

特定分野ドキュメントを対象とした 意味的連想検索のためのメタデータ空間生成方式

宮川 祥子[†] 清木 康^{††}

本論文では、特定分野のドキュメント検索において、意味的連想検索を行うために意味の数学モデルを使用する際のメタデータ空間の作成方式を提案し、その有効性を検証する。本方式は、特定の分野に関連する情報群とその分野に関する用語集、そして一般的な辞典から特定分野におけるメタデータ空間を構築する。意味的連想検索は、インターネット上で現在普及しているパターンマッチ型の検索エンジンでは扱うことのできない情報の意味を形式的に扱うことを可能にする。本方式を用いることにより、特定の分野に関する情報群に対して、文脈に応じて意味的に近い情報を検索するための検索環境を提供することができる。本論文では、また、インターネット上に実在する障害者・福祉関連のドキュメント群を対象としたメタデータ空間を構築し、パターンマッチによる検索との比較において性能評価を行う。

A Construction Method of a Metadata Space for Associative Search of Special Documents

SHOKO MIYAGAWA [†] and YASUSHI KIYOKI^{††}

In this paper we present a new construction method for a metadata space. This method enables to establish a metadata space for the mathematical model of meanings from documents related to a specific area, a lexicon related to the area, and a general dictionary. Unlike ordinary search engines employing pattern matching methods, the semantic associative search function enables to deal with the "semantics" of the information. This method makes it possible to provide a semantic information retrieval environment for documents about a specific area on the Internet. We also evaluate the performance of the semantic information retrieval based on the proposed method. We construct an information retrieval system built upon the metadata space related to information about welfare and people with disability distributed through the Internet and make a comparison with the pattern matching information retrieval system.

1. はじめに

インターネットは社会に急速に浸透し、いまや日常生活を支えるインフラストラクチャの一つとなっている¹⁾。インターネットが媒介するさまざまなサービスの中で、WWWによる情報獲得は最も人気の高いものであり、インターネット上のキラアアプリケーションとなっている。

WWWによって提供されている情報は膨大なものであり、目的の情報に迅速にアクセスするためには、リンク集や検索エンジンなどのナビゲーションシステムが欠かせない。しかし、一般的な検索エンジンは単純なパターン

マッチによる検索を基本としており、1つの単語に複数の意味がある場合や、異なる単語が同じ意味を持つといった場合に、これらの意味的情報を判断した検索を行うことができない。現在、この問題を原因とする検索効率の低さがインターネット上の情報検索における大きな問題点の1つとなっている。

一方で、インターネット上では、関心を共有するさまざまなコミュニティ（ネットワーク・コミュニティ）が活動を行っており、質の高いデータを提供している。^{2)~4)} それらは、スポーツや音楽などの趣味を同じくするグループから、インターネットのネットワーク管理について情報を交換するものまでさまざまであるが、その中のいくつかは、政府でも企業でもない第三の形態を持ち、新しい社会的役割を負う組織として注目されている⁵⁾。例えば女性関連情報、環境関連情報、障害者関連情報などを提供するネットワーク・コミュニティは、社会性の高い情報を発信、共有しているとして注目されている。特に、障害

[†] 慶應義塾大学 政策・メディア研究科

Graduate School of Media and Governance, Keio University

^{††} 慶應義塾大学 環境情報学部

Faculty of Environmental Information, Keio University

者関連のネットワーク・コミュニティは、自由に外出することが難しい障害者が社会との関わりを持つために、非常に重要な役割を果たしている。

これらのネットワーク・コミュニティは、しばしばその関心に基づく質の高い情報を、関心を共有する他の利用者のために提供している。これらの情報は、専門的な知識を用いて分類されており、リンク集の利用者は適切な分類を選択することによって必要な情報への容易なアクセスを実現している。しかし、提供される情報の量が增大するに連れて、分類のみでは効率的な情報検索ができなくなり、検索サービスの提供が必須となってくる。そして、典型的なパターンマッチに基づく検索エンジンでは、情報どうしの意味的な近さを判断することができないため、分類による情報提供の場合には同時に得られた関連性の高い情報も、パターンマッチによる検索では同時に得られないという場合も想定される。このことは、結果としてパターンマッチによる検索エンジンでは、専門家によって提供されている情報の質の高さを十分に活用することができないということを意味しており、これはネットワーク・コミュニティの活動のパフォーマンスを低下させる要因となる。

我々は、単純なパターンマッチではなく情報の意味に着目し、文脈に応じて意味的に近い情報を検索する方式として意味の数学モデル^{6)~9)}を提案している。意味の数学モデルは、従来の方式では形式的に扱うことができなかったコンテンツの「意味」を形式的に表現し、文脈に応じて変化する意味的關係を動的に計算することを可能にしている。意味の数学モデルでは、検索対象をベクトル化してメタデータ空間と呼ばれる空間に写像する。さらに、それらのベクトルをメタデータ空間の部分空間に射影して距離を計算することにより、単純なパターンマッチでは得ることのできない、意味的連想に基づく情報検索を実現している。

意味の数学モデルを用いて、先に述べた特定の関心に基づくネットワーク・コミュニティの情報を検索するためには、そのネットワーク・コミュニティの関心を表現するためのメタデータ空間を作成する必要がある。意味の数学モデルでは、メタデータ空間は、基本データとよばれる特徴つきベクトルの集合から生成される。ネットワーク・コミュニティの特徴を反映したメタデータ空間を生成するためには、このデータ行列を適切な方法で作成する必要がある。

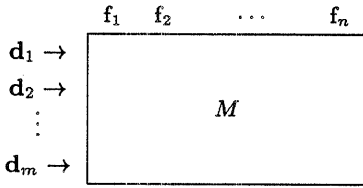
本論文では、特定分野のドキュメントを対象とした意味的連想検索に意味の数学モデルを適用する際の、メタデータ空間の生成方式を提案する。専門的な情報の集合に対して意味的検索を行う際のデータ行列の生成方式

を提案し、さらにその方式に基づいてインターネット上に実在する障害者・福祉関連のネットワーク・コミュニティが提供しているドキュメント群を用いてメタデータ空間を生成し、本論文で提案している方式の有効性を検証するための実験を行う。

本方式では、情報が提供されている特定分野そのものか、あるいはそれに近い分野の用語集が存在していることを前提とする。この用語集は、その分野の特徴を表現している専門用語を、我々が日常的に用いている一般語で解説したものである。この用語集と、ネットワーク・コミュニティによって提供されている専門性の高いドキュメント群から、反復的な手法を用いてデータ行列を生成する。この手法を用いることによって、ある言語の全体が表現可能な意味的空間を生成することなく、特定の分野に特化したコンパクトな意味的空間を生成することが可能となる。

我々の提案する方式は、文献11)に代表される多変量解析による空間生成を用いた検索手法とは次の点で本質的に異なる。提案方式では、直交空間における部分空間選択を行う演算を定義し、その演算により、言葉の意味的關係を、文脈すなわち与えられた検索要求に基づいて選択された部分空間に応じて、解釈する機構を実現している。この機構により、言葉と言葉、あるいは言葉とメディアデータ（例えばドキュメント）の間の意味的關係を、与えられた文脈に応じて動的に計算することが可能となる。本研究における実現システムでは、本モデルで表現できる部分空間すなわち文脈の相の数¹²⁶⁰は約 2^{1260} であり、ほぼ無限の文脈に対応することが可能である。また、空間生成手法に関しては、文献11)の手法がドキュメント中の単語の出現回数によるベクトル生成手法を用いているのに対し、提案方式では専門性の高い用語集の説明文から、各単語の肯定的／否定的用法に基づいた付値を行い、ベクトルを生成している。

