

## 個人文書から抽出した語彙の意味関係に基づく Web 情報検索

大 島 裕 明<sup>†</sup> 小 山 聡<sup>†</sup> 田 中 克 己<sup>†</sup>

ローカルコンピュータには非常に多くの個人的な文書が存在する。それらは人間にとっては理解可能であるが、コンピュータがそこに含まれる知識を抽出して利用することは行われていなかった。本稿では、個人的な文書として特に、これまで扱ってこなかった電子メールに着目する。我々は、これまで個人的な文書とその分類構造から、語彙の意味関係を抽出する研究を行ってきた。しかし、電子メールは、ファイルに対するディレクトリ構造のような分類構造が存在しないことが多い。そこで、電子メールが持つメタ情報を利用して階層構造を仮想的に取得することによって、これまでの手法を適応させる。そして、抽出された個人の知識を利用してウェブ情報検索の個別化を行う手法について提案を行う。

### Web Search Personalization based on Semantic Relationships between Terms Extracted from Personal Documents

HIROAKI OHSHIMA,<sup>†</sup> SATOSHI OYAMA,<sup>†</sup>  
and KATSUMI TANAKA<sup>†</sup>

This paper proposes a method to extract semantic relationships between terms which appear in documents in a personal computer, and its application to Web search personalization. Nowadays, a lot of personal information is stored in a personal computer. We have proposed the method to extract the semantic relationships from general documents classified in directory structures. In this paper, we treat email messages as personal information resources. As email messages are usually not classified, some classification structures are need to be extracted with using meta data in email messages. Then, we proposes Web search personalization using the extracted personal knowledge.

#### 1. はじめに

現在、コンピュータ上では様々な個人的な文書が取り扱われている。自分で作成した文書や、興味がある事柄に関する文書、さらに、電子メールなどによって、日々多くの文書がコンピュータ上に保存され続けている。今後ますますそのような個人的な文書が増え続ける状況においては、それらをうまく活用する必要性が出てくると考えられる。

例えば、デスクトップサーチと呼ばれるものは、ローカルコンピュータ上の個人的な文書やその他のファイルをウェブの検索エンジンのように検索できるようにするツールであり、Google、Microsoft、Yahoo!などが次々と発表している。これらは、ファイルへの便利なアクセス方法を提供し、ファイルを利用しやすくしていると言える。しかし、ファイルの中にある情報を

利用することは行われていない。

個人的な文書にはそのユーザに関する非常に多くの情報が入っていると考えられる。例えば、誰かが他人のコンピュータ上の文書を読んだ場合、その人がどのようなことに興味があるのか、どのようなことを知っているのか、ある事柄についてどのような意見を持っているのか、などを理解するのに十分な情報が存在していると考えられる。

我々はこれまで、文書とその分類構造に着目して、このような個人的な知識を抽出し利用すること研究を行ってきた<sup>1)~3)</sup>。本稿では、これまで扱ってこなかった電子メールにも着目し、Web 情報検索におけるパーソナライゼーションを実現する手法について提案を行う。

以下、2 章で関連研究について、3 章で語彙の意味関係の抽出について、4 章で電子メールのメタ情報を利用した分類構造の取得について、5 章で Web 情報検索の個別化について、6 章でまとめと今後の課題について述べる。

<sup>†</sup> 京都大学大学院情報学研究所社会情報学専攻  
Department of Social Informatics,  
Graduate School of Informatics, Kyoto University

## 2. 関連研究

すでにいくつかの研究において、個人的な情報からの知識を生成するような研究が行われている。それらについて述べる。

Haystack<sup>4),5)</sup> は MIT が開発した、個人的な情報管理システムである。扱う情報は、e-mail やカレンダー、文書、Web ページなど多岐にわたり、それらを一括して RDF で管理することができる。本研究では、Haystack と同様に、個人的な情報を管理するが、蓄積された情報そのものをより効果的に利用しようとする Haystack とは異なり、蓄積された情報から個人の知識を抽出し、それを様々なことに利用することを目的としている。

