

## 非線形問題分野データベースの構築と研究動向分析

新美 礼彦<sup>†</sup> 小西 修<sup>†</sup>

<sup>†</sup> 公立はこだて未来大学 システム情報科学部  
〒041-8655 北海道函館市亀田中野町116-2  
E-mail: †{niimi,okonishi}@fun.ac.jp

あらまし 本論文では、文献書誌情報データベースを構築し、そのデータベースからデータマイニングを用いることにより、研究分野の抽出と、研究動向分析を行う手法を提案する。提案した手法を電子情報通信学会 非線形研究会の信学技法をもとに、データベースを構築し、頻度分析、相関分析を元にキーワード解析を行い、その結果について考察した。解析結果から、ある程度の研究分野をうかがうことができることを確認した。研究動向の解析に関しては、解析途中である。

キーワード データマイニング、非線形問題、キーワード解析、頻度分析、相関分析、研究動向調査

## Construct Database and Research Trend Investigation Related with the Field of Nonlinear Problem

Ayahiko NIIMI<sup>†</sup> and Osamu KONISHI<sup>†</sup>

<sup>†</sup> School of Systems Information Science, Future University-Hakodate  
116-2 Kamedanakano-cho, Hakodate-shi, Hokkaido, 041-8655 Japan

E-mail: †{niimi,okonishi}@fun.ac.jp

**Abstract** In this paper, we propose an approach which contains with constructing a bibliography information database, extracting the fields of research, and researching trend of them, using data mining. To apply our approach to IEICE Technical Report (nonlinear problem society), the database was constructed based on its report, keywords were analyzed using the frequency analysis and the association analysis, and we discussed about the result. We could extract some field of research from the result.

**Key words** data mining, nonlinear problem, keyword analysis, frequency analysis, association analysis, research trend investigation

### 1. はじめに

本論文では、文献情報から研究動向を調査する手法について検討した。

文献情報から研究動向を調査する方法は、以前より行われている。しかし最近、論文誌や学会誌、研究会報告書などがオンライン化され、ネットワーク上から手軽に検索などの利用が行えるようになった。また、パソコンの性能の向上により個人で手軽にデータ解析が行えるようになった。このような流れから、データマイニングを用いて文献書誌情報から研究動向を探ることができないか考えた。

そこで、文献データベースの構築から調査までを含めた研究動向調査の手法について検討する。研究動向調査の手順としては、次の流れを考える。まず、文献情報から文献書誌情報データベースを構築する。構築したデータベースに対してテキスト

マイニング手法を適用し、文献情報からキーワードを抽出する。そして、抽出したキーワードとと書誌情報のキーワードからキーワード解析による研究動向調査についてを検討する。本論文では、既存の文献データベースを利用するのではなく、文献書誌情報データベースの構築から検討することにより、より幅広い分野への適用が可能となると思われる。

本論文では、対象データとして、カオス・非線形文献データベースを取り上げ、カオス・非線形文献データベースの構築、キーワードによるカオス研究分野、研究動向の調査をおこなった。なお、タイトルに研究動向調査とあるが、本論文では、実際の研究動向調査までは行っておらず、データベース構築とデータベース解析に関する検討を行い、実際にデータベースを構築し、キーワードについて、頻度と相関による解析を行うところまでしか行っていない。年ごとの研究動向解析については、解析途中である。

## 2. データベース構築

文献情報からのデータマイニングを行うために、データベースを構築する。全文をデータベースに登録するのではなく、文献書誌情報のみをデータベースに登録することにした。

データベースは構築のしやすさと構築後の拡張性のため、関連データベース (RDB) と XML (eXtensible Markup Language) を組み合わせておこなった。入出力は XML を用い、データの保存に関しては RDB を用いることにより、入出力に自由度を持たせコンピュータ、人間とも可読可能な情報にでき、かつ SQL ベースによる高速なデータ検索が可能となる。

文献データベース構築に関して、自動で論文情報を検索するシステムが提案されている。[1] XML ベースの入出力を行うことにより、これらほかのシステムとの連携も容易になる。

