

## 専門領域ドキュメント群を対象とした 意味的連想検索空間の生成方式

河本 穰<sup>†</sup> 清木 康<sup>††</sup>  
藤島 清太郎<sup>†††</sup> 相磯 貞和<sup>†††</sup>

本論文では、ベクトル空間上の特徴に関してのメタ知識を用いた意味的連想検索空間群の生成方式を提案する。本方式は、特定の専門領域を対象とした意味的連想検索の実現を目的として、すでに提案されている意味的連想検索方式を対象とし、語彙の専門性や語彙が属する対象分野のサブ領域（専門領域内におけるより細部の領域）に応じて分割し、意味的連想検索空間を生成する方式である。本方式は、幅広い観点にわたって体系的に整理された知識を用いる専門領域を対象とし、多様な利用目的に合致したドキュメント群の獲得を実現する空間の生成を可能とする。医学分野のドキュメントを対象とした検索実験により、本方式の有効性と実現可能性を示す。

### A Generation Method of Semantic Associative Search Spaces for Domain-specific Documents

MINORU KAWAMOTO,<sup>†</sup> YASUSHI KIYOKI,<sup>††</sup> SEITARO FUJISHIMA<sup>†††</sup>  
and SADAKAZU AISO<sup>†††</sup>

In this paper, we present a generation method of semantic associative search spaces for domain-specific research areas. In this method, semantic associative search spaces are created by separating a whole space into several sub-spaces by using meta-knowledge about feature words. The feature of our method is to realize several semantic associative search spaces by splitting a domain-wide search space into several application-wide search spaces. We have implemented a semantic associative search environment for domain-specific documents reflecting on application categories and significance of the feature words. We have performed several experiments by retrieving medical terms and medical documents in order to evaluate the feasibility and effectiveness of our method.

#### 1. はじめに

多面的な観点からの関連性の計量を行うことが可能なベクトル空間モデル<sup>1)</sup>による情報検索は、専門知識を用いた情報検索の実現方法の1つとして重要な研究対象である<sup>2)</sup>。体系的に整理された幅広い専門領域を扱う専門分野特有の知識を意味的連想検索空間として構成し、文書・メディアデータ群を対象とした意味的な連想検索を実現する方式として意味の数学モデルによる意味的連想検索<sup>3)-5)</sup>が提案されている。意味的な連想検索環境の実現は、専門性の高い情報源からの

適切な情報獲得を行うために重要である。意味的連想検索に用いる検索空間の実現は検索空間の情報源として用いる資料から、専門知識に関する単語間の知識を定義として記述し、意味的検索空間を実現することによって行われている<sup>6)-9)</sup>。専門性の高い分野を対象とした意味的連想検索の実現においては、専門度および利用目的によって獲得すべき文書・メディアデータ群は異なるため、専門度および利用目的に応じた意味的連想検索空間の実現が重要である。

本論文では、専門度および利用目的に応じた意味的連想検索空間の生成方式を提案する。この生成方式は、医学分野の意味的連想検索を対象とした意味的連想検索空間の具体的、かつ、実際的な実現プロセスにおいて適用する空間生成方式であり、専門性の高い分野を対象とした意味的連想検索空間生成のための実際的な方式である。本論文では、実際の医学分野の意味的連想検索空間生成および検索空間を用いた検索実験によ

<sup>†</sup> 慶應義塾大学大学院政策・メディア研究科  
Graduate School of Media and Governance, Keio University

<sup>††</sup> 慶應義塾大学環境情報学部  
Faculty of Environmental Information, Keio University

<sup>†††</sup> 慶應義塾大学医学部  
School of Medicine, Keio University



デルによる意味的連想検索方式については文献 3)~5) で詳しく述べられている。

専門分野領域を対象とした意味的連想検索によるデータ獲得のための検索空間の実現として、次の 5 プロセスの構成をとった実現が行われている<sup>7)~9)</sup>。

- (1) 検索空間を構成する単語群の設定およびデータ行列の値の決定の根拠とする情報源を決定する。
- (2) 検索空間上のベクトルに相当するキーワード(以下、基本データ語句)( $w_1, w_2, \dots, w_t$ )を決定する。基本データ語句には情報源の見出しを用いる。
- (3) 特徴語句の候補群( $f'_1, f'_2, \dots, f'_m$ )を決定する。辞書により意味的検索空間を構成する場合、特徴語句には基本データ語句を説明している語句群を用いる。
- (4) 特徴語句の候補群からの選定を行い、特徴語句群( $f_1, f_2, \dots, f_n$ )を決定する。
- (5) 特徴語句と基本データ語句の関連の強さを示すデータ行列( $M$ )を作成する。

本生成方式は、専門分野領域を対象とした意味的連想検索によるデータ獲得のための検索空間の実現を行うための方式であり、6 プロセスによって構成される。本方式では専門性の高い領域の空間を生成する方式として、すでに行われている(1)~(3)および(5)のプロセスに加え、(4')と(6')を新たに導入する。

- (4') 特徴語句の候補群に対して利用目的と専門度の各属性に対応する値を設定する。
- (5) 特徴語句と基本データ語句の関連の強さを示すデータ行列( $M$ )を作成する。
- (6') データ行列から利用目的および専門度の属性に従って複数の意味的検索空間群を生成する。

## 2.1 意味の数学モデル<sup>3)~5)</sup>の基本構成

- (1) メタデータ空間  $MDS$  の設定：  
検索対象となるメディアデータをベクトルで表現したデータをマッピングするための正規直交空間(以下、メタデータ空間  $MDS$ )を設定する。
- (2) メディアデータのメタデータをメタデータ空間  $MDS$  へ写像：  
設定されたメタデータ空間  $MDS$  へメディアデータのメタデータをベクトル化し写像する。これにより、同じ空間に検索対象データのメタデータがメタデータ  $MDS$  空間上に配置されることになり、検索対象データ間の意味的な関係を空間上での距離として計算することが可能となる。

メディアデータ  $P$  には、メタデータとして  $t$  個の印象語  $w_1, w_2, \dots, w_t$  が、以下のように付与されていることを前提としている。

$$P = \{w_1, w_2, \dots, w_t\}. \quad (1)$$

各メタデータ語句は、ベクトル表現された特徴を持っている。

$$w_i = (f_{i1}, f_{i2}, \dots, f_{in}). \quad (2)$$

各メディアデータは、メタデータとして付与されている  $t$  個のメタデータ語句が合成され、ベクトル表現された後、メタデータ空間  $MDS$  へ写像される。

- (3) メタデータ空間  $MDS$  の部分空間(意味空間)の選択

検索者は、与える文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。このコンテキストを用いて、メタデータ空間  $MDS$  に各コンテキストに対応するベクトルを写像する。これらのベクトルは、メタデータ空間  $MDS$  において合成され、意味重心を表すベクトルが生成される。意味重心から各軸への射影値を相関とし、閾値を超えた相関値(以下、重み)を持つ軸からなる部分空間(以下、意味空間)が選択される。

- (4) メタデータ空間  $MDS$  の部分空間(意味空間)における相関の定量化

選択されたメタデータ空間  $MDS$  の部分空間(意味空間)において、メディアデータベクトルのノルムを検索語列との相関として計量する。これにより、与えられたコンテキストと各メディアデータとの相関の強さを定量化する。この意味空間における検索結果は、各メディアデータを相関の強さについてソートしたリストとして与えられる。