WorkWare++<sup>6),7)</sup> は富士通研究所が開発した、会社などのグループで用いられるビジネス文書の蓄積と再利用のための情報管理システムである。さまざまな文書が登録され、その登録時には時間などのメタ情報が自動的に付加される。また、人やイベントの情報も同時に管理されている。ユーザは蓄積されたメタ情報を元に、ある研究分野に関してどのような技術が蓄積されているかや、ある事柄を知っているような人が誰であるかなどの情報を取得可能である。本研究では、文書群から知識を抽出するという点では共通しているが、WorkWare++で行っている、グループによる情報共有や、蓄積されている情報の利用を促進するのではなく、個人の知識を新たな情報獲得などに活かすことを目的としている。

Hyperclip<sup>8)</sup> は NTT が開発した、知識流通プラットフォームである。ユーザが利用した複数のコンテンツの間の関係を表現することができ、そこで作成された RDF をピア・ツー・ピアネットワークで共有することによって、ある文書と関連する文書を検索することができるようになる。Hyperclip で検索できる文書はピア・ツー・ピアネットワーク上の誰かによってメタ情報が付加されたものである。本研究では、文書群から個人の知識を抽出してそれを利用することを目的としており、ユーザが RDF という利用可能な形の知識を作成する Hyperclip とは異なる。

湯川ら<sup>9)</sup> は、個人が所有する文書に出現する単語の隣接度合いから、それぞれの単語同士の関連度合いを表す概念ベース、パーソナル・リポジトリを個人ごとに作成した。ユーザがコミュニティのピア・ツー・ピア型システムの他の人が保有する情報を検索するときには、エージェントが検索キーをパーソナル・リポジトリによって拡張し、他人のパーソナルリポジトリ

内でどのような情報が検索結果として適当であるかを判断することが可能になる。本研究と同様に、個人が所有する文書をもとに個人の知識を概念として表しているが、その目的がピア・ツー・ピア型のネットワークで文書を共有することであり、より利用範囲の広い知識を作成することを目的とする本研究とは異なる。

これらの研究は、ある特定の環境やコミュニティの中で利用可能な知識を作成しようとしており、より広い範囲で利用しようとしている本研究とは異なるものである。

## 3. 語彙の意味関係の抽出

### 3.1 意味関係抽出の考え方

我々はこれまで、ローカルコンピュータ上に保存された個人的な文書と、それが分類されているディレクトリ構造から、個人的な知識を抽出することを行ってきた。本研究においてはそれを電子メールに対して適用する。本章ではまず、文書とその分類構造から、一般的にどのように個人的な知識を抽出するかということについて、考え方とアルゴリズムについて述べる。

個人がローカルコンピュータ上に保存している文書は、ディレクトリなどの階層構造の中で分類管理されていることが多い。どのような文書を持っているかということや、どのような分類構造を使っているか、どのように実際に分類をしているかということは、個人の知識や考え方を反映したものであると考えることが可能である。

我々はそのような個人的な文書と分類構造から個人的な知識として、ユーザが文書中に現れる語彙どうしの間でどのような関係性をとらえているか、ということ抽出することを行ってきた。このような語彙の関係性を、語彙の意味関係と呼ぶ。文書群から得られる最も一般的な語彙どうしとの関係として共起関係があげられるが、文書の分類構造を利用することによってより多くの関係性が抽出できる。

そのために注目したのは、ディレクトリ構造の各階層における語彙の出現の偏りである。

図1は、ディレクトリ A における語彙の出現の偏りを模式的に表したものである。このディレクトリには、B, C という二つのサブディレクトリが存在しており、多くの文書の中には、様々な語彙が出現している。その語彙の分布を見ても、「Classic」と「Bach」は偏り無く広く分布しており、「Piano」と「Violin」は特定の場所だけに偏って分布していることがわかる。

