異分野データベース群を対象とした意味的検索空間統合方式とその実現

石原 冴子[†] 清木 康[††]

本稿では，異分野データベース群を対象とした意味的検索空間統合方式を提案する．提案方式は，分野別に構築された既存のベクトル空間のマトリクスを対象に，分野間の共通概念（共通語）を用いて，意味の解釈をともなった検索空間の統合を実現する．提案方式を用いることにより，対象とする分野を統合した視点からの意味的な検索が可能となる．本稿では，提案方式を意味的連想検索に適用する際の実現方法を示し，環境分野および医療分野を対象としたシステム実現ならびに実験により，提案方式の有効性を確認する．

An Integration Method of Semantic Retrieval Spaces for Heterogeneous Field Databases and Its Implementation

Saeko Ishihara[†] and Yasushi Kiyoki[††]

In this paper, we present an integration method of semantic retrieval spaces of heterogeneous fields. This method makes it possible to integrate semantic retrieval spaces with the interpretation of meanings by using common concepts (common terms) for matrixes of heterogeneous fields. This method realizes the information retrieval from viewpoints related to semantically integrated fields. In this paper, we also present an implementation method for applying our integration method to semantic associative search spaces. We clarify effectiveness of our method by several experiments for environmental field and medical field.

1. はじめに

現在広域ネットワーク上には膨大なメディアデータ群が存在する．それらのメディアデータから必要な情報を的確に抽出し，獲得したいというユーザの要求は高まるばかりであり，現在，様々な検索手法が研究されている．

現在，WWW 上の情報検索ツールとして，サーチエンジンやディレクトリ検索，リンク集などが普及している．しかし，サーチエンジンにおいて一般的に用いられるパターンマッチング検索は，単純なパターン照合による検索に限られるため，単語やデータ間の相関量計算といった，単語やデータの持つ意味の解釈をともなった検索は困難である．

これに対し，Latent Semantic Indexing（以下，LSI）[1),2)] や，意味の数学モデルによる意味的連想検索方式[3)~7)] など，ベクトル空間を用いて言葉の意味を扱う検索方式が提案されている．これらは，データや言葉の意味を，マトリクス上で数値表現し，直交ベクトル空間を形成することにより，データ間の意味的な関係を，ベクトル間の相関量計算によって計量している．意味の数学モデルにおける意味的連想検索方式は，文脈に応じてダイナミックにデータ間の関係を計算する点を特徴としている．

しかし，WWW 上のサーチエンジンのカテゴリ検索をはじめ，LSI，意味的連想検索方式などにおいても，現在では，静的に限定された分野において情報検索が行われている．LSI，意味的連想検索などのベクトル空間を用いた方法では，専門辞書などの専門知識によって検索空間が構築されるため，分野別に検索エンジンが実現されることになる．これらは，個々の専門分野に特化した，分野別の情報検索には有効であり，また，検索エンジン構築のプロセスを考えても，分野別のベクトル空間生成が現実的であるといえる．

しかし，実際の世界の事象は，分野ごとに独立したものではなく，互いに深く関連しあっているものであり，またそれぞれの関連性や関連度も，視点や文脈に

† 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

応じてダイナミックに決まるものである．分野ごとに限定して開発・提供される検索エンジンでは，分野を越えた統合的な視点から，ダイナミックに情報の関連性を計量し，検索することは実現できない．

静的な分野に限定されることなく，分野を越えた発想や発見のために，静的に区分された分野という枠組みを越えて，情報の相関を計算することのできる意味的連想検索エンジンを実現することは，非常に重要な研究対象である．

本稿では，それぞれに独立した複数の異分野を対象とした，意味的検索空間統合方式を提案する．提案方式では，分野別に構築された既存のベクトル空間のマトリクスを対象として，分野間の共通概念を用いて，意味の解釈をともなった複数のベクトル空間の統合を実現する．

提案方式は，分野別に構築されている既存の検索空間を統合する方法である．一般に，意味的検索のためのベクトル空間の設計段階においては，分野単位の小規模な空間を構成することの方が，分野をまたがる大規模な空間を構成するよりも容易である．

ベクトル空間による検索方式では，対象とする分野の辞書などの専門知識を反映して検索空間が構築されるが，検索空間構築の拠りどころとなる専門辞書は，一般に分野ごとに蓄積されているため，それを反映して構築される検索空間も一般には分野別に構築されることになる．そのため，複数の分野に横断的にまとめられた情報源がない限り，複数の分野をまたいだ空間を分野別に構築せずに，はじめから 1 つの空間として構築することは非常に困難である．

