)
- [6] 文部科学省ナノテクノロジー総合プロジェクト センター, <http://www.nanonet.go.jp/japanese/>

## 日本発行の科学技術雑誌数の分析

### Title Count Analysis of Journals Published in Japan

時実 象一<sup>1\*</sup>

Soichi Tokizane<sup>1\*</sup>

\*1 愛知大学文学部

\*1 Faculty of Letters, Aichi University

tokizane@aichi-u.ac.jp

要旨

#### 1 はじめに

わが国で発行されている科学技術雑誌の現状を把握するために 2 つの方法で分析した。まず多くの学術雑誌がすでに電子化されていることから電子ジャーナル・サイトに登載されている雑誌リストを調査し、重複を調べて実数を分析した。また、主要な文献データベースに収録されている雑誌を調査した。

データベースへの雑誌の収録状況についてはしばしば調査されており、日本発行雑誌についても報告がある [1,2,3,4]。ただし、これらはデータベース提供機関が公開している収録雑誌リストを基礎としており、過去に収録したが現在は収録していない雑誌もすべて含まれており、現状を必ずしも正しく反映していない。本報告では、できるだけデータベース作成機関からデータのご提供をいただき、それを基にして比較分析をおこなった。本発表は著者の別の発表 [5]

の内容を発展させたものである。

なお本調査では雑誌の同定は主として ISSN によったが、ISSN が見つからない場合は誌名と発行者名によった。

#### 2 国内電子ジャーナルへの収録

##### 状況

現在わが国では学術分野の電子ジャーナル・システムとして科学技術振興機構 (JST) が運営する J-STAGE、国立情報学研究所 (NII) が運営する電子図書館 (ELS)、株式会社メテオインターゲートが提供するメディカルオンライン、および株式会社サンメディアが提供する PierOnline がある。また海外の出版社 (日本にある子会社も含む) から出版されている雑誌も多数ある。それぞれのシステムに収録されている電子ジャーナル数を比較した (表 1)。

ここで、現行雑誌数とは現在発行され

表 1 各電子ジャーナルサイトの収録雑誌数

|                  | 雑誌リスト<br>掲載数 | 現行雑誌数 |             |     |
|------------------|--------------|-------|-------------|-----|
|                  |              | 合計    | 和文誌・混<br>載誌 | 英文誌 |
| NII-ELS          | 1003         | 395   | 317         | 78  |
| J-STAGE          | 381          | 378   | 213         | 165 |
| MedOnline        | 569          | 368   | 313         | 55  |
| PierOnline       | 5            | 0     | 0           | 0   |
| 海外出版社            | 119          | 119   | 0           | 119 |
| 電子ジャーナル<br>収録合計* |              | 1088  | 746         | 342 |

\* 重複を除いた数

ており、かつ各サイトにされている学術論文を掲載している雑誌である。NII-ELSでは大学の紀要の点数が多く、これらは除かれている。さらに雑誌の過去の号が異なる誌名により登載されている場合があるがこれも除いた。人文社会科学関係の雑誌も除いた。一方メディカルオンラインでは解説記事等からなる非論文誌が多数あるが、これも対象外とした。PierOnline は点数も少なく、非論文誌のみであった。

これによれば NII-ELS、J-STAGE、メディカルオンラインが全件数ではほぼ同等であったが、J-STAGE は英文誌の比率が高かった。

国内 3 サイトにおける国内発行 979 誌の電子ジャーナルの重複を比較したのが図 1 である。

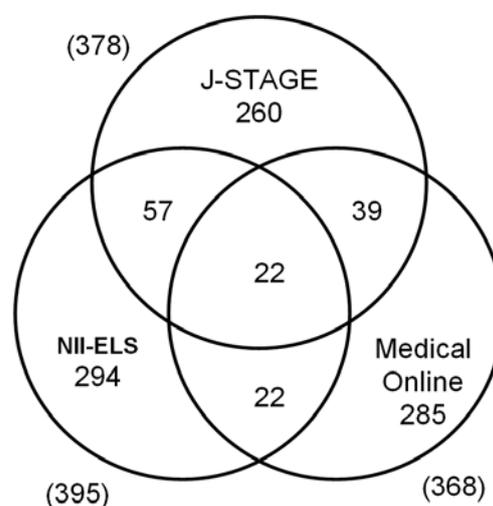


図 1 国内発行雑誌 979 誌の電子ジャーナルサイト間の重複

これから明らかなように、各サイト間の雑誌の重複は少ない。これを和文誌・和英混載誌 579 誌、英文誌 163 誌に分けて比較したのが図 2, 3 である。

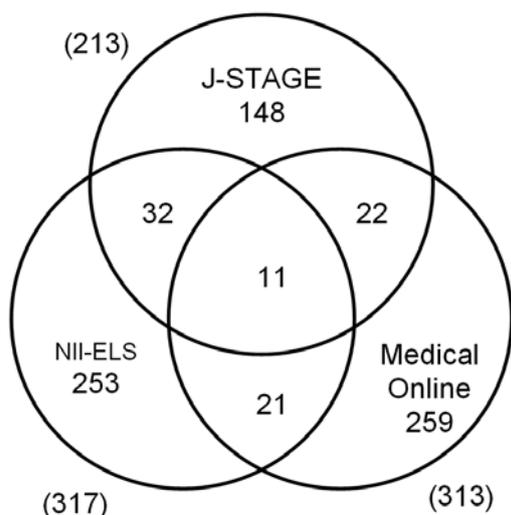


図 2 国内発行和文・和英混載誌 746 誌の電子ジャーナルサイト間の重複

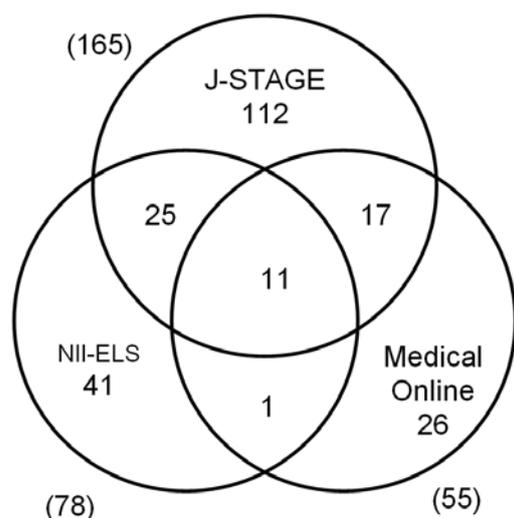


図 3 国内発行英文誌 342 誌の電子ジャーナルサイト間の重複

なお海外出版社発行 119 誌はすべて英文誌であるが、そのうち国内電子ジャーナルサイトに収録されていたのは 9 誌にすぎず、図 3 の各誌とはほとんど重複しなかった。その他若干の雑誌は学会により直接、またはその他のサイトから公開されているが、今回は集計していない。

### 3 国内文献データベースへの収

#### 録状況

今回調査対象としたデータベースは独立行政法人科学技術振興機構の JSTPlus/JMEDPlus、特定非営利活動法人(NPO)医学中央雑誌刊行会の医学中央雑誌データベース、財団法人日本医薬情報センターの JAPIC データベースである。それぞれ 2005 年の実際の収録誌のリストをいただいて分析した。これに加えて、国立情報学研究所 (NII) が作成する引用文献索引データベース (CJP) の収録雑誌リスト (ホームページより入手) も比較対象とした。