この時、2つの語彙の間にある関係性を考えると、例えば、「Classic」は「Piano」に対して広く使われる語で

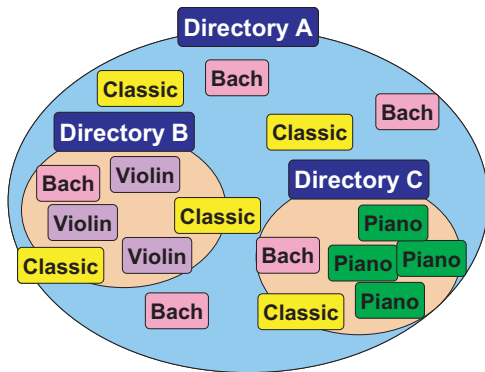


図 1 あるディレクトリにおける語彙の偏りの例

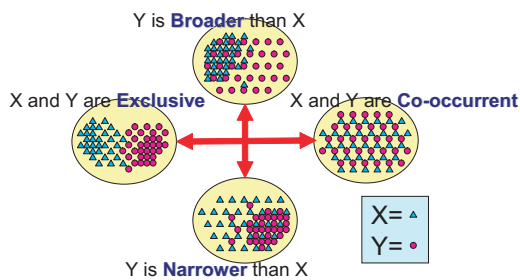


図 2 語彙 X に対する語彙 Y の関係

ある』や、『Classic と Bach は共起して使われる語である』というようなことを理解することができる。

このような関係性は、2つの語彙がどのように分布しているかによって、2つの関係軸上で考えることができる。すなわち、

- 広域的 – 狭域的
- 共起的 – 排他的関連

である。

先ほどの例では、『Classic は Piano に対して広く使われる語である』と考えられたが、この時に、Classic は Piano に対して広域的な語である、とするのである。また、この時、同時に Piano は Classic に対して狭域的な語である、といえなくてはならない。このような関係が、2つの語の分布の仕方によって求めることができるのである。

図 2 は、2つの軸における語彙の出現の偏りを表したものである。ここでは、三角で表された語彙 X に対して、丸で表された語彙 Y がどのような関係であると考えられるかを示している。

まず、2つの語彙の出現の偏りに違いがあった場合は、「広域的」な関係と「狭域的」な関係が同時に表れる。

- (1) 語彙 X の偏り大、かつ、語彙 Y 偏り小ならば、Y は X に対して広域的

- (2) 語彙 X の偏り小、かつ、語彙 Y 偏り大ならば、Y は X に対して狭域的

2つの語彙の出現の偏りが、両方とも同じような偏りの場合には、「共起的」な関係か「排他的関連」という関係を考えることができる。

- (1) 語彙 X と Y の偏り小ならば、X と Y は互いに共起的
- (2) 語彙 X と Y の偏り大ならば、X と Y は互いに排他的関連

このように、あるディレクトリ、つまり文書が分類された所における語彙の出現の偏りに着目すると、語彙の間に関係性を求めることができる。これは、文書が分類されているあらゆる所において求めることができるものである。当然、その時々によって得られる関係性も違ってくる可能性がある。例えば、あるディレクトリ D において X が Y に対して広域的と判断されたとする。しかし、別のディレクトリ D' に着目すると、X は Y に対して狭域的と判断されるかもしれない。よって、語彙の偏りから関係性を求めるときには、どの部分を情報源として用いるかと言うことが重要になる。

また、対象とするディレクトリの中でそれぞれの語彙が出現する回数が多ければ多いほど関係性が深いと言うことができる。

### 3.2 意味関係抽出のアルゴリズム

これまで述べた考え方を基に、語彙の意味関係を抽出するための基本的なアルゴリズムを考えると、以下のようになる。

- (1) 対象とするディレクトリ D における、語彙 X、Y の出現の偏りを求める
- (2) 得られた偏りをもとに、おのおのの関係においてどの程度の関係性があるかを求める。
- (3) 語彙 X、Y の出現頻度を求める。
- (4) 計算された値を基に、語彙間の意味関係を求める。

### 3.3 実装

上記アルゴリズムに従う実装には様々なものが考えられる。ここでは、現在我々が行っている実装について述べる。

まず、説明に用いる記号について説明する。

- D は対象ディレクトリとする。
- X, Y は対象語彙とし、この 2 語の間の意味関係を求める。
- $Sub_i(D)$  は D または D 以下のサブディレクトリのうちの 1 つを表す。