インタフェース部分は Java の Servlet で構築する。Servlet を通じて XML と RDB とのやり取りを行うシステムとなる。システム構成の概略を図 1 に示す。

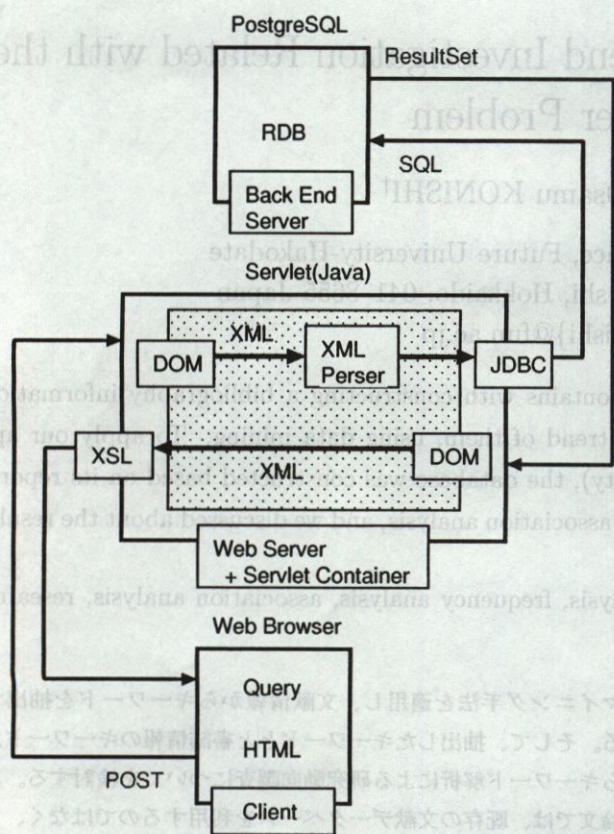


図 1 システム構成

データベースへの入力する書誌情報は、Dublin Core の 15 elements、電子情報通信学会 研究会データベース、国立情報学研究所 学会発表データベース、BiBTeXなどを参考にして決定した。入出力に合わせ、RDF (Resource Description Framework) により記述した。RDF とは、メタデータなどのリソースの相互関係を記述するための規格の 1 つである。

### 2.1 Dublin Core

文献書誌情報を検討する際、特に Dublin Core のメタデータセットを参考にした。Dublin Core のメタデータセットはリ

ソースを記述するメタデータとして、15 要素が提案されている。[2] (表 1 参照)

表 1 Dublin Core の 15 elements

Title	リソースの名前
Creator	リソース責任者
Subject	キーワード、
Description	要約、目次
Publisher	提供者
Contributor	協力、貢献者
Date	作成日、公開日
Type	カテゴリ、機能、分野
Format	メディアタイプ、量、サイズ
Identifier	リソースへの参照 (URI, ISBN)
Source	元リソースへの参照
Language	記述言語
Relation	関連リソースへの参照
Coverage	リソースの範囲、対象、場所
Rights	権利

Dublin Core の 15 要素は、RDF での記述が可能であるので、XML で扱いやすいという利点がある。今回は、15 要素を対象文献データに合うように一部変更して用いた。

## 3. データベース解析

文献書誌情報データベースからデータ解析を行う際、データマイニングが適用できる。データマイニングの代表的な分析手法として、頻度分析、相関分析、クラスタリングなどがある。これにより、研究分野のキーワード抽出や研究分野の広がり of 調査を行うことができる。また、キーワード群を時系列的に扱うことにより、研究動向の調査として捕らえることもできる。さらに、キーワードの相関などから、用語の整理を行うことができる。

タイトルや章タイトルなどは、自然言語でかかれている。そこで、タイトルや章タイトルに対して自然言語解析の手法が使える。タイトルに対する形態素解析やそれを用いてのキーワード抽出、頻度や品詞を元にした解析、相関の高いキーワードの抽出などが考えられる。