国内文献データベースに収録された

表 2 国内文献データベースの収録雑誌数 (2005 年現在)

|              | 雑誌リスト掲載数 | 純雑誌数 | 国内発行 |         |     |
|--------------|----------|------|------|---------|-----|
|              |          |      | 合計   | 和文誌・混載誌 | 英文誌 |
| JST          | 6387     | 3868 | 3839 | 3291    | 549 |
| 医学中央雑誌       | 2283     | 1879 | 1843 | 1619    | 224 |
| JAPIC        | 402      | 397  | 370  | 321     | 49  |
| CJP          | 1628     | 1304 | 1247 | 1061    | 191 |
| データベース収録合計 * |          | 4520 | 4451 | 3899    | 565 |

雑誌数 (2005 年現在) は表 2 のとおりである。雑誌リスト掲載数と純雑誌数の差は、前者には学会大会の予稿集や抄録集、大学紀要などが含まれている点である。ただしこの「純雑誌」には、論文誌と非論文誌は両方含まれている。

これによれば JSTPlus/JMEDPlus の収録誌数が一番多く、以下医学中央雑誌、NII の CJP の順であった。これら 4451 誌の重複状況を図 4 に比較した。

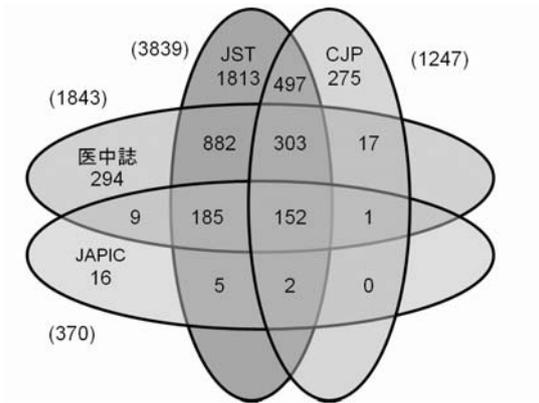


図 4 国内発行誌 4451 誌の文献データベース間での収録の重複

また和文誌・和英混載誌 3899 誌、英文誌 565 誌に分けて比較したのが図 2, 3 である。

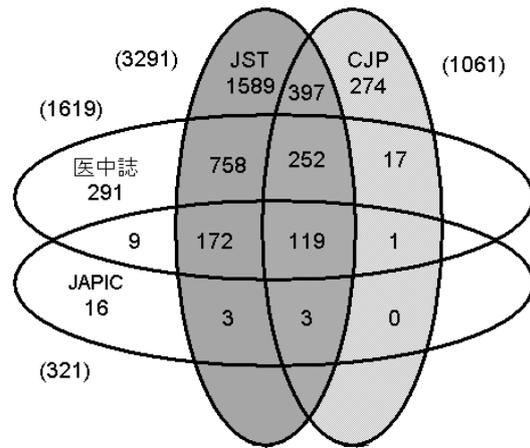


図 5 国内和文誌・和英混載誌 3899 誌の文献データベース間での収録の重複

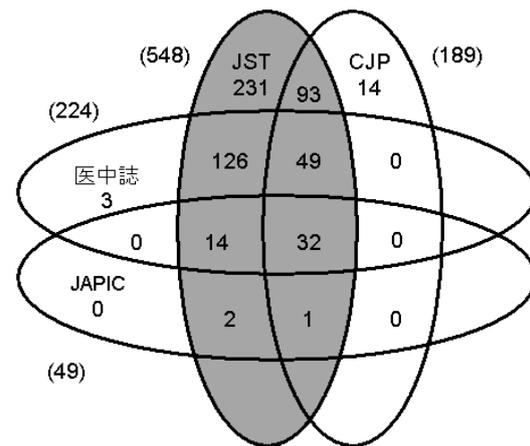


図 6 国内和文誌・和英混載誌 565 誌の文献データベース間での収録の重複

これらから、JSTPlus/JMEDPlus はその他のデータベースが収録している雑誌の多くを収録しており、他のデータベースでは独自収録誌が少ないことがわかる。

なお英文誌のうち海外出版社（国内支社も含む）から出版されているものについては図 7 のようになる。

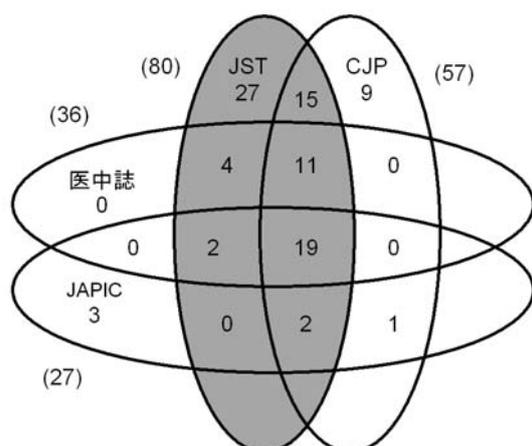


図 7 海外出版社発行 93 誌の文献データベース間での収録の重複

#### 4. 結論

データベース収録数から見て、国内で発行されている科学技術関係の雑誌の点数は約 4500 誌であるが、その中には非論文誌や小規模の研究会の雑誌なども含まれている。一方電子ジャーナルサイトに掲載されている雑誌数は約 1100 誌で、これらは概ね論文誌であり、主要な雑誌と考えられる。そのうち約 340 誌が英文誌であり、さらにその中の 119 誌が海外出版社から刊行されていることがわかる。

#### 謝辞

分析用のデータをご提供いただいた独立行政法人科学技術振興機構、特定非営利活動法人(NPO)医学中央雑誌刊行会、財団法人日本医薬情報センターに感謝いたします。

#### 参考文献

- [1] 上田修一; 出羽美子; 倉田敬子: 「海外主要データベースにおけるわが国科学技術雑誌の収録状況」, ドキュメンテーション研究, 1983, 33(9), 421-430.
- [2] Ikushima K; Tenopir C: "Availability of Japanese scientific and technical periodicals in major English language databases", Proc. Natl. Online. Meet. 1988, 115-122.
- [3] 原田 郁子: 「JST 収集・収録誌と海外主要データベース収録誌との比較」, 情報管理, 2004, 47(2), 96-101.
- [4] 殿崎正明, 「PubMed に収録されている日本の雑誌: 最近 3 年間の調査分析」, オンライン検索, 2006, 27(2), 95-105.
- [5] 時実 象一: 「日本発行雑誌の各種科学技術医学データベースへの収録状況」, 第3回情報プロフェッショナルシンポジウム予稿集, 2006, 75-78.

# 文化資源オントロジの構築とその活用 Constructing and Utilizing ontology for Cultural Resources

研谷紀夫<sup>1</sup> 馬場章<sup>2</sup>  
Togiya Norio Akira Baba

\*1 情報知識学会 所属 東京大学大学院情報学環  
Japan Society of Information and Knowledge, Interfaculty Initiative in Information Studies  
the University of Tokyo  
E-mail: togiya.norio@iii.u-tokyo.ac.jp,

2 東京大学大学院情報学環  
Interfaculty Initiative in Information Studies the University of Tokyo  
E-mail:baba@ iii.u-tokyo.ac.jp

歴史史料や様々な文化資源をデジタル化して公開するデジタルアーカイブの特長の一つは、様々な形態・内容の資料を統合的に扱えることであるが、多様な資料を1つのアーカイブに集約することは、資料を横断的に検索できるとともに、これまで資料群や形態ごとに分離していた資料相互の新たな関係性を見いだす契機ともなりうる。特に同時代に成立した資料は地理的条件や人物・社会組織、歴史的出来事などを通して、何らかの共通の情報を含んでいる。

本研究ではこれらの視点を踏まえ、オントロジを使用して歴史上の様々な事物の関連性を捉えた上で資料を統合するデジタルアーカイブを構築し、その検証を行う。

One of the characteristics of digital archives used to digitize and present historical materials and cultural resources is that they allow you to handle materials of various types and content synthetically. In addition, putting together various materials into one archive allows the possibility to find new relationships between materials that have been conventionally separated by type, as well as enable you to search materials in a cross-sector manner. Notably materials created in the same era may include common information through geological condition, figures, social organizations and historical events.

