

*Journal of Japan Society of Information and Knowledge*

# 情報知識学会誌

Vol.26 No.3 (Sep. 2016)

目 次

卷頭言	会長就任にあたって	
	長塚 隆	249
研究論文	時系列ビブリオメトリクスによる研究活動分類手法の提案	
	川村隆治, 山下泰弘, 松邑勝治	251
研究論文	FRSAD と Linked Data に基づく主題情報共有システム	
	田辺浩介, 江草由佳, 高久雅生	260
研究論文	学術文献における引用分類の観点	
	柴田大輔, 芳鐘冬樹	277
お知らせ	平成 28 年度総会議事録	297
	事務局からお知らせ	309

## トップパンの、変革と挑戦。

これまで、世界地図が幾度も刷り直されてきたように、  
私たちトップパンも、印刷の枠組みを超えて、世界の在り方の変革に貢献してきました。

その背景には、トップパンならではの「印刷テクノロジー」の存在があります。

印刷を核に挑戦を続け、体系化してきたさまざまな技術。  
社員一人ひとりに刻み込まれた知識、ノウハウ、おもい。  
これらを包含したものを、私たちは「印刷テクノロジー」と呼んでいます。

この「印刷テクノロジー」を軸に、  
分野の壁を越え、あなたのおもいに応えるパートナーに。  
人々の生活に、健康や安全、安心を届け、より心豊かなものに。  
情報やメディアの変化への対応、地球環境保全など、  
社会の課題解決の一翼を担う企業に。

私たちはお約束します。  
あなたの立場で考える、豊かで美しい感性を持つ多彩な「人財」が、  
トータルソリューションを生み出し、世界を変えていくことを。  
その変革を、決して止めないことを。

# 印刷テクノロジーで、 世界を変える。

# TOPPAN

[www.toppan.co.jp](http://www.toppan.co.jp)

凸版印刷株式会社 〒101-0024 東京都千代田区神田和泉町1番地

## 卷頭言

### 会長就任にあたって

長塚 隆

「情報と知識」についてさまざまな側面から考え、研究し、議論していく場としての「情報知識学会」は、1988年4月に設立されました。再来年の2018年には設立30周年の節目の年を迎えることになります。

30年は、それぞれの人の一生で考えても長い年月です。私自身は、微生物の研究から全く異なる領域である情報検索の仕事に転身して6年目でしたが、1988年4月の本会設立時に会員として参加することが出来ました。今では、設立時からの数少ない会員の一人ということになるかもしれません。

学会は、会員の皆様一人一人の力で支えられ、会員の積極的な参加のもとで運営されるべきと考えます。このような考えは、歴代の会長を担当して頂いてきた先輩の諸先生方のお考えとも一致しているのではないかと思います。

学会が、2018年の設立30周年を機に、新たな発展の時代を迎えられるように、会員の皆様と共に考えて行動してゆきたいと思っております。

「情報知識学会」が設立された1980年代はコンピュータの小型化と高性能化の急速な進展とコンピュータネットワークの拡大を背景に、様々な「情報」がデジタル形式で、まだその多くがテキストデータでしたが、データベースを通じて提供・利用が拡大していた時代と言えます。

その後のインターネットやWeb環境の発展の中で、「情報知識学会」に期待されることも変化し、拡大してきたと言えます。今から、8年前の設立20周年（2008年）の時の思いを振り返ることで、今後のことを考える一助に出来ればと思います。

20周年記念特別号（本会誌2008 Vol. 18, No. 5）に掲載された本学会の当時の会長細野公男氏は、式辞の中で次のように述べておられます。

本学会は、データ・情報・知識・文献などを対象とする「情報学」あるいは「情報知識学」振興の必要性に対する強い社会的ニーズに応えて、日本学術会議情報関係の3研究連絡委員会（情報学、学術文献情報、学術データ情報）を中心として設立されました。・・・米田幸夫および藤原鎮男両会長、月見里禮次郎副会長、小谷正雄、大塚明朗、湯川康秀、島内武彦の各先生、それに現在も本学会の重鎮としてご活躍いただいている方々の献身的なご努力・ご尽力、さらに凸版印刷株式会社、大日本印刷株式会社等による財政的・人的ご支援のおかげで、今日の日を迎えることができたのです。

ここで述べられておりますように、本学会の大きな特徴は、現代社会における「情報と知識」についてさまざまな側面から考え、研究し、議論していく場であるということでしょう。また、大学や研究機関だけでなく、「情報と知識」の利用・活用の場でもある企業の方との共同により、学会がより充実した内容のあるものとなっていることも本学会の大切な特徴と言えると思います。

ここで、私自身の学会とのかかわりについて、紹介したいと思います。1988年の設立時に入会し、その後、2003年に理事、2007年から常務理事、2010年から副会長を務めさせて頂いてきました。

私自身は、学会の活動のなかでも印象深い経験としては、比較的長い期間、「情報

知識学フォーラム」に関わらせて頂き、その中で自分自身も得るものが多くかったです。

「情報知識学フォーラム」は、本年（2016年）で、21回を数えることになります。情報知識学会の催しの中でも、春の研究報告会と並んで年の後半に開催される情報知識学フォーラムは大きな柱のひとつとなっていますし、「情報と知識」についてさまざまな側面から考え、研究し、議論していく場としての「情報知識学会」にとって、その時期における大切なテーマについて、様々な角度から議論を深めることができます。出来る大切な機会になっています。

私が情報知識学フォーラムの実行委員を担当していた、今から10年少し前に、前身である「SGML/XML 研修フォーラム」から、様々な議論の末に、現在の「情報知識学フォーラム」に、名称を変更し、学会の全体をカバーするフォーラムとして、継続されてきています。その当時、若手の会員が中心になって企画し、かつ発表の場となるようなフォーラムにできないかという意見も多くありました。

このような意見も反映しながら、毎年実施されてきました。情報知識学フォーラムが、その時期に注目されるトピックスを取り上げ、情報知識学の新たな展開を探るものとして発展し、今後とも、若手の会員を中心新たに課題を探る場として発展してゆくことを期待しています。

2011年の東日本大震災、本年の熊本地震など、現在の社会には様々な困難も多くあります。スマートフォンやタブレットなどの携帯端末の普及、クラウドコンピューティング、ビックデータ、オープンソース、オープンデータ、人工知能など近年の情報技術の進展は、私たちが対象とする「情報と知識」の在り方に大きな変化を引き起こしていますし、その研究対象も大きく拡大していると言えます。

このような時期に、歴代の会長を始め、会員の方々が達成された実績を踏まえた

うえで、会員の皆様と共に時代にふさわしい学会の活動とあり方を追求して行けたと考えております。

学会としては、年次大会、情報知識学フォーラムを継続して開催してゆくとともに、人文社会科学系部会、専門用語研究部会、CODATA 部会、関西部会、シニア情報知識学研究部会など部会での活動をより促進し、多様な活動の形態を進めてゆくことが大切と考えています。

また、これらの活動を進めるためには、会員間の情報を発信し、共有する手段である「メールマガジン」を、今後も定期的に発行してゆくことが大切と考えています。

2018年の設立30周年に向けて、会員の皆さまのご協力のもと、共に、より充実した学会の活動が行えるよう努力してゆきたいと考えております。会員の皆様の会の運営へのより一層の積極的なご支援・ご協力を願いいたします。

研究論文

## 時系列ビブリオメトリクスによる研究活動分類手法の提案

### Proposal of Research Activity Classification Based on Time-series Bibliometrics

川村 隆浩<sup>†\*</sup> 山下 泰弘<sup>†</sup> 松邑 勝治<sup>†</sup>

Takahiro KAWAMURA Yasuhiro YAMASHITA Katsuji MATSUMURA

<sup>†</sup> 国立研究開発法人 科学技術振興機構 情報企画部 情報分析室

Department of Information Planning, Japan Science and Technology Agency

〒102-8666 千代田区四番町 5-3 サイエンスプラザ

EMail: takahiro.kawamura@jst.go.jp

現在、論文数や被引用数といった研究業績の量に基準を設けて研究者を判断することがしばしば行われている。しかし、基準の設定は分野によって異なる上、経験則に頼らざるを得ない。また、単純な合計では年長者が有利になり易い等の問題が指摘されている。そこで、本論では業績の量に加えて時系列的変化からも特徴を見つけ出し、それに基づいて研究者を分類することを試みる。具体的には、大規模な文献データセットからビブリオメトリクスの時系列パターン群を抽出し、機械学習手法によって研究者を“優秀”かそうでないか分類するモデルを構築する。実験では、JST と Scopus のデータセットを用いて 2 分野計 114 名の研究者を対象に、F 値 80%以上の精度で分類できることを確認した。今後は、対象人数の増加や別分野での検証を進め、“優秀”な人財発見に役立てたい。

Bibliometrics such as the number of papers and times cited are often used to compare researchers based on specific criteria. The criteria, however, are different in each research domain, and are set by empirical laws. Moreover, there are arguments such that the simple sum of metric values works to the advantage of elders. Therefore, this paper attempts to constitute features from time series data of bibliometrics, and then classify the researchers according to the features. In detail, time series patterns, which correspond to knowledge of bibliometrics, are extracted from the large amount of bibliographic datasets, and then a model to classify whether the researchers are “distinguished” or not is created by machine learning techniques. The experiments achieved an F-measure of 81.0% in the classification of 42 researchers in a research domain based on the datasets of Japan Science and Technology Agency and Elsevier’s Scopus. In the future, we will conduct verification on a number of researchers in several domains, and then make use of discovering “distinguished” researchers, who are not widely known so far.

キーワード: 系列パターン, 機械学習, 科学計量学

Sequential Pattern, Machine Learning, Scientometrics

## 1 はじめに

近年、論文や特許に関するビブリオメトリクス (Bibliometrics, 書誌計量学) に基づく研究動向の分析が盛んである。科学技術振興機構（以下、JST）においても、2015 年 6 月に JST が蓄積・管理する文献、特許、研究者情報、科学技術用語、化学物質等に関する科学技術情報を Linked Data<sup>[1]</sup> として

提供する J-GLOBAL knowledge サイト<sup>1</sup> を公開し<sup>[2, 3]</sup>、論文数や被引用数、特許出願数などから国別、分野別の大学または研究者の分析を進めている。本研究は、上記分析活動の一環としてビブリオメトリクスに基づいて“優秀”な人財を発見すること目的としている。現在、研究者分析における一般的な手法は、論文数や被引用数といった研究業績の量に基準を設定し、基準に合致するかどうかで判断する

<sup>1</sup><http://stirdf.jglobal.jst.go.jp>

というものである。しかし、基準値は分野によって異なる上、多くの場合、経験則（あるいは相場感といったもの）に頼っている。加えて、単純な業績の合計では必然的に年長者が有利になったり、長期に渡って比較的少ない業績を出し続けている者と1つの飛び抜けた業績を持つ者とを区別できないといった問題がある。一方で、“優秀”な研究者の研究業績の変遷にはある種のパターンが存在することが知られている<sup>[4]</sup>。例えば、分野を少しづつ変えながら定期的にヒット（多くの被引用数を持つ論文など）を飛ばしたり、海外で先に認められてから国内で著名になるなどである。そこで、本論では従来の業績量に加えて、“優秀”な研究者の国内外の研究業績の時系列的变化から特徴的なパターンを見つけ出し、研究者を高精度に分類することを試みる。

但し、本論における“優秀”であるという意味は、研究者の能力の程度を指すのではなく、周囲から優秀と認知されているかどうかである。これは本論の目的は研究者評価ではなく、“優秀”と認知されて然るべき人財を見つけ出すことにあるからである。本論の目的、または本分類システムの使い方としては、まずファンドに応募してきた研究者を従来の単純な業績量（論文数など）で一次的にスクリーニングした後に、業績が拮抗し、単純な比較が困難な研究者群に関して、ファンドの審査員の方々に対し、時系列的变化からの知見を参考情報として提供するというものである。あくまでこのシステムで自動的に合否を判断するものではない。また、研究者個人の利用は想定していない。そこで、後述する実験では過去に大型の公的ファンドを受けているかどうかで“優秀”かどうかを判断した。いずれのファンドも各分野の複数の専門家の十分な審議を経て、優れたテーマを提案し、それを実行し得ると判断された研究者に与えられていると考えたからである。無論、ファンド採択の是非は、提案書の書き方や広い意味でのタイミングなどにも影響を受けるだろう。したがって、業績に客観的な差がないにも関わらず、これまで“優秀”と認知されてこなかった研究者を発見できれば、彼らのファンド提案に何らかの問題があるかもしれないことも推測できるだろう。

本論の主たる貢献は、従来の系列パターンマイニングの手法を改良しつつ、ビブリオメトリクスに基いて研究者を分類するという特定の問題に適用し、一定の精度を達成した点にある。

以下、2章では時系列分析のアプローチを提案し、

3章において大規模文献データセットを対象とした実験、評価について示す。4章において関連研究との相違について議論し、最後に5章にてまとめと今後の課題を述べる。

## 2 時系列分類アプローチの提案

本章では、大規模文献データからビブリオメトリクスの特徴的な時系列パターン群を抽出し、機械学習手法によって研究者を“優秀”かそうでないか分類する手法を示す。その後、人工知能とバイオインフォマティクスの分野において一定数の“優秀”な（多額の外部資金獲得実績のある）研究者を選出し、分類を試みる。提案アプローチの流れを図1に示す。

### 2.1 特徴量の生成

(時) 系列データを特徴量として表現する方法は、4章で詳述するように数値として表す方法やアイテム（属性と属性値の組）で表す方法などが存在するが<sup>[5]</sup>、本論では大規模データの圧縮を目的として実数値からなる系列データを文字列に変換する手法を用いた。系列データを文字列化する手法としては、Symbolic Aggregate Approximation (SAX)<sup>[6]</sup> が広く知られているが、我々は SAX をベースに系列データの値ではなく、変化分を文字列化した。その上で、自然言語処理における k-gram (連続する文字列) を取得し、複数の重複する系列データを表現するために時間的前後関係を持つよう緩く結合された k-gram に拡張している点に特徴がある。

図2に特徴量生成処理の流れを示す。まず個人単位で数値データを正規化 [0~100] し、時間的変化を特徴量として表現するため2年間毎の差分を6つのシンボル {U, u, S, d, D, 0} (U: 前年に比べ30ポイント以上上昇, u: 5~30ポイント未満の上昇, S: -5~+5ポイントの変動, d: -5~-30ポイントの下降, D: -30ポイント以上の下降, 0: 業績なし) に変換する。そして、上記シンボルの k-gram ( $k=1\sim 4$ ,  $k=4$  で5年分を表す) を生成し<sup>2</sup>、後述する5つの系列データに関する k-gram を開始時間  $t$  の前後関係 {+, =, -} (+: k-gram の開始時間が先頭の k-gram 以降, =: 開始時間が同年, -: 開始時間が翌年以前) で繋ぎ、特徴量となるパターンとした。

<sup>2</sup>5年分とした理由は、それ以上に長い共通パターンはほぼ見つかなかったためである。

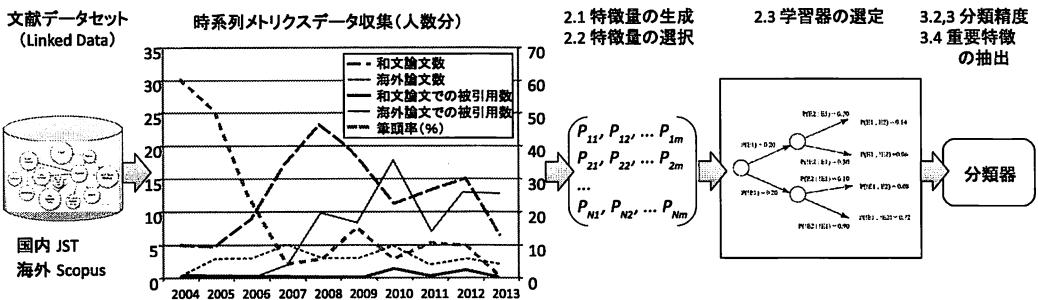


図 1: 提案アプローチの流れ

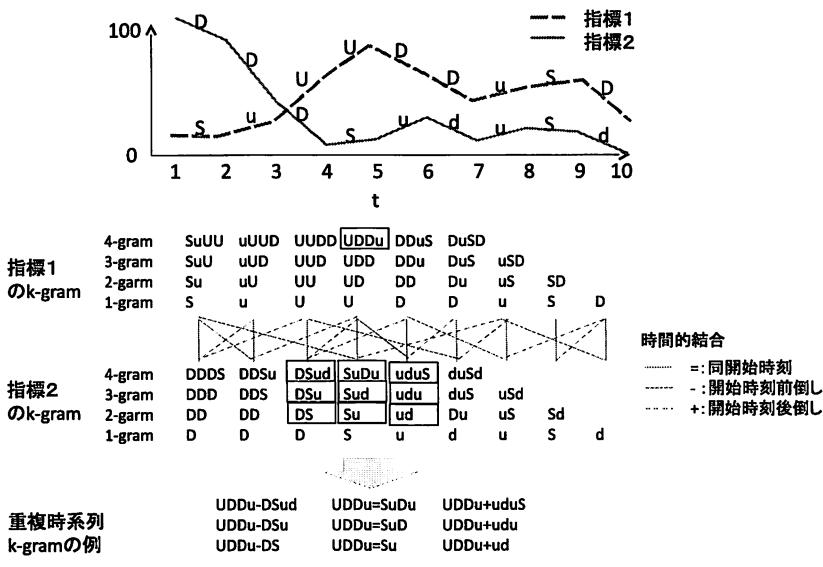


図 2: 特徴量の生成例

尚、本論では各年の研究業績の値ではなく、研究業績の変遷のパターンを特徴量とするため、変化を表す上記の指標を用いた。

## 2.2 特徴量の選択

次に、前節で生成されたパターン群の中から特徴的なパターン（後述する機械学習のプロセスにおいて重要な特徴量となるパターン）を見つけるため、各パターンに付き自然言語処理における TF・IDF (Term Frequency · Inverse Document Frequency) に相当する  $P_{ij}$  (式(1)) を算出する。 $P_{ij}$  は複数の研究者に共通する一般的なパターンをフィルタリングするものであり、多くの研究者に出現するパター

ンは重みが下がり、少数の研究者にしか出現しないパターンは重みが上がる。これは本問題設定が従来の特定カテゴリからの頻出パターンの発見という系列パターンマイニングとは異なり、機械学習プロセスにおいて自カテゴリを特徴付けるパターンを見つけるという問題になっているためである。

$$P_{ij} = paf_{ij} \times \log\left(\frac{N}{pef_j}\right) \quad (1)$$

$paf_{ij} = \# \text{occurrences of pattern } j \text{ in person } i$

$pef_j = \# \text{persons containing pattern } j$

$$N = \# \text{persons in total}$$

本手法において、理論上、可能なパターンの組み合わせは 10 年分で  $P_{max}$  (式(2)) 通りとなる。

$$\begin{aligned}
 P_t &= S^k + S^{k-1} + \dots + S \\
 &\text{if } t < 7, \text{ then } k = 4 \\
 &\text{else } k = 10 - t \\
 P_{max} &= \left( \sum_{t=1}^{10} P_t \right)^M \quad (2)
 \end{aligned}$$

$P_t = \# \text{ patterns of } k \text{ gram at time } t \ (t \leq 10)$

$S = \# \text{ symbols}$

$M = \# \text{ metrics}$

また、一人当たり生成可能なパターン数は高々  $P_{per}$  (式(3)) 通りである。したがって、合計で最大  $P_{per} \times$  人数  $N$  のパターンが生成される。実際、後述する実験では重複パターンやシンボル 0 が続くパターン、各 k-gram がいずれも短い ( $k < 3$ ) パターンなどを排除した後に、おおよそ数億～十数億程度のパターンが取得されている。

$P_{it} = k$

$\text{if } t < 7, \text{ then } k = 4$

$\text{else } k = 10 - t$

$$P_{per} = \left( \sum_{t=1}^{10} P_{it} \right)^M \quad (3)$$

$P_{it} = \# \text{ patterns in person } i \text{ at time } t \ (t \leq 10)$

$M = \# \text{ metrics}$

このことから特微量空間が非常に高次元であり、実験ではパターン合計が最大数であっても全次元数の  $(P_{per} \times N)/P_{max} = 1.23e-9\%$  となることから、スパースな特徵ベクトルが構成されることが分かる。そこで、特徵次元の絞り込みを行うため、クラスタリングを行った。ここでは文字列化した特微量に対して、バイオインフォマティクス分野で DNA 配列同士の類似性を判断するためなどにしばしば用いられるレーベンシュタイン距離 ( $d < 3$ ) を導入した。これにより類似のパターン群をクラスタリングし、1 特微量とした。最後に、1 名毎に  $P_{ij}$  値の高いものから 10 パターンを選択した。したがって、50 名であれば最大 500 パターンとなるが、重複パターンは削除した。

### 2.3 学習器の選定

今回、学習器として決定木を用いた。これは生成された木の情報ゲインを取得することで重要な特

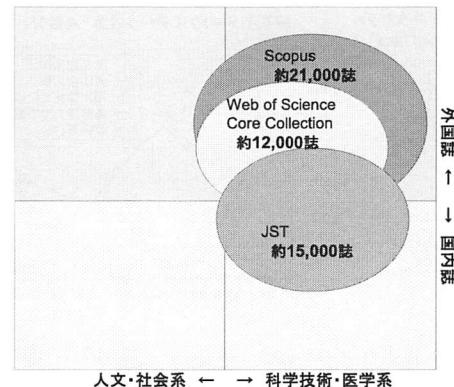


図 3: 各データセットのカバレッジ

表 1: データセットの比較

	Web of Science	Scopus	JST文献
収録範囲	海外誌 約12,000誌	約21,000誌	5,011誌
	国内誌 252誌	417誌	9,514誌
	発行年 1900～	1996～	1975～
	総記事件数 約4,500万件 (SCIEのみ)	約5,200万件	約3,600万件
	年間記事件数 約230万件	約250万件	約150万件

徴量を容易に判別できるためである。尚、学習用ライブラリとしては Java による Weka3.6.13<sup>3</sup> を使用し、決定木のアルゴリズムはよく知られた C4.5<sup>[7]</sup> をベースとして枝刈り等の改良を加えた J48 アルゴリズム<sup>[8]</sup> を用いた。枝刈りの閾値は 0.25、リーフノードの最小インスタンス数は 2、ツリーのサイズには制限を設けていない。但し、特微量が非常に高次元であるため、前処理として最良優先探索に基づく特微量選択（ラッパー法）を適用した。手法の詳細は参考文献<sup>[8]</sup> を参照してほしい。