また、提案方式における空間生成手法は、文献12)による手法とは次の意味で本質的に異なる。文献12)による手法では、空間生成のための特徴語は政治、経済など一般的な分野からあらかじめ選ばれたものであり、これらの特徴語を用いることによりほぼすべての分野のメディアデータの特徴づけを行うことを目的にしている。一方で、提案方式における空間生成手法は、あらかじめ特徴語を与えておくのではなく、特定の分野を表現するために必要な特徴語群を、特徴づけを行うプロセスの中で生成していく手法を採用している。この手法を用いることで、特定の専門分野における意味空間を表現するために必要な特徴語群をあらかじめ吟味すること無く、一般的な意味内容を扱う意味空間では表現することが難しい専門的な

図1 データ行列 M によるメタデータの表現Fig. 1 Metadata represented in data matrix M

分野の意味空間を生成することが可能となる。

また、他の意味的連想検索に関する研究として、自然言語処理を適用した手法が提案されている¹³⁾¹⁴⁾¹⁵⁾。

インターネット上の情報検索の分野では、文献16)に代表される、ユーザの嗜好などのプロフィールを利用した情報検索方式が存在する。この方式との比較における提案方式の特徴は、検索者が与える文脈を変更することにより、文脈に応じたデータ間の相関関係を求める機構を実現しているという点である。文献16)に代表される方式では、ユーザの嗜好を表現する情報は、頻繁には変化しないことを前提としており、したがってそれにもとづいたデータ間の関連づけも、ある特定のユーザから見た場合には静的であると言える。一方で、提案方式では同一のユーザであっても異なる文脈で検索要求が与えられた場合には、データ間の関連付けは文脈に応じて変化するため、ユーザに対してより柔軟な検索機能を提供することが可能となる。

2. 意味的連想検索方式

ここでは、意味的連想検索方式について概説する。詳細は、文献6)~9)に述べられている。

2.1 メタデータ空間 MDS の設定

初めに、 m 個の基本データについて各々 n 個の特徴 (f_1, f_2, \dots, f_n) を列挙した特徴付ベクトル $d_i (i = 1, \dots, m)$ が与えられているものとし、そのベクトルを並べて構成する $m \times n$ 行列を M とおく(図1)。このとき、 M は、列ごとに2ノルムで正規化されている。

- (1) データ行列 M の相関行列 $M^T M$ を計算する。
- (2) $M^T M$ を固有値分解する。

$$M^T M = Q \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_\nu & \\ & & & 0 \dots 0 \end{pmatrix} Q^T,$$

$$0 \leq \nu \leq n.$$

ここで行列 Q は、

$$Q = (q_1, q_2, \dots, q_n)$$

である。この $q_i (i = 1, \dots, n)$ は、相関行列の正規化された固有ベクトル(以下、“意味素”)である。相関行列の対称性から、この固有値は全て実数であり、その固有ベクトルは互いに直交している。

- (3) メタデータ空間 MDS を以下で定義する。

非ゼロ固有値に対応する固有ベクトル(以下、“意味素”と呼ぶ)によって形成される正規直交空間をメタデータ空間 MDS と定義する。この空間の次元 ν は、データ行列のランクに一致する。この空間は、 ν 次元ユークリッド空間となる。

$$MDS := span(q_1, q_2, \dots, q_\nu).$$

$\{q_1, \dots, q_\nu\}$ は MDS の正規直交基底である。

2.2 メディアデータのメディアデータベクトルの作成方式

ここでは、メディアデータを表現するメディアデータベクトルを形成する方法を示す。

- (1) **Step-1:** メディアデータの特徴づけ

t 個の語(あるいは、 t 個のオブジェクト)

o_1, o_2, \dots, o_t から成るメディアデータ P を次のように特徴づける。

$$P = \{o_1, o_2, \dots, o_t\}.$$

ここで、各語 o_i は、データ行列の特徴と同一の特徴を用いて表現される特徴付ベクトルである。

$$o_i = (o_{i1}, o_{i2}, \dots, o_{in})$$

- (2) **Step-2:** メディアデータ P のベクトル表現

メディアデータ P を構成する t 個の語 o_1, o_2, \dots, o_t が、それぞれ n 次元ベクトルで定義されている。オブジェクト o_1, o_2, \dots, o_t の和演算子 \oplus を次のように定義し、メディアデータのメディアデータベクトル p を形成する。

$$p = \bigoplus_{i=1}^t o_i := (\text{sign}(o_{t11}) \max_{1 \leq i \leq t} |o_{i1}|,$$

$$\text{sign}(o_{t22}) \max_{1 \leq i \leq t} |o_{i2}|,$$

$$\dots, \text{sign}(o_{tnn}) \max_{1 \leq i \leq t} |o_{in}|).$$

この和演算子 $\bigoplus_{i=1}^t$ は、 t 個のベクトルから各基底に対して絶対値最大の成分を選ぶ演算子である。

ここで $\text{sign}(a)$ は、“ a ”の符号(正, 負)を表す。また、 $l_k (k = 1, \dots, t)$ は、特徴が最大となる語を示す指標であり、次のように定義する。

$$\max_{1 \leq i \leq t} |o_{ik}| = |o_{\ell k k}|.$$

2.3 意味射影集合 Π_ν の設定

メタデータ空間 MDS から固有部分空間 (以下, 意味空間) への射影 (以下, “意味射影”) の集合 Π_ν を考える. P_{λ_i} を次の様に定義する.

$$P_{\lambda_i} := \lambda_i \text{ に対応する固有空間への射影} \\ \text{i.e. } P_{\lambda_i} : MDS \rightarrow \text{span}(\mathbf{q}_i).$$

意味射影の集合 Π_ν を次のように定義する.

$$\Pi_\nu := \{ 0, P_{\lambda_1}, P_{\lambda_2}, \dots, P_{\lambda_\nu}, \\ P_{\lambda_1} + P_{\lambda_2}, P_{\lambda_1} + P_{\lambda_3}, \dots, P_{\lambda_{\nu-1}} + P_{\lambda_\nu}, \\ \vdots \\ P_{\lambda_1} + P_{\lambda_2} + \dots + P_{\lambda_\nu} \}.$$

i 次元の意味空間は, $\frac{\nu(\nu-1)\dots(\nu-i+1)}{i!}$, ($i = 1, 2, \dots, \nu$) 個存在するので, 射影の総数は, 2^ν となる. つまり, このモデルは, 2^ν 通りの意味の様相の表現能力をもつ.

2.4 意味解釈オペレータ S_p の構成

検索者の印象やメディアデータの内容を与える文脈を表す ℓ 個の検索語列

$$s_\ell = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell)$$

と, しきい値 ε_s ($0 < \varepsilon_s < 1$) が与えられたとき, それに応じた, 意味射影 $P_{\varepsilon_s}(s_\ell)$ を構成するオペレータ (以下, “意味解釈オペレータ”) S_p を構成する. T_ℓ を長さ ℓ の検索語列の集合とすると, S_p は, 次のように定義される.

$$S_p : T_\ell \rightarrow \Pi_\nu$$

ここで, $T_\ell \ni s_\ell, \Pi_\nu \ni P_{\varepsilon_s}(s_\ell)$.