In this study, materials are integrated through an understanding of the relationship between various historical artifacts with ontology. An integrated digital archive is constructed based on these viewpoints and verified.

キーワード:デジタルアーカイブ, オントロジ, 文化資源, Digital Archive, Ontology, Cultural Resources

## 1 研究の目的とアーカイブの概要

歴史資料をデジタル化して公開するデジタルアーカイブの1つの大きな特長は様々な資料の形態やコレクションを統

合的に格納することができる点である。現物資料の保存施設においては、資料の形態や内容によって資料が分離して保存される傾向にあったがデジタルメデ

ィア上においてはこれらを統合的に扱うことが可能である。そのため、これまで分離されて扱われていた資料を複眼的な視点から結びつけ、資料の内容世界に基づいた事物の関係性から資料を閲覧することが可能となる。本研究では、歴史資料に関する様々な事物の関係性を明示する手段としてオントロジを使用し、それらがデジタルアーカイブにとって有用であるかを検証する。

現物資料においてはその形態や内容によって分離・保存される傾向があったため様々な資料をマクロ的に分析する場合には多くの課題があった。しかし、前述の2つの特質を活用することによって、より多様な資料を格納し、それらの資料を複眼的な視点で構築することが可能となる。これらの歴史研究上の課題を、解決する視点から本研究では、様々な資料を(1)統合的に格納できる汎用的なメタデータと、(2)多様な資料を様々な関係性から分類構成するためのオントロジの構築をその重要な機能として位置づけ、それらの設計を行うこととした。

これらの要素の中で重要なものは歴史資料に適応させた(2)のオントロジの構築である。歴史資料を検索する上で重要な要素は人名や地名、時代名などの固有名詞からの資料検索である。そのため本アーカイブでは、各人名と関係の深い地名や時代など固有名詞相互の関係性を複数の関係子で記述して検索の手段とした。このことによって、対象資料に関連する人名や地名を熟知しない閲覧者も、関連人物や組織、地名などの関連性を捉えながら資料の閲覧、選択を行いやすいように設計をした。そして、本研究においては、構築したオントロジを可視化することによって資料閲覧時のナビゲーションとして活用する点に重点を置いた。これらのプロジェクトの全体像につい

ては、研谷らによって述べられているが[1]、本論文においてはその中で特にオントロジの構築について詳説する。

## 2 オントロジの全体構成

### 2.1 オントロジの定義

オントロジは元来ギリシャ語で“存在”を表す語を語源としており、存在論などの思想・哲学における用語と使用されて来た。一方で 20 世紀に入り、知識工学などの研究分野において、機械に様々な知識や情報を理解させるための知識の体系化手法としてもこのオントロジという用語が使用されるようになった。

知識工学におけるオントロジにおいても複数の種類のオントロジがあり、存在論などの哲学的な議論を踏まえて、世界に存在する多くの事物の体系化を目指す上位オントロジと特定の産業や状況などに関する概念の統制を行うドメインオントロジなど様々なタイプのオントロジが存在する[2]。

またさらに、1990 年代以降のインターネットの普及により、WEBにおける情報を機械においても読み込みが可能なように知識情報を体系化する手法としてウェブオントロジが登場した [3]。これらを踏まえて本研究におけるオントロジを以下のように定義する。

- (1) 世界、あるいは特定の枠組みに存在する様々な概念・事物を対象として、その体系化を目指す。
- (2) 主述の関係や集合などの方法を用いて、様々な事物の概念・性質・意味を定義する。
- (3) 一定の規則に基づいたルール・言語で記述する。
- (4) 汎用性を目指し、再利用可能な共有化された知識とする。

これらの定義をそれぞれのオントロジにあてはめて考えると、前述の「上位オントロジ」は極めて広い世界を対象として(1)の事項を(2)から(4)の方法を用いて構築したオントロジである。またドメインオントロジにおいては特定の世界において(1)を目指して、(2)から(4)の手法を用いて構築されるオントロジである。またウェブオントロジなどは、(1)の主な対象をウェブ上のリソースと設定した上で、(2)から(4)の手法を用いて構築されるオントロジであると定義できる。但しウェブオントロジの現状に於ける重点はRDFやOWLなど(3)の記述言語の開発に、重点が置かれており、それらを使用してどのように(1)を実現していくかが大きな課題となっている。

本研究が課題とするオントロジは、主に歴史資料に掲載された世界を対象として(2)を用いて(1)を構築することを当面の目標とする。そしてこれらの結果を踏まえて、(4)を目標とするために、どのような(3)を用いて記述していくかを今後の課題とする。

## 2.2 先行研究

前項で示したように、本研究で構築するオントロジは、主に歴史資料に格納された人物・組織・地名などの固有名詞の関連性を、関係子で結びながら明示化することをその目標とする。このように歴史的な事物の体系化にオントロジを用いた先行事例は極めて少ないが、近似する例としては、メリーランド大学において、収集された口承伝承に登場する人物や地名との関係を理解するデータベースが、オントロジを用いて構築された例としてあげられる[4]。また世界の様々な歴史イベントを独自のXMLで記述しているHEML(The Historical Event Markup and Linking project)[5]などをあげることがで

きる。本プロジェクトではこれらの先行研究を踏まえながらも、対象を特定の領域の事物だけではなく日本の近世・近代期の歴史的な出来事に適応できる体系を構築することとした。

またその他のオントロジにおいては固有名詞や、実在した人物などは含まないこともあるが、本研究においては歴史上に実在した人物、組織、地名なども歴史的な“存在”としてその関係性を構築することとした。

## 2.3 オントロジの全体構成

オントロジ全体の構築にあたってはN. Guarinoの上位オントロジを基本として、主に次のような大分類を構成した[6]。

1. 場所空間、
2. 時間
3. 具体物
4. 抽象物
5. 事
6. 行為・作用
7. 社会
8. 現象

その上で各大分類以下に「isA」と「partOf」の関係子を用いて必要な事物概念を樹形列の形で配置し、これらを「概念ツリー」と名付けた。またこれらの垂直方向の系列とともに、各カテゴリに属する事物相互の横のつながりも表-1で示した関係子を用いることによって結びつけた。関係子は主に6つの部分から構成されており、本資料群における事物の関係性を構築できるように設定した。また人間関係を表現する「human relation」においては既に一般に利用されている「Relationship Vocabulary」を用いた[7]。

これらの関係子を活用して、資料に関する主要な人物の居住地や行動範囲、

生没年、社会組織や歴史的な出来事との関わりが関係づけられた。本研究で対象とした資料は明治期の人類学者で弥生式土器の発見者として知られる、坪井正五郎資料や日本の商業写真師の草分けである上野彦馬の写真資料などである。そのため各資料に関連する人間関係や様々な事物相互の関連記述が行われた。

例えば、坪井正五郎という人物名と、埼玉県比企郡吉見町という場所は、ツリー上では地名と人名の別カテゴリに入る。しかし、坪井正五郎は、たびたびこの地を訪れ、同町にある吉見百穴の発掘を行っている。そのため、坪井正五郎にとって吉見町は極めて関係性の強い土地である。この両者の関係を明示する方法としては、表-1 で示された関係子において、行動主体が特定の地を訪れたことを示す関係子として「visitTo」を用意している。これらを用いることで、坪井正五郎 (visitTo) 吉見町という関係を記述することができる。これらにより、坪井正五郎に関する各概念相互の関係性を構成することが可能である。これらを積み重ねることによって、坪井正五郎について一定の知識を持つ者しか把握しきれなかった、坪井正五郎とその他の事物概念との様々な関係をデータの中に蓄積し、閲覧者に提示することができる。