1 名当たりの学習データは前節で作成されたパターンに関する各人の  $P_{ij}$  値を並べたものである。該当するパターンを持たない場合は値 0 とした。

### 3 研究者分類実験

本章では、上述した特微量を用いて研究者群を分類し、その分類精度を示す。

<sup>3</sup><http://www.cs.waikato.ac.nz/ml/weka>

### 3.1 実験データの概要

今回対象とする JST および Scopus 文献データセットの概要を図 3, 表 1 に示す。JST 文献データセットは 1975 年以降の主に国内の科学技術文献を網羅しており、Scopus 文献データセットは 1996 年以降の主に海外の科学技術文献を網羅している。Scopus および Web of Science に関しては、下記サイト<sup>4</sup>、および JST 調べによるものである。尚、Web of Science については比較のための参考情報である。

これら 2 つのデータセットを併せて重複を除いたものから、以下の 5 種類のメトリクスに関するデータを年単位で取得した。尚、海外論文とは和文論文以外を指している。主に英語論文である。また、筆頭率とは和文論文と海外論文を区別せず合算した上で、その内の対象者が筆頭である論文の割合を求めたものである<sup>5</sup>。

- 和文論文数（学術論文誌論文および会議論文）
- 海外論文数
- 和文論文での被引用数
- 海外論文での被引用数
- 筆頭率

分類対象者は人工知能分野（AI）の日本人研究者 42 名、およびバイオサイエンス分野（Bio）の日本人研究者 72 名とし、J-GLOBAL knowledge で人工知能学会<sup>6</sup>、および日本再生医療学会<sup>7</sup> 所属の研究者を無作為に抽出した。また、科学研究費助成事業データベース<sup>8</sup> を参照し、3000 万円以上のファンドを研究代表者として取得した者を“優秀”（分類上、TRUE）としてラベル付けした。例えば、日本学術振興会の科学研究費助成事業（科研費）では基盤研究 A 以上、あるいは大型の特定領域研究、JST の CREST や ERATO などが相当する。その上で、時系列データはファンド取得前年までの 10 年分とした。これはファンドを得たことによる業績増加が無視できないためである。尚、ファンド未取得者（分類上、FALSE）は直近（2014 年）までの 10 年分とした。前述したように本分類システムは、ファンド応募者を単純な業績量などを用いて一次スクリーニ

ングした後に適用することを前提としているため、実際の利用において集合は限定され、“優秀”な方とそうでない方の比率はほぼ等しくなることを想定している。そこで、チャンスレベルとの比較を容易にするため、ファンド取得者（TRUE）と未取得者（FALSE）が丁度、同数となるように設定した。これらの対象者に対して、重複系列 k-gram による特徴的なパターンを抽出し、特徴ベクトルとして使用した。

### 3.2 業績の時系列パターンに基づく研究者分類

本節では、10 交差検定による研究者の分類（TRUE、FALSE）精度を示す。尚、機械学習プロセスには 48VCPU、98GB メモリーを用いて、各データセット毎に約 1 日を要した。

表 2: 業績の時系列パターンによる分類精度

分野	分類	適合率 (%)	再現率 (%)	F 値 (%)
AI	TRUE	85.7	28.6	42.9
	FALSE	57.1	95.2	71.4
Bio	TRUE	72.7	66.7	69.6
	FALSE	69.2	75.0	72.0

表 2 より、総じて TRUE（“優秀”な方）は適合率が、FALSE は再現率が相対的に高く、それぞれ逆は低いことが読み取れる。これは、TRUE であることの特徴は明確であり、確実に TRUE であると判別されるものの、それ以外の 0（該当パターンなし）が続く多くの方が FALSE として判別されたことによるものである。尚、メトリクスのスパース率（その年の業績が 0 の割合）は全体の 21.4% であった。但し、AI 分野における TRUE の再現率（および結果として F 値）を除けば、いずれもチャンスレベルよりは高いことを確認できる。

### 3.3 業績の時系列パターンおよび量に基づく研究者分類

1 章で述べたように本論は研究業績の量的な分析に加えて時系列的変化の分析を提案していることから、3.1 節で挙げた 5 メトリクスに関する 10 年分のデータの単純な合算（筆頭率は平均）を用いた分類との比較を実施した。特微量は、時系列変化（k-gram）ではなく業績の合計値である。但し、1 名当

<sup>4</sup>[http://hlwiki.slais.ubc.ca/index.php/Scopus\\_vs\\_Web\\_of\\_Science](http://hlwiki.slais.ubc.ca/index.php/Scopus_vs_Web_of_Science)

<sup>5</sup>研究分野によっては著者順は意味を持たない場合もあることに留意する必要がある。

<sup>6</sup><http://www.ai-gakkai.or.jp/>

<sup>7</sup><https://www.jsrm.jp/>

<sup>8</sup><https://kaken.nii.ac.jp/>

たりの特徴ベクトルが要素数 5 と少ないため、前処理（特徴量選択）は行っていない。表 3 に分類結果を示す。この結果から従来の業績量が“優秀”さと強い相関を持っていることを確認できる。

表 3: 業績量による分類精度

分野	分類	適合率 (%)	再現率 (%)	F 値 (%)
AI	TRUE	80.0	76.2	78.0
	FALSE	77.3	81.0	79.1
Bio	TRUE	64.2	89.5	74.7
	FALSE	82.6	50.0	62.3

そこで、表 2 で用いた時系列パターンからなる特徴ベクトルに、上記の量的な特徴 5 要素を追加して、前処理を掛けた上で決定木で分類した結果を表 4 に示す。

表 4: 業績の時系列パターンと量による分類精度

分野	分類	適合率 (%)	再現率 (%)	F 値 (%)
AI	TRUE	95.0	90.5	92.7
	FALSE	90.9	95.2	93.0
Bio	TRUE	87.1	75.0	80.6
	FALSE	78.0	88.9	83.1

結果として、両者を組み合わせることで時系列パターン単体、量単体よりも高精度に分類できることを確認した。尚、表 3 と表 4 における AI と Bio の正解数を合計し、 $\chi^2$  二乗検定によって 2 つの表の独立性を検定したところ、 $p = 0.014$  という結果を得た。したがって、有意水準を一般的な 5% と設定すると、業績量による分類と時系列パターンおよび量による分類とには有意な差が認められると言える。但し、AI 単体での正解数、Bio 単体での正解数で検定を行うと、それぞれ  $p = 0.061$ 、 $p = 0.084$  となり、5-10% の一般に有意傾向ありとされる範囲に低下した。しかし、これは統計的検定の問題点として指摘されている標本数の少なさに起因しているものと思われる<sup>[9]</sup>。

### 3.4 重要指標の抽出

本節では、重要な特徴（パターンおよび量）として抽出できたものの一部を紹介する。

決定木では特徴量の重要さを情報ゲインによって判断することができる。情報ゲインとは、ある特徴量でデータセットを分割した際のエントロピーの減少具合を意味し、決定木の構築は、情報ゲインが最も高い（最も同質なサブセットに分割する）特徴量

を見つけることに等しい。したがって、情報ゲインは決定木の上流に行くほど高い。表 5 に、3.3 節で得られた重要な特徴を情報ゲインが高い順に 4 つ示す。時系列パターン UU-dUUS-00-0uSS-UDDu がトップであり、次に海外論文数の合計が続いていることが確認できる。条件欄の値は、時系列パターンの場合 (#1,3,4) は  $P_{ij}$  値を、業績量の場合 (#2) は編数を表す。

表 5: 重要な特徴の例

#	特徴	条件
1	UU-dUUS-00-0uSS-UDDu	> 4.8
2	# international papers	> 62
3	Dd-UD-000-00-DU	> 16.2
4	DS-UDDU-00-00-UD	> 3.0

## 4 関連研究

本章では、系列パターンに対する機械学習手法の適用、およびビブリオメトリクスの時系列分析の 2 つ観点から関連研究と比較し、本研究の位置づけを議論する。

近年、ネットワーク環境の整備と各種センサーの普及、およびソーシャルメディアの流行などによって、さまざまな形式のデータが時々刻々と自動生成され、蓄積されるようになった。このような物理的あるいは意味的に開始点と終端点が与えられ、値または属性一値ペア間に半順序関係が成り立つデータは系列データと呼ばれ、新たなマイニング、機械学習手法を必要としている<sup>[10]</sup>。

特に、商品の購買履歴の分析に端を発して 90 年代後半から頻出系列パターン生成が盛んに研究されており<sup>[11]</sup>、代表的な手法に Apriori アルゴリズム<sup>[12]</sup>がある。しかし、本論における問題設定は、特定のカテゴリからの頻出パターンの抽出ではなく、あるカテゴリを特徴付けると思われるパターン群に絞った上で機械学習を用いて分類を行い、分類上、重要なパターンを確率的に見つけるというものである。事前にパターンを絞る理由は後述する計算量とスペースネスの問題からである。この絞り込みに Apriori アルゴリズムを適用することも考えられるが、パターンの重要度は必ずしも頻度ではないことから、本手法では Apriori アルゴリズムにおけるサポート（支持度）ではなく、自然言語処理に着想を

得て自系列内の頻度をサポートの逆数で割るという TF・IDF に相当する指標（式(1)）を用いた。

数値系列データの探索法には、時系列アクティブ探索法 (Time-Series Active Search, TAS)<sup>[13]</sup> や時間伸縮を許容する Dynamic Time Warping (DTW)<sup>[14]</sup> などがある。また、複数の系列データの類似部分を求める方法としては、ウィンドウ幅を決め、部分時系列データを生成し、TAS や DTW を繰り返す方法が挙げられる。しかし、扱うべきデータ量が膨大になるという問題が指摘されている。そこで、本論ではデータ量を削減する目的で実数值を文字列化する手法を用いた。系列データの文字列化には SAX<sup>[6]</sup> がよく用いられるが、我々は時間的変化を符号化するために SAX のように値を変換するのではなく、2年分の変化を1文字として変換している。

また、本論ではこれを重複系列の扱いに拡張している。重複系列とは、同時に複数の異なる値が記録される系列データである。SAX の重複系列への適用（多次元化）にあたっては、重点距離法を用いて次元削減を行ってから SAX を適用する方法<sup>[15]</sup> や主成分分析を用いて次元削減を行ってから SAX を適用する方法<sup>[16]</sup> などが挙げられる。また、次元削減をせずに、多次元空間を1次元空間に写像する空間充填曲線、ヒルベルト曲線を用いて多次元時系列データの情報を維持したまま SAX を適用する方法も提案されている<sup>[17]</sup>。しかし、各系列データから生成される複数のパターン間の時間間隔等の関連度合いは常に一定ではない。そのため、重複系列を符号化するには、各系列パターン間にある程度の誤差を許容するモデルが必要であるとされている<sup>[5]</sup>。そこで、本論では複数の系列パターンを開始時間のシフト情報を用いて緩く結合し、1系列パターンとして構成した。

更に、高次元特微量の選択を実施した。主に画像認識等で問題となる高次元特微量からの有用な特徴の選択方法には、フィルター法とラッパー法が存在する<sup>[18]</sup>。フィルター法は、データの一般的特性に依存して選択する高速な手法である。ラッパー法は特微量の部分集合の評価を繰り返すため、高精度であるが時間が掛かるという問題が指摘されている。そこで、本論ではまずフィルター方法に相当する大量データのクラスタリング手法 Data Squashing<sup>[19]</sup> を適用した。具体的には、Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) の考え方に基づき、互いの距離が閾値以内のデータを部分クラ

スタとして1特微量にまとめた。また、距離としては自然言語処理の手法で統一し、編集距離の1つレーベンシュタイン距離を用いた。実験では、更に最良優先探索に基づくラッパー法を適用している。尚、特微量選択に関しては計算量と精度とのトレードオフが十分に精査できていないため、今後の課題の1つとしたい。

ビブリオメトリクスデータの分析全般に関しては林ら<sup>[20], [21]</sup> が詳しいが、時系列的分析に関する事例は少なく、多くは特定の指標に関する時間的な変化を可視化するに留まっている。例えば、Prathap<sup>[22]</sup> は研究者のビブリオメトリクスの変化を評価するために、単一の学術指標として熱力学における Exergy の導入を提案し、その変化を状態図として表現している。しかし、Leydesdorff<sup>[23]</sup> は、科学の発展は再帰性や大学-企業間等の相互作用を含む、複雑で非線形なシステムであり、従来のビブリオメトリクスに基づく多変量解析が主に静的な側面にばかり焦点を当てていることに対して、より動的な発展に目を向けるべきだと主張している。また、1章で述べたように研究業績の時間的変化にはある種のパターンが存在することは、Bjork らの研究<sup>[4]</sup> でも明らかにされている。Bjork らは 1930-2005 年までのノーベル経済学賞受賞者の被引用数の時系列的変化を類型化し、イノベーション普及の Bass モデルとの整合を示している。そこで、Kajikawa らは学術論文の引用関係をネットワーク化し、トポジカル分析を通じてネットワーク内での重要論文の位置の時間的変化を中心性や媒介性といった指標を用いてデータ化し、将来の被引用回数との相関性を測っている<sup>[24], [25]</sup>。手法は異なるが、本研究と同様の目的を持つものといえるだろう。一方、特定分野に絞って主要トピックの時間的な変化を分析した研究<sup>[26]</sup> もある。ここでは核エネルギー技術に関して、国際原子力機関の文献データベースである国際原子力情報システムにおける書誌情報を分析し、研究動向の変遷を明らかにしている。しかし、これは人手による知見であり、統計的な分析は行われていない。

## 5 まとめと今後の課題

本論では、ビブリオメトリクス分析事例の1つとして、従来の量的な特徴に加えて時系列的特徴に基いて研究者を分類する手法を提案した。そして、

JST と Scopus のデータセットを用いて 2 分野 114 名の研究者の分類を試み，“優秀”な研究者をそれぞれ 92.7%, 80.6% (F 値) で分類できることを確認した。これにより提案手法に一定の有用性があることを確認した。また、分類上、いくつかの重要な特徴量を抽出した。尚、本分類において TRUE に誤分類された研究者の業績を見ていくと (False Positive, FALSE なのに TRUE と判定された群)、業績的には“優秀”と見なされて然るべき人が入っていることが確認できる。これは誤分類の詳細な分析が今後、人財発見に繋がることを示唆していると言えるだろう。

今後の課題としては、対象研究者数の増加が挙げられる。但し、データセットの大規模化は計算時間と結果精度とのトレードオフの問題を避けられない。そこで、時系列パターン分析手法の更なる改良と併せて検討していきたい。また、メトリクスの追加も検討していきたい。候補としては、共著者の広がりや被引用論文の寿命（何年に渡って引用されたか）、特にキャリアの早い時期の研究テーマの広がりや“優秀”な研究者とのコラボレーションの有無などを検討している。特に、テーマの広がりやコラボレーションなどはコントロール可能な変数であるため、その重要性が確認できれば若手研究者の支援に活用することができるだろう。尚、本内容は研究プロジェクトであり、JST が情報事業として行っているものではない。

## 参考文献

- [1] Berners-Lee, T.: “LINKED DATA”. <http://www.w3.org/standards/semanticweb/data>, (2016 年 5 月 24 日参照) .
- [2] Kimura, T.; Kawamura, T.; Watanabe, K.; Matsumoto, N.; Sato, T.; Kushida, T.; Matsumura, K.: “J-GLOBAL knowledge: Japan’s Largest Linked Data for Science and Technology”, *Proceedings of the 14th International Semantic Web Conference (ISWC 2015)*, 2015.
- [3] 木村考宏; 川村隆浩; 渡邊勝太郎; 松本尚也; 佐藤智宣; 櫛田達矢; 松邑勝治: 「J-GLOBAL knowledge: 科学技術情報に関する大規模 Linked Data セットの設計および構築」, 人工知能学会誌, Vol. 31, No. 2, 2016 (to appear).
- [4] Bjork, S.; Offer, A.; Soederberg, G.: “Time series citation data: the Nobel Prize in economics”, *Scientometrics*, Vol. 98, No. 1, pp. 185–196, 2014.
- [5] 阿部秀尚; 津本周作: 「重複系列の発生パターンに関する時系列マイニングとその医療応用」, 人工知能学会誌, Vol. 27, No. 2, pp. 154–161, 2012.
- [6] Patel, P.; Keogh, E.; Lin, J.; Lonardi, S.: “Mining Motifs in Massive Time Series Databases”, *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM)*, pp. 370–377, 2002.
- [7] Quinlan, Ross: “C4.5: Programs for Machine Learning”. Morgan Kaufmann Publishers, 1993.
- [8] Witten, Ian H.; Frank, Eibe: “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (Chapter 8)”. Morgan Kaufmann Publishers, 2000.
- [9] 水本篤: 「サンプルサイズが小さい場合の統計的検定の比較」, 統計数理研究所共同研究リポート, No. 238, pp. 1–14, 2010.
- [10] 阿部秀尚; 津本周作: 「系列パターンマイニングの最近の動向」, 人工知能学会誌, Vol. 27, No. 2, pp. 110–111, 2012.
- [11] Agrawal, R.; Srikant, R.: “Fast algorithms for mining association rules”, *Proceedings of the 20th International Conference of Very Large Data Bases (VLDB)*, pp. 487–499, 1994.
- [12] Agrawal, R.; Srikant, R.: “Mining sequential Patterns”, *Proceedings of the 11th International Conference on Data Engineering*, pp. 3–14, 1995.
- [13] Kashino, K.; Smith, G.; Murase, H.: “Time-series active search for quick retrieval of audio and video”, *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 6, pp. 2993–2996, 1999.

- [14] Keogh, E.; Ratanamahatana, C. A.: "Exact indexing of dynamic time warping", *Knowledge and Information Systems*, Vol. 7, No. 3, pp. 358–386, 2005.
- [15] Li, W.; Keogh, E.; Xiaopeng, Xi.: "SAXually Explicit Images: Finding Unusual Shapes", *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM)*, pp. 18–22, 2006.
- [16] Tanaka, Y.; Iwamoto, K.; Uehara, K.: "Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle", *Machine Learning*, Vol. 5, No. 2–3, pp. 269–300, 2005.
- [17] 大西史花; 渡辺知恵美: 「Universal SAX: 空間充填曲線を利用した SAX の多次元時系列データへの適用」, 日本データベース学会論文誌, Vol. 11, No. 1, pp. 43–48, 2012.
- [18] Kohavi, R.; John, G. H.: "Wrappers for feature subset selection", *Artificial Intelligence*, Vol. 97, No. 1–2, pp. 272–324, 1997.
- [19] DuMouchel, W.; Volinsky, C.; Johnson, T.; Cortes, C.; Pregibon, D.: "Squashing flat files flatter", *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data mining*, pp. 6–15, 1999.
- [20] 調麻佐志; 川島浩誉; 標葉隆馬; 林隆之; 吉永大祐; 山下泰弘: 「ファンディングプログラムの運営に資するエビデンス事例集」, ファンディングプログラムの運営に資する科学計量学プロジェクト資料, 2014.
- [21] 林隆之; 山下泰弘: 「ビブリオメトリクスを用いた大学の研究活動の自己分析」, 情報管理, Vol. 53, No. 12, pp. 665–679, 2011.
- [22] Prathap, G.: "The Energy-Exergy-Entropy (or EEE) sequences in bibliometric assessment", *Scientometrics*, Vol. 87, No. 3, pp. 515–524, 2011.
- [23] Leydesdorff, Loet: "Statistics for the dynamic analysis of scientometric data: the evolution of the sciences in terms of trajectories and regimes", *Scientometrics*, Vol. 96, No. 3, pp. 731–741, 2013.
- [24] Shibata, N.; Kajikawa, Y.; Matsushima, K.: "Topological Analysis of Citation Networks to Discover the Future Core Articles", *Journal of the American Society for Information Science and Technology*, Vol. 6, No. 58, pp. 872–882, 2007.
- [25] Iwami, S.; Mori, J.; Sakata, I.; Kajikawa, Y.: "Detection method of emerging leading papers using time transition", *Scientometrics*, Vol. 101, No. 2, pp. 1515–1533, 2014.
- [26] 桐山恵理子; 梶川裕矢; 藤田桂英; 岩田修一: 「核エネルギー技術の社会科学的研究に関する書誌分析」, 情報知識学会誌, Vol. 22, No. 2, pp. 132–137, 2012.

(2015年11月30日受付)

(2016年5月23日採択)

(2016年6月30日J-Stage早期公開)

研究論文

## FRSAD と Linked Data に基づく主題情報共有システム A Subject Information Sharing System Based on Linked Data and FRSAD Model

田辺浩介<sup>1\*</sup>, 江草由佳<sup>2</sup>, 高久雅生<sup>3</sup>  
Kosuke TANABE<sup>1\*</sup>, Yuka EGUSA<sup>2</sup>, Masao TAKAKU<sup>3</sup>

1 物質・材料研究機構 技術開発・共用部門科学情報プラットフォーム

National Institute for Materials Science

E-mail: TANABE.Kosuke@nims.go.jp

2 国立教育政策研究所 研究企画開発部教育研究情報推進室

National Institute for Educational Policy Research

E-mail: yuka@nier.go.jp

3 筑波大学図書館情報メディア系

University of Tsukuba

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

\*連絡先著者 Corresponding Author

本研究では、FRSAD モデルと Linked Data に基づく主題管理システムの構築モデルを提案する。本システムは、件名標目や分類標目など、異なる図書館で使用されている複数の体系の主題情報を、FRSAD モデルで定義された Thema と Nomen のふたつのエンティティを用いて、RDF として表現することができる。本研究では、教科書分類と日本十進分類法を FRSAD モデルで表現するとともに、それらの主題情報を Linked Data の技術を用いて、FRBR モデルに基づいて作成された書誌情報と共有できることを示す。

We propose a system model to develop a subject management system based on FRSAD model and Linked Data. This system can handle multiple subject schemes by using Thema and Nomen entities defined in FRSAD model. We describe "NIER Textbook Classification" and Nippon Decimal Classification 9th Edition in FRSAD model, and show that those subject information can be connected to bibliographic information based on FRBR model using Linked Data technology.