また, $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_\ell\}$ の各要素は, 特徴付ベクトルであり, データ行列 M の特徴と同一の特徴を用いて表される.

オペレータ S_p は以下の計算を行う.

- (1) \mathbf{u}_i ($i = 1, 2, \dots, \ell$) をフーリエ展開する.
検索語列 s_ℓ を構成する ℓ 個の検索語を各々メタデータ空間 MDS へ写像する.

この写像では, ℓ 個の単語を各々メタデータ空間 MDS 内でフーリエ展開し, フーリエ係数を求める. これは, 各検索語と各意味素の相関を求めることに相当する.

\mathbf{u}_i と \mathbf{q}_j の内積 u_{ij} は次のようになる.

$$u_{ij} := (\mathbf{u}_i, \mathbf{q}_j), \quad j = 1, 2, \dots, \nu.$$

ベクトル $\hat{\mathbf{u}}_i \in MDS$ を次のように定める.

$$\hat{\mathbf{u}}_i := (u_{i1}, u_{i2}, \dots, u_{i\nu}).$$

これは, 単語 \mathbf{u}_i をメタデータ空間 MDS に写像したものである.

- (2) 検索語列 s_ℓ の意味重心 $\mathbf{G}^+(s_\ell)$ を求める.

まず, 各意味素ごとに, フーリエ係数の総和を求める. これは, 検索語列 s_ℓ と各意味素との相関を求めることに相当する. このベクトルは, ν 個の意味素があるため, ν 次元ベクトルとなる. このベクトルを, 無限大ノルムによって正規化したベクトルを, 以下, 検索語列 s_ℓ の意味重心 $\mathbf{G}^+(s_\ell)$ と呼ぶ.

$$\mathbf{G}^+(s_\ell) := \frac{\left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right)}{\left\| \left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right) \right\|_\infty}.$$

ここで, $\|\cdot\|_\infty$ は無限大ノルムを示す.

- (3) 意味射影 $P_{\varepsilon_s}(s_\ell)$ を決定する.

検索語列 s_ℓ の意味重心を構成する各要素において, しきい値 ε_s を越える要素に対応する意味素を, メディアデータのメタデータを射影する意味空間の構成に用いる. 意味射影 $P_{\varepsilon_s}(s_\ell)$ を次のように決定する.

$$P_{\varepsilon_s}(s_\ell) := \sum_{i \in \Lambda_{\varepsilon_s}} P_{\lambda_i} \in \Pi_\nu.$$

ただし $\Lambda_{\varepsilon_s} := \{ i \mid \|(\mathbf{G}^+(s_\ell))_i\| > \varepsilon_s \}$ とする.

2.5 意味空間における相関の定量化

文脈 (文脈を表す検索語列) を対象として, 2.4 節で示したオペレータ S_p を用いて選択された意味空間 (部分空間) 上で, その文脈に対応したメディアデータを呼び出す意味的連想検索方式を示す.

メタデータ空間に写像されたメディアデータ群に対応する各ベクトル (メディアデータベクトル) について, 選択された意味空間 (部分空間) 上におけるノルムを求め, 文脈に相関の強いメディアデータの検索を行う. 意味空間におけるメディアデータベクトルのノルムの大きさをその文脈とメディアデータとの相関の強さとする.

文脈 s_ℓ が与えられた場合のメディアデータ \mathbf{x} のノルム $\rho(\mathbf{x}; s_\ell)$ を次のように定める.

$$\rho(\mathbf{x}; s_\ell) = \frac{\sqrt{\sum_{j \in \Lambda_{\varepsilon_s} \cap S} \{c_j(s_\ell)x_j\}^2}}{\|\mathbf{x}\|_2},$$

$$S = \{i \mid \text{sign}(c_i(s_\ell)) = \text{sign}(x_i)\},$$

$$c_j(s_\ell) := \frac{\sum_{i=1}^{\ell} u_{ij}}{\left\| \left(\sum_{i=1}^{\ell} u_{i1}, \dots, \sum_{i=1}^{\ell} u_{i\nu} \right) \right\|_\infty},$$

$$j \in \Lambda_{\varepsilon_s}.$$

ここで, 意味空間を構成する意味素 (固有ベクトル) 群において, 文脈に関係しているのは, 正と負のどちらか一方である. そこで, 意味空間を構成する意味素の符

合を考慮するため、意味空間を構成する意味素の符合と正負が逆の成分についてはノルムの計算において無視している。

また、メディアデータを特徴づける特徴の数が多いと、どのような意味空間が選ばれても、意味空間におけるメディアデータのノルムが大きくなる傾向がある。そのため、本来、文脈との相関が強いと考えられるメディアデータベクトルのノルムよりも、特徴の数が多いメディアデータベクトルのノルムが大きくなってしまい、適切な抽出が行われないことがある。そのため、メタデータ空間でのメディアデータベクトルを2ノルムで正規化している。

3. 意味的連想検索の具体例

ここでは、意味的連想検索の具体例を示す。ここでは約1260次元のメタデータ空間がすでに構築されているとする。本方式による意味的連想検索では、以下の手順で、メディアデータ群と検索者の検索語列との相関量を求める。

- (1) 検索対象および検索語列のメタデータ空間上への配置
- (2) 文脈理解
- (3) 相関量の抽出

ここでは、例として、2つのメディアデータ document1 および document2 と、検索語列 理学療法、医療 が与えられたとする。

3.1 検索語列および検索対象のメタデータ空間上への配置

検索者の与えた検索語列 理学療法、医療 は、図2のように定義されているとする。

メディアデータ document1 および document2 が、図3のように特徴づけられているとする。特徴づけに使用されている各単語は、図4のように定義されているとする。

この定義に基づき、各基本語に対応するベクトルの要素を1, -1, もしくは、0と付値して約1260次元のベクトルとする。すなわち、肯定の意味で使用されている場合は“1”、否定の意味で使用されている場合は“-1”、そして、使用されていない場合は“0”とする。

メディアデータ document1 および document2 のベクトル化については、各メディアデータの特徴づけられている3単語に対応する各ベクトルを和演算子 \oplus (2.3節)により合成する。

以上のように、検索語列とメディアデータをベクトル化し、メタデータ空間に配置する。また、メディアデータに対応するベクトルについては、2.5節で述べた通

理学療法: 障害者 身体 心理 社会 職業 経済 機能 発揮 援助 復帰 支援 医学:

医療: 生体 疾病 病気 怪我 診断 治療 予防 学問:

図2 検索語列に使用される単語の定義

Fig. 2 Definitions of the words used for the search words

document1: 障害者 リハビリ 施設

document2: 障害者 支援 機器

図3 メディアデータの特徴づけ

Fig. 3 Characterizations of media data

障害者: 身体 精神 機能 低下 - 正常 喪失 身体 一部 欠損 心身 機能 - 健全:

リハビリ: 福祉 技術 利用 障害者 社会 復帰 支援 障害 軽減:

施設: 特定 目的 建物 設備:

支援: 補助 助け 援助:

機器: 器具 仕事 技術 装置:

図4 特徴づけに使用される単語の定義

Fig. 4 Definitions of the words used in the characterizations of media data

り、各ベクトルを2ノルムで正規化する。これは、メタデータ空間上で各メディアデータに対応するベクトルのノルムを揃えておくためである。

3.2 文脈理解

検索語列 理学療法、医療 を2.4節で示したオペレータ S_p により合成し、検索語ベクトルを得る。検索語ベクトルを正規化した後、約1260の各軸に射影し、各軸での検索語ベクトルの位置があるしきい値 ε 。以上の値を持つ軸のみを選び出す。選ばれた軸のみによって構成される部分空間を得る。この部分空間選択の操作が文脈理解に対応する。

3.3 相関量の計算

得られた部分空間上にメディアデータベクトル (document1 および document2) を写像し、それぞれのノルムを計算する。このノルムは、2.5節で示した重み付きノルムである。これを、検索語列とメディアデータベクトルとの相関とする。

すべてのメディアデータベクトル (すべての検索対象) について、検索語列との相関量を求める。すべての検索対象を、相関量に従ってソートする。

4. 提案方式における行列生成方式

4.1 特定分野の日本語によるデータ行列の生成

提案方式では、ある言語が表現可能な意味的空間全体を作成するのではなく、検索対象となるデータが包含さ

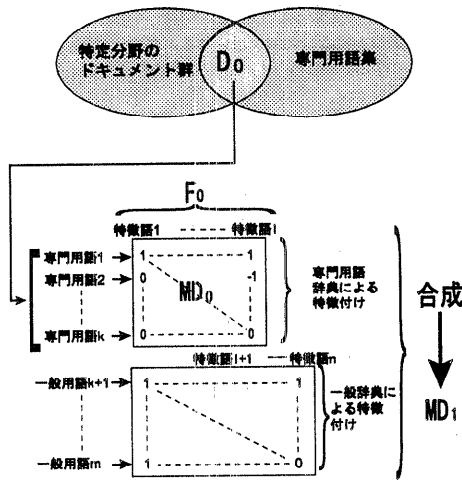


図5 特定分野のメタデータ空間の構築方式

Fig. 5 A construction method of metadata space related to specific area

れている特定の分野に関する意味的空間を生成することを目的としている。このような空間の作成を前提とすることにより、ある言語の意味的空間を作成するよりも比較的少ない労力で、特定の分野に関連する検索対象ドキュメントの意味的相関関係をより適切に表現可能なメタデータ空間を生成することができると考えられる。以下に提案する方式では、検索対象として、ある特定の分野に関するドキュメント群が存在し、またその分野と同一もしくは近い分野の用語辞典と、一般的な語を説明するための辞書が存在することを前提としている(図5)。

(1) 初期行列の設定

まず、検索対象となるドキュメント群を形態素解析し、名詞のみを抽出する。その中から抽出した名詞と用語辞典のインデックス(説明されている語)の両方に登場している語群を抽出し(この集合を D_0 とする)、これらの語群を用語辞典を用いて特徴づける。すなわち、用語辞典において各語を説明している文章中に現れる名詞群が、各語の特徴となる。各特徴に対応する値は、以下のようにして決定する。

- 特徴語が肯定的に用いられている場合: 1
- 特徴語が否定的に用いられている場合: -1
- 使用されていない場合: 0

具体的には、説明の中で「…である」という肯定

的な用法で用いられている名詞については1、「…でない」「…を伴わない」という否定的な用法で用いられ得る名詞については-1とする。例として、「弱視」について「医学的には病変がなく、あってもそれでは説明のつかない視力の低下をいう」という説明がされているとすると、この文中の名詞「病変」「説明」は否定的に、「視力」「低下」は肯定的に用いられているので、「病変」「説明」「視力」「低下」には、それぞれ-1, -1, 1, 1という値が付値され、それ以外の特徴には0が付値される。

また、検索語自身が特徴語である場合には、その特徴に対応する値を1とする。ここで作成した初期行列を MD_0 、初期行列の特徴の集合を F_0 とよぶ(図5)。

(2) 初期行列の拡張によるデータ行列の生成

(1)で作成した初期行列 MD_0 を反復的に拡張することによってメタデータ空間生成のためのデータ行列を作成する。まず、初期行列の特徴集合 F_0 の各要素 f_{0i} について、 $f_{0i} \notin D_0$ の場合には、一般的な辞書を用いて f_{0i} の特徴づけを行う。すなわち、一般的な辞書において f_{0i} を説明している文章中に現れる名詞群を f_{0i} の特徴として、初期行列と同様の方法で値を決定する。 $f_{0i} \in D_0$ の場合には何も行わない。

全ての f_{0i} についてこの操作を行った後、特徴づけられた f_{0i} と D_0 の和集合を D_1 と定義する。 D_1 の各要素を特徴づけている特徴語の集合を F_1 とする。 F_1 に含まれる新たな特徴群を対象として、 MD_0 内の各専門用語を表す特徴ベクトルについて、 F_1 に含まれる新たな特徴群に対応する要素設定を行い、これらの結果を合成したものを MD_1 とする(図5)。

メタデータ空間 MDS が正規直交空間として定義されるためには、メタデータ空間の元になるデータ行列 M が $m \times n$ 行列で与えられているとき、 $m \geq n$ であることが必要条件である。そのため、 $MD_k (k \geq 1)$ において $m < n$ の場合には、以下の操作を反復的に行う。

データ行列 MD_k の特徴集合 F_k の各要素 f_{ki} のうち、 D_k に含まれていないものについて、一般的な辞書を用いて特徴づけを行い、 MD_k の各特徴付きベクトルに加え、これを MD_{k+1} とする。ある $m \times n$ 行列 MD_k において $m \geq n$ となった時点でこの反復操作を終了する。

意味の数学モデルを用いたこれまでの意味的連想検索の

実験では、メタデータ空間生成のための基本データとして“Longman Dictionary of Contemporary English”²⁰⁾もしくは“The General Basic English Dictionary”²¹⁾における基本英単語を用いてきた。これらの辞典は、それぞれ約 2000 語、約 850 語の基本英単語のみを用いてそれぞれ約 56000 語、約 20000 語の説明をしている。これらの基本英単語を特徴とすることで、英語に関する意味的空間を実現していた。

これらの辞典は、一般的な英英辞典であるので、その基本英単語群によって作られるメタデータ空間は、一般的な英語による意味的空間であると考えられる。しかし、専門的な領域のメディアデータを扱う場合には、このような一般的な意味的空間のみでは不十分な場合がある。これは、専門的なメディアデータの特徴づけには専門的な語が用いられるため、これらの専門用語と一般的な語とのマッピングを取る必要があるからである。

例えば、社会福祉の領域では、「統合教育」という語は単なる「統合」と「教育」という 2 つの語の意味を組合せたものではなく、そこには「障害者」「健常者」という文脈が含まれている。したがって、「統合教育」という語は、「障害者」「健常者」「統合」「教育」という語群によって特徴づけられることが望ましい。しかし、このような専門性の高い語は必ずしも一般的な辞書に含まれているとは限らないため、専門性の高いメディアデータを扱う際には、この点を考慮して、専門用語を基本データとして含めたデータ行列を準備する必要がある。