### 3.1 キーワード抽出法

キーワードがつけられていない論文からキーワードを抽出するとき、タイトルや章タイトルから抽出することを考える。日本語の場合、英語のように単語の区切りが明確でないため、形態素解析などを行う必要がある。

本研究では、タイトル、章タイトルから chasen による日本語構文解析により、単語を切り出した。chasen とは日本語形態素解析システム『茶筌』に含まれている形態素解析を行うコマンドである。[3] 文章からのキーワード抽出法として、さまざまなものが提案されている。提案されているキーワード抽出法を大きく分けると、形態素解析を用いるもの、形態素解析を用いないもの、文章の構造をもとに解析するものなどがある。[4] 本論文では、主に形態素解析を用いるものを検討した。

#### 3.1.1 形態素解析

形態素解析とは、入力文を言語学的に意味をもつ最小単位で

ある形態素に分割し、各形態素の品詞を決定するとともに、活用などの語変形化をしている形態素に対しては原形を割り当てることである。[3] 例えば、「発表会を行いたい。」という文で形態素解析を行うと、表2のように分析される。

表2 形態素解析の例

要素単語	原形	品詞
発表	発表	名詞-サ変接続
会	会	名詞-接尾-一般
を	を	助詞-格助詞-一般
行い	行う	動詞-自立
たい	たい	助動詞
		記号-句点

形態素解析で分割された単語を要素単語という。要素単語に分けることにより、頻度解析や特定品詞へのフィルタリングが行えるようになる。

### 3.1.2 出現頻度による抽出

形態素解析で分割された各要素単語の出現回数(頻度)を調べる。出現頻度の高い要素単語をキーワードとして抽出する。出現頻度の高い要素単語をキーワードとして抽出するため、どんな文章からも最適なキーワードを抽出しやすい手法である。しかし、助詞などのキーワードとして適切でない語を抽出する傾向があるため、抽出後のフィルタリングが重要になる。単純な頻度を使わずに、 $tf \cdot idf$  を用いることもできる。これは、以下の式で定義される。

$$\text{スコア} = tf \times idf \quad (1)$$

ただし、

$tf$ : あるキーワードがその対象文章中に含まれる出現回数 (Term Frequency)

$idf = \log(N/n)$ : (Inverse Document Frequency)

$N$ : 全文章数

$n$ : そのキーワードを含むファイル数

$tf \cdot idf$  法を用いることにより、多数の文章に多く含まれる一般的なキーワードの重要度を下げ、特定の文章中に多く含まれるキーワードの重要度をあげることができる。つまり、あるキーワードが多くのある文章中に現れる(その文章中では一般的な)キーワードであった場合は  $idf$  が小さな値となり、逆に特定の文章中にしか現れない場合には  $idf$  が大きな値となり、出現頻度 ( $tf$ ) に重み ( $idf$ ) をつけて評価することができるようになる。

### 3.1.3 連続名詞の抽出

情報検索の世界では名詞概念をキーワードとして抽出する傾向が強い。[5] 一般的には、形態素解析を用いて名詞を抜粋し、キーワードの抽出をおこなう。「発表会を行いたい。」という表現を形態素解析を行った結果、「発表」、「会」、「を」、「行う」、「たい」の5つの要素単語に分割される。「を(助詞)」、「行う(動詞)」、「たい(助動詞)」は、名詞ではないのでキーワードとして抽出せず、この場合「発表」、「会」といった名詞をキーワードとして抽出する。ただし「発表」、「会」といった単位では、頻度は高いが具体性が低いので、「発表会」という、長い

単位で語句を抽出することにより語の具体性を上げることができ

### 3.1.4 N-グラム

構文解析を行わない方法の1つとして、N-グラム(n-gram)法がある。N-グラムは長い文字列から部分文字列を取り出す方法で、Nには2や3などの数をとることができる。N-グラムのアルゴリズムでは1文字ずつずらしながら、連続するN文字を取り出し、取り出した文字列の出現頻度を調べ、その集合の中で出現頻度の高い語をキーワードとして抽出するというものである。[5] あらかじめ文章に形態素解析による単語分けを行う必要がなく、任意の数の文字数を設定することができる。