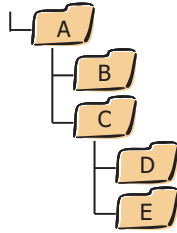


図3 ディレクトリ構造の例

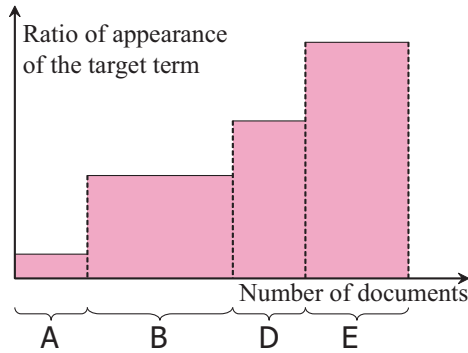


図4 各ディレクトリにおける語彙の出現割合の例

- $V_{\text{関係名}}(X, Y, D)$  は  $D$  における  $X$  と  $Y$  の各関係（広域的, 狭域的, 共起的, 排他的関連）における評価値とする。
- $Num(D)$  は  $D$  に直接存在している文書の数とする。
- $Num(X, D)$  は  $D$  に直接存在している文書のうち  $X$  を含むものの数とする。

始めに、語彙の出現の偏りの計算方法について述べる。現在は、ジニ係数という主に経済学において、富の偏在性などを表すのに用いられる指標を利用する。ジニ係数の範囲は  $[0, 1]$  であり、完全に偏りが無い場合には0になり、完全に偏っている場合には1となる。

今、着目しているディレクトリが図3におけるAというディレクトリであった場合について考えてみる。各ディレクトリにいくつかの文書が分類されていたとして、語彙  $X$  の偏りを求めるために、まず、各ディレクトリにおいて  $X$  が出現する文書がどの程度の割合で存在するのかと言うことを求める。例えば、Eというディレクトリに文書が10存在し、そのうち8の文書には  $X$  が出現する時には、割合は0.8になる。

図4は、語彙の出現の割合を縦軸に、ディレクトリの文書数を横軸にしてグラフ化したものの例である。この場合、ディレクトリAにはあまり  $X$  は出現せず、ディレクトリEにはより頻繁に出現することがわかり、ある程度偏って出現すると言うことが判断できる。

このグラフに表されたデータをジニ係数を計算するためのデータとして用いる。

縦軸の量は、ディレクトリ  $D$  以下での語彙  $X$  の出現割合として、以下のように表される。

$$P(X, Sub_i(D)) = \frac{Num(X, Sub_i(D))}{Num(Sub_i(D))}$$

横軸は各ディレクトリにおける文書数なので、 $Num(Sub_i(D))$  と表すことができる。

これらを基に、 $D$  における  $X$  のジニ係数  $GC$  を計算すると、以下のような式になる。

$$GC(X, D) = \frac{\sum_i \sum_j (|P(X, Sub_i(D)) - P(X, Sub_j(D))| N_i N_j)}{2 N_{all} (N_{all} - 1) \cdot \bar{P}}$$

ただし、 $N_i, N_j, N_{all}$  はそれぞれ以下の式で表される。

$$\begin{aligned} N_i &= Num(Sub_i(D)), \\ N_j &= Num(Sub_j(D)), \\ N_{all} &= \sum_k Num(Sub_k(D)) \end{aligned}$$

また、 $\bar{P}$  は全データの平均であり、以下の式で表される。

$$\begin{aligned} \bar{P} &= \frac{\sum_k P(X, Sub_k(D)) \cdot Num(Sub_k(D))}{N_{all}} \\ &= \frac{\sum_k Num(X, Sub_k(D))}{N_{all}} \end{aligned}$$

$\sum_k Num(X, Sub_k(D))$  は  $D$  以下の全文書の中で  $X$  を含む文書の数であり、以後、 $N_X$  と表すことにする。結果として、 $D$  における  $X$  のジニ係数は、以下のようになる。

$$GC(X, D) = \frac{\sum_i \sum_j (|P(X, Sub_i(D)) - P(X, Sub_j(D))| N_i N_j)}{2(N_{all} - 1) \cdot N_X}$$

次に、語彙の出現回数を無視し、語彙の偏りだけに着目したときの語彙どうしの関係性を  $R_{\text{関係名}}(X, Y, D)$  と表すとすると、現実装では、 $R$  を以下のように計算している。

$$\begin{aligned} R_{\text{広域的}}(X, Y, D) &= GC(X, D) \cdot (1 - GC(Y, D)) \\ R_{\text{狭域的}}(X, Y, D) &= (1 - GC(X, D)) \cdot GC(Y, D) \\ R_{\text{共起的}}(X, Y, D) &= (1 - GC(X, D)) \cdot (1 - GC(Y, D)) \\ R_{\text{排他的関連}}(X, Y, D) &= GC(X, D) \cdot GC(Y, D) \end{aligned}$$