本方式では、専門的な語を特徴づけることが可能な用語辞典を用いることによって専門用語を基本データとして含んだデータ行列の作成を可能にしている。

4.2 メディアデータの特徴づけとメタデータ変換行列の作成

検索対象となるメディアデータが最終的にメタデータ空間に正しく写像されるためには、メディアデータを特徴づける特徴ベクトルはデータ行列 M の特徴を用いたベクトルに変換可能でなければならない。そのためには、メディアデータの特徴ベクトルの各要素が、データ行列 M の特徴を用いているか、もしくはメディアデータの特徴ベクトルの各要素からデータ行列 M の特徴への変換を行う必要がある。任意の語によるメディアデータの特徴づけを実現するためには、後者の方法を採用してメディアデータの特徴とデータ行列 M の特徴とのマッピングをとる必要がある。これは、データ行列 M と同じ特徴をもつ、検索対象メディアデータのメタデータ行列 N を作ることに相当する。メディアデータのメタデータ行列 N は以下の方法で作成する。

まず、検索対象となる全てのメディアデータに対して、

任意の個数の任意の語を用いて特徴づけを行う。その後、各メディアデータの特徴ベクトルをデータ行列 M と同じ特徴をもつベクトルに変換するために、以下の操作を行う。

あるメディアデータ P を特徴づけている語の集合を O とする。 O の要素 o_i がデータ行列 M の特徴集合 F に含まれている場合には、 o_i を変換する必要はない。 o_i がデータ行列 M の基本データ集合 D の要素である場合には、データ行列 M の該当する特徴ベクトルを o_i を表す特徴ベクトル (F 内の各要素によって表現されたベクトル) と、メディアデータ P の o_i 以外の要素とを合成したベクトルを生成する。 o_i が上のどちらでもない場合には、 o_i をデータ行列 M の特徴群 F を用いて特徴づけ、メディアデータ P の o_i 以外の要素と合成したベクトルを生成する。

全ての o_i について上記の操作を行った後に得られたベクトルは、メディアデータ P をデータ行列 M の特徴集合 F によって特徴づけた特徴ベクトルとなる。

5. 本方式の具体例

5.1 メタデータ空間の生成

提案方式を用いた空間生成の具体例として、障害者・福祉関連のネットワーク・コミュニティが提供しているリンク集に掲載されている WWW ドキュメントの集合を検索対象として、メタデータ空間の生成を行った。

CHIME's SQUARE(<http://chime.vcom.or.jp/>) では、障害者・福祉関係のリンク集を提供しており、1998 年 12 月現在で 445 の日本国内の障害者・福祉関連の URL がリンク集に登録されている (図 6)。これらのリンク集は、内容別、地域別、提供者の属性別の 3 種類の属性で分類されて提供されており、利用者は自分の検索意図に適合した方法で分類されているリンク集を利用することができる。

今回の例では、CHIME's SQUARE に登録されている URL のうち、実際にコンテンツを得ることができた 371 の WWW ドキュメントを用いた。また、関連分野の用語辞典として、中央法規出版「改訂 介護福祉用語辞典」²²⁾ (以下、「介護福祉用語辞典」) を用いた。

(1) 初期行列の基本データ整備

CHIME's SQUARE に登録されている 371 の URL が示す WWW ドキュメントを形態素解析し、品詞別に分類した。形態素解析には、奈良先端科学技術大学院大学情報科学研究科の松本裕治氏らによって開発された茶釜 (ChaSen) version 1.51²³⁾ を用いた。形態素解析と品詞別分類の結果、障害者・福祉関連に特徴的な用語は名詞に集

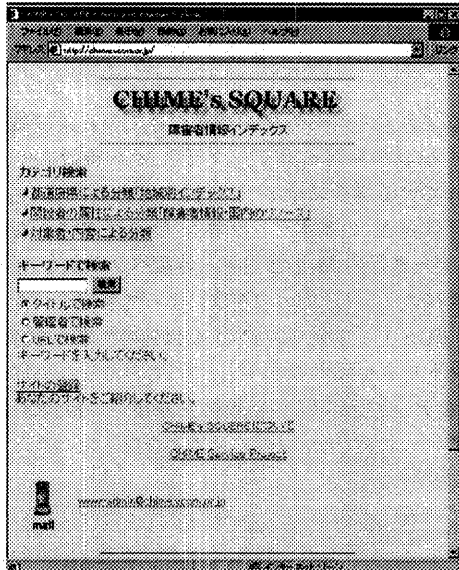


図6 CHIME's SQUARE ホームページ
Fig. 6 The homepage of CHIME's SQUARE
©CHIME's SQUARE

中しており、動詞、副詞、形容詞、記号、連体詞、接続詞には特徴的な用語が見られなかったため、名詞のみを抽出して用いることとした。

(2) 初期行列の作成

(1) で抽出した名詞のうち、介護福祉用語辞典の見出し語 3500 語中に出現する 221 語について、介護福祉用語辞典を用いて特徴づけを行った。具体的には、それぞれの語について介護福祉用語辞典による解説文から名詞のみを抜きだして特徴語とし、4.1 で示した方法によって値を決定した。この初期行列 MD_0 は、基本データ 221、特徴語 784 の 221×784 行列となった。

(3) データ行列の作成

(2) で作成した初期行列 MD_0 の特徴語 784 のうち、 MD_0 の基本データとして特徴づけられていない 704 の特徴語を広辞苑第四版²⁴⁾ (以下「広辞苑」) を用いて特徴づけを行った。特徴づけは、初期行列と同様に、広辞苑の説明文から名詞のみを抜きだして特徴語とした。この操作によって作成された行列 MD_1 は、 925×1262 行列となった。さらに、特徴語に含まれていて基本データに含まれていない語を基本データに加え、広辞苑を用いて特徴づけをするということを、基本データの数が特徴語を上回るまで繰り返した。この方式では、

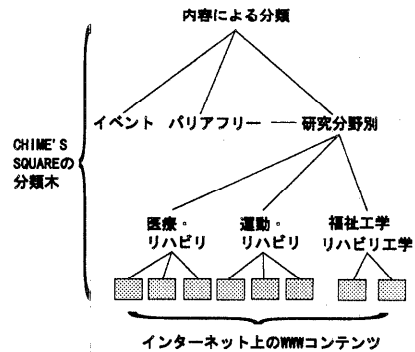


図7 CHIME's SQUARE の分類構造
Fig. 7 The classification structure of CHIME's SQUARE

繰り返し止まることは数学的には保証されていないが、経験的には数回の繰り返しで止まることがわかっている。今回の例では、3 回の繰り返しの後、基本データ数 1513、特徴語数 1392 となって終了した。

5.2 メディアデータベクトルの作成

検索対象となるメディアデータは任意の個数の語によって特徴づけることができるが、その際、各特徴語について、データ行列 M の特徴 F への写像を設定する必要がある。本方式では、メディアデータには任意の語を用いて特徴づけを行い、その後にデータ行列 M の特徴語への写像を定義している。その理由は、メディアデータへの特徴づけは、専門的な用語を用いた方がよりメディアデータの意味を的確に表現できると考えられるが、一方でデータ行列 M の特徴語、基本データには最小限の(ホームページ中の名詞と介護福祉用語辞典のインデックスの積集合)専門用語しか含まれていないため、これだけでは専門性の高いメディアデータをメタデータ空間上に写像するためには不十分であるからである。