しかし、単語分けを行わないで解析すると、単語の一部分を含んだ文字列を大量にキーワードとして抽出する恐れがある。これを改善するために、本論文では形態素解析を行い、要素単語に分けた後で、その要素単語の連続を調べる手法を検討した。

例えば、「非線形問題分野」という表現から2-gram(2文字連続)を全て取り出すと、(非線)(線形)(形問)(問題)(題分)(分野)を取り出すことができる。しかし、キーワードに適さない語が多いので、この作業を形態素解析により(非線形)(問題)(分野)と単語分けした後で行い、連続する2語を取り出すと、(非線形問題)(問題分野)が得られる。

### 3.1.5 関連ルール

文章中に現れる文字や単語の関連から、キーワードを抽出することが考えられる。これを関連ルールと呼び、ルールはいくつかの文字(または単語)からなり、どれだけ同時に現れやすいのか(関連があるか)が評価対象となる。関連ルールを高速に抽出する手法として、aprioriアルゴリズムがある。[6] 関連ルールの探索では、支持度(support value)と確信度(confidence value)という2つの指標を用いて関連ルールを評価する。本論文では、関連ルールの支持度( $sup$ )は全データに対する構成要素が含まれる割合、確信度( $conf$ )はある構成要素が含まれたときに他の構成要素が含まれる割合の平均であると定義する。

関連ルール探索は、N-グラムを用いたアルゴリズムと同様に、形態素解析を行わなくてもキーワードを抽出することが可能である。しかしこれも、単語の一部分のみを抽出する可能性を減らすため、本論文では形態素解析を行った後に要素単語間の関連ルールからキーワードを作成することを考える。

例えば、「カオス」と「分岐」に注目したとすると、全データの中で「カオス」と「分岐」というキーワードを同時に含む割合が支持度となり、「カオス」を含むときに「分岐」を含む割合と「分岐」を含むときに「カオス」を含む割合の平均が確信度となる。なお、ルールを構成している単語が1語の場合は支持度と確信度は等しくなる。

### 3.2 フィルタリング

単に文章からキーワードを切り出しただけでは、「てにをは」や「数字」などキーワードに適さない語が含まれてしまう可能性がある。また、「行う」、「行い」など、活用によって語尾が変化する語もある。これらを除去したり、適切な形に替える必要があり、本論文ではこの作業をフィルタリングと定義している。フィルタリングでは、このほかにも同じ意味の違う単語を

統一するなどの作業を行うことがある。フィルタリングを行うことにより、キーワードの質を高めることができる。

### 3.3 研究動向調査

文献書誌情報として入力されているキーワードやタイトルなどから抽出したキーワードを用いて、キーワードの出現頻度に着目して解析することを考える。よく使われるキーワードはその分野の中心的なキーワードとして考えることができる。

また、1つの論文中のキーワードに対して、相関ルールを考へることにより、同時に使われやすい単語を抽出することも考えられる。例えば、カオスというキーワードとニューラルネットワークというキーワードが同時に使われることが多ければ、この2つのキーワード間に深いつながりがあると考えることができる。

さらに、抽出したキーワードを用いて、研究動向を調査することを考える。まず、年ごとの頻度分析の結果から、研究会としての研究動向を調査する参考になると考えられる。また、特定のキーワードについて、年ごとの頻度分析をおこなうことにより、そのキーワードに関係する研究分野の研究動向を調査する参考になると考えられる。

また、キーワードをクラスタリングすることにより、研究分野の広がりを検討することが可能となる。年ごとのクラスタリング結果から研究動向の広がりを把握することが可能であると考えられる。

これらの解析をシステムに組み込むことにより、現在までの研究動向の把握や、新しい研究分野の開拓の支援が容易にできるのではないかと考えている。研究を進めるための道具として、このような研究動向の解析ツールは非常に有用であると考えられる。