語彙の出現回数による重み  $T$  は、ディレクトリ  $D$  以下の文書で対象とする語彙  $X, Y$  がどれだけ出現したか、を表す  $DF$  によって表される。しかし、一般的な語彙は出現回数が多くなるので、 $DF$  に全文書を

対象とした  $IDF$  を掛ける．結果として，以下の式を出現回数による重みとする．

$T(X, Y, D) = DF(X, D) \cdot DF(Y, D) \cdot IDF(X) \cdot IDF(Y)$  かし，上記のような分類だけでは，メール数に対して階層構造が小さすぎ，語彙の意味関係を求めるには十分ではない．何らかの方法で階層構造を取得する必要がある．

$$DF(X, D) = \frac{\sum_k Num(X, Sub_k(D))}{\sum_k Num(Sub_k(D))}$$

$$IDF(X) = \log \left( \frac{1}{DF(X, \text{ルートディレクトリ})} \right)$$

$R$  と  $T$  を掛け合わせたものが，関係性  $V$  となる．すなわち，以下のような式で，それぞれの意味関係の式が表される．

$$V_{\text{関係名}}(X, Y, D) = R_{\text{関係名}}(X, Y, D) \cdot T(X, Y, D)$$

この式は，例えば，あるディレクトリにおいて，広域的な関係が最も深い語彙のペアを見つける，ということにはそのまま用いることができる．

しかし，実際のアプリケーションを考えると，ある語が始めに与えられ，その語に対して他の語彙がどのような関係性を持っているか，ということを知りたい時があり，そのような時にはこのままでは問題があることがわかる．例えば， $X$  は同じ語彙のまま， $Y$  にほかの様々な語を入れたとき， $V_{\text{広域的}}(X, Y, D)$  と  $V_{\text{共起的}}(X, Y, D)$  の式に着目すると，それらが比例関係にあることがわかる．同様に， $V_{\text{狭域的}}(X, Y, D)$  と  $V_{\text{排他的関連}}(X, Y, D)$  も比例し，例えば，広域的な関係として最も値の大きな語彙は，同時に，共起的な関係としても最も値の大きな語彙となってしまうのである．

そこで，ある語が与えられて，その語に対する他の様々な語の関係性を調べる際には，対象とするディレクトリで意味関係を求めて，さらに，そのサブディレクトリにおいても意味関係を求め，最終的にそれらに重み付けして足し合わせるということを行う．この部分に関しては，以前<sup>2)</sup>と同じ手法であるため省略する．

#### 4. 電子メールのメタ情報を利用した分類構造の取得

これまで語彙の意味関係を求める手法について述べたが，語彙の意味関係を求めるためには，文書群とそれらが属する階層構造が必要である．一般的な文書はディレクトリ階層で主に内容によって分類されていることが多いが，電子メールでは，一般的な文書ほど分類がされないことが多い．

当然，メールが分類される場合もあり，例えば，送受信者による自動分類，メーリングリストの自動判別，タスクごとの手動分類などがあげられる．これらの場

合は，ある程度の階層構造が存在することもあり，利用することが可能である．

しかし，上記のような分類だけでは，メール数に対して階層構造が小さすぎ，語彙の意味関係を求めるには十分ではない．何らかの方法で階層構造を取得する必要がある．

メールは通常の文書以上よりも，定型的なメタ情報を持っている．例えば，以下のようなメタ情報を取得することが可能である．

- 送受信者
- 日時
- メッセージ ID，リプライメッセージ ID

ここでは，これらから，分類に用いることが可能な階層構造を求める手法について考える．

送受信者の情報は潜在的に階層構造を持っている場合がある．例えば，所属するグループや学年などであり，それらの情報は，アドレス帳などには含まれている場合がある．それらを利用可能な状態にすれば，メールに対して自動的に階層構造を与えることができる．

メッセージ ID は各メールにユニークに付けられる ID であり，メールのヘッダにおいて Message-Id という名前で付加されている．また，あるメールに対してリプライした場合には，返信される対象になったメールのメッセージ ID が In-Reply-To という名前で付加される．他にも，References というメタ情報があり，参照するメールのメッセージ ID を付加することができるが，これは，あまり効果的には使われていないため，考慮しない．