## 2.2 専門分野を対象とした意味的連想検索空間

本節では、意味的連想検索空間の生成方式における空間構築のための情報源の選定プロセス、および、その情報源を用いた空間生成のプロセスを示す。

### 2.2.1 専門分野を対象とした意味的連想検索空間のための情報源

専門分野を対象とした意味的連想検索空間を構成するためのデータ行列作成の際に参照する情報源となる資料の選定プロセスを示す。これは 2 章で示した本方式を構成するプロセスの(1)に相当する。

意味的検索空間のための情報源は以下の複数の観点を検討して決定する。

- (1) 幅広い概念に包括的に言及している  
検索空間の語彙数はそのまま検索空間が把握可能なアトミックな概念の量となる。すなわち、語彙数の多い空間は多くの概念を間接的な計量ではなく直接に取り扱うことができる。
- (2) 対象とする専門分野において広く認知されている  
記述内容に関して信頼性が高い。
- (3) 文書が構造的に組織されている  
半自動的な処理を行うことで人手による作業の支援が行いやすい。
- (4) 幅広い記述が行われている  
専門分野における概念間の関係を記述するには語義の説明のみでは十分ではなく、応用的な側面に踏み込んだ記述が行われるのが望ましい。

### 2.2.2 専門分野を対象とした意味的連想検索空間の実現プロセス

専門分野ドキュメント群を対象とした意味的連想検索空間の実現は、次のプロセスによって行われる。これは、2章で示した本方式を構成するプロセスの(2)、(3)、(4)、(5)に相当する。

- (2-1) 専門知識の根拠とする文献資料中の対象専門領域に関連する章(以下、本文)を抜粋する。
- (2-2) 本文の文章中より章、節、小見出し等からキーワードとなりうる語句を選び、基本データ語句として抜粋する。HTML等の構造が抽出可能な文書を対象として扱う場合はソフトウェアによる作業支援を適用することができるが、最終的な決定は人手による監修を経ることが望ましい。
- (3) 本文の記述を参照し、基本データ語句群の説明に用いられている語句を抜粋し、特徴語句群の候補語句群とする。対象文書によっては、特定の文書構造に従って検出できる場合もあるが、不可能な場合は手作業により行う。
- (4) 該当分野の用語集等から特徴語句とする語句を選び、検索空間の特徴として選択する。
- (5) 本文の記述を参照し、基本データ語句と特徴語句の関連性を記述した検索空間を作成する。ここで、関連の強さは通常  $-1, 0, 1$  の値から、肯定的な関連の場合に  $1$ 、無関係の場合に  $0$ 、否定的な関連(たとえば、低血圧に対して高血圧等の関係)の場合に  $-1$  と決定する。具体的には本文を走査し基本データ語句の説明文中に出現している特徴語句に  $1$  を、その他の語句に  $0$  を付与することで下地となる候補のデータを作成することが可能である。記述されている概念

群を直接関係づけることに重点を置いて編集されている専門用語辞典類等ではこのような下地のデータをそのまま利用可能である場合もある。生成された下地となる候補のデータの質が十分でない場合は人手による修正を行う。

### 2.3 単語レベルのメタ知識による検索空間の分割

本節では、特定の専門分野を対象とした検索空間用のデータ行列から語句に関するメタ知識を用いた空間を生成するためのデータ行列群を生成するプロセスについて示す。本プロセスは、すでに実現されている専門分野ドキュメントを対象とした意味的連想検索空間のためのデータ行列に対する分割プロセスである。2章で示した本方式を構成するプロセスの(6')に相当する。

本プロセスでは、ベースレベル検索空間からの検索空間の生成を行う。

- (1) 検索空間  $M_U$  の特徴群  $f_1, f_2, \dots, f_n$  に対してメタ知識を利用した属性  $a_1, a_2, \dots, a_l$  を付与する。
- (2) サブ領域別のカテゴリに属する特徴によってまとめたサブ空間  $M_{a_1}, M_{a_2}, \dots, M_{a_l}$  を設定する。
- (3)  $M_{a_1}, M_{a_2}, \dots, M_{a_l}$  それぞれを対象として、メタデータ空間  $MDS$  を生成する。

## 3. 医学分野における実現

専門性の高い分野における本生成方式の実現として、実際の医学・肺・呼吸器領域を対象とした意味的連想検索空間の実現を行った。検索空間の実現の際に利用する専門知識の情報源として、Cecil Textbook of Medicine<sup>10)</sup>(以下 Cecil)を用いて実現を行った。

- (1) 肺・呼吸器分野に関係している辞典である Cecil の文章中(具体的には CHAPTER 81 から CHAPTER 99 に相当)より章、節、小見出しから病名に相当するものを基本データ語句の候補群(132個、 $t'_1, t'_2, \dots, t'_{132}$ )として抜粋する。
- (2) Cecil の記述を参照し、基本データ語句群の説明に用いられている語句を抜き出し、Cecil の索引に掲載されている語句と医学の辞典である Stedman's Medical Dictionary<sup>11)</sup> および循環器・呼吸器関連の用語集である Stedman's Cardiovascular & Pulmonary Words<sup>12)</sup> に掲載されている語句を特徴語句の候補語句群( $f'_1, f'_2, \dots, f'_m$ )として選択する。
- (3) 特徴語句の候補語句群に対して専門度および語

表 1 サブ領域の一覧

Table 1 List of subdomains.

| 名称                                  | 日本語訳     | 略称  |
|-------------------------------------|----------|-----|
| Clinical Manifestations             | 症候       | CM  |
| Definitions and Diagnosis           | 定義と診断    | DD  |
| Prevention, Treatment and Prognosis | 予防・治療・予後 | PTP |
| Pathology and Pathophysiology       | 病理・病理生理  | PP  |
| Epidemiology                        | 疫学       | EP  |
| Etiology                            | 病因       | ET  |

表 2 サブ領域と見出しの対応表

Table 2 Mappings between Subdomains and Headlines.

| 見出し名                                | サブ領域 |
|-------------------------------------|------|
| clinical presentation               | CM   |
| clinical manifestation              | CM   |
| signs and symptoms                  | CM   |
| definition                          | DD   |
| diagnosis                           | DD   |
| definitions                         | DD   |
| prevention, treatment and prognosis | PTP  |
| treatment and prognosis             | PTP  |
| thrombolytic therapy                | PTP  |
| pathology and pathophysiology       | PP   |
| pathology of asthma                 | PP   |
| epidemiology                        | EP   |
| epidemiology and statistics         | EP   |
| etiology                            | ET   |
| incidence and prevalence            | ET   |

句が属するサブ領域の 2 属性を付与する。

- (4) 前段階で得られた専門度および所属サブ領域を反映した、各データ行列を作成し、それぞれを検索空間として設定する。

本方式による意味的連想検索空間の構築は、医学分野の肺・呼吸器に関連する用語・ドキュメントを対象とした検索空間の実現として行った。本章では、医学分野における意味的連想検索空間の実現およびサブ領域と専門度を用いた検索空間の分割の具体例を示す。

サブ領域  $M_{c_1}, M_{c_2}, \dots, M_{c_6}$  として表 1 に示す 6 サブ領域を採用した。6 サブ領域の選定は、Cecil の保有する章、節等の文書構造の見出し文字列および記載内容を考慮し、慶應義塾大学医学部において肺・呼吸器系の専門家の判断により表 2 のように決定されている。

特徴語の選定の基準となる専門度の高低の判定については慶應義塾大学医学部において肺・呼吸器系の専門家の手作業によって行った。ここで、医学的な見地における専門度は次の基準に従った。