## 4. カオス文献情報データベース

解析対象のデータベースとして、カオス文献データベースを構築した。文献書誌情報データベースとして、RDBをベースとし、入出力をXMLベースにした。入力する文献書誌情報はDublin Coreを参考に決定した。データベースの入出力はServletを用いて実装した。データとして、電子情報通信学会非線形研究会(信学技法)の1959年から2001までを用いた。電子情報通信学会ではすでに、文献データベースがインターネット上で公開されているが、今回の解析に使うには、必要な項目が少ないこと、登録されている論文数が少ないことなどから、独自に構築した。

古い論文でアブストラクト、キーワードがついていないものが多数あったため、アブストラクトの代わりに各章のタイトルを入力することにした。解析では、タイトル、章タイトルからキーワードを抽出し、書誌情報のキーワードと合わせて解析を行った。データベースに登録する際、「はじめに」、「結論」などのあきらかにキーワードにならない章タイトルは入力していない。データ入力の際、いくつかの論文でキーワードも章タイトルもないものがあつた。これに関しては、図・表見出しを章タイトルの代わりに入力した。

データベースのサイズについては表3、登録した項目につい

ては表4参照。

表3 非線形研究会データベース

電子情報通信学会 非線形研究会 (信学技法)	1959年から2001年
論文数	2315
キーワード(日本語)	5881
キーワード(英語)	5953
章タイトル	14395
切り出しキーワード	9439

表4 入力した文献書誌情報

項目	対応する Dublin Core Element
タイトル(日・英)	title
キーワード(日・英)	subject
章タイトル	description
著者名(日・英)	creator
所属(日・英)	creator
文献番号	identifier
雑誌名	source
雑誌番号(Vol, No)	source
ページ数	source
研究会名	contributor
学会名	publisher, rights
発表日	date
発表言語	language
分類(Proceedings)	type

データマイニングをを利用して、論文を整理する方法として、バイオインフォマティクス分野をはじめとして、様々な分野で行われている。[8]しかし、カオス・非線形分野では分野の広がりがあり、電気、数学、物理、神経系、画像、信号処理などの複数の分野にまたがっているため、解析しにくいと考えられる。本論文では、幅広い発表が行われている非線形研究会を取り上げた。この研究会では、多くの異なる分野の研究が含まれており、理論中心や、応用中心、メカニズム、見方、利用、式、理論中心、シミュレーション中心、実世界指向など様々な切口で捕らえることができる。

## 5. 解析結果と考察

タイトル、章タイトルからchasenによる日本語構文解析により、単語を切り出した。抽出したキーワードと書誌情報のキーワードを合わせて、9439のキーワードを対象に解析を行った。解析には、日本語キーワード、英語キーワード、日本語タイトルからのキーワード、英語タイトルからのキーワード、章タイトルからのキーワードのそれぞれの組み合わせをつくりおこなった。キーワードには簡単なフィルタリングを行い、キーワードになりにくいものははずしてある。

頻度解析では、上位にカオス、ニューラルネットワーク、分岐、回路、モデルなどのカオス・非線形の研究分野を表しているキーワードが抽出された。(頻度解析による日本語キーワードの解析結果の一部を表5に示す。)

連続名詞の頻度解析による結果を表6に示す。ほぼ、頻度解析による結果と同じような結果になっている。

N-グラムに関して、単語をベースにnを取りうる最大まで増

やしながら解析した。ここでは、n-gram とは連続する n 語の組み合わせの解析である。抽出キーワードは、ほぼ連続名詞頻度による解析と同じものが抽出された。(n-gram 解析による日本語キーワードの解析結果の一部を表 7 に示す。)

表 7 N-グラム解析による結果 (上位 20)

tf.idf	keyword	tf.idf	keyword
452.27	カオス	127.94	同期
323.94	ニューラルネットワーク	126.29	フラクタル
301.90	分岐	122.57	結合発振器
194.26	法	116.36	非線形回路
171.39	方程式	105.70	回路
143.39	制御	102.77	同期現象
143.39	ヒステリシス	99.80	連想記憶
143.39	カオス制御	99.80	非線形
142.93	発振器	94.75	ニューラルネット
135.78	分岐現象	93.73	学習