メッセージ ID とリプライメッセージ ID によって，メールのツリーを生成することができ，いくつかのメールクライアントでは視覚的にこのツリーを表示するものがある．このツリー構造は小さいながらも利用できる可能性がある．

日時は細分化することによって，いくらかでも構造を作ることが可能である．例えば，まず，一週間ごとにメールをまとめて最下層の分類を作り，その上位層として一ヶ月ごと，半年ごと，一年ごとなどとまとめていくのである．その上で，最近一ヶ月に着目したときに，語彙の意味関係を求めることによって，最近ではどういう考えを持っているのか，ということを表すことができると考えられる．

このようにして，いくつかの分類構造を作ることができれば，それらを組み合わせることによってさらに深い構造を作ることにも可能である．つまり，例えば図 5 のように，全体を送受信者によって自動的に分類し，各人のまとまりの中でさらに日時によって階層化する

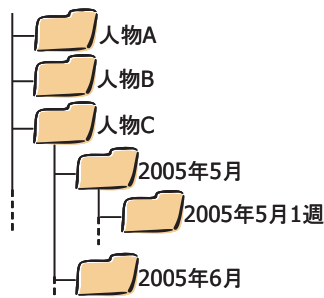


図5 電子メールに対する分類構造の構築

のである。これによって、十分な階層構造を得ることができるので、あとは3章で述べた方法で語彙の意味関係を取得することが可能となる。

### 5. Web 情報検索の個別化

我々はこれまで、一般的な文書から得た語彙の意味関係を用い、Web 情報検索のパーソナライゼーションが可能なシステムを作成してきた<sup>2),3)</sup>。そのシステムでは、下記の2つの機能を持っている。

- クエリ拡張
- 検索結果の再ランキング

まず、現在対象としているディレクトリをユーザに選択してもらい、それを基に語彙の意味関係を取得する。ユーザはクエリを入力するが、そこで関連が深い語をクエリに追加するなどのクエリ拡張を行うことができる。そのクエリをもとにシステムは既存の検索エンジンを利用して検索結果を求める。求められた検索結果は、検索エンジンによって順序づけがなされているが、それを、クエリで用いた語彙に対して関連が深い語が多く含まれるものが上位に来るように再ランキングを行うことができる。

今回は、一般的な文書ではなく、電子メールから得た語彙の意味階層を用いて、どのようにウェブ情報検索のパーソナライゼーションが行えるかということについて考える。

メールに対する階層構造としては、4章の最後に述べたように、送受信者によって自動分類したあと日時によってさらに階層化したものを用いた。

そのような、図5のような分類構造においては、これまで行ってきた Web 情報検索のパーソナライゼーションの操作に加えて、

- 誰を念頭において検索を行うのか
- いつのことを念頭において検索を行うのか

ということ、ユーザが指示することによって、これまでの単なる語彙マッチングの検索よりも、より様々な意味をもった検索を行えるようになると考えられる。

表1 各意味関係において「研究」という語と関連する上位の語彙

広域的	狭域的	共起的	排他的関連
検索	学部	内容	アブストラクト
個人	博士	概念	セマンティック
場合	計画	考え	範囲
自分	修士	利用	学生

実験のために用意したメールは、私の2003年4月から2005年6月までの27ヵ月分のすべての送信メール744通である。送信メールにはそのコンピュータのユーザが書いた文書が存在し、その人の知識を表すものとして非常に有用であると考えられる。

まず始めに、メールを送信先アドレスごとに分類した。これは、メールヘッダの To を用いた。今回は実験のため、同報メールの場合は To の最初に記載されているアドレスへのメールとして分類を行った。

次に、送信アドレスごとの分類の中で日時を用いて、以下のようにしてさらに下位の階層構造を作成した。

- (1) 最小の単位として一ヶ月ごとの分類を作成する。
- (2) それらを3つまとめて四半期の分類を作成する。
- (3) さらに4つまとめて一年の分類を作成する。

これにより、人ごと一年ごと四半期ごと一ヶ月ごと、という4階層の分類構造が作成された。

まず、ある程度長期間にわたって継続的にメールを送っている教官のメールアドレスを対象とした。ここで「研究」という語に対して、どのような関係性を持つ語が現れるかを調べた結果が表1である。