- (1) 記述している対象の病名を特徴づけるのに必要な語句を最も専門的な語句群とする（専門度：

表 3 特徴語句のサブ領域

Table 3 Subdomains of feature words

| 特徴語句             | サブ領域                |
|------------------|---------------------|
| asthma           | PTP, DD, PP         |
| hyperventilation | DD, ET, CM          |
| tuberculosis     | PTP, DD, ET, CM     |
| terbutaline      | PTP                 |
| fevers           | PTP                 |
| chest pain       | PTP, DD, PP, ET, CM |
| lung disease     | PTP, DD, PP, ET, CM |
| poliomyelitis    | (該当なし)              |
| echocardiogram   | PTP, DD             |
| physiotherapy    | PTP                 |
| rehabilitation   | PTP                 |

表 4 特徴語句の専門度

Table 4 Significance values of feature words.

| 特徴語句             | 訳         | 専門度 |
|------------------|-----------|-----|
| asthma           | 喘息        | A   |
| hyperventilation | 過換気       | A   |
| tuberculosis     | 結核        | A   |
| terbutaline      | テルブタリン    | B   |
| fevers           | 熱         | B   |
| chest pain       | 胸痛        | B   |
| lung disease     | 肺病        | C   |
| poliomyelitis    | 灰白髄炎      | C   |
| echocardiogram   | 心エコー図     | C   |
| physiotherapy    | 物理療法      | D   |
| rehabilitation   | リハビリテーション | D   |

A)。

- (2) 肺・呼吸器の分野に関連している医学分野の語句を 2 番目に専門的な語句群とする（同：B）。
- (3) 上記に含まれない医学分野の語句を 3 番目に専門的な語句群とする（同：C）。
- (4) 一般的な語句群（同：D）。

各特徴語句とサブ領域の対応関係を表 3 に、各特徴語句と専門度の対応関係を表 4 に示した。

#### 4. 関連研究

文書検索やクラスタリング等に用いる多次元ベクトル空間の次元数を縮減する手法には LSI<sup>13)</sup>、Random projection を用いる手法<sup>14)</sup>、LPI<sup>15)</sup> 等が知られている。これらの手法は、データ行列からの計算によって直接縮減されたベクトル空間を生成しており、また、応用分野に対応したパラメータを用いない。これに対して、本方式は、次元縮減の操作の情報源として専門家による専門知識に基づく操作を行っている点、次元数の縮減操作について特定の利用目的や規模をパラメータとして指定する点が大きく異なっている。特定の知識を用いて次元の縮減を行う手法として、索引語間のクラス・メンバシップ関係を用いて次元の縮減を

表 5 生成されたデータ行列と空間の構成要素, 名称, 規模

Table 5 The elements, expressions and sizes of generated data matrices and search spaces.

| 本文中の表記                 | カテゴリ | 専門度        | 図凡例中の表記 | データ行列             | 直交空間の次元数 |
|------------------------|------|------------|---------|-------------------|----------|
| $Space_{A+B+C+D}$      | 指定なし | A, B, C, D | w-abcd  | $131 \times 3954$ | 3136     |
| $Space_{A+B+C}$        | 指定なし | A, B, C    | w-abc   | $131 \times 2163$ | 1658     |
| $Space_{A+B}$          | 指定なし | A, B       | w-ab    | $131 \times 1166$ | 853      |
| $Space_A$              | 指定なし | A          | w-a     | $131 \times 903$  | 530      |
| $Space_{A+B+C+D}^{DD}$ | DD   | A, B, C, D | p-abcd  | $131 \times 1668$ | 1204     |
| $Space_{A+B+C}^{DD}$   | DD   | A, B, C    | p-abc   | $131 \times 855$  | 627      |
| $Space_{A+B}^{DD}$     | DD   | A, B       | p-ab    | $131 \times 512$  | 328      |
| $Space_A^{DD}$         | DD   | A          | p-a     | $131 \times 358$  | 223      |

行う教師付き次元縮減アルゴリズムを用いる手法<sup>16)</sup>が提案されている。

専門分野のドキュメント群を対象とした意味的連想検索空間の実現を行った方式として, 文献 6)~9) 等がある。これらの方式が, 主に専門分野に関する辞書や教科書を用いて専門知識を反映させた検索空間を実現したのに対し, 本方式はこのようにして実現された専門分野検索空間に対する縮減の操作を行うことにより規模を縮小している点が特徴であり, これによってデータ行列の修正の負荷を軽減し, 検索者の目的をよりの確に反映させ, 検索性能を向上させることが可能となる。

## 5. 実験

本方式の実現可能性および有効性を示すために行った実験およびその結果について示す。

### 5.1 実験環境

医学分野のドキュメントを対象とする検索空間を実現し, 実験を行った。検索空間の設定については, 3章で述べた検索空間を使用した。本実験で用いているサブ領域は, 本実験が想定している応用目的である Problem-based learning<sup>17)</sup> (以下, PBL) の学習における有効性の高いサブ領域という基準により “DD” (Definitions and Diagnosis; 定義と診断) を選定し, 比較対象としてサブ領域による分割を使用しない検索空間を用いた。専門度  $p$  を選択して用いた空間を  $Space_p$  とし, サブ領域  $s$  を選択して用いた空間を  $Space^s$  とする。本実験において, 表 5 に示す各検索空間をそれぞれ実現し, 比較を行った。

実験 A では, 特定の利用目的において, 本方式により実現された検索空間が適切な文書を応答可能であることを検証した。実験 A における検索空間の利用目的は, PBL による学習を支援する教科書検索の教育システムと位置づけ, PBL で学習を行う医学部の学生が症例から連想するキーワードを用いて, 教科書を対象とした検索を行った。実現された検索空間を用

いて検索を行い, 関連する病気の情報を獲得するシステムとしての有用性という観点からの評価を行った。

本実験は, 提案方式の検索精度に着目した従来方式との比較実験として位置づけることができる。

$Space_{A+B}$  は, 検索空間として, 従来方式によって実現されたデータを保持している。すなわち, 利用目的による絞り込みを行っておらず, 専門分野の特徴となる単語群 ( $A+B$ ) として該当する単語群を特徴語の選択基準として用いている空間である。検索精度の比較を行う実験である実験 A および実験 B において, 従来の検索空間の生成方法との比較として,  $Space_{A+B}$  を従来方式によって実現された空間として,  $Space_{A+B}^{DD}$  および  $Space_{A+B+C}^{DD}$  を提案方式によって実現された空間として, 評価を行った。

実験 B では, 本方式により実現した意味的検索空間群が, 有用かつ発見的な検索を行うことが可能であることを, 適合と評価された検索結果文書群に付与されたメタデータ語句群と文脈語句との照合を行うことによって評価を行った。

実験 C では, 本方式によって実現した検索空間による発見的な検索の作用がどのように実現されているかについて検証し, 考察を行った。

### 5.2 検索対象文書群のメタデータの生成

本章における文書の検索実験において用いられている検索対象文書のメタデータ作成方法について述べる。Cecil<sup>10)</sup> から肺・呼吸器に関する章を抜き出し, 検索対象とした。抜き出した文章を, 特定の病気についての記述が単位となるよう分割し, それぞれを検索対象文書とした。

このようにして設定した検索対象文書群を対象として, 単語の出現頻度 (tf × idf 法<sup>18)</sup>) によるメタデータ抽出方式<sup>19)</sup> を用いて各文書について tf × idf の値が大きい特徴語のうち, 最大上位 10 個までのメタデータを抽出し, それぞれ検索対象文書のメタデータとして設定した。

表 6 実験 A の結果 (集計値によるもの)  
Table 6 Result of Experiment-A (in total).