表 5 頻度解析による結果 (上位 20)

tf.idf	keyword	tf.idf	keyword
580.94	カオス	310.43	同期
448.39	回路	287.57	振動
415.14	分岐	267.88	モデル
412.44	非線形	266.80	関数
374.83	結合	263.61	現象
332.89	問題	256.35	周期
331.17	方程式	255.72	写像
323.94	ニューラルネットワーク	247.72	線形
316.30	制御	236.31	システム
314.53	発振器	226.83	安定

章タイトルから抽出したキーワードを用いた頻度解析の結果を表 8 に示す。章タイトルにつけやすい単語が抽出されているが、キーワードとして使えるものが埋もれてしまっている。論文のスタイルのようなものはうかがえるが、研究分野に直結したようなキーワードを抽出するためには、フィルタリングを工夫する必要がある。

表 8 章タイトルからの頻度解析による結果 (上位 20)

tf.idf	keyword	tf.idf	keyword
303.94	シミュレーション	97.96	circuit model
223.10	回路モデル	96.38	ポアンカレ写像
185.24	数値例	90.02	学習アルゴリズム
178.72	回路方程式	85.02	計算機シミュレーション
164.84	モデル	78.37	安定性
135.69	基礎方程式	75.02	エネルギー関数
130.26	アルゴリズム	70.98	分岐図
116.75	基本方程式	69.05	平衡点
106.22	問題の記述	69.05	システム
98.99	応用例	66.88	分岐現象

タイトル、章タイトルなどからの抽出キーワードを含んだキーワードによる頻度解析による結果を表 9 に示す。日本語、英語をまとめて解析しているため、日本語の用語を英語に言い換えたものが比較的近いスコアで抽出されている。研究分野を調べる際には、日英の言い換えなどをフィルタリングしてまとめる必要がある。

表 9 抽出キーワードを含んだ頻度解析による結果 (上位 20)

tf.idf	keyword	tf.idf	keyword
564.07	chaos	186.89	数値例
554.89	カオス	183.20	回路方程式
377.79	ニューラルネットワーク	180.09	neural networks
359.74	bifurcation	178.41	モデル
318.01	分岐	177.17	hysteresis
315.79	シミュレーション	165.24	ヒステリシス
294.19	neural network	143.39	associative memory
227.84	回路モデル	140.16	結合発振器
191.58	synchronization	138.96	非線形回路
187.92	分岐現象	136.76	基礎方程式

全体的に、書誌情報からの解析のほうが、キーワード抽出

処理を含んでいないぶん、きれいなキーワードが抽出された。キーワード抽出法や、その後のフィルタリングなどを検討する必要がある。特に、回路や分岐など、ほかのキーワードの一部として含まれているキーワードが多数出力されたので、これの扱いをどうするか検討する必要もある。

抽出したキーワードに関して、ほとんど同じ意味の用語 (return map, return plot etc.) や、分野による用語の違いなど単純に頻度によるフィルタリングで処理できないものがあることが確認できた。解析には、その分野の専門家との連携が不可欠であるといえる。

また、データベース中でタイトルなどに数式が多く使われていた。カオス・非線形分野では数式はキーワードとして重要であると考えられるので、ここまでの解析では、数式は TeX 表記を用いてデータベースに登録し、1 単語として扱った。似たような数式があったことから、数式の取り扱いについても検討する必要がある。

次に、関連ルールによる分析を行った。解析データに簡単なフィルタをかけて、キーワードになりにくい単語をはずして解析した。また、数式に関しては、他のキーワードと関連が取りにくいいため解析データからははずしてある。日本語キーワードに対する関連分析の結果を表 10 に示す。(表中の *sup* は全てのキーワードを含む割合、*conf* はキーワードの組み合わせによってできるルールの確信度の平均を元にしてしている。)

表 10 キーワード群の関連分析による結果 (上位 20)