まず、共起的という所を参照してみると、ユーザである私が行っている研究に必ず出てくるような語が現れていることがわかる。次に、広域的という所を見てみると、もう少し一般的な時にも使うような語が出現している。狭域的では、「学部」「博士」「修士」という関連するような語が出現しており、それぞれの時期における研究の話が別々に出現していたことが容易に予想される。排他的関連では「セマンティック」という語が現れており、これは、本研究がセマンティックウェブと近いが異なるやり方でやっていることが現れているともとらえることができる。この結果では、クエリ拡張に使うには少し一般的すぎる語が出現してしまっていて効果的ではないかもしれないが、再ランキングでは効果的に働くと考えられる。

現在興味があることや、一年前に興味があったことなどによって再ランキングを行い、自分の興味がどのように変化したかというようなことも、ある時期のメールを対象とすることでとらえることができると考えられるが、今回は行うことができなかった。他にも、得られた関係性の違いを活かして、より概要的な文書

や,より詳細な文書を求めるといったような,意味づけを伴った検索を行えると考えられるので,そのようなことも含めて今後の課題としたい.

## 6. まとめと今後の課題

本稿では,個人的な文書から,語彙の間にある意味関係を抽出する手法について提案を行った.文書として電子メールを用いたが,電子メールが持っているメタ情報から構造を抽出し,それを利用することによって意味関係抽出の手法が適用できるようにした.そして,抽出された個人の知識を利用してウェブ情報検索の個別化を行う手法について提案を行い,実験による検証を行い,検索に意味を持たせることができることがわかった.

今後は,さらに多くの実験を行い,一般的な文書から得られる意味関係と電子メールから得られる人や時間を考慮したような意味関係が,それぞれどのような場面で役立つのかということを検証していく.また,語彙の関係性も現在の4つばかりでなく他の関係性なども考慮したい.同時に,別のアルゴリズムや実装を検討し,精度を高めていくつもりである.

## 謝 辞

本研究の一部は,文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」(代表:田中克己),および,平成17年度科研費特定領域研究(2)「Webの意味構造発見に基づく新しいWeb検索サービス方式に関する研究」(課題番号:16016247,代表:田中克己),および,21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」によるものです.ここに記して謝意を表すものとします.

## 参 考 文 献

- 1) 田中克己 大島裕明, 小山聡. 個人コンテンツからの概念体系の生成とこれに基づく web 検索のパーソナライゼーション. 情報処理学会 DBS 研究会技術報告, (Vol.2004, No.72, 2004-DBS-134):345-351, 2004.
- 2) 田中克己 大島裕明, 小山聡. 文書と保存ディレクトリ構造からの個人的な概念・語彙階層木の生成と web 検索の個別化への利用. 電子情報通信学会技術研究報告, 第 16 回データ工学ワークショップ (DEWS 2005), (6C-i6), 2005.
- 3) KatsumiTanaka HiroakiOhshima, SatoshiOyama. Extracting personalconceptual structures frompc documents and itsapplication to web search personalization. *Proceedings of International*

*Special Workshop on Databases For Next Generation Researchers (SWOD 2005)*, pages 32-35, 2005.

- 4) D.Karger E.Adar and L.Stein. Haystack: Per-user information environment. *Proc.1999 Conference on Information and Knowledge Management*, pages 413-422, 1999.
- 5) HaystackWeb Site. <http://haystack.lcs.mit.edu/>.
- 6) 井形伸之 渡部勇 津田宏 片山佳則, 小櫻文彦. セマンティックグループウェア workware++と knowwho 検索への応用. 情報処理学会 研究報告「情報学基礎」, (No.071), 2003.
- 7) 松井くにお 内野寛治, 津田宏. Workware:web を用いた文書の時間順整理の試み. 情報処理学会 研究報告「情報学基礎」, (No.051):380-385, 1998.
- 8) Yutaka Abe Hiroyuki Sato and Atsushi Kanai. Hyperclip: a tool for gathering and sharing meta-data on users' activities by using peer-to-peer technology. *WWW2002 Workshop on Real world RDF and Semantic Web applications*, 2002.
- 9) 桑原和宏 湯川高志, 吉田仙. パーソナル・レポジトリに対するピア・ツー・ピア型協調検索機構の提案. 電子情報通信学会 信学技報 AI2001-48, 2001.