| 文脈                | 検索空間                 | 評価値 | 文脈                | 検索空間                 | 評価値 |
|-------------------|----------------------|-----|-------------------|----------------------|-----|
| Rheumatoid-factor | $Space_{A+B}^{DD}$   | 643 | dyspnea           | $Space_{A+B}^{DD}$   | 840 |
| Rheumatoid-factor | $Space_{A+B+C}^{DD}$ | 619 | dyspnea           | $Space_{A+B+C}^{DD}$ | 810 |
| Rheumatoid-factor | $Space_{A+B}$        | 554 | dyspnea           | $Space_{A+B}$        | 812 |
| Rheumatoid-factor | $Space_{A+B+C}$      | 601 | dyspnea           | $Space_{A+B+C}$      | 793 |
| Cough             | $Space_{A+B}^{DD}$   | 753 | Smoking           | $Space_{A+B}^{DD}$   | 586 |
| Cough             | $Space_{A+B+C}^{DD}$ | 722 | Smoking           | $Space_{A+B+C}^{DD}$ | 650 |
| Cough             | $Space_{A+B}$        | 726 | Smoking           | $Space_{A+B}$        | 570 |
| Cough             | $Space_{A+B+C}$      | 756 | Smoking           | $Space_{A+B+C}$      | 492 |
| FEV1              | $Space_{A+B}^{DD}$   | 521 | Pleural-effusions | $Space_{A+B}^{DD}$   | 800 |
| FEV1              | $Space_{A+B+C}^{DD}$ | 499 | Pleural-effusions | $Space_{A+B+C}^{DD}$ | 730 |
| FEV1              | $Space_{A+B}$        | 434 | Pleural-effusions | $Space_{A+B}$        | 664 |
| FEV1              | $Space_{A+B+C}$      | 497 | Pleural-effusions | $Space_{A+B+C}$      | 670 |
| Asbestos          | $Space_{A+B}^{DD}$   | 629 | chest-pain        | $Space_{A+B}^{DD}$   | 551 |
| Asbestos          | $Space_{A+B+C}^{DD}$ | 549 | chest-pain        | $Space_{A+B+C}^{DD}$ | 576 |
| Asbestos          | $Space_{A+B}$        | 505 | chest-pain        | $Space_{A+B}$        | 538 |
| Asbestos          | $Space_{A+B+C}$      | 494 | chest-pain        | $Space_{A+B+C}$      | 573 |

### 5.3 検索文脈群の設定

実験で用いた 8 つの文脈は、「呼吸器疾患ケーススタディ」<sup>20)</sup>より、無作為に抽出した症例を選び、症例中に出現するキーワード群から、医学の専門家の監修の下、学習環境の利用者として想定される医学の学生によって選ばれたキーワード 8 つを選ぶことによって設定した。

実際の医学分野の教科書 (Cecil<sup>10)</sup>) を検索対象とした検索を行い、実現した 4 つの検索空間の検索精度を比較することにより、実現した検索空間の有効性を検証する。検索結果のうち  $k$  位 ( $k = 20$  とした) までの文書に医学の専門家が目を通し、検索語で入力された事象から強く可能性が疑われる病名にあたるものに適合度 5 を、可能性があると判断されるものに適合度 3 を、可能性なしと判断されるものに適合度 1 を、適合度 5 と 3 の中間に位置するものに適合度 4 を、適合度 3 と 1 の中間に位置するものに適合度 2 を、という基準で適合度を決定した ( $s_r$ :  $r$  位に検索された病名の適合度)。

### 5.4 実験 A

各空間を用いて検索を行った結果による適合度を集計した。検索精度の評価は以下の基準で行った ( $v_\alpha$  は空間  $\alpha$  の検索性能の集計値)。

$$v_\alpha = \sum_{i=1}^k s_i(k-i) \quad (3)$$

$s_r(k-r)$  を  $r$  位に検索された文書の検索精度評価値 (“評価値”) と呼ぶ。実験 A の評価値の分布を図 2~9 に、集計値を表 6 に示す。表中の凡例における記号と空間の対応関係は表 5 に示すとおりである。

#### 5.4.1 考察：実験 A

実験 A の結果、表 6 に示すとおり 8 つの文脈中 5 つの文脈で空間  $Space_{A+B}^{DD}$  が最も高い評価値を得た。また、文脈 “Cough” で検索を行った場合には、最も高い評価値を示したのは空間  $Space_{A+B+C}$  であったが、空間  $Space_{A+B+C}^{DD}$  もほぼ遜色ない評価値を得た。文脈 “Smoking” と文脈 “chest-pain” による検索を行った場合の評価値は、ともに空間  $Space_{A+B+C}^{DD}$  が最も高い評価値を得た。

FEV1 (図 5)、Cough (図 6)、chest-pain (図 7)、dyspnea (図 8) を文脈として指定した検索を行った場合には、実験を行った各空間間における比較において、非常に似通った検索結果を返している。FEV1 を文脈とした検索においては、少数の例外を除いて、ほとんどの検索結果における適合度が 2 (多くの検索結果が不適合) であり、これは、FEV1 に適合する検索対象が少数しか存在しなかったことに起因している。同様に Cough を文脈とした検索においては、ほとんどの検索結果が適合度 3 もしくは 4、chest-pain を文脈とした検索においては、ほとんどの検索結果が適合度 2 もしくは 3、dyspnea を文脈とした検索においては、ほとんどの検索結果が適合度 4 となっており、これらの検索結果を表した各図は線形に集約した傾向を示す。Smoking (図 2)、Asbestos (図 3)、Pleural-effusions (図 4)、Rheumatoid-factor (図 9) を文脈として指定した検索を行った場合には、適合度が高い文書と低い文書が混在して含まれており、これらの文脈の指定においては各空間間の比較においてそれぞれ大きく異なった性質を示している。

一般に、医学の見地において Smoking (図 2)、

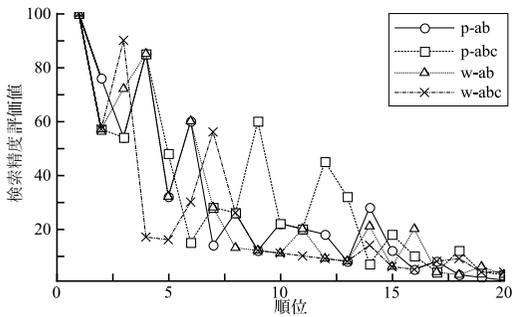


図 2 実験 A の結果 (文脈: Smoking)

Fig. 2 Result of Experiment A (Context: Smoking).

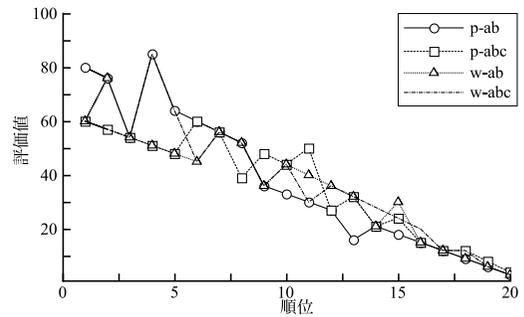


図 6 実験 A の結果 (文脈: Cough)

Fig. 6 Result of Experiment A (Context: Cough).

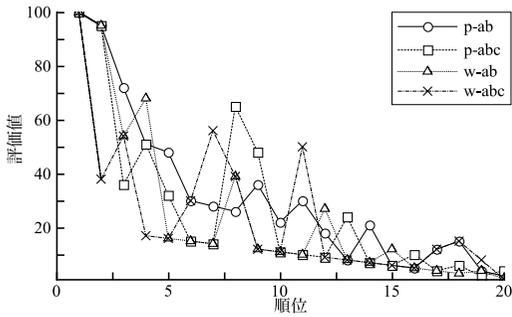


図 3 実験 A の結果 (文脈: Asbestos)

Fig. 3 Result of Experiment A (Context: Asbestos).