キーワード: FRBR, FRSAD, Linked Data, 主題管理

Keywords: FRBR, FRSAD, Linked Data, subject management

### 1 はじめに

図書館で扱う資料が Web 上で流通するようになるにつれて、図書館で提供する情報と Web 上の情報を結びつけるための手法として、書誌情報を RDF と Linked Data で表現するための取り組みが進められている。主題典拠についても、

dewey.info<sup>[1]</sup> や Web NDLA<sup>[2]</sup> のように、Linked Data として表現された件名標目表や著者名典拠を公開している例がある。2015 年には日本図書館協会と国立国会図書館の共同研究により、日本十進分類法 (NDC) の Linked Data ファイルが試験公開される<sup>[3]</sup>など、図書館での件名標目表や著者名典拠の Linked Data での公開の動きが

急速に進んでいる。

一方、図書館の目録では以前から、主題アクセスの需要に応えるための手段として、件名標目表や分類表、シソーラスなどの主題典拠が作成されてきた。これらを構成する主題典拠レコードが備えるべき機能要件を定めることを目的として、FRBR(Functional Requirements for Bibliographic Records) を拡張したモデルである FRSAD(Functional Requirements for Subject Authority Data) が開発されている<sup>[4]</sup>。また、FRBR モデルを採用した目録規則として普及の進んでいる RDA (Resource Description and Access)<sup>[5]</sup>においても、FRSAD モデルに基づく主題情報の記述の提案が承認されている<sup>[6][7]</sup>。

しかし、FRSAD のモデルを図書館で利用されている既存の分類法に適用したり、FRSAD のモデルに基づいて複数の体系の分類を関連づけたりする研究は行われている<sup>[8]</sup>ものの、実際に FRSAD に基づく主題情報の作成や共有、また他の図書館やコミュニティによって作成された典拠情報を用いた検索を行うことのできる図書館目録システムの実装例は見当たらない。また、FRSAD モデルによる主題典拠の情報が Linked Data として公開されている事例も見当たらない。

本研究の目的は、専門的な知識を持っている個々の図書館やコミュニティが作成する主題情報の共有を行い、資料検索時の主題情報によるナビゲーションを改善する手法について示すことである。筆者たちはこれまで、FRBR の Work と Expression のエンティティを別システムで扱う、疎結合モデルによる FRBR の実装を提案してきた<sup>[9]</sup>。本研究では、その実装モデルを拡張した、FRBR モデルと FRSAD モデルに基づく主題情報の作成と共有の方法として、WE システム (Work/Expression 管理システム)・SC システム (Thema/Nomen 管理システム)・SPARQL エンジンを組み合わせた「Enju Root モデル」を

提案する。また、実際の書誌データおよび 2 種類の異なる体系の分類を用いて、FRBR および FRSAD モデルによる主題情報共有システムの構築を行えること、ならびにそれらのモデルに基づく RDF と Linked Data を用いて、既存の書誌情報システムの検索やナビゲーションの機能の拡張を行えることを示す。

## 2 FRBR と FRSAD

FRBR と FRSAD の特徴は、書誌情報をエンティティ・リレーションシップモデルで表現する点である。以下にそれぞれのモデルの詳細を示す。

### 2.1 FRBR とは

FRBR は、書誌レコードが備えるべき機能的要件を検討するために開発されたモデルである。FRBR の目標は、多メディア化する資料に対する利用者本位の図書館目録を実現するための、書誌記述の新しい枠組みを構成することである<sup>[10]</sup>。FRBR は書誌レコードを表すために、エンティティ・リレーションシップモデルを採用している。FRBR モデルでは 10 個のエンティティが定義されている。まず、以下の 4 つが資料について記述するためのエンティティである。これらは FRBR において Group 1 として定義されている。

- **Work (著作)**  
ひとつの作品を表す
- **Expression (表現形)**  
ある Work の文字や音声として表現を表す
- **Manifestation (体現形)**  
ある Expression が本や CD、コンピュータ上のファイルのように流通するための形状を表す
- **Item (個別資料)**  
個人や図書館によって所有されているある Manifestation の複製のうちの 1 部や 1 枚を表す

次に、以下の 2 つが人物・団体について記述するためのエンティティである。これらは FRBRにおいて Group 2 として定義されている。

- Person (個人) :  
個人を表す
- CorporateBody (団体) :  
組織や団体を表す

最後に、以下の 4 つが主題について記述するためのエンティティである。これらは FRBRにおいて Group 3 として定義されている。

- Concept (概念) :  
抽象的な概念を表す
- Event (出来事) :  
時代や出来事を表す
- Object (物) :  
物質を表す
- Place (場所) :  
場所を表す

FRBRにおいて、Group3 の各エンティティは、Group1 の Work エンティティに対してのみ関連を作成できるようになっている。Group1 のエンティティと Group3 のエンティティの関連の例を図 1 に示す。四角形は FRBR の各エンティティ、矢印はエンティティ間の関連を表す。図 1 では、「ハリー・ポッターと賢者の石」という Work に対して、「魔法使い」という Concept、「クィディッチワールドカップ」という Event、「ほうき」という Object、「イングランド」という Place が、それぞれ主題として関連づけられている。

このように FRBRにおいても、主題を表現するためのエンティティは Group3(Concept・Event・Object・Place) として定義されていたものの、記述の中心は Group1 のエンティティ(Work・Expression・Manifestation・Item)であり、主題典拠レコードについては FRBR を拡張した

モデルで扱うことが想定されていた。それが具體化されたのが、FRSAR Working Group によって 2010 年に発表された FRSAD<sup>[4]</sup> である。

## 2.2 FRSAD とは

FRSAD (Functional Requirements for Subject Authority Data, 「主題典拠データの機能要件」) は、ユーザが主題検索をより効果的に行えるようにするために、主題典拠データが備えるべき要件として策定された概念モデルである。

FRSAD は、以下の二つのエンティティを用いて主題典拠データを表現する。

- Thema:  
主題の概念を表す
- Nomen:  
主題の名称や表記方法を表す

FRSAD モデルでは、Thema エンティティと Nomen エンティティとの関連を記述することによって、主題典拠レコードを表現する。また、主題典拠レコードと書誌レコードの関連は、Thema エンティティと FRBR の Work エンティティとの関連を記述することで表現する。

図 2 に、Thema と Nomen、さらに Work との関連の例を示す。Thema1 は、日本十進分類法の分類記号 “007” が示す概念を表したものである。この Thema に対して、日本十進分類法の “007” という分類番号を持つ Nomen1.1 が関連づけられている。また、この Nomen1.1 にはさらに日本語の分類標目名「情報科学」を表す Nomen1.2 が関連づけられている。

さらに、Thema1 の下位概念として Thema2 と Thema3 が存在する。Thema2 には、日本十進分類法第 9 版の “007.1” という分類記号を表す Nomen2.1 が関連づけられており、さらに Nomen2.1 に対して「情報理論」という分類標目名を表す Nomen2.2 が関連づけられている。同様に Thema3 には、日本十進分類法第 9 版の “007.2” という分類記号を表す Nomen3.1 が関連づけられており、さらに Nomen3.1 に対

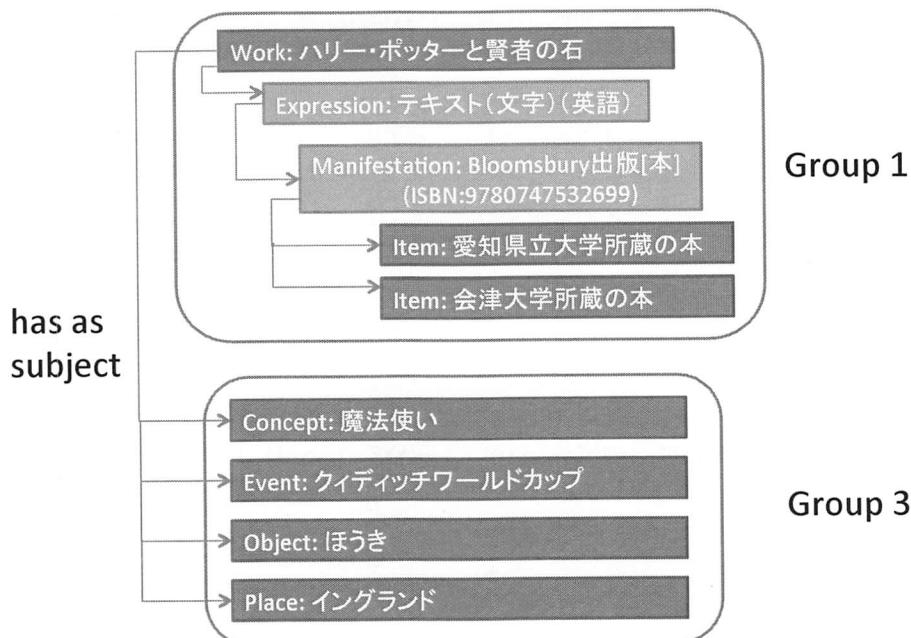


図1 FRBR の Group3 エンティティの関連

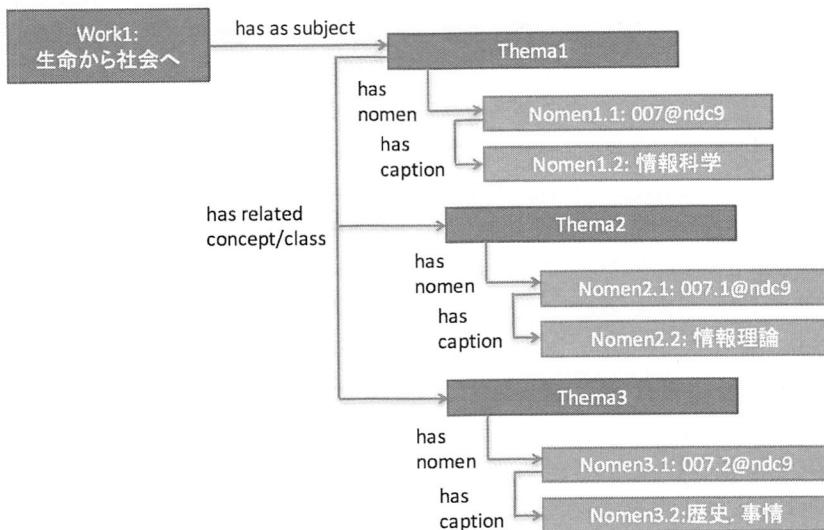


図2 FRSAD による日本十進分類法の表現

して「歴史, 事情」という分類標目名を表す  
Nomen3.2 が関連づけられている。

### 3 対象データ

本研究の目的は、専門的な知識を持っている  
個々の図書館やコミュニティが作成する主題情  
報を他の図書館と共有し、日本十進分類法のよ

うな一般的な主題情報とともに資料の検索に使用することで、主題情報によるナビゲーションを改善する手法について示すことである。このため、対象のデータの要件は、2種類以上の主題情報が付与される書誌情報となる。

本研究では、提案する手法が効果的に働くと考えられるユースケースとして、専門図書館が利用している分類と、多くの図書館が利用している一般的な分類、ならびにそれら両方の主題情報が付与されている書誌情報を対象とした。

### 3.1 教科書の書誌データ

本研究では、教育図書館が明治初期から現在使用されている教科書まで網羅的に収集している約10万冊の教科書コレクション<sup>[11]</sup>のうち、教育図書館がNACSIS-CATに登録している戦前の教科書14488件の書誌データを使用した。

### 3.2 教科書の分類データ

また、分類データとして、教育図書館が維持管理している教科書のための分類「教育図書館教科書分類」を使用した。この分類は複数の図書館の教科書の所蔵目録をまとめた明治以降教科書総合目録<sup>[12]</sup>が刊行された1967年に登場することから、少なくとも48年以上使用され維持され続けている分類である。これは、学校種別、教科書の採用制度、教科書・教授書、教科を元にした特殊分類であり、分類記号の先頭の桁が教育図書館でのコレクション種別（教科書分類は必ずKから始まる）、2桁目から3桁目が採用制度と学校種別、4桁目が教科書・教授書の区分、小数点以下が科目を示している。例えば、K120.1であれば、明治検定教科書の時代の小学校の教科書で、科目が修身であることを表し、K121.1であれば、明治検定教科書の時代の小学校の教授書で、科目が修身であることを表す。また、K130.1であれば、国定教科書の時代の小学校の教科書で、科目が修身であることを表し、K131.1であれば、国定教科書の時代の小学校の教授書で、科目が修身であることを表す。

「教科書分類」は教育図書館によってXMLファイルを用いて維持管理が行われており、2015年5月現在1192件の分類標目がある。教科書分類を用いたFRSADのエンティティの作成例を図3に示す。図中の四角形はFRSADの各エンティティを、矢印はエンティティ間の関連を示している。

### 3.3 日本十進分類法第9版

さらに本研究では、異なる体系の主題典拠を用いたシステムの運用が行えることを示すため、日本十進分類法第9版のデータファイル<sup>[13]</sup>を用いて、FRSADのエンティティを作成した。エンティティの作成例は2.2節で示している。

## 4 Enju Root モデルによるシステムの構成

Enju Root モデルによるシステムの概要を図4に示す。本モデルは、FRBRのWork/Expressionエンティティを管理するWEシステム、FRBRのManifestationエンティティのパーマリンクを提供するMIハブ、Manifestation/Itemエンティティを管理するMIシステム、FRSADのThema/Nomenエンティティを管理するSCシステム、WEシステム・SCシステムから出力されるRDFの検索機能を提供するSPARQLエンジンによって構成される。WEシステム・SCシステム・SPARQLエンジンはそれぞれ独立したシステムであり、パーマリンクとWEシステム・SCシステム独自のWebAPI、ならびにSPARQLクエリを用いて連携を行う。

### 4.1 WEシステム

WEシステムは、既存の書誌情報（Manifestation/Itemエンティティに相当）に対して、Work/Expressionエンティティを管理するシステムである。WEシステムは、システム内のWork/Expressionエンティティに対するパーマリンクを提供する。このパーマリンクと、後述するMIハブから提供されるManifestationのパーマリンクを用いて、WEシステム上のWork

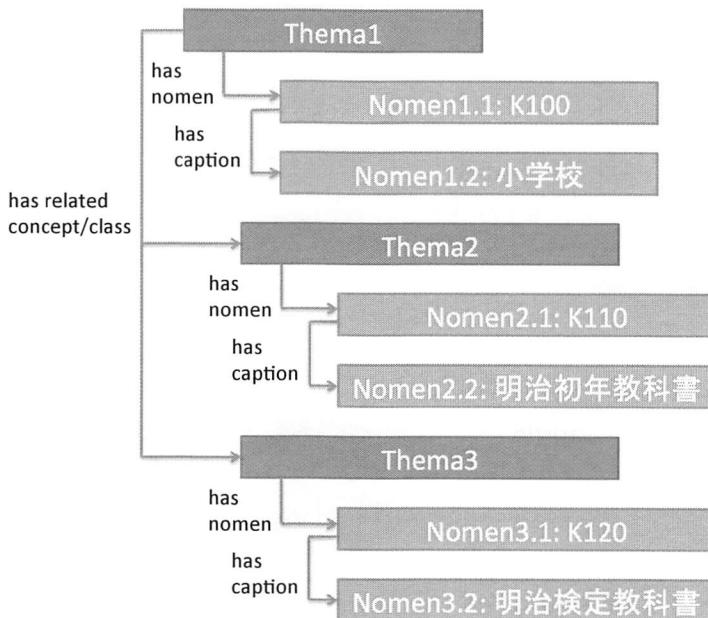


図3 FRSADによる教育図書館教科書分類の表現

や Expression の情報と Manifestation の情報を関連づけることができる。また、WE システム上の Work と、後述する SC システム上の Thema や Nomen の情報を関連づける機能も備える。

WE システムで作成された FRBR・FRSAD モデルに基づく Work・Expression・Manifestation・Thema・Nomen の関連情報は、RDF の Turtle ファイルのダンプとして出力することができる。

#### 4.2 MI ハブ

MI ハブは、Manifestation に対して ID を付与し、他の書誌情報システムに対して Manifestation のパーマリンクを提供するためのシステムである。日本国内では CiNii Books や NDL サーチが相当し、後述する MI システムへのリンクを提供している。

#### 4.3 MI システム

MI システムは、各図書館において、ある Manifestation に対する Item の所蔵情報を管理するためのシステムである。MI システムは MI ハブ

で提供される ID やパーマリンクを用いて、書誌情報と所蔵情報の関連を作成する。MI システムは、現在一般に広く図書館で利用されている蔵書検索システムに相当する。

#### 4.4 SC システム

SC システムは、FRBR の Group3 と FRSAD、つまり主題情報に相当するエンティティの管理を行うためのシステムである。SC システムは、システム内の Thema/Nomen エンティティに対するパーマリンクを提供する。このパーマリンクと、WE システムから提供される Work エンティティのパーマリンクを用いて、SC システム上の Thema の情報と Work の情報を関連づけることができる。本研究では、教科書分類、ならびに日本十進分類法の情報を Thema エンティティ・Nomen エンティティとして SC システムに保存している。SC システムは、Thema・Nomen（各主題情報レコード）のパーマリンクと、Thema・Nomen 間の関連を表す RDF の Turtle ファイルのダンプを提供する。

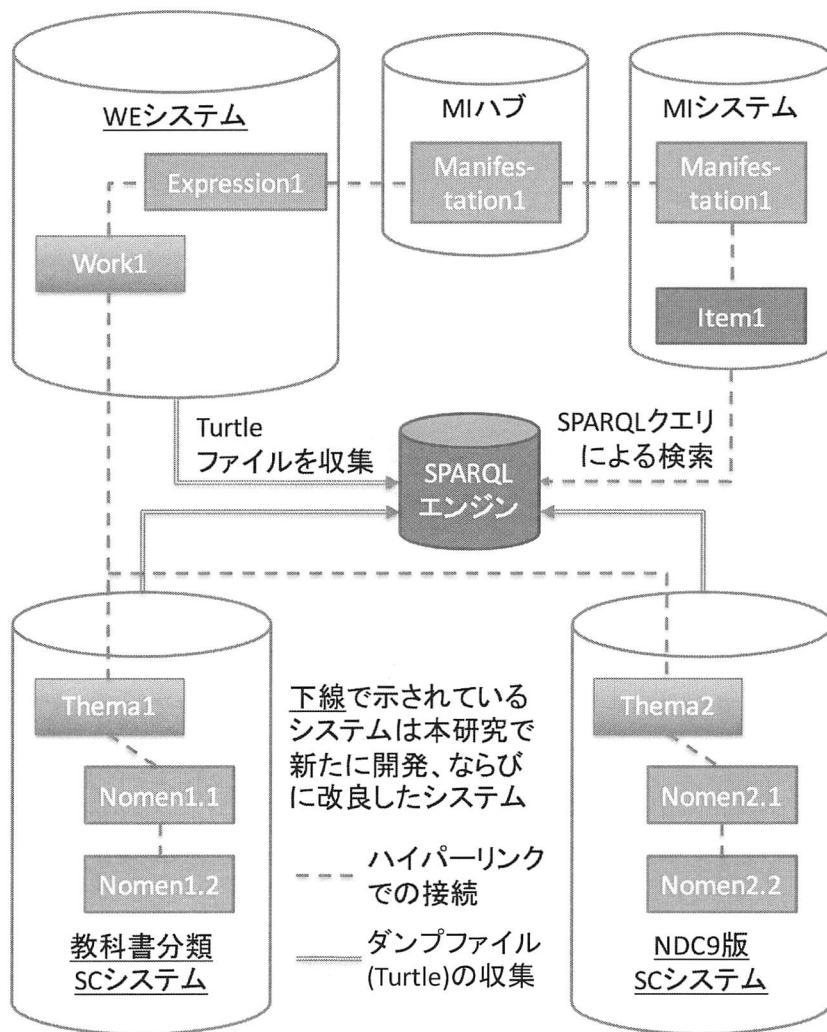


図4 Enju Root モデルによるシステムの構成

#### 4.5 SPARQL エンジン

SPARQL エンジンは、WE システムと SC システムから出力される RDF の Turtle ファイルを読み込んで保存する。また、MI システムなどの他のシステムから SPARQL クエリを受け付け、ある Work に対する Thema を持つ別の Work の一覧の取得など、WE システム上の Work や SC システム上の Thema の関連情報を検索結果として出力する。

#### 4.6 WE システム・SC システムでの Linked Data の実装

先述のとおり、WE システム・SC システムで作成されたエンティティの関連付けの情報は、各システムから RDF の Turtle ファイルのダンプとして出力できるようになっている。出力される RDF の Turtle ファイルの例を図 5 に示す。

本 RDF では、Work・Thema・Nomen の関連を示すボキャブラリに、IFLA(International Federation of Library Associations, 国際図書館連盟)によって定義されたボキャブラリ<sup>[14]</sup>を使用し

```

<http://we-textbook.next-l.jp/works/5014>
  dc:title "東北教科用かなづかい / 松尾捨次郎著" .
<http://we-textbook.next-l.jp/works/5014>
  <http://iflastandards.info/ns/fr/frsad/P2001>
  <http://sc-textbook.next-l.jp/themata/K120.82>.
<http://we-textbook.next-l.jp/works/5014>
  <http://iflastandards.info/ns/fr/frbr/frbrer/P2001>
  <http://we-textbook.next-l.jp/expressions/5014> .