また、メタデータの「Subject」や「Coverage」のエレメントに概念ツリー上の適応する語を格納し、メタデータと概念ツリーとの間を接続し、「Subject」欄に示された各語をクリックするとそれと対応する相関図が表示可能となるように設計した。

## 2.4 アーカイブの全体構成

前項のような機能を基本として、本プロジェクトでは文化資源統合アーカイブ

表 1 主要な関係子

|                       |                                                                                                                                                                      |                                                                     |
|-----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|
| (Basic)               | isA<br>temporaryIsA<br>partOf<br>temporarypartOf                                                                                                                     | である<br>である (一時的)<br>の一部<br>の一部 (一時的)                                |
| (physical relation)   | consistOf<br>connectedTo<br>contain<br>own<br>interconnect<br>branchOf<br>ingredientOf                                                                               | から成り立つ<br>連結する<br>を含む<br>を所有<br>と相互関係<br>の支部<br>の成分・材料              |
| (spatial relation)    | locatedIn<br>locationOf<br>adjacentTo<br>surround<br>traverse                                                                                                        | にある<br>がある<br>近隣<br>に囲まれる<br>横切る                                    |
| (functional relation) | treat<br>interactWith<br>visitTo<br>succeedTo<br>researchIn<br>majorIn<br>liveIn<br>resultsOf<br>co-occursWith                                                       | 扱う・処遇<br>相互作用<br>訪れる<br>へ継承<br>研究・調査する<br>専攻する<br>住む<br>の結果<br>共起する |
| (conceptual relation) | analyzes<br>derivativeOf<br>developmentalform Of<br>methodOf<br>issueIn<br>relatedTo                                                                                 | 分析する<br>派生<br>より発展<br>の方法<br>論じる<br>関係する                            |
| (human relation)      | rel:worksWith<br>rel:spouseOf<br>rel:siblingOf<br>rel:parentOf<br>rel:knowsOf<br>rel:knowsInPassing<br>rel:knowsByReputation<br>rel:hasMet<br>rel:is a competitor of | 共働<br>配偶者<br>兄弟<br>の親<br>知人<br>少し知っている<br>噂を聞く<br>面識がある<br>ライバル     |

を構築した[8]。アーカイブの基本機能は主に、資料一覧、事物つながり検索、談話室などの機能が主要機能となっている。資料一覧では古写真、建築資料、人類学資料、古地図など、違うコレクションの資料を同一のビューで閲覧することが可能である。これらの資料には、紙資料、写真、彫像、絵画、映像、音、3Dオブジェクトなど多様な資料が格納されている。また、事物つながり検索においては、人、組織、地名などの様々な要素の関係性が関係子を使用して構築されており、こ



「良いが改善が必要」と答えた評価者から指摘された具体的な改善内容としては「インターフェイスを使いやすくする」「操作性を軽くする」などの改善の指摘が大部分を占めた。そのため、オントロジを用いた事物つながり検索に関しては利用しやすいインターフェイスを開発していくことが必要である。

### 3.2 課題と展望

今回構築したオントロジは、主にアーカイブの中で独立的に用いられる、歴史上実在する人物・組織などを中心とするオントロジである。これらのオントロジを構築することによって、資料内容を熟知していないユーザにおいても、資料に関連する人物や土地、組織などの関係性から資料を一覧し、選択することが可能となった。そのため、様々な歴史的事実と関連する資料を閲覧し検索ツールとして歴史的な事物に関するオントロジを構築することは極めて有効であると考えられる。

しかし、その一方でさらにユーザの利便性を向上させるためにインターフェイスの改善を行う必要がある。また、本研究で設定した関係子は主に今回対象となった資料群を対象としている。そのため、その他の資料にも適応できる汎用的関係子を体系化する必要がある。

また構築したオントロジは、本アーカイブのデータベースのみで完結するものである。そのため今後これらをウェブオントロジ言語などによって公開し、ウェブ上で様々な人々が使用できるリソースとする必要がある。さらに、メタデータについても順次ウェブオントロジ言語で記述し公開することにより歴史情報に関する情報と資料に関するメタデータの両者が共

通のオントロジ言語によって記述され、資料とそれを読み解く情報の両者が一体となった知識情報をウェブ上で形成することが可能となる。

さらに、本プロジェクトではオントロジを可視化して主にナビゲーションとして活用したが、今後は記述した様々な事物の関係性を活用して、歴史的な多様な文脈より資料の検索を行えるような検索機能を充実させる必要がある。

### 参考文献

- [1] 研谷紀夫 他, 「オントロジとコミュニティを用いた統合型デジタルアーカイブの構築の場合」, 情報処理学会シンポジウム じんもんこん 2006 論文集, 情報処理学会, pp.57-62, 2006
- [2] 溝口 理一郎, 「オントロジ — 工学知の科学」, pp.3-10, オーム社, 2005
- [3] Tim Berners-Lee, “ Semantic Web ” ,Scientific American. 284(5), pp.34-44, 2001
- [4] Ryen.W.White, Hyunyoung Song, and Jay Liu, “Concept Maps to Support Oral History Search and Use” ,Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries ,pp.192-194, 2006
- [5] <http://www.heml.org/>
- [6] Nicolas Guarino, “Some organizing principles for a unified top-level ontology”, Working Notes of AAAI Spring Symposium on Ontological Engineering, Stanford,
- [7] <http://vocab.org/relationship/>
- [8] <http://cr-arch.chi.iii.u-tokyo.ac.jp/>

## ターミノロジー論における階層関係の概念について

### The concept of the hierarchical relation in terminological theories

山本昭

Akira YAMAMOTO

愛知大学文学部

Faculty of Letters, Aichi University

**Abstract:** In terminology theories, hierarchical relationship between concepts is regarded to be central. After reviewing the classical theories and the criticism against them, the discussions and the points on the hierarchical relation are described, focusing on the individual concept, the partitive relation, and the context dependency of the hierarchy. The followings are concluded. In terminology, the hierarchical relations are subsequent relations resulting from the construction of a concept system, and therefore make few serious confusion in the terminology workplace even though they are inappropriately defined, compared to thesauri for information retrieval. The disciplinary view is reflected on the selection of essential characteristics. The arbitrariness and the intuitiveness are hardly avoidable in hierarchical relations. The exploitation of the viewpoint of hierarchical relation may be of use during the course of exploring the 'concept of concept' in terminology science, which is indispensable for the wider and multicultural application of the terminology theory.

**Keywords;** terminology, concept of concept, hierarchical relation, partitive relation, delimiting characteristics

#### 1. はじめに

ターミノロジー論 (terminology: 専門用語論) は、専門用語を扱う分野である。ターミノロジー論においては、「用語」と並んで「概念」は、中心的概念である。辞書学 (lexicology) が「語」を中心に扱うのに対しターミノロジーでは「概念」に注目し、概念間の関係をも取り扱う。概念間の関係の中で、概念の階層関係は中心的なものである。

Wüster により基礎を作られた近代ターミノロジーは実践、理論の両面から確立していった。実践においては、当初の適用範囲は

科学技術、工学分野を前提としてきた。この分野においては、社会科学、人文学分野と比較して「概念」や「特性」は明確に定義可能である。この前提のもとにターミノロジーの手法は確立し、標準化されてきた<sup>[1],[2],[3]</sup>。ターミノロジーにおいては、「概念」を中心として行われてきた。われわれは、「概念」を直接取り扱うことは不可能なので、やむを得ず「物的補助」<sup>[4]</sup>として「用語」を介して間接的に概念を扱う、という立場であった。