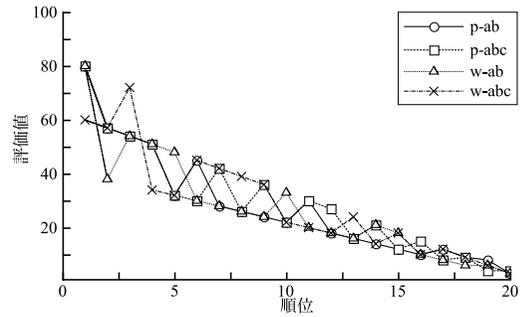


図 7 実験 A の結果 (文脈: chest-pain)

Fig. 7 Result of Experiment A (Context: chest-pain).

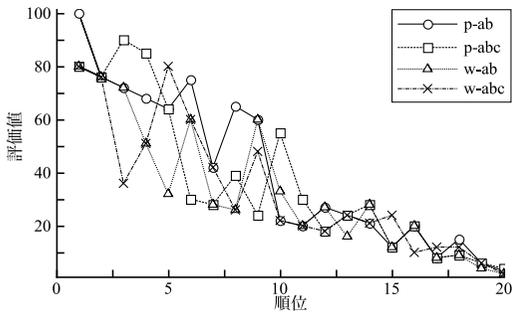


図 4 実験 A の結果 (文脈: Pleural-effusions)

Fig. 4 Result of Experiment A (Context: Pleural-effusions).

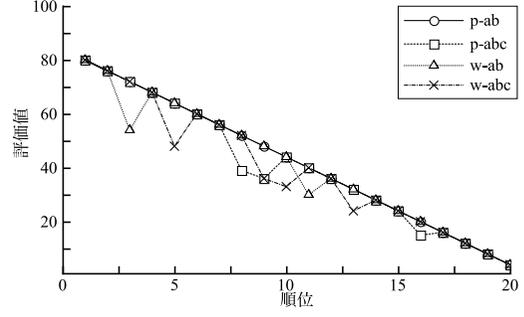


図 8 実験 A の結果 (文脈: dyspnea)

Fig. 8 Result of Experiment A (Context: dyspnea).

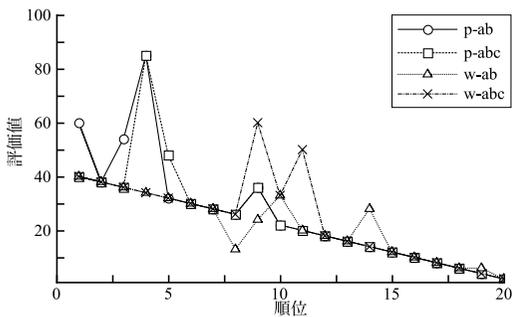


図 5 実験 A の結果 (文脈: FEV1)

Fig. 5 Result of Experiment A (Context: FEV1).

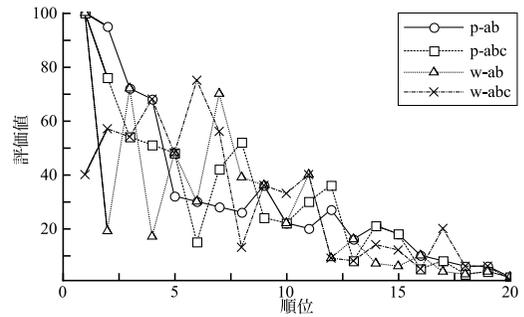


図 9 実験 A の結果 (文脈: Rheumatoid-factor)

Fig. 9 Result of Experiment A (Context: Rheumatoid-factor).

Cough (図6), chest-pain (図7), といった多種の病気に共通して関連する症状を単体の語句として検索の文脈に用いた場合, 専門家による診断を行う場合でも情報の不足が生じており, このような場合には本方式による有意な改善が見られない結果となっている. これらの語句を検索の文脈として用いる検索は, より文脈を明確にするための別の語句を加え, 検索を行うのが望ましいと考えられる. dyspnea (図8) を検索の文脈に用いた場合, 提案方式によって生成された空間における評価値が比較的高く, 検索性能が向上する傾向が見られる. Rheumatoid factor, Asbestos, Pleural effusions といった語句による文脈の指定を行った場合, 検索結果の評価値から判断して提案方式による空間生成により検索精度の改善を実現できたことが分かる. これら3つの文脈を用いて検索を行った場合について, 特に提案方式によって実現した空間である空間  $Space_{A+B}^{DD}$  (p-ab), および, これに次いで空間  $Space_{A+B+C}^{DD}$  (p-abc) は高い評価値を示している (図10における太線がこれらの3文脈を用いた検索結果の評価値に対応する).

以上の結果により, 医学の専門的な観点から実用的に利用される可能性が高い文脈語句群を用いて検索を行った場合において, これまでに行われてきた意味的検索空間の実現手法による検索空間と比較して, 提案方式により実現した空間  $Space_{A+B}^{DD}$  の検索性能の高さが示された.

### 5.5 実験 B

本方式により実現した  $Space_{A+B}^{DD}$ ,  $Space_{A+B+C}^{DD}$ ,  $Space_{A+B}$ ,  $Space_{A+B+C}$  の各空間上において, 文脈 Pleural-effusions を指定して検索を行った際の検索結果を, 各空間が実現している発見性の比較を目的として表7に示す. ここで, 発見性とは上位 (本実験においては上位20件) に検索された文書  $P_j$  のメタデータ中において文脈に用いた語  $u$  が出現しておらず, かつ適合度が4以上と判定された文書の数によって求めることとする. 表7には, 検索に対する問合せ結果として上位20件までの文書と, 対応する適合度が示されている. また, 適合度が4以上と判定された適合性が高い文書について, 文書のメタデータ中に文脈と一致する語句が含まれていたかどうかを示している. 文書のメタデータと文脈との一致が見られない検索結果文書群のうち, 適合度が4以上の値を持つ文書の件数を適合度ごとに集計して比較した (表8).

#### 5.5.1 考察: 実験 B

実験Bの結果, 空間  $Space_{A+B}^{DD}$  での検索において, 文書のメタデータ中に文脈と同じ語句を含まず, かつ

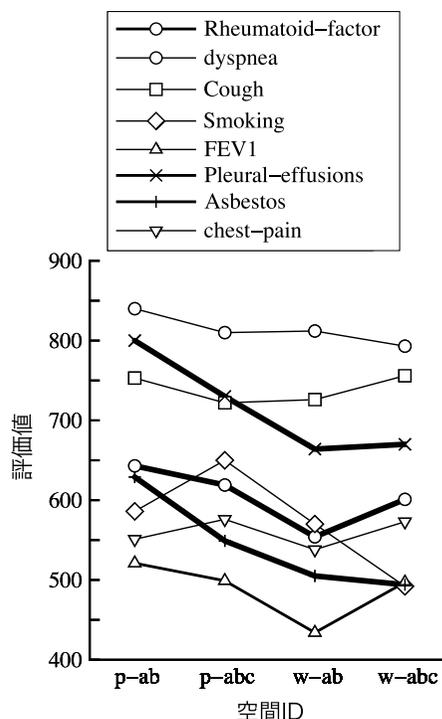


図10 実験Aの結果

Fig. 10 Result of Experiment A.

適合度が5と判定された文書が3件得られた. 空間  $Space_{A+B+C}^{DD}$  の検索結果においては, 文書のメタデータ中に文脈と同じ語句を含まず, かつ適合度が5と判定された文書は1件であり, 空間  $Space_{A+B}$  と空間  $Space_{A+B+C}$  では文書のメタデータ中に文脈と同じ語句を含まず, かつ適合度が5と判定された文書は検索することができなかった.