```

図5 WEシステムから出力されるRDF (Turtleファイル) の例

ている。

これらのTurtleファイルをSPARQLエンジンにインポートすることで、WEシステムやSCシステムの主題情報をMIシステム（各図書館の蔵書目録システム）から呼び出し、資料の検索やナビゲーションに用いることができる。

#### 4.7 システムの構成例

Enju Rootモデルによる主題情報共有システムは、4.1節から4.6節で述べた機能を持つソフトウェアやサービスを用いて実装を行うことができる。本研究では、Enju Rootモデルの実装として、以下のソフトウェアやサービスを使用している。

- WEシステム・SCシステム:  
Next-L Enju Root<sup>[15]</sup>
- MIハブ:  
CiNii Books<sup>[16]</sup>
- MIシステム:  
Next-L Enju Leaf<sup>[17]</sup>
- SPARQLエンジン:  
Apache Jena Fuseki<sup>[18]</sup>

### 5 システムの実行例

#### 5.1 使用するシステムとデータ

WEシステム上のWorkに対して、SCシステム上のThemaを関連づけるための実行例を示

す。なお、以下の例では、システム全体は図4のとおり、WEシステム(<http://we-textbook.next-l.jp>)、教科書分類を扱う教科書分類SCシステム(<http://sc-textbook.next-l.jp>)、NDC9版を扱うNDC9版SCシステム(<http://sc-ndc9.next-l.jp>)で構成されるものとする。

また、教育図書館の所蔵する戦前の教科書の書誌データ14488件を使用して、WEシステム上に10212件のWorkエンティティを作成した。Workエンティティの作成には、国立教育政策研究所教育図書館の所蔵資料の請求記号の付与ルールを使用した。具体的には、請求記号の2段目までの文字列から末尾のアルファベットを除いた文字列が同一の場合は、同一の著作であるというルールを使用している。例えば、”K120.8||51||1”<sup>[19]</sup>と”K120.8||51a||1”<sup>[20]</sup>は、”K120.8||51”という識別子をもつ同一の著作とみなすことができる。

次に、教科書分類ファイルを用いて、教科書分類SCシステム上に1175件のThemaエンティティと2350件のNomenエンティティを作成した。さらに、日本十進分類法第9版のデータファイル<sup>[13]</sup>を用いて、NDC9版SCシステム上に、9843件のThemaエンティティと19686件のNomenエンティティを作成した。また、これらのエンティティの情報をもとに、130588件のWork・Expression・Manifestation・Thema・Nomenの関連を示すRDFトリプルを作成した。

## 5.2 WE システムでの分類の登録

図 6 は、WE システム上で Work 「萬國政體論」の詳細画面を開いている状態である。

Work の詳細画面にある「Add Thema」リンクを選ぶと、Thema の URL を入力する画面（図 7）に遷移する。

図 6 WE システムでの Work の詳細画面

図 7 WE システムでの Thema の URL の入力画面。教科書分類 SC システムの Thema の URL を入力している

この画面で、付与したい Thema の SC システム上での URL を入力する。図 7 では、教科書分類 SC システムの Thema の URL として、<http://sc-textbook.next-l.jp/themata/K110.3> を入力している。

「Thema を登録する」ボタンを押すと、教科書

分類 SC システムの Thema の URL が Work に登録される。

## 5.3 WE システムでの分類を用いた検索

図 8 は、WE システムを教科書分類 SC システムの Thema の URL(<http://sc-textbook.next-l.jp/themata/K110.3>) で検索した画面である。このとき、WE システムは SC システムの WebAPI を用いて、Thema とその Thema に関連づけられている Nomen のラベル、さらにその Nomen に関連づけられている Nomen のラベルを取得して表示する。図 8 では、教科書分類 SC システムの Thema(<http://sc-textbook.next-l.jp/themata/K110.3>) から、その Thema に関連づけられている Nomen のラベル”K110.3”と“公民”，ならびに分類体系の名称である”nier”が表示されている。

図 8 WE システムの Work を Thema の URL で検索した結果の一覧画面

## 5.4 WE システムでの分類の表示

図 8 で Thema の URL(<http://sc-textbook.next-l.jp/themata/K110.3>) のリンクをたどると、NDC9 版 SC システムでの Thema の詳細画面に遷移し（図 9），その Thema に関連づけられている Nomen の一覧が表示される。Thema に親 Thema と子 Thema がある場合は、さらにそれ

らに関連づけられている Nomen が表示される。

**Themaの表示**

Node: K110.3  
Nomen:  
• K110.3 (nier)  
◦ 公民  
Parent:  
• K110 K110 政教書  
Child:

図 9 SC システムでの Thema の詳細画面

## 5.5 複数の SC システムからの分類の登録

**New tema**

**Work**  
萬國政體論

**Thema** (e.g. <http://sc.next-i.jp/themata/1>)  
<http://sc-ndc9.next-i.jp/subject/310>

図 10 WE システムでの Thema の URL の入力画面。NDC9 版 SC システムの Thema の URL を入力している

WE システムは複数の SC システムの Thema を登録することができる。登録には Thema の URL を用いる。図 10 は図 6 の Work (萬國政體論) に対して、NDC9 版 SC システムの Thema (Nomen: 310, 政治) の URL(<http://sc-ndc9.next-i.jp/subject/310>) を入力している画面である。

URL の登録が完了すると、WE システム上の Work に NDC9 版 SC システムの Thema が関連づけられ、WE システム上で Work (萬國政體論) の詳細画面の”Subject”に、教科書分類 SC システム上の Thema (Nomen: 310, 政治) と、NDC9 版 SC システム上の Thema (Nomen: K110.3, 公民) の両方が表示される (図 11)。

## 5.6 SC システムの WebAPI

SC システムは、登録されている Thema と Nomen を JSON ファイルに整形して出力し、WebAPI として他のシステムに提供することができる。たとえば、図 10 で追加した Thema

と Nomen は、SC システムから図 12 のような JSON として出力される。

## 5.7 WE システムでの SC システム上の文字列を用いた検索

WE システムは、Work に関連づけられた Thema について、前節で述べた WebAPI から提供される JSON ファイルを取得し、その Thema に関連づけられる Nomen の文字列表現を検索対象に追加することができる。図 13 は、教科書

**[W1476] Workの表示**

タイトル: [W1476] 萬國政體論  
(Creator: カスバル・ホブキンス著 ; 箕作譲祥訳述)  
他のタイトル:  
**Form Of Work:**  
**Date Of Work:**  
**Parent:**  
**Child:**  
**Expression:**

- [E1476] Japanese text
- [M1464] 萬國政體論 Publisher: 中外堂

**Subject:**

- 310
- K110.3

新たに追加された Thema と Nomen

**Other WE system:**  
**relationship graph:**

```

graph TD
    W1476[W1476 萬國政體論] --- E1476[E1476 Japanese text]
    E1476 --- M1464[M1464 萬國政體論]
  
```

図 11 WE システムの詳細表示画面。教科書分類 SC システムと NDC9 版 SC システムの Thema が表示されている

```
{
  "thema": {
    {
      "id": "Thema1",
      "nomen": {
        {
          "term": "310",
          "scheme": "ndc9",
          "label": "政治"
        }
      }
    }
  }
}
```

図 12 連携先の SC システムから出力される Thema エンティティの JSON ファイル

分類 SC システム上の Thema に関連づけられる Nomen である「公民」という語を用いて、WE システムを検索した結果である。先ほど Thema を追加した Work である「萬國政體論」が、検索結果一覧に表示されている。

Listing works			
公民	○ 検索		
24 件ヒットしました。			
タイトル		権限	削除
[W] 國際法－一名萬國公法		権限	削除
[W] 識識者：一名萬國公法		権限	削除
[W] 萬國政體論		権限	削除
[W] 生產進捗内		権限	削除
[W] The compendium of political economy : from the lesson book		権限	削除
[W] 法律概要：幼學讀本		権限	削除
[W] 食育講政治小學		権限	削除
[W] 學者文心論		権限	削除
[W] 葵西國法論		権限	削除
[W] 國體新論：全		権限	削除

図 13 WE システムを検索語「公民」で検索した結果の画面

## 5.8 SPARQL での検索の実行例

Enju Root モデルによって構成されるシステムは、SPARQL エンジンを用いて、Work や Thema の関連情報を取得することができる。図 14 は、MI システム上で、ある資料 (Manifestation) の関連 Work を、WE・SC システムを

用いて表示している画面である。この資料には Thema に相当する情報として、NDC9 版の分類”375.35”が付与されており、以下の手順を用いて関連資料の Manifestation を取得している。

1. NDC9 版の分類番号”375.35”の Thema に教科書 WE システムで関連づけられている Work の一覧を取得（図 15）
2. 1 で取得した Work の一覧に教科書分類 WE システムで関連づけられている教科書分類の Thema の一覧を取得（図 16）
3. 2 で取得した Thema の一覧に教科書 WE システムで関連づけられている Work の一覧を取得（図 17）
4. 3 で取得した Work の一覧に教科書 WE システムで関連づけられている Manifestation の一覧を取得（図 18）

この例では、NDC9 版”375.35”の主題に関連する Thema として、教科書分類”K110.1”を取得し、さらにその Thema に関連づけられた Work をもとに、「対話と共同を育てる道徳教育」の関連資料として Manifestation 「斯邁爾斯(スマイルス) 自助論」を取得している。

これにより、MI システムの利用者は、複数の WE システムや SC システム上にある主題情報を用いた資料のブラウジングを行うことができる。

## 6 考察

### 6.1 本構成の特徴

FRSAD の実装として、本研究で提案する方法のほかに、Work-Thema 間の関連を单一のシステムで集中的に管理する方法がある。この方法は、Work-Thema-Nomen の関連が单一システム内に保存されているため、関連情報の取得を比較的容易かつ高速に行うことができるという利点がある。

一方で、集中管理での実装と運用では、单一システムをどの組織の責任において運用するか

The screenshot shows a detailed view of a document entry in the MI system. At the top, there are search and filter options. Below that, the title '[M] 対話と共同を育てる道徳教育：現代の教育課題と集団づくり' is displayed. Underneath the title, the author '大和久勝, 今関和子 著' and publisher information '著者: 今関, 和子, 1928- 大和久, 勝, 1945- 出版者: クリエイツかもがわ (出版日: 2014-07)' are shown. A '詳細' button is visible. To the right, a thumbnail image of the book cover is displayed, which features a desk and chairs. A sidebar on the right lists several actions: '画像ファイルの新規作成', '書誌情報を電子メールで送信する', '購入リクエスト', '編集', and '所蔵を編集 (新規作成)'.

図 14 MI システムの資料の詳細画面：WE・SC システムから取得した関連情報を用いて MI ハザの書誌情報の一覧を表示する例

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?work ?title
WHERE {
  # S: WE システム上の個々の Work のパーマリンクの URL
  # V: Work と Thema 間の"has as subject"の関連を示すパーマリンクの URL
  # O: SC システム上の個々の Thema のパーマリンクの URL
  ?work <http://iflastandards.info/ns/fr/frsad/P2001>
    <http://sc-ndc9.next-1.jp/subject/375.35>.
  ?work dc:title ?title .
}
```

図 15 NDC9 版 SC システム上の Thema の URL をもとに、教科書 WE システム上の Work の URL とタイトルを取得するための SPARQL クエリ

という問題が存在する。この問題は、単にアプリケーションの運用だけでなく、システムの運用にかかる費用の分担や、システムへのデータ登録資格をどうするかという点も含まれる。

本研究で提案する構成は、Work-Thema-Nomen の作成を各図書館で所有する WE システム・SC システムで行うが、共有を行うのはデータのみであり、システム自体は共有しない。また、WE システムや SC システムは、各図書館の既存の書誌情報システムとは独立して存在する。このため、Work-Thema-Nomen のデータの公開を行いたい各図書館が、自館の必要に応じて、WE・SC システムで書誌情報を公開すること

ができる。また、検索を行うユーザは、検索の対象としている図書館の WE・SC システムのみを選んで、SPARQL と Linked Data による書誌情報の取得を行うことができる。

## 6.2 複数の主題情報の連携

複数の主題典拠を結びつけ、主題間の関連を示すサービスとして、VIAF(Virtual International Authority File)<sup>[21]</sup> がある。VIAF は世界各国の国立図書館が管理する件名・人名典拠レコードを用いて、仮想の巨大な典拠データベースを構築する取り組みである。VIAF の特徴は、世界各国の国立図書館で共通の典拠レコードを用いた单一の典拠データベースを構築するのではなく、各国の管理する典拠レコードを収集し、各

```

SELECT ?thema
WHERE {
  # S: WE システム上の個々の Work のパーマリンクの URL
  # ここでは http://we-textbook.next-1.jp/works/5990 を指定
  # V: Work と Thema 間の"has as subject"の関連を示すパーマリンクの URL
  # O: SC システム上の個々の Thema のパーマリンクの URL
  <http://we-textbook.next-1.jp/works/5990>
  <http://iflastandards.info/ns/fr/frsad/P2001> ?thema .
  FILTER contains(str(?thema), 'http://sc-textbook.next-1.jp') .
}

```

図 16 WE システム上の Work の URL をもとに、教科書分類 SC システム上の Thema を取得するための SPARQL クエリ

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT ?work ?title ?nomen ?nomen_name ?nomen_label ?nomen_name_label
# ここでは http://sc-textbook.next-1.jp/themata/K220.1 を指定
WHERE {
  ?work <http://iflastandards.info/ns/fr/frsad/P2001>
    <http://sc-textbook.next-1.jp/themata/K220.1>.
  ?work dc:title ?title .
  <http://sc-textbook.next-1.jp/themata/K220.1>
    <http://iflastandards.info/ns/fr/frsad/P2003> ?nomen .
  ?nomen skos:related ?nomen_name .
  ?nomen rdfs:label ?nomen_label .
  ?nomen_name rdfs:label ?nomen_name_label .
}

```

図 17 教科書分類 SC システム上の Thema の URL をもとに、教科書 WE システム上の別の Work の URL とタイトル、ならびに教科書分類 SC システムの Thema に紐付けられた Nomen を取得するための SPARQL クエリ

レコードの間に”sameAs”で表現される関連を作成することによって、分散型・疎結合構成の典拠データベースを構築している点にある。その一方で、VIAF は現時点での収録対象が国立図書館の典拠レコードに限られており、専門図書館や大学図書館などが作成したレコードを直接 VIAF で公開することはできない。また、2015 年 4 月以降、VIAF で表現される主題典拠間の関連は sameAs のみとなっており、seeAlso や上位・下位の関連は VIAF には含まれていない<sup>[22]</sup>。分類表においても、dewey.info は分類標目間の上下関係、ならびに 同一の分類標目に対

する複数の言語による表現を提供している。しかし、dewey.info で提供されるのは DDC 内での関連や表現であり、DDC に含まれていない分類標目との関連を表すことはできない。

一方、先述のとおり、SC システムは Thema エンティティの間の関連を、Thema の URL を用いて行う。つまり、本研究において開発した SC システムに限らず、Thema を URL として表現できるシステムであれば、そのシステム上の Thema に対する関連を、SC システムを用いて作成し、SPARQL エンジンを用いて、MI シ

```

PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?expression ?manifestation
WHERE {
  # ここでは http://we-textbook.next-1.jp/works/152 を指定
  <http://we-textbook.next-1.jp/works/152>
  <http://iflastandards.info/ns/fr/frbr/frbrer/P2001> ?expression .
  ?expression <http://iflastandards.info/ns/fr/frbr/frbrer/P2003>
  ?manifestation .
}

```

図 18 教科書 WE システム上の別の Work をもとに、Manifestation を取得するための SPARQL クエリ

システムなど他のシステムから検索することができる。

SC システム上にある独自の Thema と、VIAFなどの外部のシステム上に存在する Thema との関連を作成する際、SC システムは外部の Thema の URL を属性値とする Thema を SC システム内部に作成し、その Thema に対して SC システム上の独自 Thema との関連を作成する。これらの Thema は URL を用いて関連づけられているため、Web ブラウザ上に表示される Web ページや、SPARQL エンジンからの検索結果に含まれるハイパーリンクを用いてたどることができる。図 19 は、dewey.info 上の DDC の分類標目(641, Food and drink)を表す Thema を、NDC9 版 SC システム上の NDC の分類標目(383.8, 飲食物(民俗))を表す Thema と関連づけている例である。

また、それらの Thema 間の関連を用いて、Thema に関連づけられている Work/Expression、さらにそれらに関連づけられている Manifestation/Item へのリンクをたどることができるようになる。たとえば、5 節の実行例により、教科書分類 SC システムと NDC9 版 SC システムは、図 20 のように関連づけられる。

さらに、SC システム上の独自 Thema には、その Thema を表現するラベル(Nomen)と、Thema の対象となる著作(Work)が紐付けられており、

この関連も URL を用いて表現することができる。

なお、明示的に指定された Thema-Thema 間の関連だけでなく、SC システム上で明示されていないが、複数の Work を介して関連づけられる Thema-Thema 間の関連を用いた検索を行うこともできる。たとえば 5.8 項での検索例において、教科書分類の Thema と NDC9 版の Thema には、明示的な関連は指定されていないが、仮想的な関連があるとみなして検索を行うことができる。このような検索ができる点も本方式のメリットである。

## 7 おわりに

本研究では、教科書分類と日本十進分類法の分類体系を、FRSAD で示された Thema と Nomen の 2 つのエンティティを用いて表現することで、異なる体系の主題情報サービスの関連づけと共有を行えることを示した。また、FRBR・FRSAD モデルに基づく書誌情報とそれらの主題情報の関連づけと共有を、WE システム・SC システム・SPARQL エンジンが備える Linked Data の技術によって行えることを示した。

## 謝辞

本研究は JSPS 科研費 26330362, 258030, 268019 の助成を受けたものです。

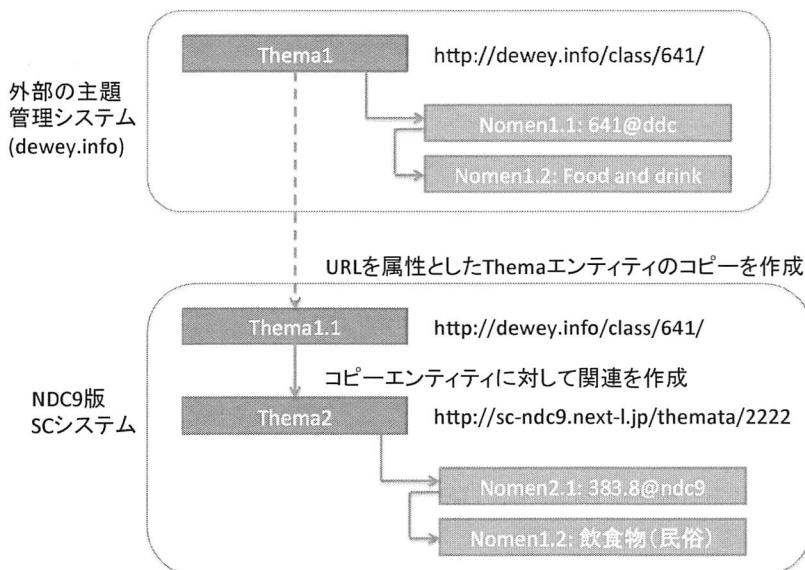


図 19 外部システムの Thema エンティティに対する SC システムでの関連の作成

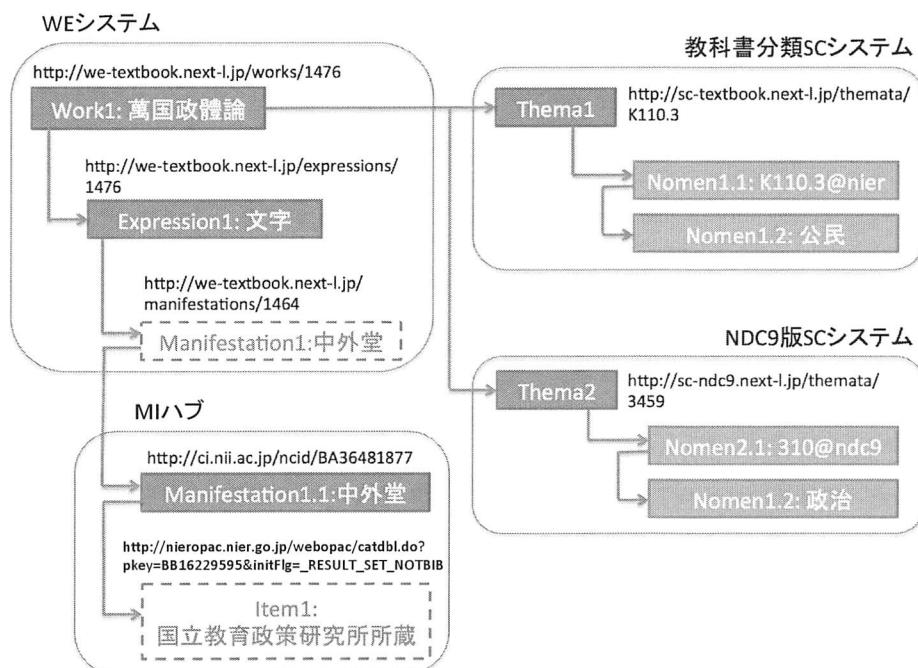


図 20 5 節で実行した Thema/Nomen と Work/Expression/Manifestation/Item の関連

## 参考文献

- [1] “dewey.info”. <http://dewey.info> (2014年12月16日参照).
- [2] “Web NDL Authorities”. <http://id.ndl.go.jp/auth/ndla> (2015年4月17日参照).
- [3] 「日本十進分類法のLinked Data形式化に係る国立国会図書館との共同研究について」. <http://www.jla.or.jp/committees/bunrui/tabid/578/Default.aspx> (2015年12月8日参照).
- [4] IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR): “Functional Requirements for Subject Authority Data (FRSAD) A Conceptual Model”, 2010. <http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf> (2016年4月12日参照).
- [5] Joint Steering Committee for Development of RDA: “RDA: Resource Description and Access”, 2010.
- [6] “Subject Relationship Element in RDA Chapter 23”. <http://www.rda-jsc.org/archivedsite/docs/6JSC-ALA-31-rev.pdf> (2015年12月23日参照).
- [7] “RE: Referential Relationships: RDA Chapters 24-28 and Appendix J”. <http://alcts.ala.org/ccdablog/wp-content/uploads/2015/07/RBMS-2015-1-rev.pdf> (2015年12月23日参照).
- [8] Mitchell, Joan S.; Zeng, Marcia Lei; Zumer, Maja: “Modeling Classification Systems in Multicultural and Multilingual Contexts”, *Cataloging & Classification Quarterly*, Vol. 52, No. 1, pp. 90–101, 2014.
- [9] 田辺浩介; 常川真央; 高久雅生; 江草由佳: 「疎結合構成によるFRBRモデルに基づく書誌情報システム」, 情報知識学会誌, Vol. 24, No. 3, pp. 321–341, 2014. [http://doi.org/10.2964/jzik\\_2014\\_031](http://doi.org/10.2964/jzik_2014_031) (2015年4月17日参照).
- [10] IFLA Study Group on the Functional Requirements for Bibliographic Records: “Functional requirements for bibliographic records : final report”, 1998. [http://archive.ifla.org/VII/s13/frbr/frbr\\_current\\_toc.htm](http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm) (2015年4月18日参照).
- [11] 江草由佳: 「戦前期教科書の電子化・保存とその応用」, 情報知識学会誌, Vol. 17, No. 4, pp. 225–234, 2007. [http://doi.org/10.2964/jzik.17\\_4\\_225](http://doi.org/10.2964/jzik.17_4_225) (2015年4月17日参照).
- [12] 烏居美和子: 「明治以降教科書総合目録」. 教育文献総合目録 / 国立教育研究所編, 第3集 1,2. 小宮山書店, 1967.
- [13] 「電子データの提供（日本図書館協会）」. <http://www.jla.or.jp/publications/publish/tabid/291/Default.aspx> (2015年2月2日参照).
- [14] “FRSAD model”. <http://iflastandards.info/ns/fr/frsad.rdf> (2015年12月8日参照).
- [15] “enju\_root”. [https://github.com/next-1/enju\\_root](https://github.com/next-1/enju_root) (2016年4月12日参照).
- [16] “CiNii Books”. <http://ci.nii.ac.jp/books/> (2016年4月12日参照).

- [17] “enju\_leaf”. [https://github.com/next-l/enju\\_leaf](https://github.com/next-l/enju_leaf) (2015年4月17日参照).
- [18] “Fuseki: serving RDF data over HTTP”. [https://jena.apache.org/documentation/serving\\_data/](https://jena.apache.org/documentation/serving_data/) (2015年12月8日参照) .
- [19] 「帝國讀本」.<http://nieropac.nier.go.jp/webopac/BB16234267> (2016年4月3日参照) .
- [20] 「帝國讀本」.<http://nieropac.nier.go.jp/webopac/BB16234353> (2016年4月3日参照) .
- [21] “VIAF”. <http://viaf.org> (2015年4月5日参照) .
- [22] “April 2015 VIAF Changes”. <http://www.oclc.org/developer/news/2015/april-2015-viaf-changes.en.html> (2015年4月28日参照) .  
(2015年12月23日受付)  
(2016年6月13日採択)

研究論文

## 学術文献における引用分類の観点

### Study of Citation Classification Scheme on Academic Articles

柴田大輔<sup>1\*</sup>, 芳鐘冬樹<sup>1</sup>

Daisuke SHIBATA<sup>1\*</sup>, Fuyuki YOSHIKANE<sup>1</sup>

1 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: daisuke.shibata@gmail.com, fuyuki@slis.tsukuba.ac.jp

\*連絡先著者 Corresponding Author

本研究では、研究評価指標や情報検索技術の改善に利用される学術論文の引用情報について概念的な整理を試みた。第一段階として、引用をタイプ別に分類した様々な先行研究から分類スキーマを抽出し、その観点を(1)意義的観点、(2)評価的観点、(3)機能的観点、(4)形態的観点、(5)位置的観点、(6)社会的観点の6観点に分けて再定義した。第二段階として、第一段階で得られた各観点について尺度水準、推奨される分類スキーマ、活用時の注意点を考察した。従来複合的な観点から構築されていた区分を分離し、異なる観点として位置づけられてきた区分を統合することで、引用分類についての基礎的なスキーマを作成できた。

In this research, we attempt to classify and arraign citations in academic articles. They were variable classification type by each previous research, and were not organized has been a problem. First stage, we reorganized their by perspective, (1) significance type , (2) evaluative type, (3) functional type, (4) morphologic type, (5) location type, (6) sociological type. Second stage, we discussed scaling level, recommended classification scheme, and point to be noted, for perspectives obtained in the first stage. A division that has been built from the complex perspective to separate, by integrating the division, which has been positioned as a different perspective, we create a basic scheme for the citation classification.