近年はこの古典的な理論の弱点が指摘されるようになり、新たな視点からの理論が提唱されてきた<sup>[5]</sup>。特に、社会科学、人文学な

どへ、あるいは日常生活との接点においては限界が指摘され、Socioterminologie 等の手法も導入されている<sup>[6],[7]</sup>。

ターミノロジーにおいては「同一の特性を持ち」つつ「異なる特性を持つ」概念を、階層関係にあると位置づける<sup>[3]</sup>。この階層関係は概念体系(concept system)を形成する上で重要な役割を果たす。

いっぽう、概念の階層関係を類似の方法で扱ってきたものとして、情報検索のためのシソーラスがある。(「シソーラス」には、同義語展開用の同義語辞書も含まれることがあるが、本論文では「シソーラス」と言った場合、索引言語としての伝統的シソーラスのみををさすものとする。)シソーラスにおいては、階層関係はより実用的に作成される。つまり文献で扱われている概念自体の階層関係を中心としながらも、時には文献の階層関係、さらには当該文献が解決するであろう情報要求の階層関係を考慮している<sup>[8]</sup>。

また、近年のオントロジー研究では、階層関係を当該主題領域における主たる関心から扱うことにより、混乱を回避しようとしている<sup>[9]</sup>。

本論文では、階層関係の概念の再検討のために、ターミノロジー論における階層関係に関する論点を整理する。

## 2. ターミノロジー論における「階層関係」の範囲

前述のように、ターミノロジーの理論は、科学技術、工学分野を前提としてきた。これらの分野においては、「より詳しく」「より特定の」という要請から、内包を増加させてい

く方法で記述するのと考えが常識的に行われてきた。すなわち、「上位から下位へ」の下位区分という視点である。ターミノロジーの原則および方法に関する国際規格である ISO 704<sup>[3]</sup>においても、この考え方は基本的に踏襲されている。

In a hierarchical relation, concepts are organized into levels where the superordinate concept is subdivided into at least one subordinate concept. Subordinate concepts at the same level and having the same criterion of subdivision are called coordinate concepts. Concepts are super-ordinate, subordinate or coordinate, not on their own, but always in relation to each other in a hierarchy. (ISO 704:2000)

階層関係においては、概念はレベルの考え方で組織される。上位概念は、少なくとも一つの下位概念に細分される。同一の区分基準によって区分され、かつ同じレベルにある(二つ以上の)概念を「同位概念」と呼ぶ。概念は、階層関係の中での相互の関係から上位概念、下位概念、同位概念となるのであり、それらの概念が自ら本来的に上位概念、下位概念、同位概念であるわけでない。(発表者訳)

「概念はアプリオリに存在し、それらを整理していく過程で概念間の関係が発生する」という解釈が行われている。

いっぽう、ターミノロジーの作業は実際には概念を直接取り扱うことはなく、それを指し示す表現(=用語および定義)を扱うことで行われる。実務的には、すでにある概念と、それに対応づけられた用語の関係を整

理していく作業が行われる。この過程においては複数の用語をその指し示す概念の類似性や共通特性からグループ化していく。「類似性」によるグループ化の一実現形態として「ある区分特性から見て共通の特性を持つ」グループ化が「同位関係」となり、そのグループに対応する概念が「上位概念」となる。ここでは「下位から上位へ」の視点が中心となる。

なお、「上位概念が少なくとも一つの下位概念に区分される」という箇所については、ISO/TC37(国際標準化機構 第37技術部会(用語及び他の言語資源における原則と調整))における、上記ISO 704 およびISO 1087-1<sup>[10]</sup>の審議過程で、日本と幹事国カナダとの間で論争になった。日本は、「…少なくとも二つの下位概念に…」とすべきと主張した。外延が同じなら内包も区別されるべきでないという主張である。この主張は、日本の委員には直感的に理解しやすいものであった。しかしながら、内包(備えるべき特性)が増えても外延(対象)は変わらないこともあり得る、ということで上記のような記述となった。

これは、概念に対する直感的理解が、文化に依存することを暗示している。

### 3. 階層関係の問題 (1)個別概念

「個別概念」とは、「単一の対象と結びついた概念」と定義される<sup>[3]</sup>。必ずしも、階層関係のなかで最下位に位置するものではない。「国際機関>国際連合(UN)>国際連合教育科学文化機関(UNESCO)」のように全体部分関係を含むことがあるためである。

仲本は「概念」の概念を「包括」という観点

からとらえ、「個々の対象に対する表象は包括性をもちえない」ために「個別概念を認めない」という立場をとっている<sup>[11]</sup>。「概念」の概念を「集合」として解釈している。この解釈は伝統的な理論とは異なるものである。

実際、ターミノロジーにおける「概念」の概念は、かなり直感的であり、その明示的定義もさまざまである。「概念」の概念についてのいくつかのアプローチは Pozzi により包括的に検討されている。そこでは、この基本的概念がかなり直感的に定義されていることが指摘されている<sup>[4]</sup>。このような直感的理解に基づく概念は、その解釈に文化依存性が高く、たとえば、東洋の学生は西欧の学生に比べて、習得時に「概念」の概念を直感的に理解しにくいと指摘されている<sup>[12]</sup>。ターミノロジーのロシア学派では、“notion”と“concept”を区別し、前者を論理的帰結、後者を個人的解釈の結果としている。概念(concept)を心理的(mental construct)と考えれば、それが包括的であろうがなかろうが関係ない。一方で、包括性を要求することは、論理的な規定で、なおかつ「共通の」特性をもつ、と論理的に規定している<sup>[13]</sup>。

### 4. 階層関係の問題 (2)全体部分関係

Wüster は概念間の関係を「存在論的關係(ontologische Beziehung)」と「論理的關係(logische Beziehung)」に区別している<sup>[1]</sup>。種類関係等の階層関係は前者に属し、全体部分関係は後者に属する。主として、機械全体と、それを構成する部品との関係を前提に考えていたようである。細野は階層関係における全体部分関係における部分概念

の把握の困難さを指摘している<sup>[14]</sup>。全体部分関係を階層関係として扱う利点は、検索用シソーラスにおいて認められる。特に地理的概念、組織関係、などにおいては包括的検索のためにきわめて有効である<sup>[15]</sup>。いっぽう、製品－構成部品、などの関係を持ち込むことは好ましくないとされている。ターミノロジーにおいては、全体部分関係を規定することにはシソーラスと比して有効性が見出しにくい。

## 5. 階層関係の問題 (3)文脈依存の

### 階層性

多ファセット(multi-facet)的な階層関係がターミノロジー論の基本になっている。これは異なる区分原理から見た特性の組み合わせにより概念を規定できるためである。

同様に、シソーラスも多ファセット的階層関係を持つ。情報検索においてはシソーラスの階層関係について、利用者は不都合を感じる。一つは、その場に適した階層関係が設定されていないとき、もう一つはその場に不適切な階層関係が設定されている場合である。前者は包括的検索の失敗を、後者はノイズの発生に結びつく。実際には熟練した検索者は、シソーラスに規定された以外の階層関係を補完し、検索語に加えることによってこれらの失敗から逃れている。また、「エタノール(下位)－燃料(上位)」のような階層関係は、検索および情報要求のある文脈では有効であり、他の文脈では層ではない。情報検索を目的とするシソーラスにおいては想定された利用者の想定された情報要求に対して、最大の利益をもたらすという観点で階層関係を設定することが

できる。しかしながら、ターミノロジーにおける階層関係の目的は、概念体系を構築することにより概念を明確にしていくことが主である。上記の例では、エネルギー分野において、「エタノール」は「飲料」の下位概念ではなく、「燃料」の下位概念であるべきであろう。これは、エネルギー分野という視点を表しているものであって、「エタノール」の本質が何か、ということとは関係しない。