今回の例では、CHIME's SQUARE に登録されている 371 の URL をメディアデータとして、それらに特徴づけを行った。特徴づけは、CHIME's SQUARE のメンバーが手作業で行った。メディアデータに付加された語の数は 314 で、そのうちデータ行列 M の基本語にも特徴にも含まれていない語は 153 であった。これらの 153

の語に対して、データ行列 M の基本データ D とデータ行列 M の特徴 F をドメインとする要素によって特徴づけを行った。その後、メディアデータに付加された語のうち D の要素であるものについては、データ行列 M を用いて F への写像を行い、最終的に F をドメインとするメタデータ変換行列 N を生成した。

6. 実験

本論文で提案した方式に基づいて構築されたメタデータ空間を用いた情報検索の実験結果について、CHIME'S SQUARE で現在提供されているリンク集を正解の基準として、パターンマッチによる検索との再現率の比較において評価を行った。具体的には、以下の2つのケースにおいて実験を行った。

実験1 CHIME's SQUARE では複数のページにわたっているカテゴリの検索

実験2 CHIME's SQUARE で用いられているカテゴリ名を検索語として用いた検索

CHIME's SQUARE では、インターネット上に存在する障害者・福祉に関する WWW ページを、独自の分類を用いて整理して提供している。この分類は、木構造になっておりユーザは木構造をたどることによって目的の情報に到達できるが、一方で木構造が細分化しすぎると、一般的なテーマで検索を行う際には、複数のリーフノードにアクセスする必要がある場合もある。例としてリハビリテーションという語/概念を挙げる。リハビリテーションは障害者の社会復帰にあたって非常に重要な役割を担っているため、CHIME's SQUARE にもリハビリテーションに関する情報が多数登録されている。これらの情報は、CHIME's SQUARE 上では「医療・リハビリテーション」「運動・リハビリテーション」「福祉工学・リハビリテーション工学」の3つに分類されている(図7)。そのため、ユーザがリハビリテーション全般に関する情報を得たい場合には、この3つのカテゴリをすべて見る必要があり、非効率的である。

このような状況を前提にして、以下の実験により提案方式の有効性について実験を行った。

6.1 実験環境

実験に用いるメタデータ空間とメディアデータを5章で示した方法によって実現した。使用した計算機は Sun Enterprise30, OSは Solaris5.5.1 である。CHIME's SQUARE の提供するリンク集とパターンマッチによる検索エンジンは本実験環境とは異なるサイト上に構築されている。パターンマッチ型検索に用いられている検索エンジンと日本語形態素解析エンジンは、それぞれ Sony SGSE-1.1, 茶釜 (ChaSen) version 1.51²³⁾ を用

いている。検索に必要なデータのサイズは、意味的連想検索の場合はメタデータ空間が約 64,085kbyte, メディアデータベクトルが約 16,610kbyte で合計するとおよそ 80,965kbyte で、パターンマッチ検索の場合には転置ファイルやファイルリストなど合計で約 811kbyte である。

意味的連想検索系におけるメディアデータベクトル作成のために用いられた特徴づけの一部は表1のとおりである。

6.2 実験1

本論文で提案する方式で作成したメタデータ空間とメディアデータを用いて、「リハビリテーション」を文脈語として意味的連想検索を行った。一方で、CHIME's SQUARE で提供されている全文検索エンジンを用いて「リハビリテーション」という文脈語で検索を行い、この両者の比較を行った。比較の基準として、CHIME's SQUARE リンク集の中で、リハビリテーションに関連している「医療・リハビリテーション」「運動・リハビリテーション」「福祉工学・リハビリテーション工学」という3つのカテゴリに登録されている情報を統合したリストを用い、このリストに対する再現率を比較することで評価を行った。この3つのカテゴリを統合したリストには30個のURLが含まれているので、それらを正解として設定し、意味的連想検索とパターンマッチにおける検索結果のうち上位30位を対象として評価を行った。

検索は、意味的連想検索、パターンマッチともに「リハビリテーション」とその同義語である「理学療法」という文脈語を用いて行った。その結果を表2に示す。ここで、意味的連想検索欄、パターンマッチ欄の数字は、それぞれ意味的連想検索、パターンマッチ検索を行った結果中で、該当するメディアデータが出現した順位である。今回生成したメタデータ空間においては、「リハビリテーション」と「理学療法」は等しいベクトルとして定義されており、どちらを文脈語として用いても同じ結果となるため、「リハビリテーション」の結果のみを示した。

また、同様の実験を「視覚・聴覚障害」「コンピュータ」という2つの検索要求においても行った。CHIME's SQUARE のリンク集では「視覚・聴覚障害」は視覚障害、聴覚障害という2つのカテゴリに、「コンピュータ」は「コンピュータ(ソフト・ハード)」「コンピュータ(サポート)」という2つのカテゴリに分かれており、検索要求に該当するドキュメント群を得るためには、2つの別のカテゴリを見なければならぬものである。なお、「視覚・聴覚障害」のパターンマッチ検索については、「視覚」「聴覚」「障害」の3つの語を全て含むものを検索した。

表 1 実験で用いた特徴づけの一部

Table 1 A part of definitions used in the experiment.

メディアデータ ID	特徴語群
www.kyoto.xaxon-net.or.jp/ matuki	障害者 自立 生活 センター
www.linkclub.com/ t-satake	言語 発達 障害 研究
www.lwc-eirec.go.jp	労災 リハビリテーション 福祉工学 義肢 義足 装具開発
www.lwc-eirec.go.jp/kenkyuu	肢体障害 ワープロ 情報 提供
www.mahoroba.or.jp/ keison/toukou	障害者 旅行 手記 情報
www.mars.dti.ne.jp/ shigerus	リハビリテーション 情報 提供
www.mech.t.u-tokyo.ac.jp/ t60258/chosakai	視覚障害 聴覚障害 大学 調査

表 2 実験 1 の結果

Table 2 The result of the experiment 1.

数値: 検索結果中の順位
 文脈語: リハビリテーション, 理学療法

メディアデータ ID	意味的連想検索 リハビリテーション	パターンマッチ	
		リハビリテーション	理学療法
huri.sfc.keio.ac.jp/ kadota/	19		
keisys.tutpse.tut.ac.jp/ imura/http/			
tokyoweb.or.jp/tokyo-reha/	2	1	2
www.a.tsukuba-tech.ac.jp/docs/			
www.asahi-net.or.jp/ fy6k-kky/tube/	7		
www.asahi-net.or.jp/ JK8S-YMST/	6	7	3
www.bekkoame.ne.jp/ souei/			
www.bekkoame.ne.jp/ yamori/	30		
www.crl.go.jp/st/st821/			
www.fuji-mt.or.jp/ sifc/	20		
www.fukusi.softopia.pref.gifu.jp/reha-13/	12		
www.ijnet.or.jp/sekison/	17		
www.ipc.hiroshima-u.ac.jp/ exercise/			
www.japan-net.or.jp/ nrc/	4	19	
www.kanagawa-rehab.or.jp/ATAC/	5	13	
www.lwc-eirec.go.jp/	15		
www.mars.dti.ne.jp/ shigerus/	3	3	1
www.nise.go.jp/research/kogaku/hiro/	14		
www.portnet.co.jp/fukushi/	18	2	
www.pref.aichi.jp/hsc/inst/			
www.pref.ishikawa.jp/eisei/rihabiri/	1	17	
www.rehab.go.jp/	9	10	
www.ricoh.co.jp/net-messena/ACADEMIA/			
www.ritsumei.ac.jp/bkc/ tejima/index.html			
www.saigata-nh.go.jp/	22	4	5
www.sfc.keio.ac.jp/ e-ito/rehab/	10	9	
www.thx.inst.keio.ac.jp/ moon/	21		
www.uitec.ac.jp:8080/rehb/welcomej.html		8	
www1.fukui-ut.ac.jp/ENTRANCE/establish/	8		
www-hio.dj.kit.ac.jp/			

表 3 実験 1 の再現率

Table 3 The recall of the experiment 1.