キーワード：引用分析、引用の分類、計量書誌学、研究評価

Citation Analysis, Citation Classification, Bibliometrics, Research Evaluation

## 1 はじめに

本研究では、学術文献の引用を対象に、それらを実際に客観的に分類することが可能なスキーマの基盤構築を目指して、分類観点の設定を行う。本研究では、引用のタイプを観点ごとに系統的に整理した枠組みを、引用の分類スキーマと呼ぶ。分類スキーマにおいて、各観点は区分として表現され、個々の引用のタイプは区分肢として定義・列挙される。

学術的なコミュニティにおいて文献を執筆する際、他者の文献や研究を活用したことを明示するため参考文献を記述する。このようにして発生する引用した／されたという関係は、定量的かつ客観的なデータとして扱えることから学術的なコミュニティの動向調査や雑誌のインパクトを測定するためなどに用いられる。近年では、学術雑誌出版社の PLOS が引用元文献（引用をした文献）と被引用文献（引用された文献）間の引用関係をグルーピングすることにより論文同士の関係性を詳細に示すことが可能なデータベースを開発中<sup>1</sup>であったり、効率的な電子出版を目指す Semantic Publishing の動きが起きたある中で引用関係を明示したオントロジー、CiTO（The Citation Typing Ontology）が開発される[1]など、引用をタイプ別に分けることによって情報検索方面への応用に役立つ可能性も見えていく。しかし、このような引用をタイプ別に分ける行為については多くの研究があるにもかかわらず、その目的、手法、分

類スキーマ作成の観点や尺度、判断基準などが極めて多様なまま形成されており、引用という行為を統合的・俯瞰的に捉えることができない状態にある。Small[2]によって、“分類スキーマの作成において累積的な努力がなくそれぞれの研究がユニークな存在になってしまっている”という指摘があるように、分類スキーマの構造や用いられる区分についての検証を俯瞰的に行った研究が少なく、実際に引用を分類して研究へ応用するためには分類やリスト化が行われる際の観点を整理することが必要であると言える。

引用の分類スキーマを活用するためには、主に 2 点の課題がある。1 点目は、分類スキーマにおける区分および区分肢に関する課題であり、2 点目は、分類スキーマそのものではなく、分類作業に関する課題である。社会科学とその関連分野で実証的に引用分類スキーマを用いた Peritz[3]は、複数の米国科学雑誌から論文および引用を抽出し、各区分肢の構成比率を調査した。8 区分肢で構成された分類スキーマは、定義の段階で「歴史的」という区分肢がしばしば「ステージのセット」という区分肢と不可分になりがちであることに触れていた。これは分類スキーマにおける個々の区分肢の定義が相互排他的ではないことを表しており、「ステージのセット」という区分肢の割合が最多（33.4-49.5%）となる傾向が調査対象全誌に共通し「歴史的」という区分肢の割合が非常に少ない（0.3-4.0%）結果であったことを示されても、その結果を信じてよいか疑問が生じる。一方、分類作業とは分類スキーマに個々の引用

<sup>1</sup><http://www.plos.org/rich-citations-open-data-about-the-network-of-research/>

をあてはめる作業、または、個々の引用にラベルを付与する作業を指す。分類作業の面からは、現実的な判断可能性に関する問題がある。例えば、引用元文献の文章を読み、ある引用が分類スキーマのある区分肢に該当するか判断する際、記述部分が極端に短く判断材料が含まれていないければ判定は極めて難しくなる。これでは、分類スキーマにおける各区分肢の定義が明確であっても分類ができないことになる。これら 2 点の課題は分類について検討する際、本質的に避けられず、完全な解決は難しい。しかし、今まで先行研究で用いられてきた分類スキーマの区分や区分肢を、引用という現象の側面から分野や用途を限定せず俯瞰的に統合整理することで、これまで網羅できなかった区分肢の補完や、区分肢の定義を議論しやすくするための基盤を作ることは可能である。

そこで、本研究では、先行研究が提案・利用している様々な引用分類について紹介した後、それらに含まれている要素を分類の観点から俯瞰的に整理・再定義し、各観点について客觀性・実用性のある区分や判断基準を検討する。

## 2 先行研究

### 2.1 代表的な引用分類研究とスキーマ

引用分類に関連する研究は 1960 年代から行われている。Garfield[4]は、「機械的に引用索引を作るためにはそもそも文献の執筆者が適切な引用を明示的に行っていなければ処理できない」と考え、引用が行われる理由を 15 種類列挙した(表 1)。同時代に Liepetz[5]も引用索引の改善を

目的として引用理由を挙げている。The Physical Review 誌、Soviet Physics 誌 (JETP) から 60 文献 750 引用を抽出して考察し、4 グループ 29 区分肢の引用理由を挙げた(表 2)。ただし、これらの論文は考えられる理由を列挙しており、統計的・実証的に分類することを目的としたわけではない。

当初は引用が行われる理由などの行動学的な侧面に焦点があてられ理由のリストアップが中心だったが、次第に作成した引用分類スキーマを用いて実際の文献で行われている引用を分類する実証的な研究が始まり、どのようなタイプの引用が多いのか明らかにされた。

表 1 Garfield[4]が列挙した引用理由

1. 先行研究者への献辞
2. 関連業績に対する敬意(同輩への敬意)
3. 方法論や装置等の確認
4. 背景説明の提示
5. 著者自身の研究の修正
6. 他の研究者の研究の修正
7. 先行研究の批判
8. 主張の立証
9. 引き続く研究のアラート
10. あまり知られていない、引用もされていない研究の紹介
11. データや一群の事実の認証(物理定数など)
12. アイディアや概念が議論された当初の文献を確認
13. もともとの概念や用語が著された文献の確認
14. 他人の著作や考え方に対する否認
15. 他人の先行する主張に反論

表 2 Liepetz[5]における引用分類

グループ 1: オリジナルな科学的貢献または引用文献の意図 観測された現象の記述、データの変換、説明、仮説や理論、理論の計算、予測、定義、実験技術の宣言
グループ 2: オリジナルな科学的貢献以外の引用文献の貢献 レビュー論文、書誌、データの累積
グループ 3: 引用文献が被引用文献との違いや継続関係を示すため 共通の著者、同じテキスト、アブストラクトや要約、エラー訂正、継続、先駆者、包括
グループ 4: 引用文献における被引用文献の科学的貢献の性質 注釈されただけ、区別された、レビューまたは比較された、応用された、改善・修正された、置き換えられた、精度を変えた、適用される範囲を変えた、疑問視した、肯定した、反対した

Murugesan と Moravcsik[6]は物理学の論文における引用を分類するため、二項対立的にとらえられる観点を持った分類スキーマを作成し、実際の文献における引用を対象として分類作業を行った。1968 年から 1972 年の間に Physical Review 誌に掲載された 30 文献 575 引用を対象とし、各引用が各観点のどちらに当たるかを調査し集計した（表 3）。結果からは、確認的な引用が否定的な引用に比べ多いことが明らかとなり、どちらでもない・どちらにも該当するような引用が多くないことから一定の評価を得て、多くの引用分類スキーマで同様の観点が組み込まれた[7, 8, 9, 10]。

Vinkler[11]は、引用を行う動機には、他の研究との関係を持たせる「専門的な動機」と科学コミュニティにおける社会的な関係を構築する「接続的な動機」の 2 種類あると考え、それぞれ複数の選択肢を挙げて化学分野の研究者 20 人を対

象にアンケートを行った。それら 20 人が執筆した 20 文献 484 引用についてどの選択肢が該当するか複数回答可として調査した。調査結果からは、81% の引用は専門的動機のみであり、17% の引用は専門的動機・接続的動機の両方あり、2% の引用が接続的動機のみであることがわかった。つまり、大多数の引用は「論文の序論の完成度を高めるため」や「被引用文献の理論や手法の一部を参考にした」

表 3 Moravcsik と Murugesan[6]の分類スキーマ

観点	比率		
	概念的（概念や理論の引用）か、操作的（ツールや物理的技術の引用）か	操作的	どちらでもない
	53%	43%	7%
有機的（引用文献の理解に必要な引用）か、おざなり（謝意を示すための引用）か	有機的	おざなり	どちらでもない
	60%	41%	1%
進化的（引用文献が被引用文献を基礎にして積み上げている）か、並列的（引用文献が被引用文献の代わりになる）か	進化的	並列的	どちらでもない
	59%	40%	2%
確認的（被引用文献が正しいと確認された）か、否定的（被引用文献の正しさに異議を唱える）か	確認的	否定的	どちらでもない
	87%	14%	5%

N = 575 ※各観点における区分肢ごとの比率を足しても 100%にならないのは、どちらにも該当する引用が存在するため

などの専門的動機により行われつつも、一部の引用には「自己の文献を広めるため」などの接続的動機も含まれていた。それまで行われていた引用分類に関する研究では、専門的動機により生起する引用を中心と想定されていたが、実際には読者やコミュニティに所属する人間とつながることを目的に含んで引用することが明らかになった。

Small[12]は Citation Cube とよぶモデルで引用行動について整理した。Citation Cube では、引用しているコミュニティにおける賛成の度合いを表す次元である“総意度”，引用した文脈と被引用文献との適合度を表す“固執度”，自己引用の度合いを表す“公平度”と名付けた 3 つの次元を直交させることにより 8 区分で引用を分ける。それぞれの次元には高低が存在し、総意度と固執度の 2 次元を交差させた場合は表 4 のようになる。表 4 に記載されていない公平度は、上記 2 次元から作られる 4 象限に対して自己引用か否かという点で分割するための次元となっている。この Citation Cube によるモデルは引用という行為について 3 次元から判別できるが、あくまで引用行動についての概念的な整理方法であり実証的な分類方法でないためか、活用はされていない。

Harwood[13]は社会学分野および計算機科学分野の研究者それぞれ 6 人ずつを対象に、引用の機能を調査した。従来の研究とは異なり、あらかじめ引用理由や分類スキーマを作成してその比率などを調査する手法ではなく、インタビューに基づく談話分析を行い、その結果として引用のタイプを 11 種類ほど見出した（表

5）。さらに、インタビュー結果における各区分肢の出現頻度を調査したところ、「立場」「補足」「信頼」といった機能についての言及は社会学分野と計算機科学分野共通の傾向として見られたのに対し、社会学者では「約束」が、計算機科学では「道標」がそれぞれ多く見られた。このことから、分野によって引用を行う目的が慣習的に異なることも示唆された。

表 4 Small[12]による Citation Cube

		総意度	
		低	高
固執度	低	【構成的】 誤帰属、再解釈、特異な否定、革命的	【構成的】 慣習的否定、共通の誤帰属、因習的な転換、結合による消滅
	高	【規範的】 本質的、有機的、創造的なつながり、普通ではない引用	【規範的】 おざなり、儀式的、共通の直接引用、従来の解釈、模範的

## 2.2 引用分類研究の整理

引用を分類する研究には歴史と積み重ねがあり、すでに先行研究を整理した文献が存在する。

Bornmann と Daniel[14]は、引用回数を数えることの意義について考察するため、1960 年代初頭から 2005 年中盤までに出版された約 30 編の引用行動に関する研究結果をまとめ、その過程で引用を分類する実証的な研究について言及した。その際、(1)文脈分析や内容分析、(2)引用行動のトピックについての郵便調査や対面インタビューといった調査手法ごとに整理した。(1)は文献間の関係を発見するために用いられる。その中でも文脈分析

表 5 Harwood[13]による引用の理由

道標	詳しくない読者への補助、興味を引く 議論の足跡を保持 場所の節約
補助	トピック、手法、主張の正当化
信頼	著者の債務(著者のオリジナリティへ敬意を払う) 信頼と自己防衛 信頼と評価的形容詞
立場	立場の手本 立場や結果の詳細な説明 時間をかけた立場のトレース
約束	称賛した後情報源の問題を識別 情報源の立場における矛盾の識別
積み上げ	自己引用による継続的な積み上げ
連結	他の手法、学派、議論と抱きあわせる
報知	自らの著作を読者へ知らせる 他者の著作を知らせる
未来	著者が描く未来の地図
能力	文献の知識を示す 今後の研究を実現する可能性を示す
話題性	研究が新しい課題であることを示す

は引用元文献の引用周辺テキストに基づいて分析を行う手法であり、内容分析は引用元文献の意味内容から被引用文献を特徴づける手法とした。作業者によって判断が異なる場合もあり、分野による引用行動の差や文献タイプによる差が生じる可能性もあり、全文を読んで判断できるエキスパートが必要になることが弱点でもある。ただし、研究対象の量を増やしやすく統計的な分析を行いやすい。対照的に(2)の手法は引用の動機や目的を調査する研究であり、研究者を対象に科学行動学的な側面から行う分析手法である。他者が推測するより引用の理由が正確とも考えられるが、著者への質問は真の理由を明らかにできないこともあり(正直に回答するとは限らない)、著者本

人も理由を忘れてしまう、共著者が執筆した部分である等)、全面的に正確であるとは言えないとした。Bornmann と Daniel の論文は引用回数を数える意味について迫るものではあるが、引用を分類する場合に見られる観点やその性質について明らかにするところまでは踏み込んでいない。

Ying ら[15]は、従来の引用分析を超えた“内容に基づく引用分析(Content-based citation analysis: CCA)”を推進するため、(1)理論的基礎、(2)アプローチ、(3)アプリケーションの各段階から、引用を分類する研究について整理した。(1)理論的基礎としては、①構造的な分析方法と②意味的な分析方法の 2 つを挙げた。①は introduction, related works, methods, results, discussion, conclusions といったセクション構造を活用する方法であり、②は引用によって関係づけられる文献間の意味的な分析を行う方法である。(2)アプローチについては、①手作業による内容分析と②自然言語処理による半自動的な分析があるとし、複数の研究[3, 5, 16, 17, 18, 19, 20, 21]で用いられた分類スキーマの区分肢について、Background information, Theoretical framework, Prior empirical/experimental evidence, Negative distinction, Methodology explanation の 5 区分にグループ化し直している。(3)アプリケーション、すなわち研究の適用としては①引用動機の分類のため、②自動文章要約のため、③検索のため、④引用の推薦および予測、⑤ナレッジグラフマイニングの 5 つを挙げた。Ying らの研究では引用分類研究が行われる背景やその方向性について整理されたが、個々の研究

で行われた分類スキーマの妥当性については言及されておらず、区分肢の整理が不十分であると言える。そのため引用分類研究を概観することはできるものの、引用分類自体を概観できるとは言えない。

### 2.3 引用分類スキーマの統合

先行研究で用いられている区分肢を統合して分類スキーマを作成する研究も存在する。

Brooks はそれまでの研究で用いられている区分肢のうちいくつかを用いて引用動機の分類スキーマを作成し [7]、それらの関連性を調査した[22]。様々な分野から選定した 20 人の著者に対し、各自の著作内の引用について、あらかじめ用意した 7 区分肢のうちいずれに当てはまるか回答を求めた。7 区分肢からなるスキーマは、(1)Line と Sandison[23]が表現した”新しさ”，(2)Garfield[4]が表現した”ネガティブな信用”，(3)Murugesan と Moravcsik [6]で用いられた”操作的な情報”，(4)Gilbert[24]が引用とは読者を説得するために行っていると提唱した”説得性”，(5)Garfield[4]が定義した”ポジティブな信用”，同じく(6)Garfield[4]が定義した”読者へのアラート”，(7) Bavelas[25]が研究分野で慣習的に行われる不特定かつ曖昧な動機で行われる引用を称した”社会的総意”である。1 つの引用に対し複数の区分肢を付与することも可能として分析した結果、29.3% の引用は 1 つの理由、70.7% は複数の理由が付与された。また、これらの区分肢に対して主成分分析を行った結果、(1) (4) (5) (7)からなる”新しく・重要で・必要と判断される引用のグループ”と、(2)だけのネガティブな引用の

グループ、(3) (6)からなり読者への背景説明の意味を込めた”サービス的な動機のグループ”の 3 グループに分けられるとした。しかし、研究対象とした分野が多岐にわたりサンプル数があまり多くない。また、大多数の引用に複数のカテゴリが付与されることから、分類スキーマとしては整理されていないと言える。

Garzone と Mercer[8]は Citation Index に引用の機能的な区分を自動的に付与するための手法を提案した。分類スキーマは複数の研究[3, 6, 20, 26, 27, 28, 29, 30]に基づいて Garzone[31]が作成した 10 カテゴリ 35 区分肢で構成されたスキーマを用い、Garzone が物理学 8 論文と生化学 6 論文を分析して開発した構文解析ルールに基づいて自動分類を行った。10 カテゴリとは、(1)否定的タイプ、(2)肯定的タイプ、(3)仮説的タイプ、(4)暫定的タイプ、(5)手法的タイプ、(6)解釈的／開発的タイプ、(7)将来的研究タイプ、(8)概念的物質使用タイプ、(9)比較タイプ、(10)読者へのアラートである。物理学 3 論文、生化学 3 論文を実験対象とし、人手によって付与されたカテゴリと自動分類によって付与されたカテゴリを比較したところ、物理学は 41% が正しく、21% が部分的に正しく、38% が間違いであり、生化学は 61% が正しく、12% が部分的に正しく、27% が間違いであるという結果になった。この結果は引用の方法や記述形式が分野によって異なる可能性とともに機械的な判定の難しさを示している。また、例えば重複するとは考えにくい「否定的タイプ」と「肯定的タイプ」が並列的に扱われているなど、Moravcsik と Murugesan[6]の研究で用いられた「軸」

という概念は見られない。

以上のように引用を分類する研究においては、引用分類研究の手法と引用分類スキーマの区分肢についてのグループ化がなされているものの、分類スキーマの区分肢を観点から整理し直し、妥当性を考察した研究は見当たらない。そこで本研究では、引用を分類する際に使用される観点を整理・再定義したのち、個々の観点の特性について述べる。

### 3 観点の再定義

本研究の目的は引用を分類する観点について俯瞰的に整理・再定義することにある。そのため研究目的（研究者の引用行動を明らかにする、評価指標への応用、引用索引や検索精度の改善等）、調査手法（内容分析、アンケート等）および分析対象分野を限定せずに、学術文献の引用をタイプ別に分けた先行研究・先行事例を収集し調査対象とした（付表1）。これらで用いられた分類スキーマ中の区分肢や、質問紙中の質問項目等について、同じ観点と考えるのが妥当なものを同一観点として集約した。以下ではその結果として得られた観点について区分肢等の実例を記載し、観点の範囲を再定義している。引用を分類する場合には「時間」や「分野」により分けた分析が行われることがあるが、それらは周知の事実であり再整理の必要がないため本研究の範囲には含めなかつた。

#### 3.1 意義的観点

引用にどれほどの意味があるか（有意性）から分類する観点を意義的観点とよぶ。有意性は「周辺的・中程度 - 重要

- 不可欠」[10]といった段階的な区分や「有機的 vs おざなり」[6]や「中心的 vs 周辺的」[21]といった2項対立的な区分肢だけでなく、「被引用文献が中心的な扱いをされている」[16]、「レファレンスを増やす必要がある（実際は必要ない）」[11]などのように質問項目が設定されることもある。それぞれ表現に異なりはあるものの引用の必要性・有意性という概念で定義することができる。引用元文献により決定される被引用文献を引用する意義により重みをつけることで、単純な被引用数よりも改善した研究評価方法が可能となる。

#### 3.2 評価的観点

被引用文献に対する引用元文献の評価から分類する観点を評価的観点とよぶ。著者が被引用文献へ行う評価には賛成・肯定といった「ポジティブな評価」や否定・批判といった「ネガティブな評価」のように「賛否」という分け方も存在する。「確認的 vs 否定的」[6]という対立的な見方だけでなく、「弱点を指摘した引用、アプローチの取り方や問題の所在を肯定的にとらえた引用、中立的な記述」[32]などのように肯定や否定の対象とセットで区分肢として扱われる場合や、「部分否定、完全否定」[16]のように強弱が意識される場合もある。被引用数によって被引用文献を評価する際にネガティブな引用を除外する措置を行うことで、被引用数による単純なインパクトではなく著者の評価的スタンスを含んだ評価指標を作成することが可能になる。引用元文献は被引用文献に対して、「研究動機を語る上で必要ではあるが、賛成はして

いない」ということもあり、必ずしも有意性と評価は一致しない。そのため主な用途として研究評価指標の改善という点で類似しているものの概念的な分類の観点としては分けて取り扱うべきであると言える。

### 3.3 機能的観点

引用の機能や役割から分類する観点を機能的観点と呼ぶ。引用行動や文章の意図を調査する場合、この観点から分類されることが多い。ある被引用文献が引用元文献中で引用される場合、その目的は一様ではない。同一の文献であっても研究方法を参照するために引用される場合もあれば、自らの研究結果を考察する際の比較対照として引用される場合もある。そのため著者が引用した用途や目的の情報を引用索引に付加できれば、引用した／されたという関係以上の情報を引用索引の利用者へ提供できることになる。

機能的な観点に該当する区分肢は先行研究でも様々な表現で用いられている。被引用文献の理論や概念を引用したか、あるいは実験手法や操作方法を引用したか[6, 11, 19, 22]という捉え方は、「引用対象の具体性」から判断した分類となる。

先行研究からの継続性から分類することも可能である。「継続、包括、改善、修正された、応用された、置き換えられた、精度を変えた、適用される範囲を変えた」[5]などの区分肢や「被引用情報源がリサーチクエスチョンの出発点になる、引用文献の結果で被引用情報源のデータの新解釈や説明を追加」[18]という区分肢は先行研究の継承や適用範囲拡大といった積み上げ方を表しているともいえる。「未

来への言及」[4, 13]も将来的な発展・継続の可能性を示す区分と捉えることができるため機能的観点における継続性を表す区分として扱うことができる。ライデン大学の Centre for Science and Technology Studies (CWTS)が開発している CitNet Explore<sup>2</sup>のように、引用関係から研究者の系譜をわかりやすく整理することによりコミュニティの全体像をつかむ際により精度が上げる効果が期待できる。