このような、階層関係の文脈依存性に関して、シソーラスに比してターミノロジー論においては問題とされることが多くないようである。これは、シソーラス中に出現する階層関係を持った概念が、必ずしも、そこで検索のテーマとなっている、そのシソーラスやデータベースが対象としている分野における中心的な概念ばかりではないということである。たとえば、一手段として用いられている文献等の索引検索に用いる場合である。

それに対し、ターミノロジーは、分野を限定して行われる活動であり、扱う語数(概念数)も少ないことが多い。その意味では、文脈、視点を共有していることが前提になっている。当該分野において周知的な概念は、高度に階層化、構造化されている必要はない。文脈依存性に階層関係の設定は当該概念に対する当該分野の視点を提供するものであると言っても過言ではない。しかし、分野を越えたコミュニケーションのために、用語および概念のハーモナイゼーションを考える際、このような狭い了解に基づいた文脈依存性が障害になることは考えられる。

なお、オントロジー構築においては、「人間(上位)－教師(下位)」のような、「やめることで変化する関係」を「ロール」概念として「階層関係」と区別することにより、文脈や状

況の変化により可変な関係を排除する試みが行われている<sup>[9]</sup>。

## 6. 結論:階層関係を見る視点

情報検索用シソーラスにおける階層関係は、よりよい検索性能という明確な目的を持っている。このため、その論理性よりも検索上の都合で設定される。それに対してターミノロジーにおいては、用語および概念を組織化していく過程で発生するものであり、階層関係の形成自体が目的ではない。その点から、ターミノロジーにおいては階層関係は副次的に発生するものであり、重要度は低い。

また、狭い分野を対象としたターミノロジーでは、その目的は当該分野の専門家間のコミュニケーションの円滑化である。故に、当該分野の視点が、明確に階層関係に反映する。

個別概念に関する議論は「概念」の概念の議論に発展する。概念の区別特性として、どの特性を本質的なものとするかは、合目的的に決定せざるを得ない。

ターミノロジーの適用範囲の拡大に伴い、「概念」の概念の再検討は重要な課題となろう。「概念」の概念が、直感的に規定されている限り、階層関係の直感性、恣意性は避けられないであろう。逆に、「概念」の概念を研究する過程において、階層関係からのアプローチが有効性を持つことは可能である。

## 参考文献

[1] Wüster, E. "Einführung in die allgemeine Terminologielehre und

*terminologische Lexikographie*"

ヴュスター 中村幸雄、荒木啓介(訳): 『一般用語学入門および用語辞書編纂法』 東京. 情報科学技術協会. 1998 (ISBN 4-88951-032-X)

[2] 岡谷大, 尾関周: 『ターミノロジー学の理論と応用 : 情報学・工学・図書館学』 東京. 東京大学出版会. 2003 (ISBN 4-13-001037-9)

[3] ISO 704:2000(E) Terminology work - principles and methods

[4] Pozzi, M.: "The concept of 'concept' in terminology: a need for a new approach", Proceedings of the Fifth International Congress on Terminology and Knowledge Engineering (TKE' 99), Sandrini, P. ed., Innsbruck, August 1999, Vienna, TermNet, 1999, p. 28-42 (ISBN 3-901010-24-6)

[5] Cabré, M.T.: "Theories of Terminology" Terminology 9(2) p.163-199 (2003)

[6] 山本昭: 「SOCIOTERMINOLOGIE の方法」第 17 回専門用語シンポジウム 2004 年 12 月東京、東京. 情報知識学会専門用語研究部会. 2004 p.26-31

[7] Gaudin, F.: "Socioterminologie: Une approche sociolinguistique de la terminologie" Bruxelles, Edition Duculot, 2003 (ISBN 2-8011-1319-0)

[8] 武田宣之: 「シソーラスの階層構造における問題点: 情報検索の結果に及ぼすその影響」 オンライン検索, 15(4) p.183-186 (1994)

[9] 溝口理一郎(編): 『オントロジー構築入門』 東京. オーム社. 2006 (ISBN 4-274-20292-5)

[10] ISO 1087-1:2000(E) Terminology work

- vocabulary
- [11] 仲本秀四郎:『専門用語論:ターミノロジー』東京. 日本図書館協会. 2006 (ISBN 4-8204-0523-3)
- [12] Wright, S.E.: Private Communication (2006)
- [13] Leitchik, V.M.; Shelov, S.D.: "Some Basic Concepts of Terminology: Traditions and Innovations", Journal of the International Institute for Terminology Research, 14, p.86-101 (2003)
- [14] 細野公男:「概念間の関係に関する一考案:全体-部分関係を中心に」情報知識学会誌, 10(1) p.28-39 (2000)
- [15] Aitchison, J., Gilchrist, A.: "Thesaurus construction: a practical manual" 内藤衛亮 (訳)『シソーラス構築法』東京. 丸善. 1989 (ISBN 4-621-03385-9)

## シンポジウム

【テーマ】 Web2.0 時代の情報システム

【テーマ解説】

世の中では Web2.0 という言葉が各種の雑誌にも取り上げられ、新しいコンセプトに基づく情報システムの時代が花開いているかのように言われている。しかし、実際には図書館や教育機関をはじめとした多くの機関で旧態依然としたシステムが使われ続けているのが現状である。この背景には、情報システムの設計者や利用者の多くが Web2.0 の概念についてはある程度理解していても、実際にどのようなサービスができるかなどという具体的なイメージを持つことができていないために、実際的な提案が困難であることなどがあげられよう。すなわち、概念としての Web2.0 の理解と現実的な応用との乖離である。このシンポジウムでは、既存のシステムと Web2.0 時代時代のシステムとの違いなどについて具体的な例を含めて検討することで、その考え方と有効性などを探る。

【パネリスト】

岡本 真 (おかもと まこと)

1973 年東京都生まれ。国際基督教大学卒業。1998 年より、インターネットの学術利用をテーマにした専門サイトである"ACDEMIC RESOURCE GUIDE"を運営。メールマガジンを発行するかたわら、主に研究環境の電子化、電子図書館化、蔵書検索 (OPAC) の高度化、デジタルアーカイブ、ウェブアーカイブについて執筆・講演活動を精力的に行う。また IT 企業でウェブプロデューサーも担当。

兼宗 進 (かねむねすすむ)

1963 年東京都生まれ。1987 年千葉大学工学部電子工学科卒業、1989 年筑波大学大学院理工学研究科修士課程修了。15 年間の企業勤務ののち、2004 年筑波大学大学院ビジネス科学研究科博士課程修了。博士(システムズ・マネジメント)。2004 年より一橋大学総合情報処理センター助教授(現、准教授)。主な研究対象分野はコンピューター教育など。自ら教育用プログラミング言語「ドリトル」の開発も手がける。また、図書館システムなどにも造詣が深い。

【司会】

宇陀 則彦(筑波大学大学院図書館情報メディア研究科 准教授)

## 論文賞の投票方式の変更について

情報知識学会では、情報知識学のさらなる発展に貢献するとともに、本学会の社会的評価を高め、かつ本学会誌のレベルアップをはかることを意図して、2004年より論文賞を設けてきました。これまでは、論文賞にふさわしい優れた論文を学会誌編集委員の投票によって決定してきましたが、編集委員の中から「選考に関してもっと会員の声が届くようにできないか」という意見がでています。このため、編集委員会で改善策を協議し、次のように変更いたします。