<i>sup</i> (%), <i>conf</i> (%)	keyword
(1.2, 63.3)	法, ホモトピー
(1.2, 24.1)	カオス, 分岐現象
(1.1, 22.6)	カオス, 発振器
(1.3, 25.4)	カオス, ヒステリシス
(1.6, 34.0)	カオス, 同期
(2.1, 40.0)	カオス, 制御
(1.8, 15.1)	カオス, 分岐
(2.8, 20.3)	カオス, ニューラルネットワーク
(1.2, 1.2)	bvp
(3.1, 3.1)	制御
(1.0, 1.0)	線形計画法
(3.0, 3.0)	ヒステリシス
(1.3, 1.3)	巡回セールスマン問題
(2.0, 2.0)	回路
(1.0, 1.0)	カオス同期
(1.8, 1.8)	連想記憶
(1.6, 1.6)	モデル
(1.6, 1.6)	対称性
(1.3, 1.3)	krawczyk
(1.6, 1.6)	ダイナミクス

表 10 は、解析結果のうち、ルールの長いものを抽出したものである。高頻度のカオスというキーワードと、分岐現象、発振器、ヒステリシス、同期、制御、分岐、ニューラルネットワークを含むルールが高い相関をしめす結果となった。これにより、対象の研究会がカオスというキーワードを中心に分岐現象や制御、ニューラルネットワーク等を研究対象にしていることがうかがえる。

年ごとの研究動向解析については、解析途中である。

## 6. おわりに

本論文では、文献情報から研究動向を調査する手法について検討し、検討した手法をカオス・非線形関係分野の文献書誌情報データベースに適用した。文献から、文献書誌情報データベースを構築した。つぎに、構築したデータベースのタイトル、章タイトルからキーワードを抽出した。抽出したキーワードと書誌情報のキーワードから頻度分析と相関分析を用いてキーワードベースの解析を行い、それに対する考察を行った。研究動向の解析に関しては、解析途中である。

解析結果から、ある程度の研究分野をうかがうことができることが確認できた。また、解析には用語の統一が重要であるということが確認できた。

現在、データベースのクリーニングを行っており、それと並行して、他のデータマイニングの適用を検討している。クリーニングに関しては、用語の統一、キーワードとして不適切な用語の削除などである。キーワードレベルではなく、キーワード群や、研究分野、研究動向に関して、考察できるように実験を進めている。

### 謝辞

本研究に対して、データ提供と助言をいただいた公立はこだて未来大学の田目完亮先生、高橋信行先生、小西啓治先生、斎藤朝輝先生に感謝致します。

### 文 献

- [1] 高田 伸彦, 田村 武志, 大沢 一彦: XML による Web 上の論文検索システムの構築, 電子情報通信学会論文誌 D-I, Vol.J84-D-I, No.6, pp.650-657, 2001.
- [2] 永田 昌明, 平 博順: テキスト分類 - 学習理論の「見本市」, 情報処理 Vol.42 No.1, pp.32-37, 2001.
- [3] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 浅原 正幸: 日本語形態素解析システム『茶釜』version 2.0 使用説明書第二版, 1999.
- [4] 市村 由美, 長谷川 隆明, 渡部 勇, 佐藤 光弘: テキストマイニング - 事例紹介, 人工知能学会誌 Vol.16 No.2, pp.192-200, 2001.
- [5] 那須川 哲哉, 河野 浩之, 有村 博樹: テキストマイニング基盤技術, 人工知能学会誌 Vol.16, No.2, pp.201-211, 2001.
- [6] Rakesh Agrawal, Ramakrishnan Srikant: Fast Algorithms for Mining Association Rules, the 20th International Conference on Very Large Databases, Santiago, Chile, September 1994: 32 pages, 1994.
- [7] Dublin Core Metadata Initiative (DCMI), <http://dublincore.org/>
- [8] 辻井 潤一: ゲノム情報学と言語処理, 情報処理 Vol.43 No.1, pp.29-35, 2002.