「具体性」や「継続性」のように一つの軸を形成しているわけではないが、機能面から引用を分類する場合には、被引用文献によって「背景説明」を行うため[4, 33, 34]、自研究において検証をおこなう「仮説」として位置づけるため[3, 5, 8]、先行研究における検証方法との同一性を宣言したり定義を参照することにより研究トピック・採用手法・自著の主張の説得性・正当性を強化するため[4, 5, 8, 13, 18, 32, 33, 34, 35, 36, 37]、先行研究の方法や目的、結論などを对比・比較するため[18, 32, 33]、先行研究や関連研究の認知度を上げ被引用文献を読者へ広めるため[4, 7, 11]という目的も機能的な観点に含まれる区分である。

### 3.4 形態的観点

被引用対象の形態的な側面から分類する観点を形態的観点と呼ぶ。被引用対象はデータの場合もあるため必ずしも文献とはならないが、例えばレターなどの速報性が高い文献がよく引用される分野を見出し分野内の引用行動を把握すること

<sup>2</sup> <http://www.citnetexplorer.nl/Home>

で、より適切な図書館のコレクション管理を行うことが可能になる。原著論文、レビュー論文などの文献タイプから分類される[5]が、被引用対象の入手先といった所在情報による分類もコレクション管理には役立つ観点となる。近年では学術情報の電子化によってトムソン・ロイター社がデータ引用に基づくツールである **Data Citation Index** を公開するなど被引用対象の捉え方は旧来の「文献タイプ」という分類だけではおさまらないと言える。そのため、従来の引用分類では行われてこなかった「被引用データ」も含めるべきである。

ジャーナルインパクトファクターを算出する場合は原著論文とレビュー論文のみが集計対象となるなど、文献タイプの区分は研究評価指標へ応用する際にも不可欠な区分となっているため、評価的観点を生かすためにも有用な観点となる。

### 3.5 位置的観点

引用が生起している箇所の位置に基づいて分類する観点を位置的観点と呼ぶ。引用分析を行う際には全ての引用を等しく扱うのではなく文献のどこで引用されたかを含めるべきであるとする考え方に対応する観点である。学術的な文献を記述する際には、「はじめに」や「方法」のように文の内容ごとに見出しが設けられ、文の意味内容により章や節として区切られている。異なる章や節で行われる引用は、その理由や目的、機能が異なると考えられたために、分類することが提案された[38]。機能的観点で分類する場合には本文の記述内容を基に判断するが、位置的観点で分類する場合はその必要がなく、提唱された

当初は機能的観点の代替としての側面が強かった。3.2で示した機能的観点の中でも「背景説明」などは論文のはじめの部分に多く出現する可能性が高いなど、文章中の出現位置から引用の機能を推測可能な場合はある。しかしながらオープンなフルテキストが使用しやすくなっていることから位置的観点で機能的観点を代用する必要は少くなり、明確に両者を区別すべきであるとの立場から本研究では異なる観点として扱う。

引用の位置は引用元文献の「構造的な位置」から捉えられる。例えば情報行動的な側面として機能的観点と組み合わせて引用の分布を調べる場合には、「序論 - 方法 - 考察 - 結論」[38]や **IMRaD** 構造のように本文中のセクションから分類する場合や、文単位 [39]、単語単位 [40]、文字単位[10]で引用元文献の長さを表し引用の出現位置を定義する場合がある。また、引用された位置の近さから被引用文献同士の類似度評価に用いる場合は、「章」単位に加え「節、段落、文、列挙（列挙とは文末などで複数の引用を並べる形を指す）」[41, 42]単位で位置を捉える場合もある。

### 3.6 社会的観点

著者が引用を行う社会的な背景に基づいて分類する観点を社会的観点と呼ぶ。科学社会学的、行動科学的な側面から情報行動や引用行動などの特徴を把握する際に有用な観点である。ここで述べる社会的とは、研究コミュニティとの関わりに起因する引用を指し、「コミュニティの人的つながりによる引用、コミュニティの慣習による引用」などがある。

「コミュニティの人的つながりによる引用」は「被引用文献の著者と専門的なつながりを作りたい・維持したい」[11]や「知人の論文であるため引用した」など、人的関係の構築や人的影響により発生する引用を含む。自己引用の場合も再帰的ではあるが「過去の自己とのつながりによる引用」と考えることもできるため、人的つながりによる引用に含む。自己引用と他者引用を行った際の動機には大きな差が見られない[35]ことから、評価的観点では分けて考えるべきであるが、社会的観点においては自己引用か他者引用かといった違いはあまり重要ではない。

「コミュニティの慣習による引用」は「政治的な圧力で引用しなければならなかつた」[35]、「研究トピックの先駆者・研究領域の権威者を引用する」[4, 5, 7, 12, 34]や「他の研究者が引用していたため」[11]などコミュニティ内で何らかの妥当性を持って行われた引用を指す。

これらの社会的観点で捉えられる引用は、著者が研究コミュニティの構成員としての役割を踏まえて執筆を行うために発生する。必ずしも被引用文献の評価を行ってはいないため評価的観点からすれば「おざなり」とも見られ有用性は低く扱われる場合があるものの、引用関係からコミュニティの全体像や傾向を把握するためには十分に活用できる情報である。

以上のように整理した結果を表6にまとめた。

#### 4 各観点の特性

4章では、3章で設定した各観点に該当する区分について、尺度水準、推奨される分

表 6 引用分類の観点と定義

観点	観点の定義
(1)意義的観点	行われた引用の有意性から分類する観点
(2)評価的観点	被引用文献に対する引用元文献の評価により分類する観点
(3)機能的観点	引用の機能や役割から分類する観点
(4)形態的観点	被引用対象の形態的な側面から分類する観点
(5)位置的観点	引用が生じている箇所の位置に基づいて分類する観点
(6)社会的観点	著者が引用を行う社会的な背景に基づいて分類する観点

類スキーマ、活用時に注意すべき点について述べる。

#### 4.1 意義的観点の特性

意義的観点は評価指標の改善に有用であり、引用の有意性により分類する観点である。「有意性」による区分は定性的な順序尺度としてスキーマを構成することができる。著者にとって全く不要な引用というのは極めて少数である[11]一方、60%以上の著者が引用元文献中の不可欠な引用は20-80%であると回答している[43]ため、不可欠な引用ばかりではないと仮定する。Hanneyら[44]の調査では、「周辺的-限定的-重要-不可欠」4段階に引用を分けた場合はそれぞれ35%, 56%, 8%, 1%であり、不可欠なほどの引用はあまり多くないと考えられる。しかし、Hanneyらは「重要」と「限定的」を「中心的議論として扱われているか」で分けており、その判断は文章の読解力や主題への精通といった専門的な経験に左右されやすく、判断基準が主観的になりやすい。評価を行うための観

点としては「おざなりに行われた引用を除く」事がまず重要であるため、複雑化はさせず「論文の執筆には有用だが研究そのものには必要ない - 研究の基盤や前提となつており必要不可欠」のように単純で互いに排他的に構成するのが望ましい。ただし、低被引用論文が引用された場合より、高被引用論文が引用された場合の方が有意と扱われる傾向もある[45]ため、分析対象によるバイアスには注意が必要である。

#### 4.2 評価的観点の特性

評価的観点は研究評価指標の改善に有用であり、被引用文献に対する引用元文献の評価により分類する観点である。被引用文献に対する評価を直接的に表現している「賛否」を分類する場合は「賛成・肯定 - 反対・否定」のように対極的な名義尺度として扱うことも可能だが、「肯定的一否定的一客観的／抽象的（肯定的・否定的のどちらでもない、判定できない等）」に分けた場合「客観的／抽象的」の割合が70%以上ある[46]など、肯定的・否定的だけでは捉えきれない引用が大きな比率を占める。そのため、「賛成・肯定 - 中立 - 反対・否定」のような3段階の順序尺度として活用する方が適切と考えられる。否定の場合は、先行研究の完全否定ではなく主に部分否定であり[16]、反対・否定より賛成・肯定に該当する引用が圧倒的に多く見られることから「賛成・肯定」は「被引用対象を研究の前提として用いるなど明確に肯定的スタンスが読み取れる」、「反対・否定」は「先行研究の部分否定や不足・弱点の指摘を行った」、「中立」は「そのどちらでもない」のように定義するべきである。

本研究ではあくまで「引用」を分類することを目的としている。そのため「被引用文献についての評価」は異なる問題として考えなければならない。例えば、1つの被引用文献に対してある文脈では不十分であることを指摘した引用を行いながら、別の文脈では肯定することも考えられる。被引用文献に対する評価を行うのであれば、複数個所に存在する引用を全て含め、量的に賛成・肯定と反対・否定の回数から総合評価を行うか、手法は肯定しているが結果には疑問を持っているなど質的な評価を行う必要がある。

#### 4.3 機能的観点の特性

機能的観点は主に引用索引や情報検索のような場面で情報を付加するために有用な観点である。そのため分類スキーマを作成する場合には名義尺度として扱い、一つの引用に付与する複数のラベルとして扱うのが妥当である。従来の分類スキーマであれば「概念的／操作的」「進歩的／並列的」の2種類は対立的に扱われてきたが、それらと他の機能的観点で見られる区分肢を対等に扱い、「概念や理論を参照するため、操作方法を参照するため、継続性を示す、背景情報を提供する、仮説・推測する、定義や前提として参照し補足・正当化するため、比較するため、広報するため」のように再構成した。「並列的」という区分は「継続的ではない」という意味が強いためスキーマには含めなかった。情報探索上有用であれば、さらに他の区分肢を付け加えることが望ましい。実際に区分を行う際、引用が行われた周辺のテキストに出現する用語を判断基準とすることで分類が可能であるため、文章中の動詞や構文解析

などにより客観的に判定しやすく、機械的な分類も行いやすい。

概念的／操作的という区分は科学技術分野以外に適用されることを想定していない[47]と指摘されるように、研究領域によって記述方法や内容が異なる可能性は高い。そのため、活用時には全てのラベルが使用されるとは限らないことに注意しなければならない。

#### 4.4 形態的観点の特性

形態的観点は引用行動の傾向を掴むことで図書館のコレクション管理に役立てることが可能な観点である。

被引用対象には「文献タイプ」と「データタイプ」がある。文献は「書籍、原著論文、レビュー論文、レター論文、会議録、見解、コメント・意見、書評、エディトリアル、エッセイ」などのタイプからなる主義尺度として分類が可能である。一つの引用が複数の区分肢に当てはまるることは考えにくく、客観的に判断しやすい。データの引用については発展の途上であり、未だデータの標準的なタイプは定まっていない。あくまで例ではあるが、Nature Publishing Groupが創刊したデータジャーナルのScientific Data誌では「オミックス、摂動スクリーニング、体系的顕微鏡データ、生態データ・環境データ、行動データ、人口データ<sup>3</sup>」といったタイプが挙げられている。自然科学分野の実験データだけでなく、行動データや人口データなどの社会科学分野で活用されやすいタイプも存在することから、本研究ではひとまずこれらのタイプを採用する。

<sup>3</sup><http://www.natureasia.com/ja-jp/scientificdata/for-authors>

被引用対象の形態に関する情報は、引用元文献におけるレファレンス部分の記述方式から判定可能である。しかし、データセットの場合、Dataverse、Dryadなどのデータリポジトリでは引用の推奨記述形式が存在するものの、2016年3月現在、データセットの引用形式が示されているスタイル(APA<sup>4</sup>など)と記述形式が定まっていないスタイル(MLAやChicagoスタイル)が混在しており、レファレンスの記述形式だけでは判断しにくい環境であると言える。

また、活用に際しては分野の慣習や雑誌の投稿規程によって文献タイプの定義や表記が異なるため、分類後のデータを集計・活用する際には調査対象に合わせた区分の設定が必要となる。

#### 4.5 位置的観点の特性

位置的観点は他の観点と掛け合わせて引用の性質を調べるためや、引用元文献内の位置関係から被引用文献の類似度を評価するためなどに活用される。構造を捉えるための単位としては「章、節、段落、文、列挙」といった単位が存在した。章の単位は、学術雑誌のXML記述仕様であるJournal Article Tag Suite (JATS)<sup>5</sup>で、<sec>タグを使用して本文のセクションを構造化する場合、intro, methods, materials, results, discussion, conclusions, cases, subjects, supplementary-materialなどの値を記述することが推奨されているように、「序論、方法、結果、考察、結論、事例、材料」などと分けることができる。一見すると客観的

<sup>4</sup><http://blog.apastyle.org/apastyle/2013/12/how-to-cite-a-data-set-in-apa-style.html>

<sup>5</sup> <http://dtd.nlm.nih.gov/>

ではあるが、文献の構造は分野の慣習や雑誌の投稿規程、文献タイプ、著者個人の考え方などに左右される。例えば医学分野ではほぼIMRaD構造であるとされている[48]が、他分野では必ずしもそのような慣習は存在しない。また、「研究方法、研究手法、分析手法と対象」のように表記や範囲の揺れも多く見られるため、統計的な分析を行うためのカテゴリとして使用する際には、あらかじめ調査対象でどのような区分が行われているか明らかにしてから活用することが望ましい。

活用時のバイアスとして、他分野の文献はintroductionやfindingsの箇所で行われやすく同分野の文献はそれに加えMethodologyでもよく引用されることや[49]、高被引用文献は引用元文献の前方で引用されやすい[40]といった点にも注意する必要がある。

#### 4.6 社会的観点の特性

社会的観点は科学社会学的、行動科学的な分析を行う際に有用であり、「人的つながりによる引用、慣習による引用」が存在する。社会的観点は研究コミュニティとの関わり方を観察する観点である。そのため引用元文献のみだけではなく、コミュニティや他の研究との関係から考えなければならない。

「人的つながりによる引用」は著者のつながりに着目するため、引用元文献および被引用文献の著者情報から分類される。近年では1論文あたりの著者数が増加しているため<sup>6</sup>共著論文同士でも取り扱うことができるよう区分肢としなけ

ればならない。同時に分類として排他的な区分とするため「自己引用-隠れ自己引用-他者引用」からなる名義尺度として扱う。引用元文献の第一著者が被引用文献の著者の一人である場合を「自己引用」、引用元文献の共著者（第一著者以外の著者）が被引用文献の第一著者または共著者になっている引用を「隠れ自己引用」[50]、引用元文献と被引用文献の著者に重なりがない場合を「他者引用」と捉える。あくまで社会的観点におけるつながり方の種類を定義しているだけであり、「研究評価指標へ自己引用を含めるべきか」といった評価的側面は考慮しない。分類を活用する場合には、被引用回数が少ない論文ほど自己引用の割合が高い傾向がある[51]、隠れ自己引用は近年増加傾向にある[52]などのバイアスを意識する必要がある。

「慣習による引用」は、従来、評価的な意味での「おざなり」な引用を含む概念として扱われてきた。しかし、本研究では慣習的に行われる引用は社会的な意味があるとの立場をとる。「慣習的な引用」は、「先駆者への敬意」や「他の著者も引用していた」など、コミュニティ内の研究者が同じ文献を引用していることにより引用が生起する。そのため、引用元文献と同じ研究領域の論文を集め、同じ被引用文献があれば、それは慣習として行われた引用であるとみなすことができる。「同じ研究領域の論文で既に引用された-同じ研究領域の論文では引用されていない」に分けるのが妥当である。ただし、対象となる引用元文献の分野が非常に先進的で同じ研究領域の論文が見当たらない場合、この判定方法は適さない

<sup>6</sup><https://jipsti.jst.go.jp/foresight/dataranking/collaboration/authors/>

ことには注意が必要である。

## 5 再定義の効果

本研究では、従来の研究で行われた引用分類の観点を整理・再定義したうえで、それぞれの特性について述べた。分類の観点を、(1)意義的観点、(2)評価的観点、(3)機能的観点、(4)形態的観点、(5)位置的観点、(6)社会的観点の6つにわけ活用目的に合わせた概念的な分類スキーマを再構築した。意義的観点は「論文の執筆には有用だが研究そのものには必要ない - 研究の基盤や前提となっており必要不可欠」、価値的観点は「賛成・肯定 - 中立 - 反対・否定」、機能的観点は「概念や理論を参照するため、操作方法を参照するため、継続性を示す、背景情報を提供する、仮説・推測する、定義や前提として参考し補足・正当化するため、比較するため、広報するため」、形態的観点は「文献タイプ：書籍、原著論文、レビュー論文、レター論文、会議録、見解、コメント・意見、書評、エディトリアル、エッセイ」「データ：オミックス、撮動スクリーニング、体系的顕微鏡データ、生態データ・環境データ、行動データ、人口データ」、位置的観点は「章（序論、方法、結果、考察、結論、事例、材料）、節、段落、文、列挙」、社会的観点は「人とのつながり：自己引用-隠れ自己引用-他者引用」「慣習：同じ研究領域の論文で既に引用された-同じ研究領域の論文では引用されていない」のように分けた。調査対象とする文献の研究分野や論文タイプなどによって、どの観点が特に重視されるかといった違いはあるだろう。し

かしながら、引用という現象を統合的に整理し、分野に依存せず適用できる分類スキーマを設定したことは、分野間比較を可能にする共通基盤の構築につながるものと考える。

次に、従来の区分肢を統合できたことと分離できたことが効果として挙げられる。例えばGarzoneとMercer[8]が実験で用いたカテゴリでは「肯定的タイプ」と「否定的タイプ」は単純に異なる区分肢として扱われていた。しかし、本研究では評価的観点である「賛否」という観点で見られる区分肢としてそれぞれを扱い、実証的な先行研究の研究結果から「中立」という区分肢もまた現実的には必要であることを明らかにした。引用理由を分析することや機械的な分類可能性を模索するために実際的な区分を用いるだけでは得られない、整理による効果であったと言える。また、本研究では研究評価指標で「おざなり」とされて不要とされやすい「慣習的に行われる引用」を明確に評価的観点と社会的観点として概念的に異なる観点から生成される区分として分離したこと、引用が分類される観点に応じて必要とされる区分のあり方も議論できる準備が整ったと言える。

## 6 おわりに

本研究では引用を分類する際に用いる観点を体系的に整理・再定義し、各観点の特性と分類スキーマについて述べた。引用を分類する場合に用いられる観点を整理することで、複合的な観点から用いられていた区分肢を分離することができた。実際に引用分類スキーマを用いて分

類作業を行う場合には、作業者の判断にブレがでない客観的な判定基準や引用部分の機械的な検出方法も検討し、その精度や問題点を検証する必要がある。また、今回は引用を分類する際の観点を整理したため「引用」単位で分類することを前提に考察しているが、特に研究評価指標へ適用するために計量する際は1つの被引用文献が引用元文献内で何度も引用される場合を考慮する必要がある。今後の研究では、今回整理した引用分類の観点に対しどのように区分肢を設定し、被引用文献の重みづけの方法を構築するか検討していく必要がある。

## 謝辞

本研究に際して種々ご指導を頂きました筑波大学の小野寺夏生名誉教授、緑川信之教授に深謝致します。また日常の議論を通じて多くのアドバイスをいただいた筑波大学大学院の学生の皆様に感謝致します。

## 参考文献

- [1] Shotton, David: "CiTO, the Citation Typing Ontology", *Journal of Biomedical Semantics*, vol.1(Suppl .1):S6, pp.1–18, 2010.
- [2] Small, Henry: "Citation context analysis", Dervin, B. J. & Voigt, M. J. (eds.), *Progress in Communication Sciences*,3, Norwood, NJ: Ablex, pp. 287–310, 1982.
- [3] Peritz, Buma C.: "A classification of citation roles for the social sciences and related fields", *Scientometrics*, Vol.5, pp.303–312, 1983.
- [4] Garfield, E. "Can citation indexing be automated?" In: *Essays of an Information Scientist*, Vol.1, pp. 84–90 Philadelphia, PA: ISI Press; 1977
- [5] Lipetz, ben-ami: "Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators", *American Documentation*, Vol.16, pp.81–90, 1965.
- [6] Moravcsik, Michael J.; Murugesan, Poovanalingam: "Some results on the function and quality of citations", *Social Studies of Science*, Vol.5, No.1, pp.86–92, 1975.
- [7] Brooks, Terence A.: "Private acts and public objects: an investigation of citer motivations", *Journal of the American Society for Information Science*, Vol.36, pp.223–229, 1985.
- [8] Garzone, Mark ; Mercer, Robert E.: "Towards an Automated Citation Classifier", *Advances in Artificial Intelligence 2000*, Hamilton, H. and Yang, Q. (Eds.), Montréal, Québec, Canada, Springer, pp.337–346, 2000.
- [9] White, Howard D.: "Citation analysis and discourse analysis revisited", *Applied linguistics*, vol.25, iss.4, pp.89–116, 2004.
- [10] Cano, V.: "Citation behavior: Classification, Utility, and location", *Journal of the American Society for Information Science*, Vol.40, pp.284–290, 1989.
- [11] Vinkler, P.: "A quasi-quantitative citation model", *Scientometrics*, Vol.12, pp.47–72, 1987.
- [12] Small, Henry: "On the shoulders of Robert Merton: Towards a normative theory of citation", *Scientometrics*, Vol.60, No.1, pp.71–79, 2004.