1. 学会誌編集委員会とは別に論文賞推薦委員会（委員長：根岸副会長）を設ける。
2. 本学会員は、この委員会に対して論文賞にふさわしいと思われる論文をその理由をつけて推薦することができる。
3. 論文賞推薦委員会は会員から推薦された論文の総数が4件以上の場合には、一次選考を行って3論文程度を論文賞の候補とする。また会員からの推薦論文の総数が2件以下の場合には、推薦委員会自らが理由をつけて候補論文を追加推薦し、候補を3論文程度にする。
4. 3件程度の論文賞候補論文とその推薦理由を学会ホームページおよび学会誌に掲載し、会員に周知する。
5. 会員による投票の結果、最多得票を得た候補論文を論文賞とする。ただし、論文賞推薦委員会は得票数や論文内容などを勘案し、得票数が第2位の論文についても論文賞とすることができる。

このように、論文賞選考において会員の意見を直接反映させることによって、会員の論文投稿へのインセンティブを高めるとともに、「学会員の選ぶ論文賞」という他では例のない新しい試みを導入することで、社会の注目をより集めることを期待しています。

日程としては、上記2の会員による推薦はNo.4が刊行された後、すなわち論文賞対象年の12月ごろからの1か月間、上記4の候補論文とその推薦理由の提示は翌年1月ごろ、会員による投票は3月ごろを想定しています。

以上のような改正案について、ご意見がございましたら編集委員会や事務局あるいは会長・副会長あてにお寄せください。論文賞の変更につきましては、決定次第順次学会ホームページや学会誌に掲載していきます。

会長  
細野公男  
編集委員長  
国沢 隆

「情報知識学会誌」投稿規定

2002年8月27日 制定

2003年3月19日 一部改定

2006年8月 1日 一部改定

0. 情報知識学会誌編集規程による本会機関誌「情報知識学会誌（以下、会誌という）」への投稿に関する事項は、この規定の定めるところによる。

1. 投稿資格

投稿者の少なくとも1人は本会員でなければならない。ただし、編集委員会による依頼原稿の場合にはこの限りではない。

2. 投稿原稿

2.1 広い意味での情報知識学に関連し、またその発展に貢献するもの（情報／知識の収集、整理、蓄積、検索および各種解析、利用などに関するもの）とする。刊行時において未発表の原著でなければならない。本会誌の記事の種類を以下に示す。

2.2 投稿者は会誌記事の種類を明記して投稿しなければならない。ただし、編集委員会で変更することがある。

(1) 研究論文 (Research Paper) : オリジナルな研究論文で、内容の主要な部分が学術論文として他に公表されていないもの。

(2) 事例／調査報告 (Report) : 情報知識学に関連したシステムなどの開発、利用、調査に関するもの。資料も含む。

(3) 解説／展望 (Review) : 情報知識に関連した特定分野の論文や学説などを総括、解説、紹介、あるいは技術動向などを展望したもの。技術、研究上の処理、解析方法などに関する解説。

(4) 論談 (Proposal Paper) : 情報知識学に関連した新たな意見の表明、提案など。

(5) 討論 (Discussion) : 本会誌に掲載された論文についての学術的な討論。

(6) 研究速報 (Notes) : 技術、手法、新事実などの簡単な報告。

(7) 講座 (Lecture) : 情報知識学の各分野に関する基礎理論、技術の適用などについて、テーマを定めて系統的に説明するもの。

(8) 学会記事 (News) : 本会の事業、運営などの報告、記事、資料など。

(9) ニュース、お知らせ (News) : ニュース、お知らせ。最近刊行された単行本やモノグラフの紹介

(10) 講演 (Lecture) : 特別号などにおける講演資料。

(11) その他 : 編集委員会が適当と判断したもの。

2.3 会誌記事の種類のうち、(1)から(6)までは査読を行う。その他については編集委員会で編集を行う。

3. 投稿原稿

3.1 原稿の形式

(1) 最初の投稿時

原則として、以下の体裁で作成された電子ファイル (PDF形式) を電子メールに添付し

た投稿とする。体裁は、刷り上り原稿を想定したレイアウト（A4判，2段組，20字×46行×2段）にして、図，表は希望の位置に配置すること。その他の執筆に関する詳細は「執筆要領」を参照のこと。

## (2) 採択決定後の原稿

PDFとその元になったファイル（Word fileなどで編集可能なもの）。

### 3.2 原稿の制限

#### (1) 原稿の長さを原則として次のように制限する。

研究論文，事例／調査報告，解説／展望，論談：刷り上がり20ページ以内

討論，研究速報，講座：刷り上がり6ページ以内

ニュース他：刷り上がり2ページ以内

#### (2) 冊子体の図原稿（原図）の大きさはA3判を越えないものとする。

(3) 原則として，図版も含めてモノクロ印刷とする。ただし，カラーでなければならぬ図版を使用する場合は，別途編集委員会と相談する。なお，カラーページやページを超過する分については，印刷費を著者の全額負担とする。本学会誌はJ-STAGEから電子ジャーナルとしても公開するので，カラーの図をWeb上の電子付録とすることができる。また，冊子体よりもより詳細な図表やさらには動画も電子付録とすることが可能である。電子付録はすべて無料で利用できる。

#### (4) 使用言語は日本語または英語とする。

### 4. 原稿の採否

投稿原稿の採否は，専門家による査読の後，編集委員会において決定する。

### 5. 査読のプロセス

学会員の中から編集委員会が指名した査読者2名によって査読を行う。内容によっては，編集委員会は著者に照会し，原稿の修正を求めたうえで，再査読を行うことがある。

### 6. 校正のプロセス

採択が決定した投稿原稿は，掲載原稿として著者に校正を依頼する。著者による校正は原則として1回とする。その際，字句の修正以外は原則として認めない。

### 7. 別刷

別刷（抜刷）は著者の実費負担とする。希望部数を事務局に申し出ること。

### 8. 投稿の手続き

最初の原稿投稿時には下記のファイルを電子メールに添付する。

#### 8.1 必要ファイル

a. 投稿原稿整理カード：ホームページからコピーして，必要事項を記入したテキストファイル。

b. 論文原稿のPDF形式ファイル（図，表を含む）

#### 8.2 原稿の送付先

学会誌編集委員会委員長 E-mail: kunisawa@rs.noda.tus.ac.jp

なお，以下の2つのアドレスにもCCメールとして送ること。

学会誌編集委員会副委員長 E-mail: ashino@acm.org

情報知識学会事務局 E-mail: jsik@nifty.com

#### 8.3 原稿の受付

事務局が原稿を受け取った日を受付日とする。受付の確認を1週間以内に投稿者の連絡先にE-mailで通知する。不備のある投稿原稿は返送し，再提出するものとする。

#### 8.4 著者は査読候補者リスト（5名程度の住所，所属，電子メールアドレスを記入した

もの)を提出できるものとする。

9. 原稿提出期日

投稿は随時とする。ただし、特集号などは除く。

10. 著作権

10.1 機関誌『情報知識学会誌』に掲載された論文(電子版を含む)の著作権(著作財産権, copyright)は情報知識学会に帰属する。

10.2 掲載論文は冊子による出版の他, 電子的に蓄積し, 本会が行う情報提供サービスなどを通じて公開する。

10.3 本学会誌に掲載された執筆内容が第三者の著作権を侵害するなどの指摘がなされた場合には, 執筆者がその責任を負う。

11. 規定の改訂

11.1 本規定の改訂は, 編集委員会の議を経て, 理事会の承認を得なければならない。

12. 施行

12.1 本規定は2006年7月1日より施行する。

12.2 本規定の施行により, 現行規定(第5版(暫定板)2003年3月)は廃止する。

13. 改訂履歴

2003年3月19日一部改訂。「10.著作権」に, 10.3項を追加。

2006年8月1日一部改訂。投稿手段を郵送から電子メールに変更。

## [事務局からのお知らせ]

### 平成 19 年度年会費の納入をお願いします

平成 19 年度(2007 年 4 月 1 日～2008 年 3 月末日)の年会費を、郵便局または銀行の下記口座へお振込願います。1 年分の年会費は正会員 8 千円、学生会員 4 千円です。前年度分未納のかたは合計額を納入してください。請求書が必要なかたは、その旨、事務局へお知らせ願います。