数値: 再現率 (%: 小数点以下四捨五入)

問い合わせ	意味的連想検索	パターンマッチ
リハビリテーション	67	37
理学療法	67	17
視覚・聴覚障害	72	8
コンピュータ	63	14

表 4 実験 2-1 の結果

Table 4 The result of the experiment 2-1.

数値: 検索結果中の順位
文脈語: 医療・リハビリテーション, 医療・理学療法

メディアデータ ID	意味的連想検索 医療・リハビリ	パターンマッチ	
		医療・リハビリ	医療・理学療法
tokyoweb.or.jp/tokyo-reha/	1	2	2
www.mars.dti.ne.jp/shigerus/	3	1	1
www.rehab.go.jp/	11	9	
www.sfc.keio.ac.jp/e-ito/rehab/	4	17	
www.japan-net.or.jp/nrc/	6	16	
www.bekkoame.ne.jp/yamori/			
www.fuji-mt.or.jp/sifc/			
www.bekkoame.ne.jp/souei/			
www.asahi-net.or.jp/JK8S-YMST/	7	5	3
www.asahi-net.or.jp/fy6k-kky/tube/	10		
www.saigata-nh.go.jp/			
www.thx.inst.keio.ac.jp/moon/			
www.portnet.co.jp/fukushi/			
www.ipc.hiroshima-u.ac.jp/exercise/			
www.pref.aichi.jp/hsc/inst/			

表 5 実験 2-2 の結果

Table 5 The result of the experiment 2-2.

数値: 検索結果中の順位
文脈語: 運動・リハビリテーション, 運動・理学療法

メディアデータ ID	意味的連想検索 運動・リハビリ	パターンマッチ	
		運動・リハビリ	運動・理学療法
www.pref.ishikawa.jp/eisei/rihabiri/	1		

表 6 実験 2-3 の結果

Table 6 The result of the experiment 2-3.

数値: 検索結果中の順位
文脈語: 福祉工学・リハビリテーション工学, 福祉工学・理学療法

メディアデータ ID	意味的連想検索 福祉工学・リハ工	パターンマッチ	
		福祉工学・リハ工	福祉工学・理学療法
www.hio.dj.kit.ac.jp/			
www.ricoh.co.jp/net-messena/ACADEMIA/			
www.a.tsukuba-tech.ac.jp/docs/			
www.ijnet.or.jp/sekison/	4		
www.uitec.ac.jp:8080/rehb/		1	
www.ritsumei.ac.jp/bkc/tejima/			
www.fukusi.softopia.pref.gifu.jp/reha-13/	3		
www1.fukui-ut.ac.jp/ENTRANCE/establish/	5		
www.nise.go.jp/research/kogaku/hiro/	6		
keisys.tutpse.tut.ac.jp/imura/http/		6	
www.lwc-eirec.go.jp/	2	12	
www.crl.go.jp/st/st821/			
huri.sfc.keio.ac.jp/kadota/	8		
www.kanagawa-rehab.or.jp/ATAC/	1		

それぞれの実験結果から得られた再現率を表 3 に示す。

6.3 実験 2

ここでは, CHIME's SQUARE のリハビリテーションに関する 3 つのカテゴリ「医療・リハビリテーシ

ン」「運動・リハビリテーション」「福祉工学・リハビリテーション工学」のカテゴリ名を文脈語として意味的連想検索を行い, パターンマッチ検索との比較を行った。比較の基準として, CHIME's SQUARE の該当する

表7 実験2の再現率
Table 7 The recall of the experiment 2.

数値: 再現率 (%: 小数点以下四捨五入)

問い合わせ	意味的連想検索	パターンマッチ
医療・リハビリ	47	27
運動・リハビリ	100	0
福祉工学・リハビリ工学	50	21
障害児・教育	56	44
福祉・介護・機器	63	25
バリアフリー・車椅子	63	13
就労・支援	50	18

カテゴリとを正解とした時の再現率を用いた。各カテゴリに含まれる URL の数はそれぞれ 15, 1, 14 であったので、対象とする検索結果は、意味的連想検索、パターンマッチ検索ともにそれぞれ上位 15 位, 1 位, 14 位までを対象とした。なお、パターンマッチ検索では、上記の文脈語はそれぞれに含まれる単語の and 検索を行ない、その結果が対象とする数に満たないものについては不足分を or 検索の結果で補った。

実験 2-1 では、「医療・リハビリテーション」を問い合わせとして用いた。この結果を表 4 に示す。

実験 2-2 では、「運動・リハビリテーション」を問い合わせとして用いた。この結果を表 5 に示す。

実験 2-3 では、「福祉工学・リハビリテーション工学」を問い合わせとして用いた。この結果を表 6 に示す。

この他にも、「障害児・教育」「福祉・介護・機器」「バリアフリー・車椅子」「就労・支援」による問い合わせを行い、再現率を計測した。その結果を表 7 に示す。

6.4 考察

全ての実験を通じて、意味的連想検索では「リハビリテーション」という文脈を与えることによって、その同義語である「理学療法」に関連する WWW ドキュメントをも検索されている。

実験 1 では、すべての場合において、意味的連想検索の再現率がパターンマッチ検索の再現率よりも高い。また、「リハビリテーション」という文脈を与えることによって、その同義語である「理学療法」に関連する WWW ドキュメントも正しく検索されている。これは、意味的連想検索がパターンマッチ検索では検索できない、意味的相関関係の強いドキュメントをも検索できることを示している。また、パターンマッチによる検索結果よりも意味的連想検索の結果の方が専門家による分類により近い結果となっている。これにより、特定の専門分野において、下位の分類を持つやや抽象的な文脈に対して、意味的連想検索はパターンマッチよりも優れた結果を出すことがわかった。

実験 2 では、より具体的な下位概念に対して検索を

行っている。全ての実験例において、意味的連想検索の再現率がパターンマッチ検索の再現率よりも高い値を示している。この実験結果より、特定の専門分野における具体的な文脈に対しても、意味的連想検索はパターンマッチよりも優れた結果を出すことがわかった。

これらの実験結果は、特定分野におけるドキュメント検索において、ドキュメント群と専門用語集、一般的な辞書を用いてメタデータ空間を作成し、意味的連想検索を行う際の本方式の有効性を示している。

他方で、特に具体的な文脈を問い合わせとして与えた場合には、パターンマッチ検索との比較においては優れているものの、全体的には再現率にはまだ改善の余地があると言える。これは、メタデータ空間が、抽象度の高い部分では CHIME's SQUARE の専門分野と合致しているが、具体的な部分においてはやや不整合があるということを示している。この不整合を解消するためには、学習プロセスなどを用いて、メタデータ空間を専門分野により近づけていくプロセスが必要である。