- [13] Harwood, Nigel : “An interview-based study of the functions of citations in academic writing across two disciplines”. *Journal of Pragmatics*, vol.41, pp.497–518. 2009.
- [14] Bornmann, Lutz; Daniel, Hans - Dieter : “What do citation counts measure? A review of studies on citing behavior” *Journal of Documentation*, vol.64, Iss.1 , pp.45–80, 2008.
- [15] Ding, Ying; Zhang, Guo ; Chambers, Tammy ; Song, Min. ; Wang, Xiaolong ; Zhai, Chengxiang: “Content-based citation analysis: The next generation of citation analysis”, *Journal of the Association for Information Science and Technology*, vol.65, Iss.9, pp.1820–1833, 2014.
- [16] Chubin, Daryl E.; Moitra, S. D.: “Content analysis of references: Adjunct or alternative to citation counting?”, *Social Studies of Science*, Vol.5, No.4, pp.423–441, 1975.
- [17] Murugesan, Poovanalingam; Moravcsik, Michael J.: “Variation of the nature of citation measures with journals and scientific specialties”, *Journal of the American Society for Information Science*, Vol.29, Iss.3, pp.141–147, 1978.
- [18] Spiegel-Rösing, Ina: “Bibliometric and content analysis”, *Social Studies of Science*, Vol.7, No.1, pp.97–113, 1977.
- [19] Oppenheim, Charles; Renn, Susan P: “Highly cited papers and the reasons why continue to be cited”, *Journal of the American Society for Information Science*, Vol.29, pp.225–231, 1978.
- [20] Frost, Carolyn O.: “The literature of online public access catalogs, 1980-1985 : an analysis of citation patterns”, *Library Resources and Technical Services*, vol.33, No.4, pp.344–357, 1989.
- [21] McCain, Katherine W.; Turner, Kathleen: “Citation context analysis and aging patterns of journal articles in molecular genetics”, *Scientometrics*, vol.17, Iss.1, pp.127–163, 1989.
- [22] Brooks, Terence A.: “Evidence of complex citer motivations”, *Journal of the American Society for Information Science*, Vol.37, pp.34–36, 1986.
- [23] Line, Maurice B.; Sandison, A.: “PROGRESS IN DOCUMENTATION: ‘obsolescence’ and changes in the use of literature with time”, *Journal of Documentation*, Vol. 30, Iss.3, pp.283–350, 1974.
- [24] Gilbert, G Nigel: “Referencing as persuasion”, *Social Studies of Science*, vol.7, No.1, pp.113–122, 1977.
- [25] Bavelas, Janet B.:“The social psychology of citations”, *Canadian Psychological Review*, vol.19,Iss.2, pp.158–163, 1978.
- [26] Cole, Stephen The growth of scientific knowledge: Theories of deviance as a case study. In *The Idea of Social Structure; Papers in Honor of Robert K. Merton*, pages. 175–220. New York: Harcourt, Brace Jovanovich., 1975.
- [27] Duncan, E. B.; Anderson, F.D.; McAleese, R.: Qualified citation indexing: its relevance to educational technology. In *Information retrieval in educational technology: proceedings of the first symposium on information retrieval in educational technology.*, pp70–79. Aberdeen: University of Aberdeen., 1981.
- [28] Finney, B. “The reference characteristics of scientific texts” Master’s thesis, The City

- University of London., 1979.
- [29] Small, Henry G. "Cited documents as concept symbols". *Social Studies of Science*, vol.8, No.3, pp.327–340, 1978.
- [30] Weinstock, Melvin: *Citation indexes*. In *Encyclopedia of Library and Information Science*, vol.5, Dekker, New York, 1971.
- [31] Garzone, Mark. : "Automated classification of citation using linguistic semantic grammars". Master's. thesis, Dept. of Computer Science, The University of Western Ontario, 1996.
- [32] Teufel, Simone; Siddharthan ,Advaith; Tidhar, Dan: "Automatic Classification of citation function" *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing(EMNLP 2006)*, 2006, Sydney, Australia, Association for Computational Linguistics, pp.103–110, 2006.
- [33] Chang, Yu-Wei: "A comparison of citation contexts between natural sciences and social sciences and humanities", *Scientometrics*, Vol.96, No.2, pp.535–553, 2013.
- [34] Case, Donald O., Higgins, Georgeann M.: "How can we investigate citation behavior? A study of reasons for citing literature in communication", *Journal of the American Society for Information Science*, Vol.51, pp.635–645, 2000.
- [35] Bonzi, Susan; Synder, H.W.: "Motivations for citation: a comparison of self citation and citation to others", *Scientometrics*, Vol.21, pp.245–254, 1991.
- [36] Shadish, William R.; Tolliver, Donna; Gray, Maria; Sen Gupta, Sunil K.: "Author Judgements about Works They Cite: Three Studies from Psychology Journals", *Social Studies of Science*, Vol. 25, no. 3, pp.477–498, 1995.
- [37] Willett, Peter: "Readers' perceptions of authors' citation behavior", *Journal of Documentation*, Vol.69, No.1, pp.145–156, 2013.
- [38] Voos, H.; Dagaev, K. S.: "Are all citations equal? Or. Did we Op. Cit. Your Idem?", *Journal of Academic Librarianship*, Vol.1, pp.19–21, 1976.
- [39] Bertin, Marc ; Atanassova, Iana; Gingras, Yves; Larivière, Vincent: "The invariant distribution of references in scientific articles", *Journal of the Association for Information Science and Technology*, Vol.67, Iss.1, pp.164–177, 2016.
- [40] Hu, Zhigang; Chen, Chaomei; Liu, Zeyuan: "Where are citations located in the body of scientific articles? A study of the distributions of citation locations", *Journal of Informetrics*, Vol.7, Iss4, pp.887–896, 2013.
- [41] Callahan, Alison; Hockema, Stephen ; Eysenbach, Gunther: "Contextual cocitation: Augmenting cocitation analysis and its applications", *Journal of the American Society for Information Science and Technology*, Vol.61, No.6, pp.1130–1143, 2010.
- [42] 江藤正己. 「引用論文における引用箇所間の近さをとらえる尺度」, 第15回(2007年度)研究報告会論文集, 情報知識学会誌, Vol. 17, No. 2, pp. 65–68, 2007.
- [43] Liu, Mengxiong: "A study of citing motivation of Chinese scientists". *Journal of Information Science*, Vol.19, No.1, pp.13–23, 1993.
- [44] Hanney, Steve; Frame, Iain; Grant,

- Jonathan; Buxton, Martin; Young, Tracey ; Lewison, Grant: "Using categorisations of citations when assessing the outcomes from health research", *Scientometrics*, Vol.65, Iss.3, pp.357–379, 2005.
- [45] Hooten, Patricia A.: "Frequency and functional use of cited documents in information science", *Journal of the American Society for Information Science*, Vol.42, Iss.6, pp.397–404, 1991.
- [46] Athar, Awais ; Teufel, Simone: "Context-enhanced citation sentiment detection", *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT '12)*, pp.597–601, 2012.
- [47] Swales, John : "Research articles in English" , *Genre Analysis: English in Academic and Research Settings*, Cambridge, UK, Cambridge University Press, pp.110–176, 1990.
- [48] 青木仕 : 「医学における情報ソースとしてのStructured Abstracts (構造化抄録) の研究」, 総ページ数p. 65, 2005. 博士論文.
- [49] Halevi, Gail; Moed, Henk F. : "The thematic and conceptual flow of disciplinary research: a citation context analysis of the journal of informetrics", *Journal of the American society for information science and technology*, Vol.64, No.9, pp.1903–1913, 2013.
- [50] Zmaić, Ljerka; Maričić , Siniša; Simeon, Vladimir: "Visibility of peripheral journals through the science citation index", *Information Processing & Management*, Vol. 25, Iss.6, pp.713–719, 1989.
- [51] Aksnes, Dag W.: "A macro study of self-citation", *Scientometrics*, Vol.56, no.2, pp.235-246, 2003.
- [52] Costas, Rodrigo; van Leeuwen, Thed N.; Bordons, Maria : "Self-citations at the meso and individual levels: effects of different calculation methods", *Scientometrics*, vol.82, no.3, pp.517–537, 2010.

(2016年6月28日受付)  
(2016年9月18日採択)

付表 1 調査対象文献一覧

著者	研究目的	調査手法	対象分野
Garfield, 1962	引用理由の解明	-	-
Lipetz, 1965	引用索引の改善のため	対象論文を読みスキーマ構築	物理学
Murugesan and Moravcsik, 1975	文脈における引用の質の測定	文脈分析	物理学
Chubin and Moitra, 1975	引用する文献の採用理由の解明	インタビュー	物理学
Voos and Dagaev, 1976	引用分析を行う際に位置を考慮する必要性の明示	被引用文献が掲載された位置を集計	主に生物学、医学、物理学（上記分野の文献を引用した文献が対象）
Ina Spiegel-Rusing, 1977	学際的な雑誌の傾向の解明	内容分析	科学社会学
Oppenheim and Renn, 1978	古い文献が引用される理由の解明	文脈分析	主に物理学と物理化学（※物理学と物理化学の論文を引用した文献が対象）
Cano, 1989	引用のタイプ、重要性、位置の関係の解明	アンケート	工学
Peritz, 1983	引用の役割を実証的に分類するためのスキーマの提案	内容分析	社会科学関連領域
Brooks, 1986	引用分類の評価としての有用性の検証、区分肢同士の相関の解明	インタビュー	解剖学、人類学など 15 分野
Vinkler, 1987	引用理由の解明	アンケート（複数回答可）	化学関連領域
Bonzi and Snyder, 1991	自己引用と他者引用の違いの解明	アンケート	化学、生物学、地学、物理学
Shadish et al., 1991	著者が引用を行う判断基準の解明	アンケート	心理学
Case and Higgins, 2000	引用動機の解明	アンケート	人文社会系
Garzone and Mercer, 2000	引用機能の分類の自動付与	自動分類	生化学、物理学
Hanney et al., 2005	リサーチインパクトの評価	文脈分析	糖尿病
Teufel, 2006	自動引用索引の改善	自動分類	コンピュータ科学
Harwood, 2009	引用行動や記述方式の比較	インタビュー	コンピュータ科学、社会科学
Chang, 2013	引用機能の分野間比較	自動分類	自然科学、社会科学
Hu Z, Chen C, Liu Z. , 2013	引用の位置と分布の関係の	自動分類	計量情報学

情報知識学会 平成 28 年度総会 議事録

1. 開催日 平成 28 年 5 月 14 日(土)11:55 ~ 12:33  
 2. 会 場 筑波大学筑波キャンパス春日地区(茨城県つくば市春日)  
 3. 議 長 石塚会長  
 4. 議 事 1) 総会有効成立確認 [資料 1]  
     出席者 21 名、委任状 50 通、計 71 名  
     従って定足数(正会員の 10 分の 1)を満たし、総会成立。  
     2) 平成 27 年度事業報告：田良島常務理事 [資料 2]  
     3) 平成 27 年度決算報告：田良島常務理事 [資料 3]  
     4) 平成 27 年度監査結果報告：細野監事 [資料 4]  
     5) 平成 28~29 年度 役員選挙結果報告：松村選挙管理委員 [資料 5]  
     6) 平成 28 年度事業計画説明：長塚副会長 [資料 6]  
     7) 平成 28 年度予算説明：長塚副会長 [資料 7]  
     8) その他の報告  
       ・ 2016 年度論文賞は会員投票の結果、児玉閥・小野寺夏生 論文および  
         荒井俊介・辻慶太 論文に決定。表彰式・記念講演は 15 日 14 時から  
       ・ 2016 年度永年会員表彰対象者 12 名。15 日 15 時から表彰式 [資料 8]  
 上記の報告および事業計画・予算案はいずれも原案どおり承認された。以上

平成 28 年度総会資料

## [資料 1] 会員数

	平成 27 年 3 月 31 日現在	平成 27 年度 入 会	平成 27 年度 退 会	平成 28 年 3 月 31 日現在
正会員	234名	9 (1)†	13 (2)†	230
学生会員	43	2	10 (1)†	35
ユース会員	8	2	1	9
シニア会員	21	2 (1)†	1	22
名誉会員	1	1 (1)†	0	2
賛助会員 *	7 (24)	0	1	6 (21)
合 計	314	16	26	304

† 入会者、退会者の()は会員種別変更者数で内数。

学生会員→正会員、正会員→シニア会員／名誉会員

初代副会長藤原譲 正会員が平成 27 年度総会において名誉会員に推戴されました。

\* 賛助会員は全て団体。()内数字は口数。賛助会員 1 口の年会費¥30,000。

## [資料2] 平成27年度事業報告

### (1) 会議

- ・平成27年4月27日 18:30～20:40 第1回常務理事会 於：凸版印刷(株)西館
  - ・平成27年5月11日 16:00～18:00 監査会 於：(株)アドスリー内
  - ・平成27年5月11日 18:30～20:25 第1回理事会 於：(株)アドスリー内
  - ・平成27年5月23日 15:10～15:45 第1回総会 於：東洋大学白山内
  - ・平成27年5月23日 15:45～16:10 第2回理事会 於：東洋大学白山内
  - ・平成27年7月17日～7月21日 第3回理事会 メールによる持ち回り
  - ・平成27年11月16日 18:30～20:30 第6回常務理事会 於：凸版印刷(株)西館
  - ・平成27年11月24日～11月28日 第4回理事会 メールによる持ち回り
  - ・平成27年12月15日～12月17日 第1回選挙管理委員会 メールによる持ち回り
  - ・平成28年1月15日 19:20～20:40 第2回選挙管理委員会 於：(株)アドスリー内
  - ・平成28年2月18日 19:05～20:30 第3回選挙管理委員会 於：(株)アドスリー内
  - ・平成28年2月29日 18:30～20:40 第8回常務理事会 於：凸版印刷(株)西館
- なお、常務理事会の第2回、第3回、第4回、第5回、第7回はメールによる持ち回り会議で、それぞれ平成27年5月、7月、8月、8月、12月に行われた。

### (2) 事業

- ・情報知識学会／第24回年次大会（総会・研究報告会） 主催：情報知識学会  
平成27年5月23～24日 於：東洋大学白山キャンパス
- ・アート・ドキュメンテーション学会／2015年度年次大会 後援：情報知識学会ほか  
平成27年6月6～7日 於：国立西洋美術館
- ・Code4Lib Japan 2015 カンファレンス 後援：情報知識学会ほか  
平成27年9月5～6日 於：リクルート本社
- ・図書館総合展運営委員会／第17回図書館総合展・学術情報オープンサミット2015 後援：情報知識学会ほか  
平成27年11月10～12日 於：パシフィコ横浜
- ・JATS-Con Asia 国際会議 主催：科学技術振興機構、学術情報 XML 推進協議会  
平成27年10月19日 於：科学技術振興機構 後援：情報知識学会ほか
- ・第20回情報知識学フォーラム 主催：情報知識学会  
平成27年12月12日 於：同志社大学今出川キャンパス
- ・INFOSTA／2015 ジョイント懇話会 後援：情報知識学会ほか  
平成27年12月18日 於：大阪市中央公会堂
- ・情報処理学会／じんもんこん2015 後援：情報知識学会ほか  
平成27年12月19～20日 於：同志社大学京田辺校地

### (3) 学会誌編集委員会（芦野俊宏 編集委員長）

1. 学会誌発行

- ・情報知識学会誌 25巻2号 (研究報告会論文集) 平成27年5月23日発行
- ・情報知識学会誌 25巻3号 通常論文 平成27年9月28日発行
- ・情報知識学会誌 25巻4号 (フォーラム特別号) 平成27年12月12日発行
- ・情報知識学会誌 26巻1号 通常論文 平成28年2月25日発行

## 2. 学会誌を cover-to-cover のデジタルデータとして保存するプロジェクト

第18巻以降、印刷時のデータ収集、欠落部分については印刷物のスキャンなどを行い、第26巻第1号までのデジタル化を完了した。今後は各巻号発行後にスキャンデータを作成、随時公開する。

## 3. 第12回(2015)情報知識学会論文賞選考 (論文賞推薦委員長:長塚隆 副会長)

### (4) 広報関連

1. メールマガジン (岡本由起子 メールマガジン編集長、高橋幸雄 副編集長)  
2015年度も、各理事の方々の多大なご協力を得て、メールマガジンの発刊を継続することができたことを感謝いたします。
  - ・昨年度に引き続き、編集室は二体制。配信はアドスリーに委託。
  - ・第90号から99号まで、10回刊行しました(4月、8月 休刊)。
  - ・編集:5月 岡本・堀、6月 村川、7月 高久、9月 田良島、10月 田窪、11月 長田、12月 孫、1月 梶川、2月 村井、3月 大槻、の各理事の方々) (敬称略)。
  - ・学会誌Vol.26, No.1「メールマガジン・アーカイブ」に、2015年1月号~12月号を収録。
2. ホームページ (江草由佳 常務理事)  
平成27年4月~平成28年3月 随時更新。

### (5) 第23回年次大会 (芦野俊宏 実行委員長)

平成27年5月23-24日の両日、東洋大学白山キャンパスにおいて総会、研究報告会を開催した。その中で第12回(2015)情報知識学会論文賞授与・記念講演及び永年会員表彰式を実施した。都内での開催にも関わらず参加者は49名に留まり、広報面での課題が残った。

### (6) 第20回情報知識学フォーラム (原 正一郎 実行委員長)

第20回情報知識学フォーラムは、2015年12月12日(土)に同志社大学今出川キャンパスで開催した。実行委員は原正一郎(委員長)、原田隆史、田良島哲で、「地域情報学における知識情報基盤の構築と活用」をテーマに6本の講演を行った。約40名の参加があり、多面的な討論が行われて、最新の研究動向を共有する有意義な機会となつた。

### (7) 部会活動

#### ・CODATA部会(岩田修一 部会長)

オープンアクセス、オープンデータ、オープンサイエンスへの大きなトレンドを踏まえて、日本学術会議における活動への協力を行った。

・人文・社会科学系部会（田良島 哲 部会長）

2015年12月に開催した情報知識学フォーラム「地域情報学における知識情報基盤の構築と活用」を実行委員長の原正一郎理事、原田隆史理事とともに企画し、情報学の人文学への応用について幅広い知見を得るとともに、議論を行った。

・専門用語研究部会（長田孝治 理事）

1) 専門用語シンポジウムの開催

2012年にISO/TC37(専門用語及び他の言語、情報内容の資源)に新しいサブコミッティとしてSC5(翻訳・通訳関連)が新設され活発に活動しています。TC37の総会が松江で開催されたのに伴い、その参加者を中心に『『翻訳の品質管理』を考える—翻訳業界への黒船?』とのタイトルで6月27日に文京シビックホールでシンポジウムを開催いたしました。当日は30人強ほどの参加者でしたが、情報知識学会の会員のみならず翻訳・通訳の業務に従事されている方、多くの参加がありました。

2) 国際協力

ISO/TC37総会が6月21日から松江で開催されました。長年の課題であった日本開催が実現でき、部会として可能な限りの協力を行いました。参加者は、およそ国外80名、国内20名といった状況で盛会でした。EATerm(東アジア専門用語フォーラム)の開催はありませんでした。

3) 国内協力

ISO/TC37の国内事務局が規格協会から情報科学技術協会に移動したので、専門用語研究部会として協力をを行っています。

・関西部会（村川猛彦 部会長）

第5回 知識・芸術・文化情報学研究会（平成28年2月6日、於：立命館大阪梅田キャンパス）を、アート・ドキュメンテーション学会関西地区部会と共に催した。

第21回公開シンポジウム「人文科学とデータベース」（平成28年2月27日、於：同志社大学）を後援した。

・シニア情報知識学研究部会（松村多美子 代表世話人）

シニア情報知識学研究部会では、卓話会を年間数回開催するものとしてきており、本年度は下記の2回を開催した。

1) 2015年度第1回卓話会（通算第9回）

講師：飯塚健氏（群馬大学名誉教授）、テーマ：「化学でのネーミングで思い出すことと、その後」、2015年10月2日（金）18:00-19:30、鶴見大学記念館。

2) 2015年度第2回卓話会（通算第10回）

講師：斯波義信氏（東洋文庫文庫長、日本学士院会員）、テーマ：「東洋文庫の沿革とこれから」、2016年2月3日（水）17:00-18:30、東洋文庫。なお、卓話に先立ち、斯波先生のご案内にて、東洋文庫所蔵の貴重書を実地に観覧し、解説して頂いた。

卓話会講師には、卓話会後に演題に関連した論稿を学会誌に寄稿して頂くようお願いしており、上記本年度第1回卓話会については、講師より寄稿を得て、飯塚健「化学の境界領域およびナノスケール構造のネーミング問題」、情報知識学会誌、26(1)、p29-34(2016年2月)が「部会報告、事始めシリーズ(4)」として既に刊行された。

#### (8) 学会事務局の外部委託 (石塚英弘 会長)

本学会の学会事務の外部委託の検討が始まったのは平成25年の4月である。それ以来、3年間を掛けて、外部委託を慎重かつ段階的に進めてきた結果、平成28年3月末を以て、学会事務の外部委託が完成した。お蔭さまで、外部委託は成功した。これは、関係各位のご尽力の賜物である。

外部委託を進めてきた手続き、本学会としての方針決定などについては、平成25年度、26年度の事業報告で述べ、また、常務理事会、理事会、総会の審議・決定についてはメールマガジンで報告してきた。そこで本稿では、関係各位が如何に考えて進めてきたかについて報告する。

外部委託を検討し、本学会のためのソリューションを策定した組織は常務理事会であった。前会長根岸先生は東京大学大学院修了の経営学の専門家として、全体を見通した上で、原田副会長をソリューション立案者として推薦した。原田先生は文化系の某学会で会員担当を、次いで会計担当を体験し、その実務を熟知するとともに、その実務を改善する明確な案を考えていた。また、田良島常務理事は文化系の某学会で学会事務の外部委託を実施して成功させた経験者であり、外部委託の際に注意すべき点は何かを熟知していた。本学会は、様々な専門分野から、個々の領域を超える意識を持った多様な方々が集まっている点が特長である。この常務理事会の構成は正に、その特長を顕していた。また、殆どのメンバーが前会長の時からのメンバーであるため、気心が知れていて気兼ねなく議論ができた点も有効・有益であった。

本学会設立の3年後の平成3年8月から26年1月まで、実に22年半の長きに亘って、五所氏には事務局長として学会事務の全てについて木目細かなサービスを提供していただいた。その五所氏から学会事務の実務について聴いたことは外部委託の検討に誠に有益であった。そもそも、凸版印刷㈱内の部屋の一角に学会事務所の机が置けた理由は、同社で活躍された五所氏が事務局長に就いてくださったお蔭であった。

外部委託のソリューションを実現され、他の学会の仕事もある中、時間を割いて実行してくださる方々は㈱アドスリーの福田氏ほかの社員である。その着実な仕事には心から敬意を表したい。

ところで、石塚は何をした?との質問があろう。石塚はソリューションを実現するために、㈱アドスリーの福田氏と詳細な打ち合わせを幾度となく行ってきた。表に立っていたと言えよう。今まで、ソリューションの立案者の名前は伏せていたが、それは外部委託が失敗した際の責任が石塚一人で止まるようにするためにであった。今や成功は明らかである。従って、立案者の名前を感謝を以て示し、その功に報いたい。

## [資料3] 平成27年度決算

平成28年3月31日現在

(単位:円)

## 収入の部

科目	細目	27年度予算	予算細目	27年度実績	実績細目	備考
会費	正会員	2,445,000	1,572,000	2,617,000	1,742,000	年会費¥8,000
	学生会員		127,000		122,000	年会費¥4,000
	ユース会員		32,000		41,000	年会費¥4,000
	シニア会員		84,000		82,000	年会費¥4,000
	賛助会員		630,000		630,000	一口¥30,000
参加費	年次大会	150,000	80,000	186,000	81,000	資料代、原稿超過頁代、情報交流会参加費
	フォーラム		70,000		105,000	資料代、情報交流会参加費
	セミナー		0		0	
売上金	学会誌	338,423	150,000	314,024	135,434	
	別刷り		150,000		65,000	
	著作権料		38,423		113,590	学術著作権協会:著作権料(国内・国外)、著作権管理・啓蒙補助
その他の収入	預金利息	1,000	1,000	200	200	
	寄付金・その他		0		0	
小計		2,934,423	2,934,423	3,117,224	3,117,224	
前年度繰越金		2,158,187	2,158,187	2,158,187	2,158,187	
合計		5,092,610	5,092,610	5,275,411	5,275,411	
特別会計		0	0	0	0	
特別会計 積立金		2,400,000	2,400,000	2,400,480	2,400,480	
総計		7,492,610	7,492,610	7,675,891	7,675,891	