#### 1. 振込先 (振込手数料はご本人負担でお願いします)

- a. 郵便振替口座 00150-8-706543 情報知識学会 (代表 細野公男)
- b. 三菱東京UFJ銀行 秋葉原駅前支店 普通預金 3586133 情報知識学会 (会長 細野公男)

#### 2. 納入した年月日の確認方法

情報知識学会から送られた学会誌など、封筒の宛名ラベルをご覧ください。[ ] 内に過去 4 年間、ご自分の納入日が印字されているので確認できます。納入年(西暦の下 2 桁)、月(2 桁)、日(2 桁)の 6 桁です。年会費を滞納している場合は、[未納]と表示してあります。金融機関へ振り込まれてから事務局へ通知が届き、宛名ラベルに印字、発送するまで 10 日ほどかかりますので、ご了承ください。

情報知識学会事務局

〒110-8560 東京都台東区台東 1-5 凸版印刷隣内

TEL:03-3835-5692 FAX:03-3837-0368

E-mail:jsik@nifty.com URL:http://www.jsik.jp

**第15回(2007年度)情報知識学会研究大会実行委員会**

|        |       |           |
|--------|-------|-----------|
| 実行委員長  | 原田 隆史 | 慶應義塾大学    |
| 担当常任理事 | 宇陀 則彦 | 筑波大学      |
| 実行委員   | 春山 暁美 |           |
|        | 江草 由佳 | 国立教育政策研究所 |

**情報知識学会誌 編集委員会**

|        |       |        |
|--------|-------|--------|
| 編集委員長  | 国沢 隆  | 東京理科大学 |
| 副編集委員長 | 芦野 俊宏 | 東洋大学   |

**編集委員**

|        |            |        |          |
|--------|------------|--------|----------|
| 相田 満   | 国文学研究資料館   | 石井 守   | 情報通信研究機構 |
| 石塚 英弘  | 筑波大学       | 岩田 覚   | 東京大学     |
| 内田 努   | 北海道大学      | 宇陀 則彦  | 筑波大学     |
| 江草 由佳  | 国立教育政策研究所  | 大久保 公策 | 国立遺伝学研究所 |
| 岡本 由起子 | 東京家政学院大学   | 小川 恵司  | 凸版印刷(株)  |
| 神立 孝一  | 創価大学       | 五島 敏芳  | 国文学研究資料館 |
| 阪口 哲男  | 筑波大学       | 菅原 秀明  | 国立遺伝学研究所 |
| 太原 育夫  | 東京理科大学     | 田良島 哲  | 東京国立博物館  |
| 時実 象一  | 愛知大学       | 中川 優   | 和歌山大学    |
| 長田 孝治  | (株)カテナ     | 長塚 隆   | 鶴見大学     |
| 中山 堯   | 神奈川大学      | 中山 伸一  | 筑波大学     |
| 西川 信孝  | みずほ情報総研(株) | 西澤 正巳  | 国立情報学研究所 |
| 西脇 二一  | 奈良大学       | 根岸 正光  | 国立情報学研究所 |
| 原 正一郎  | 京都大学       | 原田 隆史  | 慶應義塾大学   |
| 藤井 賢一  | 産業技術総合研究所  | 藤原 譲   |          |
| 細野 公男  |            | 村川 毅彦  | 和歌山大学    |
| 安永 尚志  | 国文学研究資料館   | 山本 毅雄  |          |
| 山本 昭   | 愛知大学       |        |          |

■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外は、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。

著作物の転載、翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107-0052 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会

TEL: 03-3475-5618 FAX: 03-3475-5619 E-mail: naka-atsu@muj.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA. 01923, USA

TEL: 978-750-8400 FAX: 978-750-4744 URL: <http://www.copyright.com/>

情報知識学会誌 Vol.17, No.2 2007年 5月 25日発行 編集・発行 情報知識学会

頒布価格 3000 円

**情報知識学会(JSIK: Japan Society of Information and Knowledge)**

会長 細野 公男

事務局

〒110-8560 東京都台東区台東 1-5-1 凸版印刷(株)内

TEL: 03(3835)5692 FAX: 03(3837)0368 E-mail: [jsik@nifty.com](mailto:jsik@nifty.com)

URL: <http://www.jsik.jp>

# *Journal of Japan Society of Information and Knowledge*

## Contents

|                                                                                                                                                                                                   |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Program</b>                                                                                                                                                                                    | 59  |
| Automatically Classifying the Reference Records into NDC<br>Takashi HARADA, Masaki ETO, Minako ONISHI                                                                                             | 61  |
| Measuring Distance between Cited Papers in a Citing Paper<br>Masaki ETO                                                                                                                           | 65  |
| A SRU/SRW-based Information Retrieval System for Bibliographic Data<br>at Library of Education<br>Yuka EGUSA, Masao TAKAKU                                                                        | 69  |
| A Prototype for a Research Area Browsing System focusing on Author-Keyword Relationship<br>Noriko LEBOWITZ, Atsushi MATSUMURA, Norihiko UDA                                                       | 75  |
| Realizing Abduction, and Making up the structure of Lattice of behaviors<br>Masao FUKUNAGA                                                                                                        | 81  |
| The History of the Music Copyright in France between from the Revolution to the<br>19th Century.<br>Daisuke ISHII                                                                                 | 85  |
| Usage of a Scientific Dictionary for Analysis and Visualization for Scientific and<br>Technological Document<br>Akira KOUDA, Satoru BANNAI                                                        | 91  |
| Improving Web Retrieval Precision base on User Web browsing behavior<br>Yukio HORI Yoshiro IMAI Takashi NAKAYAMA                                                                                  | 95  |
| Extraction of Domain Knowledge from the Explanations of Japanese Classics<br>Takashi NAGATSUKA, Noriko KANDO                                                                                      | 101 |
| Database Systems for Humanities —Scripture Retrieval and Genealogical Visualization—<br>Ming ZHE PIAO, Masashi MORIMOTO, Junji TACHIBANA,<br>Takehiko MURAKAWA, Keigo UTSUNOMIYA, Masaru NAKAGAWA | 105 |
| The Effect of Grants-in-Aid for Scientific Research:<br>its analysis through researchers'career stages<br>Sumio KAKINUMA, Masaki NISHIZAWA, Yuan SUN, Masamitsu NEGISHI                           | 111 |
| Investigation into Genome-related and Nanotechnology Research at Grants-in-Aid in JAPAN<br>Masaki NISHIZAWA and Yuan SUN                                                                          | 117 |
| Title Count Analysis of Journals Published in Japan<br>Soichi TOKIZANE                                                                                                                            | 123 |
| Constructing and Utilizing ontology for Cultural Resources<br>Norio TOGIYA, Akira BABA                                                                                                            | 129 |
| The concept of the hierarchical relation in terminological theories<br>Akira YAMAMOTO                                                                                                             | 135 |
| <b>Symposium</b>                                                                                                                                                                                  |     |
| How to build information systems now and future in the time of the Web2.0                                                                                                                         | 141 |
| Information for members, Instructions for Author                                                                                                                                                  | 142 |

情報知識学会誌 第17巻2号 2007年5月25日発行

編集兼発行人 情報知識学会 〒110-8560 東京都台東区台東1-5-1 凸版印刷(株)内

TEL:03(3835)5692 FAX:03(3837)0368 E-mail:jsik@nifty.com

URL: <http://www.jsik.jp/> (振替 : 00150-8-706543)

学術刊行物 ISSN0917-1436