空間  $Space_{A+B}^{DD}$  および  $Space_{A+B+C}^{DD}$  における検索結果中に発見的な検索結果が出現し, 対照的に  $Space_{A+B+C}$  および  $Space_{A+B+C}$  における検索結果中に発見的な検索結果が出現しなかった事象は2つの要因が重なっていることによるものと考えられる.

1つの要因として, 空間  $Space_{A+B}^{DD}$  および  $Space_{A+B+C}^{DD}$  の実現の過程において, 利用目的に合致した特徴を選択することによって文脈と検索対象文書の意味的な関係の把握をより正確に行うことが可能となったことがあげられる. 医学における単語間の意味的な関係は利用目的等の文脈に応じて多様に存在する. 空間  $Space_{A+B}^{DD}$  および  $Space_{A+B+C}^{DD}$  の構成要素となっている特徴語句群が本実験で応用対象として決定されているPBL用の教育用途という利用目的に対応する「定義・診断」カテゴリを用いて選択されていることによって雑音の少ない意味的な関係を検出する

表 7 実験 B：Pleural-effusions (胸水) を文脈として用いた検索結果  
Table 7 Experiment-B: Retrieval results (context phrase: pleural effusions).

| 空間ID | p-ab     |              | p-abc |             | w-ab |        | w-abc          |             |    |
|------|----------|--------------|-------|-------------|------|--------|----------------|-------------|----|
| 専門度  | A+B      |              | A+B+C |             | A+B  |        | A+B+C          |             |    |
| 利用目的 | 定義: 診断のみ |              |       | 定義: 診断のみ    |      |        | 利用目的による絞り込みなし  |             |    |
|      | 順位       | 検索対象ID       | 適合度   | パターン<br>の一致 | 順位   | 検索対象ID | 適合度            | パターン<br>の一致 |    |
|      | 1        | 114_OTHER.IN | 5     | あり          |      | 1      | 112_OTHER.INFE | 4           | あり |
|      | 2        | 36_ANTINEOP  | 4     | あり          |      | 2      | 118_OTHER.CON  | 4           | あり |
|      | 3        | 118_OTHER.C  | 4     | あり          |      | 3      | 36_ANTINEOPLA  | 4           | あり |
|      | 4        | 65_RADIATION | 4     | あり          |      | 4      | 114_OTHER.INFE | 5           | あり |
|      | 5        | 112_OTHER.IN | 4     | あり          |      | 5      | 85_OTHER.EXTF  | 4           |    |
|      | 6        | 120_MALIGNA  | 5     |             |      | 6      | 50_AUTOSOMAL   | 2           |    |
|      | 7        | 21_LYMPHOC   | 3     |             |      | 7      | 9_EMPHYSEMA    | 2           |    |
|      | 8        | 115_HEMOTHK  | 5     | あり          |      | 8      | 34_ANTI-INFLAN | 3           |    |
|      | 9        | 105_THORACK  | 5     |             |      | 9      | 19_DESQAMAT    | 2           |    |
|      | 10       | 2_HYPOVENTI  | 2     |             |      | 10     | 120_MALIGNANT  | 5           |    |
|      | 11       | 102_KYPHOSC  | 2     |             |      | 11     | 43_INTERSTITIA | 3           |    |
|      | 12       | 124_Lymphore | 3     |             |      | 12     | 7_CHRONIC.BRC  | 2           |    |
|      | 13       | 32_DRUG-IND  | 3     |             |      | 13     | 68_HYDROCARB   | 3           |    |
|      | 14       | 98_THE_DIAPH | 3     |             |      | 14     | 117_IMMUNOLOI  | 4           |    |
|      | 15       | 7_CHRONIC.B  | 2     |             |      | 15     | 55_Silicosis   | 2           |    |
|      | 16       | 106_FIBROTHI | 4     | あり          |      | 16     | 53_ASBESTOSIS  | 4           |    |
|      | 17       | 67_ASPIRATIC | 2     |             |      | 17     | 61_THERMAL.IN  | 2           |    |
|      | 18       | 108_TRANSUC  | 5     |             |      | 18     | 81_SKIN.DISEAS | 3           |    |
|      | 19       | 68_HYDROGAI  | 3     |             |      | 19     | 80_RESPIRATOF  | 3           |    |
|      | 20       | 60_PHYSICAL  | 3     |             |      | 20     | 119_MALIGNANC  | 4           |    |

表 8 有効かつ発見的と判定された検索結果文書群の件数 (文脈: pleural-effusions)

Table 8 Statistics of effective results of heuristic documents (context: pleural-effusions).

| 空間                   | 適合度 5 の件数 | 適合度 4 の件数 |
|----------------------|-----------|-----------|
| $Space_{A+B}^{DD}$   | 3         | 1         |
| $Space_{A+B+C}^{DD}$ | 1         | 4         |
| $Space_{A+B}$        | 0         | 2         |
| $Space_{A+B+C}$      | 0         | 3         |

ことが可能となっていると考えられる。

もう 1 つの要因は、多次元ベクトル空間モデルによる検索アルゴリズムそのものの性質にある。一般に多様な関係を集約して計量するベクトル空間モデルによる情報検索において、次元の数がより多いほど、次元を構成する特徴語句どうしの直接的な関係による相関量を反映しやすいという性質を帯びることとなる。本方式の適用によって、より少ない次元によって構成されたベクトル空間を用いた検索を行った結果、より潜在的な語句間の関係による相関量を反映しやすいという性質が、実現した各空間に生じていると考えられる。

### 5.6 実験 C

5.4 節において、最も評価値の高い空間が実現されていることが確認された検索空間  $Space_{A+B}^{DD}$  上において、5.4 節で設定した 8 つの文脈を用いて検索を行った。上位 20 件に検索された文書のうち、検索対象の文書中に文脈に用いた語句を含んでいる問合せ結果文書群をキーワードのパターンマッチングによる検索手法によって検索可能な組合せととらえ除外する。検索対象の文書のうち、文脈に用いた語句をメタデータとして含んでいない問合せ結果の文書群中、検索結果としての適合性が 4 以上と判定された文書群について、

表 9 文書群  $D_c$  および文書のメタデータ中に含まれる、関連があると考えられるメタデータ

Table 9 Relevant metadata within the result documents while a context-phrase is "Rheumatoid-factor" and the search space is " $Space_{A+B}^{DD}$ ".

| 順位 | 文書 ID                          | 適合度 | メタデータ       |
|----|--------------------------------|-----|-------------|
| 2  | 16 SARCOIDOSIS                 | 5   | ILD         |
| 3  | 81 SKIN DISEASE                | 4   | Lupus       |
| 4  | 85 OTHER EXTRATHORACIC DISEASE | 4   | sarcoidosis |

検索空間におけるメタデータの特徴による記述がどのように寄与しているか調べた。

具体的には、検索空間:  $Space_{A+B}^{DD}$ 、および、文脈 "Rheumatoid-factor" を用いて検索された上位の文書のうち、適合度の評価が 4 以上であり、メタデータの語句中に文脈に用いた語句を含んでいない問合せ結果文書群 ( $D_c$ ) のメタデータについて、文脈と関連のあるメタデータが文書メタデータ中に含まれているかを調べた。また、これらのメタデータ間の関連がどの程度検索結果における相関度に寄与しているかを *Semantic Spectrum Analyzer* による分析手法<sup>21)</sup> を用いて調べた。このとき、文書群 ( $D_c$ ) に含まれる文書のメタデータのうち、文脈として用いた語句と関係があるとあらかじめ医学の専門家によって指摘が行われたいくつかの文書とメタデータ (表 9) を対象として、分析を行った。