## 支出の部

科目	細目	27年度予算	予算細目	27年度実績	実績細目	備考
印刷費	学会誌	1,185,000	1,100,000	750,384	670,464	25巻1~4号、別刷り含む
	選挙		22,000		14,904	役員選挙:投票案内状、投票用紙
	その他		63,000		65,016	角4号・長3号封筒、総会案内・葉書
人件費	事務局	85,000	0	63,936	0	
	編集事務局		54,000		34,560	J-STAGE登載作業
	HP管理		31,000		29,376	Webページ更新作業、学会誌目次更新作業
事業費	第24回年次大会	320,000	100,000	356,154	0	アルバイト代、情報交流会経費その他
	第23回年次大会		100,000		134,470	アルバイト代、情報交流会経費その他
	第22回年次大会		0		0	
	フォーラム		100,000		181,684	予稿集印刷・製本代、講師交通費、アルバイト代、情報交流会経費
	部会補助金		20,000		30,000	専門用語研究部会
会議費	理事会	87,000	54,000	64,266	16,926	理事会1回分
	常務理事会		23,000		35,100	常務理事会3回分
	その他		10,000		12,240	その他会議2回分
業務委託費		806,760	806,760	806,760	806,760	(株)アドスリー @67,230 * 12ヶ月=¥806,760円
賃借料			0		0	
通信費	学会誌発送費	274,000	120,000	236,081	106,210	
	編集通信費		0		0	
	総会		43,000		42,210	論文賞投票および総会案内発送費
	選挙		31,000		31,020	役員選挙投票一式発送費
	電話代		0		0	
	インターネット		50,000		34,617	ドメイン登録、さくらネット、BizFax、Nifty維持料
	その他		30,000		22,024	郵便切手代、宅急便代
交通費	役員旅費	58,000	0	35,232	0	
	事務局通勤費		0		0	
	その他交通費		58,000		35,232	郵便物引き取り:
消耗品費		10,000	10,000	6,894	6,894	ラベル代
維持修繕費		50,000	50,000	14,040	14,040	業務委託先への会計データ、定期購読者管理データの移行
雑費	手数料	30,000	30,000	32,678	32,678	振込手数料:個人会員年会費郵便振替料学会負担む
事務所借上経費		0	0	0	0	
小計		2,905,760	2,905,760	2,366,425	2,366,425	
予備費		2,186,850	2,186,850	0	0	
次年度繰越金		0	0	2,908,986	2,908,986	
合計		5,092,610	5,092,610	5,275,411	5,275,411	
特別会計		0	0	0	0	
特別会計 積立金		2,400,000	2,400,000	2,400,480	2,400,480	
総計		7,492,610	7,492,610	7,675,891	7,675,891	

**貸借対照表**  
平成28年3月31日現在

(単位:円)

科目	借方	貸方	備考
1.資産の部			
現金	0		
普通預金	815,400		三菱東京UFJ銀行
定期預金	2,400,480		三菱東京UFJ銀行
振替貯金	2,008,305		振替00150-8-706543
振替貯金	129,281		振替00130-9-706558
2.負債の部			
前受金		44,000	次年度以降の年会費7件
3.特別会計			
積立金		2,400,480	
4.正味財産		2,908,986	前年度は2,158,187 (750,799) 増
合計	5,353,466	5,353,466	

## [資料4] 監査報告書

### 監査報告書

情報知識学会  
会長 石塚 英弘 殿

作成日 平成28年4月20日

情報知識学会 監事 山本 敏雄

情報知識学会 監事 細野 公男

当監事は、情報知識学会の平成27年4月1日から平成28年3月31日までの財務諸表について、監査を行いました。この監査にあたって、当監事は一般に公正妥当と認められる監査基準に準拠し、通常実施すべき監査手続きを実施致しました。

監査の結果、当監事は上記財務諸表が平成28年3月31日現在の情報知識学会の財務状況を適正に表示しているものと認めます。

[資料 5] 平成 28～29 年度 役員選挙結果

平成 28 年 3 月 2 日、木村美実子役員選挙管理委員長より常務理事会に報告

[会 長] 長塚 隆 鶴見大学寄付講座教授

[理 事] 芦野 俊宏 東洋大学教授  
石塚 英弘 筑波大学名誉教授  
岩田 修一 事業構想大学院大学教授  
宇陀 則彦 筑波大学准教授  
江草 由佳 国立教育政策研究所総括研究官  
大槻 明 日本大学准教授  
岡本 由起子 欧州連合情報協会研究員  
長田 孝治 ロゴヴィスタ(株)  
梶川 裕矢 東京工業大学准教授  
阪口 哲男 筑波大学准教授  
孫 媛 国立情報学研究所准教授  
高久 雅生 筑波大学准教授  
田窪 直規 近畿大学教授  
田良島 哲 東京国立博物館 博物館情報課長  
根岸 正光 国立情報学研究所名誉教授  
原 正一郎 京都大学教授  
原田 隆史 同志社大学教授  
堀 幸雄 (株)ホワイトベース  
村井 源 東京工業大学助教  
村川 猛彦 和歌山大学准教授

[監 事] 細野 公男 慶應義塾大学名誉教授  
山本 肇雄 図書館情報大学名誉教授

補足

・役員選挙によって選出される理事のほかに、会長の指名により、特別賛助会員（凸版印刷(株)と大日本印刷(株)）からの指定者、各 1 名計 2 名を理事にしている。また、年次大会実行委員長、情報知識学フォーラム実行委員長が理事でない場合は、会長が実行委員長を理事会の議を経て理事としている。

・所属・肩書は 3 月 2 日時点に本学会・事務局が有する情報に依る。

## [資料 6] 平成 28 年度事業計画

### (1) 会議

- ・総会 平成 28 年 5 月 14 日開催 於:筑波大学筑波キャンパス春日地区
- ・理事会 年 2 回以上開催:平成 28 年 4 月 20 日開催、5 月 15 日開催予定
- ・常務理事会 隨時開催
- ・論文賞推薦委員会 年 2 回開催
- ・監査会 平成 29 年 4 月開催

### (2) 第 24 回年次大会 (高久雅生 実行委員長)

平成 28 年 5 月 14-15 日、筑波大学筑波キャンパス春日地区において、総会、研究報告会、第 13 回(2016)情報知識学会論文賞授与・記念講演及び永年員表彰式を実施する。

研究発表会では従来の一般セッションに加えて、学生セッションと学生奨励賞を復活し、ポスターセッションを新設する。また、情報処理学会人文科学とコンピュータ研究会と連携して、合同企画セッションを開催し、同研究会のポスターセッションも併催する。

### (3) 学会誌編集委員会 (芦野俊宏 編集委員長)

#### 1. 学会誌発行

例年どおり 1 年間に 4 号を発行する。

- ・情報知識学会誌 26 卷 2 号 (研究報告会論文集) 平成 28 年 5 月 14 日発行
- ・情報知識学会誌 26 卷 3 号 通常論文 平成 28 年 9 月末発行
- ・情報知識学会誌 26 卷 4 号 (フォーラム特別号) 平成 28 年 11 月末発行
- ・情報知識学会誌 27 卷 1 号 通常論文 平成 29 年 2 月末発行

#### 2. NII-ELS から J-Stage へのデータ移行対応

国立情報学研究所電子図書館(NII-ELS)事業の終了に伴い、ここに収録されている情報知識学会誌、情報知識学会研究報告会講演論文集の J-Stage へのデータ移行に対応する。

#### 3. 第 13 回(2016)情報知識学会論文賞選考

### (4) 広報関連

#### 1. メールマガジン (メールマガジン編集室 岡本由起子、高橋幸雄)

この 4 月でメールマガジンは第 100 号となり、記念号を発刊します。それを期に編集室の刷新を図って行きたいと思います(今年度を編集長交代の準備期間とする。編集をご担当下さる方を増やす。その他、読者との交流を目指す新企画の構想等。)

・各号の発刊については、昨年と同じような体制で臨みます(8 月を除く、毎月の刊行を予定。編集室は二人体制で、配信はアドスリーに依頼)。

・各号の編集は、今年度も、理事各位にご協力をお願いして進めさせて頂きます。

今年もまた、メールマガジンをなにとぞよろしくお願い申し上げます。

#### 2. ホームページ (江草由佳 常務理事)

平成 28 年 4 月～平成 29 年 3 月 随時更新

## (5) 第 21 回情報知識学フォーラム

従来は、数件の招待講演、最後に総合討論を行う構成で実施していたが、平成 28 年度からは、一般発表の募集を行い、必要に応じて招待講演を加える構成に変更する。この変更により、会員が発表する機会を増やすことができる。近々、実行委員長を決定する。

## (6) 部会活動

### ・ CODATA 部会（岩田修一 部会長）

1) 日本学術会議における活動への協力を継続する。特に、オープンアクセス、オープンデータ、オープンサイエンスへの大きなトレンド、そして IoT、Industry 4.0、FinTech、RDA などのデータ関連のアプリケーションの動向を踏まえて、第 23 期における提言作成への準備作業に参加する。

2) 新たなスタイルのデータジャーナル発刊、国際データ庁設立に向けた概念構築を行う。

### ・ 人文・社会科学系部会（田良島 哲 部会長）

1) 研究会を開催して、情報学と人文学との情報交換、連携の機会を設ける。

2) 行事の共催などを通じて、関連分野の学会との連携を図る。

### ・ 専門用語研究部会（長田孝治 理事）

#### 1) 専門用語シンポジウムの開催

翻訳・通訳に関連する用語の調整が TC37 全体の課題として検討されています。この動きを見ながら、必要に応じて紹介や意見交換を中心としたシンポジウムを検討します。EAFTerm（東アジア専門用語フォーラム）については本年度は各国の動きを見ながら対応を行います。

#### 2) 国際協力 ISO/TC37（用語）総会への参加

2016 年 6 月 27 日からデンマーク・コペンハーゲンにて ISO/TC37 総会が開催されるので、部会として協力を行います。

#### 3) 国内協力

ISO/TC37 の国内事務局である情報科学技術協会に対して専門用語研究部会として協力を行います。

### ・ 関西部会（村川猛彦 部会長）

知識・芸術・文化情報学研究会を今年度も実施し、関西部会の大会的な位置づけの会としたい。その他、関連学協会と連携しながら、学会員が興味を持ちそうな行事の共催・後援を行う予定である。

### ・ シニア情報知識学研究部会（松村多美子 代表世話人）

シニア情報知識学研究部会では、卓話会の企画を継続し、情報知識学各分野における温故知新と高齢情報化社会の動向を主たるテーマとして、相応の講師により数回の卓話会の開催を予定する。講師には学会誌の【事始めシリーズ】等への原稿執筆を依頼し、講演内容の定着を図る。その他活動についても世話人会にて随時検討し、実現を図る。

## [資料7] 平成28年度予算(案)

(単位:円)

科目	細目	27年度実績	実績細目	28年度予算	予算細目	備考
会費	正会員	2,617,000	1,742,000	2,378,000	1,576,000	年会費¥8,000
	学生会員		122,000		112,000	年会費¥4,000
	ニース会員		41,000		36,000	年会費¥4,000
	シニア会員		82,000		84,000	年会費¥4,000
	賛助会員		630,000		570,000	一口¥30,000
参加費	年次大会	186,000	81,000	120,000	80,000	資料代、原稿超過貢代、情報交流会参加費
	フォーラム		105,000		40,000	資料代、情報交流会参加費
	セミナー		0		0	
売上金	学会誌	314,024	135,434	231,410	135,434	
	別刷り		65,000		65,000	
	著作権料		113,590		30,976	学術著作権協会:国内著作権料
その他の収入	預金利息	200	200	60	60	
	寄付金		0		0	
小計		3,117,224	3,117,224	2,729,470	2,729,470	
前年度繰越金		2,158,187	2,158,187	2,908,986	2,908,986	
合計		5,275,411	5,275,411	5,638,456	5,638,456	
特別会計		0	0	0	0	
特別会計 積立金		2,400,480	2,400,480	2,400,480	2,400,480	
総計		7,675,891	7,675,891	8,038,936	8,038,936	

## 支出の部

(単位:円)

科目	細目	27年度実績	実績細目	28年度予算	予算細目	備考
印刷費	学会誌	750,384	670,464	1,165,000	1,100,000	26巻1~4号、別刷り含む
	選挙		14,904		0	役員選挙:投票案内状、投票用紙
	その他		65,016		65,000	角4号・長3号封筒、総会案内・葉書、請求書、宛名ラベル
人件費	事務局	63,936	0	105,000	0	
	編集事務局		34,560		55,000	J-STAGE登載作業
	HP管理		29,376		50,000	Webページ更新作業、学会誌目次更新作業
事業費	第25回年次大会	356,154	0	360,000	100,000	アルバイト代、情報交流会経費その他
	第24回年次大会		0		100,000	アルバイト代、情報交流会経費その他
	第23回年次大会		134,470		0	平成27年度年次大会:支払い完了
	フォーラム		181,684		100,000	アルバイト代、講師交通費、情報交流会経費
	部会補助金		40,000		60,000	CODATA部会、人文・社会科学系部会、専門用語研究部会、関西部会、シニア情報知識学研究部会
会議費	理事会	64,266	16,926	87,000	54,000	理事会2回分
	常務理事会		35,100		23,000	常務理事会2回分
	その他		12,240		10,000	その他会議
業務委託費		806,760	806,760	806,760	806,760	(株)アドスリー @67,230 * 12ヶ月=¥806,760円
賃借料		0			0	
通信費	学会誌発送費	236,081	106,210	243,000	120,000	別刷り発送費含む
	編集通信費		0		0	
	総会		42,210		43,000	論文賞投票および総会案内発送費
	選挙		31,020		0	役員選挙投票一式発送費
	電話代		0		0	
	インターネット		34,617		50,000	ドメイン登録、さくらネット、BizFax、メルマガ、Nifty維持料
交通費	その他		22,024		30,000	郵便切手代、宅急便代
交通費	役員旅費	35,232	0	14,000	0	
	事務局通勤費		0		0	
	その他交通費		35,232		14,000	旧事務所への郵便物の引取:27年度精算分、月1回4~9月
消耗品費		6,894	6,894	10,000	10,000	文房具、用紙、ラベル代など
維持修繕費		14,040	14,040	0	0	業務委託先へのデータ移行
雑費	手数料	32,678	32,678	35,000	35,000	振込手数料:個人会員年会費郵便振替料学会負担含む
30周年記念事業資金積立		0	0	200,000	200,000	20周年:記念号¥312,395、レセプション支出¥533,400収入¥130,000
小計		2,366,425	2,366,425	3,025,760	3,025,760	
予備費		0	0	2,612,696	2,612,696	
次年度繰越金		2,908,986	2,908,986	0	0	
合計		5,275,411	5,275,411	5,638,456	5,638,456	
特別会計		0	0	0	0	
特別会計 積立金		2,400,480	2,400,480	2,400,000	2,400,000	
総計		7,675,891	7,675,891	8,038,456	8,038,456	

## [資料 8] 平成 28 年度永年会員表彰対象者

### 永年会員表彰対象者

石井 大輔, 小田切 夕子, 河手 太士, 川原 亜希世,  
斎藤 ひとみ, 福島 丞, 福永 征夫, 古屋 保,  
松村 敦, 村井 源, 安平 哲太郎, 渡邊 隆弘  
(敬称略, 会員期間 10 年以上)

情報知識学会 第 24 回(2016 年度)年次大会 永年会員表彰式  
2016 年 5 月 15 日(日)14:00 論文賞表彰式・記念講演の終了後, 引き続き,  
筑波大学筑波キャンパス春日地区 情報メディアユニオン 2 階にて行う予定。  
表彰対象者が当日欠席の場合は後日, 表彰状を郵送。

### 第 24 回(2016 年度)年次大会

日程 : 2016 年 5 月 14 日(土) — 15 日(日)

会場 : 筑波大学筑波キャンパス春日地区 情報メディアユニオン 内

全体プログラム

14 日(土)	15 日(日)
9:00— 9:30 受付	9:00— 9:30 受付
9:30—10:45 一般セッション 1, 3	9:30—10:45 学生セッション 1
11:00—11:50 一般セッション 2, 4	10:55—12:35 学生セッション 2
11:55—12:35 総会	12:35—14:00 お昼休み
12:35—13:45 お昼休み／理事会	14:00—15:05 論文賞表彰式・記念講演
13:45—15:45 ポスターセッション	15:05—15:15 永年会員表彰式
	15:30—16:45 一般セッション 5
15:50—17:00 合同企画セッション	16:45—16:55 学生奨励賞表彰
17:30—19:30 情報交流会	16:55—17:00 閉会挨拶

## 事務局からのお知らせ

### [1] 個人会員の皆様へ、年会費納入のお願い

平成28年度の年会費未納のかたは下記の学会口座へ至急お振り込みください。

1年分の年会費は正会員8千円、学生会員・ユース会員・シニア会員は4千円です。  
過去数年分未納のかたは合計額を納入くださるようお願いいたします。

退会希望のかたは、事務局宛（[office@jsik.jp](mailto:office@jsik.jp)）メールにてご連絡ください。4月1日から、退会日までの四半期単位で計算した金額を事務局よりお知らせいたしますので、下記へ早急にお振り込みをお願いいたします。

#### (1) 振込先

- 1) 郵便振替口座 00150-8-706543 情報知識学会
- 2) ゆうちょ銀行 ○一九店(セロイチキョウ店) 当座 0706543 情報知識学会

#### (2) 会費の納入年月の確認方法

お手元に届いた学会誌の封筒の宛名ラベルには、ご自分の年会費の納入日が年度毎に西暦下2桁、月(2桁)、日(2桁)の6桁の数字で印字されています。会費未納年度には[未納]と表示されております。

お振り込みの後、事務局に通知が届き、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。

請求書が必要な方はその旨、情報知識学会事務局にメールでお知らせください。

### [2] 学会誌送付先、会員種別、メールアドレスの変更について

会員種別、学会誌送付先、メールアドレスが変わられ、変更手続きがお済でないかたはご連絡ください。年会費を納入していただいているのに学会誌やメールマガジンが届かないのでは申し訳ありませんので、変更の情報を事務局：[office@jsik.jp](mailto:office@jsik.jp)まで、メールでお知らせください。新・旧の情報を並べてお書きいただけすると確認できるので助かります。

### [3] 新規入会申込方法

入会ご希望の方は情報知識学会ホームページ<http://www.jsik.jp>から「本会について」→「入会案内」→「入会申込フォーム」に必要事項を入力・送信してください。

あるいは申込用紙をpdf形式、doc形式でダウンロードし、ご記入のうえ下記の事務局へ電子メール・FAX送信または郵送などでお願いいたします。

#### 情報知識学会事務局

〒164-0003 東京都中野区東中野4-27-37 (株)アドスリー内

FAX:050-3730-8956 E-Mail:[office@jsik.jp](mailto:office@jsik.jp) URL:<http://www.jsik.jp/>

## 情報知識学会誌 編集委員会

編集委員長	芦野 俊宏	東洋大学	天野 晃	物質・材料研究機構
副編集委員長	梶川 裕矢	東京工業大学	岩田 覚	東京大学
編集委員			江草 由佳	国立教育政策研究所
相田 満	国文学研究資料館	岡 伸人	東北大	東北大
石塚 英弘	筑波大学名誉教授	小川 恵司	凸版印刷(株)	筑波大学
宇陀 則彦	筑波大学	阪口 哲男	筑波大学	筑波大学
大槻 明	日本大学	高久 雅生	愛知大学	愛知大学
岡本 由起子	欧州情報協会	時実 象一	ロゴヴィスタ(株)	ロゴヴィスタ(株)
五島 敏芳	京都大学	長田 孝治	神奈川大学	神奈川大学
孫 媛	国立情報学研究所	中山 勇	国立情報学研究所	国立情報学研究所
田良島 哲	東京国立博物館	西澤 正己	国立情報学研究所名譽教授	国立情報学研究所名譽教授
中川 修	大日本印刷(株)	根岸 正光	同志社大学	同志社大学
長塚 隆	鶴見大学	原田 隆史	慶應義塾大学名誉教授	慶應義塾大学名誉教授
中山 伸一	筑波大学	細野 公男	和歌山大学	和歌山大学
西脇 二一	奈良大学	村川 猛彦	東京大学	東京大学
原 正一郎	京都大学	森 純一郎	産業技術総合研究所	産業技術総合研究所
藤田 桂英	東京農工大学	山下 雄一郎		
村井 源	東京工業大学			
村田 健史	情報通信研究機構			
安永 尚志	人間文化研究機構			
山本 昭	愛知大学			

(五十音順)

### ■複写される方に

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. 26, No.3 2016 年 9 月 30 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 石塚 英弘

事務局 〒164-0003 東京都中野区東中野 4-27-37 (株)アドスリー内

FAX: 050-3730-8956

E-mail: [office@jsik.jp](mailto:office@jsik.jp) URL: <http://www.jsik.jp/>

# JSTの情報サービスが、あなたの研究をサポートします

JSTは研究者情報や科学技術文献情報、特許等の研究成果情報といった、研究開発に必要とされる情報資源を提供しています。

科学技術情報の効果的、効率的な流通と活用を支える基盤を整備し、「基礎研究」から「産業化」に至る、様々なシーンの問題解決に向けたヒントを提供します。

- 新規研究テーマの探索に ■連携先・共同研究先の発掘に ■研究、技術動向の把握に
- 保有技術の新たな用途開拓に ■競合他社の分析に

豊富なコンテンツでイノベーション創出をサポートします。



J-GLOBAL は、文献、特許、研究者情報等、10種類の科学技術情報を一度に検索できるサービスです。また、これらの情報の間を関連する項目でつなぎ、そのつながりから次々と情報を取り出すことができます。新たな気づきや、今まで見つからなかつた情報の発見に役立ちます。



J-STAGE は国内の学協会等の電子ジャーナルを公開する国内最大級のプラットフォームです。約9割の論文が全文無料で閲覧できます。



研究人材のキャリア形成・能力開発に役立つ情報を提供するポータルサイトです。産学官の求人情報、eラーニング教材等のキャリア支援コンテンツやイベント情報を提供します。



researchmap は研究者の双方向コミュニケーションを実現した日本の研究者総覧です。研究分野や所属機関、キーワードで研究者を検索し、プロフィールや業績などの情報を閲覧できるため、産学連携や共同研究の候補となる研究者を見つけることができます。



# *Journal of Japan Society of Information and Knowledge*

## ~~~~~ **Contents** ~~~~~

### **Foreword**

Takashi NAGATSUKA ..... 249

### **Research Papers**

Proposal of Research Activity Classification Based on Time-series Bibliometrics

Takahiro KAWAMURA, Yasuhiro YAMASHITA, Katsuji MATSUMURA ..... 251

A Subject Information Sharing System Based on Linked Data and FRSAD Model

Kosuke TANABE, Yuka EGUSA, Masao TAKAKU ..... 260

Study of Citation Classification Scheme on Academic Articles

Daisuke SHIBATA, Fuyuki YOSHIKANE ..... 277

### **Information**

Minutes of General Assembly 2016 ..... 297

Others ..... 309

**情報知識学会誌** 第 26 卷 3 号 2016 年 9 月 30 日発行

編集兼発行人 情報知識学会 〒 164-0003 東京都中野区東中野 4-27-37 (株)アドスリー内

E-mail : office@jsik.jp

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

(振替 : 00150-8-706543)

学術刊行物 ISSN 0917-1436