7. 結論

本論文では、特定分野のドキュメント検索において、意味的連想検索を行う際のメタデータ空間の作成方式を提案し、その有効性についての検証を行った。本方式を用いることにより、障害者・福祉の分野に限らず、さまざまな分野でドキュメントの意味的連想検索を行うためのメタデータ空間を生成することが可能となる。

今日、インターネット上で、質の高い情報を探し出すことは必ずしも容易なことではない。一方で、人間がフィルタリングと分類によって質の高い情報を提供するためには、専門家が多くの労力と時間をデータの整備に割かなければならない。本方式を用いることにより、情報の登録の際に、コンテンツを表現する特徴づけを行うことで、さまざまな文脈に対応可能な質の良い情報フィルタリングサービスを、これまでより少ない労力によって提供することが可能となる。

意味的連想検索は、メディアオブジェクトをベクトル

化してメタデータ空間上に射影するため、語群による検索だけでなく、あるメディアオブジェクトと意味的に関係が強いメディアオブジェクトを検索することも可能である。この方式を用いて、意味的連想検索の結果のなかで、検索者の関心に最も近いメディアオブジェクトを指定して関連フィードバックを実現できる。

本研究では、意味的連想検索を行う際のメタデータ空間の作成手法について述べたが、検索対象となるメディアオブジェクトをどのようなメタデータによって特徴づけるかという手法は未だ確立されていない。今後は、メディアオブジェクトの特徴づけに関する研究に加えて、デジタル証明書などの認証情報や与信情報、アクセス許可情報といった、より複雑な構造を持つメディアデータを扱う方式について研究を行い、インターネット上のコミュニティの活動基盤整備を目指す。

謝辞

本研究において貴重な御助言を頂いた、慶應義塾大学 政策・メディア研究科 金子 郁容教授、同環境情報学部 村井 純教授、筑波大学 電子・情報工学系 北川 高嗣助教授に感謝致します。障害者・福祉関連の WWW ドキュメントを利用するにあたって、CHIME's SQUARE の 井村 保氏 (中部学院大学) および 藤田 桂氏 (慶應義塾大学) に御協力を頂きここに感謝致します。

参考文献

- 1) 郵政省編: 平成 10 年版 通信白書, 大蔵省印刷局, 1998.
- 2) Miyagawa, S., Kaneko, I. and Murai, J.: Making Network Communities Work, Proceedings of The Internet Global Summit INET'98, 1998.
- 3) Miyagawa, S., Kaneko, I., Inaoka, G. and Shimizu, K.: Constructing and Operating an Internet Site Supporting Network Communities, Proceedings of The International Workshop on Asia-Pacific Area advanced research information sharing technology, Internet Workshop '98(IWS'98), 1998.
- 4) 池田健一編: ネットワーキング・コミュニティ, 東京大学出版会, 1997.
- 5) 金子郁容, 松岡正剛, 下河辺淳 他: ボランティア経済の誕生, 実業之日本社, 1998.
- 6) Kitagawa, T. and Kiyoki, Y.: The mathematical model of meaning and its application to multidatabase systems, Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems, pp.130-135, April 1993.
- 7) Kiyoki, Y., Kitagawa, T. and Hayama, T.: A metadatabase system for semantic image search by a mathematical model of meaning, ACM SIGMOD Record, vol. 23, no. 4, pp.34-41, 1994.
- 8) Kiyoki, Y., Kitagawa, T. and Hitomi, Y.: A fundamental framework for realizing semantic interoperability in a multidatabase environment, Journal of Integrated Computer-Aided Engineering, Vol.2, No.1, pp.3-20, John Wiley & Sons, Jan. 1995.
- 9) 清木 康, 金子 昌史, 北川 高嗣: 意味の数学モデルによる画像データベース探索方式とその学習機構, 電子情報通信学会論文誌, D-II, Vol. J79-D-II, No.4, pp.509-519, 1996.
- 10) 北村 康彦: インターネット上での知的情報統合, 情報処理学会 アドバンスト・データベース・シンポジウム'98, 1998.
- 11) Deerwester, S., Dumais, S. T., Landauer, T. K., Furnas, G.W. and Harshman, R. A.: Indexing by Latent Semantic Analysis, Journal of the Society for Information Science, Vol.41, No.6, pp.391-407, 1990.
- 12) 芥子 育雄, 池内 洋, 小淵 保司: 意味ベクトルによる百科事典テキストデータベースの構築, 情報処理学会 アドバンスト・データベース・シンポジウム'93, 1993.
- 13) Salton, G., Allan, J. and Buckley, C.: Automatic structuring and retrieval of large text files, Communications of the ACM, Vol.37, No.2, pp.97-108, 1994.
- 14) Evans, D. A. et al.: Automatic indexing using selective NLP and first order thesauri, Proceedings of RIAO '91, pp.624-643, 1991.
- 15) Lewis, D.D. and Spark Jones, K.: Natural Language Processing for Information Retrieval, Communications of the ACM, Vol.39, No.1, pp.92-101, 1996.
- 16) Resnick, P. and Varian, H. R.: Recommender systems, Commun. ACM, vol.40, no.3, pp.56-58, 1997.
- 17) Paul, R., Neophytos, I., Mitesh, S., Peter, B. and John, R.: GroupLens: An Architecture for Collaborative Filtering of NetNews, Proceedings of CSCW'94, 1994.
- 18) Terveen, L., Hill, W., Amento, B., McDonald, D. and Creter, J.: PHOAKS a system for sharing recommendations Communications of The ACM, vol.40, no.3, 1997.
- 19) Kautz, H., Selman, B. and Shah, S.: Referral Web combining social networks and collaborative filtering, Communications of The ACM, vol.40, no.3, 1997.
- 20) Longman Dictionary of Contemporary English, Longman, 1987.

- 21) Ogden, C.K.: The General Basic English Dictionary, Evans Brothers Limited, 1940.
- 22) 改訂 介護福祉用語辞典, 中央法規出版, 1993.
- 23) 松本裕治, 北内啓, 山下達雄, 今一修, 今村友明: 日本語形態素解析システム「茶釜」 version1.0 使用説明書, NAIST Technical Report, NAIST-IS-TR97007, 1997.
- 24) 新村出編: 広辞苑第四版, 岩波書店, 1997.
 (平成10年12月20日受付)
 (平成11年3月27日採録)

(担当編集委員 小川泰嗣)



宮川 祥子 (学生会員)

1969年生。1994年一橋大学商学研究科経営学および会計学専攻修士課程修了。現在、慶應義塾大学政策・メディア研究科博士課程在学中。ネットワーク上のコミュニティ活動を支援するための情報システムに関する研究に従事。



清木 康 (正会員)

1978年慶應義塾大学工学部電気工学科卒業。1983年同大学院工学研究科博士課程修了。工学博士。同年、日本電信電話公社武蔵野電気通信研究所入所。1984年～1995年筑波大学電子・情報工学系講師、助教授。1996年～1998年慶應義塾大学環境情報学部助教授を経て、現在、慶應義塾大学環境情報学部教授。データベースシステム、知識ベースシステム、マルチメディアシステムの研究に従事。ACM, IEEE, 電子情報通信学会, 日本ソフトウェア科学会各会員。