#### 5.6.1 考察: 実験 C

表 9 の検索結果に含まれる文書のうち、2 位と 4 位に出現している各文書、すなわち、文書 16 SARCOIDOSIS および文書 85 Other EXTRATHORACIC DISEASE の各メタデータを持つ各軸上における重みを、文脈の持つ重みが大きいものから順に 20 件を表 10、

表 10 メタデータの寄与度ごとに分割した，検索対象文書メタデータの各意味素における重み  
(16 SARCOIDOSIS)

Table 10 Weight values of metadata of the search target document divided according to *Semantic Elements* (16 SARCOIDOSIS).

| 意味素ID | ILD            | granulomatous | granulomas | noncaseating-granulomas | sarcoidosis |
|-------|----------------|---------------|------------|-------------------------|-------------|
| 40    | 0.0022         | -0.1610       | -0.0789    | -0.1537                 | -0.1647     |
| 136   | 0.0011         | 0.0124        | -0.0104    | -0.0005                 | 0.0000      |
| 222   | -0.0189        | -0.0136       | 0.0114     | 0.0076                  | 0.0000      |
| 64    | <b>-0.1042</b> | 0.0276        | -0.0947    | 0.0120                  | -0.1181     |
| 202   | -0.0164        | 0.0199        | -0.0166    | 0.0066                  | -0.0000     |
| 138   | -0.0187        | -0.0582       | 0.0487     | 0.0075                  | 0.0000      |
| 172   | 0.0746         | -0.0357       | 0.0299     | -0.0300                 | 0.0000      |
| 233   | <b>0.0378</b>  | 0.0163        | -0.0136    | -0.0152                 | -0.0000     |
| 46    | 0.0258         | -0.0520       | 0.0319     | -0.0010                 | 0.1256      |
| 275   | <b>0.0064</b>  | -0.0038       | 0.0032     | -0.0026                 | 0.0000      |
| 2     | -0.0190        | -0.0190       | -0.0219    | -0.0261                 | -0.0743     |
| 110   | 0.0801         | -0.0196       | 0.0006     | 0.1698                  | -0.1121     |
| 248   | 0.0126         | -0.0102       | 0.0086     | -0.0051                 | 0.0000      |
| 152   | -0.0010        | -0.0097       | 0.0081     | 0.0004                  | 0.0000      |
| 141   | <b>0.0462</b>  | 0.0097        | -0.0081    | -0.0185                 | 0.0000      |
| 246   | -0.0107        | 0.0448        | -0.0375    | 0.0043                  | -0.0000     |
| 105   | -0.0069        | 0.0050        | 0.0507     | -0.1028                 | -0.0261     |
| 59    | 0.0052         | -0.0810       | -0.0937    | -0.1667                 | -0.0012     |
| 35    | 0.0201         | 0.0202        | 0.1156     | 0.0993                  | 0.0767      |
| 99    | <b>0.1575</b>  | 0.0653        | -0.0140    | 0.2488                  | 0.0839      |

表 11 メタデータの寄与度ごとに分割した，検索対象文書メタデータの各意味素における重み  
(85 OTHER EXTRATHORACIC DISEASE)

Table 11 Weight values of metadata of the search target document divided according to *Semantic Elements* (85 OTHER EXTRATHORACIC DISEASE).

| 意味素ID | granulomatous | sarcoidosis    | granulomas |
|-------|---------------|----------------|------------|
| 40    | -0.1610       | <b>-0.1647</b> | -0.0789    |
| 136   | 0.0124        | 0.0000         | -0.0104    |
| 222   | -0.0136       | 0.0000         | 0.0114     |
| 64    | 0.0276        | <b>-0.1181</b> | -0.0947    |
| 202   | 0.0199        | -0.0000        | -0.0166    |
| 138   | -0.0582       | 0.0000         | 0.0487     |
| 172   | -0.0357       | 0.0000         | 0.0299     |
| 233   | 0.0163        | -0.0000        | -0.0136    |
| 46    | -0.0520       | <b>0.1256</b>  | 0.0319     |
| 275   | -0.0038       | 0.0000         | 0.0032     |
| 2     | -0.0190       | <b>-0.0743</b> | -0.0219    |
| 110   | -0.0196       | <b>-0.1121</b> | 0.0006     |
| 248   | -0.0102       | 0.0000         | 0.0086     |
| 152   | -0.0097       | 0.0000         | 0.0081     |
| 141   | 0.0097        | 0.0000         | -0.0081    |
| 246   | 0.0448        | -0.0000        | -0.0375    |
| 105   | 0.0050        | -0.0261        | 0.0507     |
| 59    | -0.0810       | -0.0012        | -0.0937    |
| 35    | 0.0202        | 0.0767         | 0.1156     |
| 99    | 0.0653        | <b>0.0839</b>  | -0.0140    |

表 11 に，これらの 20 件の重みの値を反映した意味のスペクトル<sup>21)</sup>を図 11，図 12 に示した。

図 11，12 におけるグラフは 2.1 節で示した意味的連想検索の計算過程の一部を可視化したものである。図中における直交軸番号は，クエリによって選択された部分空間上にマッピングされた意味素に対応する。

図中におけるフーリエ係数とは，選択された部分空間上にマッピングされた検索対象および意味重心の，各意味素に対応したフーリエ係数を示している。検索対象に対応するフーリエ係数は，それぞれ検索対象を構成している '1' の値を持ったメタデータごとに寄与度に応じて縦に分割されている。

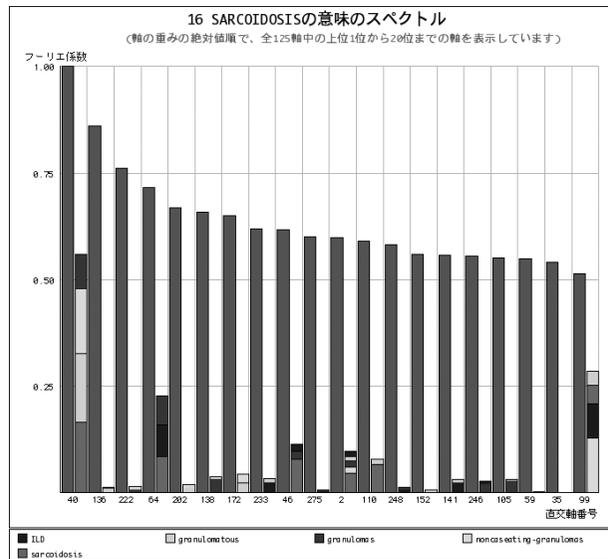


図 11 文脈：Rheumatoid-factor の検索結果 2 位のドキュメントのスペクトル

Fig. 11 Spectrum of the 2nd-ranked document while search context is 'Rheumatoid-factor'.

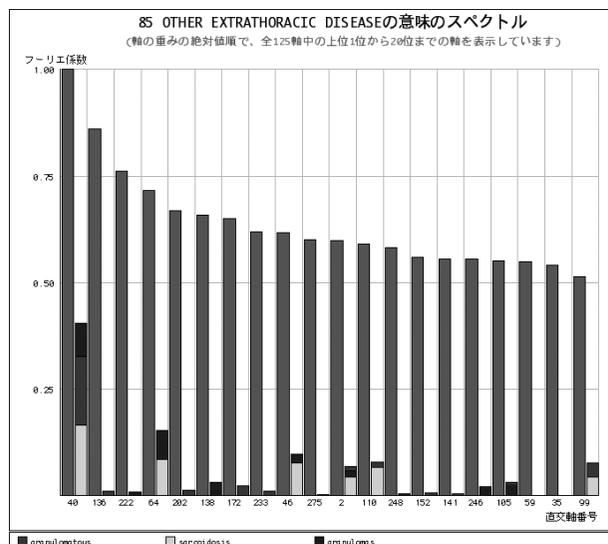


図 12 文脈：Rheumatoid-factor の検索結果 4 位のドキュメントのスペクトル

Fig. 12 Spectrum of the 4th-ranked document while search context is 'Rheumatoid-factor'.

図 11 と表 10 の組、および図 12 と表 11 の組において見られるように、表 9 で専門家によって分析の対象としてあげられた“ILD”および“sarcoidosis”について、検索結果文書の示した相関量に大きく寄与していることが確認できる。

本方式が実現する意味的な検索空間の実現により、検索対象の文書に付与されたメタデータ語句と、文脈に用いられた語句の間に直接の一致関係を持たない検索結果の中に、専門家の指摘によって指定された特定の重要なメタデータが寄与していることにより、上

位に検索されるメカニズムが働いていることが確認できた。

## 6. ま と め

本論文では、特定の専門領域を扱うための意味的検索空間を対象とし、専門家によって示される語句に関するメタ知識を用いて意味的検索空間群を実現する意味的検索空間の生成方式を示した。

本方式は、体系的に整理された幅広い専門領域を扱う、専門分野特有の知識に関する文書・メディアデータ

群を対象とした意味的連想検索において、利用目的および特徴語句の専門度に対応した意味的検索空間の生成を実現するものであり、専門性の高い分野を対象とした、新しい意味的連想検索空間方式として位置付けられる。

また、実験 A の結果が異なる文脈によって精度の差異を示したことから、専門分野を対象とした意味的な検索空間の実現により、特に明確な文脈を与えることが可能な専門性の高い検索語を用いた検索を行った場合において、より有効な検索を行うことを可能とした。

今後の課題として、第 1 に各利用目的に応じて生成した意味的検索空間の実現、および、これらの意味的検索空間を用いた応用の実現があげられる。

謝辞 医学分野の意味的検索空間の実現にあたり多大なご協力をいただいた慶應義塾大学医学部若泉謙太氏ならびに中村真弓氏に、つつしんで感謝の意を表する。

### 参 考 文 献

- 1) Salton, G., Wong, A. and Yang, C.S.: A vector space model for automatic indexing, *Comm. ACM*, Vol.18, No.11, pp.613–620 (1975).
- 2) Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, Addison Wesley (1999).
- 3) Kitagawa, T. and Kiyoki, Y.: A mathematical model of meaning and its application to multidatabase systems, *Proc. 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp.130–135, IEEE (1993).
- 4) 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). (refereed as the invited paper for special issue on metadata for digital media).
- 5) 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, Vol.J79-D-II, No.4, pp.509–519 (1996).
- 6) 宮川祥子, 清木 康: 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, 情報処理学会論文誌: データベース, Vol.40, No.SIG5 (TOD2), pp.15–28 (1999).
- 7) 河本 穰, 清木 康, 吉田尚史, 藤島清太郎, 相磯貞和: 医療分野ドキュメント群を対象とした意味的連想検索空間の実現方式, 日本データベース学会 Letters, Vol.1, No.2, pp.12–15 (2003).
- 8) Kawamoto, M., Kiyoki, Y., Yoshida, N., Fujishima, S. and Aiso, S.: An Implementation of a Semantic Associative Search Space for Medical Document Databases, *IEEE International Symposium on Applications and the Internet (SAINT 2004) — the International Workshop on Cyberspace Technologies and Societies (IWCTS 2004)*, pp.488–493, IEEE (2004).
- 9) 佐々木史織, 清木 康, 薬師寺泰蔵: 国際関係分野ドキュメント群を対象とした意味的連想検索のための空間生成方式, 日本データベース学会 Letters, Vol.2, No.1, pp.39–42 (2003).
- 10) Goldman, L. and Ausiello, D. (Eds.): *Cecil Textbook of Medicine*, 22nd edition, W.B. Saunders Company (2003).
- 11) Stedmans: *Stedman's Electronic Medical Dictionary*, version 6.0 edition, Lippincott Williams & Wilkins (2003).
- 12) Stedmans: *Stedman's Cardiovascular & Pulmonary Words: Electronic Word Book Includes Respiratory*, 4th CD-ROM edition, Lippincott Williams & Wilkins (2004).
- 13) Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.: Indexing by Latent Semantic Analysis, *Journal of the American Society of Information Science*, Vol.41, No.6, pp.391–407 (1990).
- 14) Bingham, E. and Mannila, H.: Random projection in dimensionality reduction: Applications to image and text data, *KDD '01: Proc. 7th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, pp.245–250, ACM Press (2001).
- 15) He, X., Cai, D., Liu, H. and Ma, W.-Y.: Locality preserving indexing for document representation, *SIGIR '04: Proc. 27th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp.96–103, ACM Press (2004).
- 16) Karypis, G. and Han, E.-H. : Fast Supervised Dimensionality Reduction Algorithm with Applications to Document Categorization & Retrieval, *CIKM '00: Proc. 9th international conference on Information and knowledge management*, New York, NY, USA, pp.12–19, ACM Press (2000).
- 17) 吉田一郎, 大西弘高: 実践 PBL テュートリアルガイド, 南山堂 (2004).
- 18) Salton, G. and Buckley, C.: Term Weighting Approaches in Automatic Text Retrieval, Technical report, Cornell University, Ithaca, NY, USA (1987).
- 19) 河本 穰, 図子泰三, 清木 康: 単語の出現頻度を用いたドキュメントデータベースのメタデータ自動生成, 第 13 回データ工学ワークショップ

(DEWS2002)論文集,電子情報通信学会(2002).

- 20) 泉 孝英(編):呼吸器疾患ケーススタディ:最新の呼吸器病を学ぶために,南江堂(2003).
- 21) Kiyoki, Y., Chen, X. and Ohashi, H.: A Semantic Spectrum Analyzer for Realizing Semantic Learning in a Semantic Associative Search Space, *Information Modelling and Knowledge Bases*, Vol.XVII, pp.50-67 (2006).

(平成 18 年 6 月 20 日受付)

(平成 18 年 10 月 2 日採録)

(担当編集委員 灘本 明代)



河本 穰(学生会員)

慶應義塾大学大学院政策・メディア研究科博士課程在学中。2002年慶應義塾大学総合政策学部卒業。2004年慶應義塾大学大学院政策・メディア研究科修士課程修了。データベース

システムの研究に従事。日本データベース学会学生会員。



清木 康(正会員)

慶應義塾大学環境情報学部教授。1983年慶應義塾大学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。1984~1996年筑波大学電子・情報

工学系講師,助教授を経て,1996年慶應義塾大学環境情報学部助教授,1998年同教授。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,日本データベース学会各会員。



藤島清太郎

1982年慶應義塾大学医学部卒業,同大学内科研修。1988年~1991年米国Stanford大学留学。1992年博士(医学)取得。1997年より同大学救急部講師。同大学病院で診療,教育に従事する傍ら,炎症性肺疾患,感染症等の研究に従事。日本救急医学会,日本内科学会,日本呼吸器学会等各会員。



相磯 貞和

慶應義塾大学医学部教授。1976年慶應義塾大学医学部卒業,1980年慶應義塾大学大学院医学研究科博士課程修了,医学博士。慶應義塾大学医学部助手,専任講師,Stanford大学医学部微生物学免疫学 Post-doctoral Fellow を経て,1993年より現職。形態形成学の研究に従事。日本解剖学会,日本内科学会,日本消化器病学会各会員。