提案方式は，設計段階において分野ごとに小規模な空間が構成されている状況において，それらの小規模な空間群を統合する機構により，動的に，分野をまたがる大規模な統合空間を構築する方式として位置付けられる．この方式により，設計時に分野をまたがる知識を有さない場合においても，それらの分野を統合的に扱う空間，ひいては，検索環境を実現することが可能となる．

本稿においては，意味的検索空間統合方式のモデルを示し，さらに，提案方式を意味的連想検索へ適用する際の実現方法を示し，環境分野，および医療分野を対象としたシステム実現ならびに実験により，その有効性を示す．

2. 意味的検索空間統合方式

ここでは，本稿において提案する，複数の意味的検索空間を対象とした意味的検索空間統合方式を示す．

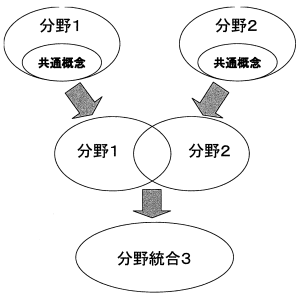

図 1 共通概念を介した分野の統合
Fig. 1 Integrated fields by using common concepts.

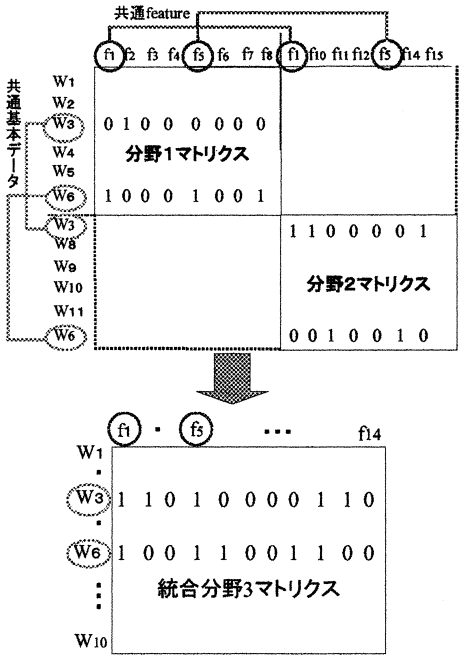

図 2 分野別マトリクスの統合
Fig. 2 Integrated field matrix.

なお，本稿における「特徴語（feature）」とは，意味的検索空間生成のためのマトリクスにおいて，横軸にあたる単語・用語を指し，「基本データ」とは，マトリクスの縦軸にあたるメディアデータ，あるいは単語などのデータを指す．このマトリクスに対して直交化を行うことにより，意味的検索空間が生成される．

提案方式の主要な特徴を以下にまとめる．概念図を**図 1**，**図 2** に示す．

(1) 共通語を介した複数マトリクス統合

分野ごとに構築されている意味的検索空間生成のためのマトリクス生成段階において，分野間の共通概念（＝共通 feature，共通基本データ）を介して，2 つのマトリクスを統合し，新たな統合マトリクスを生成する．

(2) 既存マトリクス統合による統合マトリクス生成

あらためて第3のマトリクスを構築するのではなく，分野別に構築された既存の複数マトリクスを統合することで，1つの分野統合マトリクスを生成する．これにより，既存のマトリクスを活用することが可能となり，また半自動的に，統合的な視点からの検索が可能な統合意味的検索空間の生成を実現する．

(3) 相関量計算を行う検索空間の統合

単なる検索結果の合成ではなく，相関量を計算する意味的検索空間を統合する点が特徴的である．これにより，統合対象の分野間において，分野統合的な視点から計算された相関量に基づいた意味的検索結果を得ることが可能となる．

提案方式では，あらかじめ分野別に生成されたマトリクスに注目し，既存の複数のマトリクス間における共通 feature および共通基本データを，分野間に内在する共通概念ととらえる．その共通部分を介して複数のマトリクスを統合し，1つの統合マトリクスを生成する．

提案方式である意味的検索空間統合方式は，以下のようなコンセプトによる．

異分野間の統合を考える場合，統合の対象とする分野間には共通概念が含まれているものとし，共通概念は共通の feature・基本データで表現されているものと考える．その共通概念を統合の際に合成し，すなわち，各分野のマトリクス間における共通の feature と基本データを合成し，各分野の持つ “意味” を統合的な視点から特徴付け直すことで，意味の解釈をともなった，複数分野を対象とした意味的検索空間の統合が可能となる．

本方式では，異分野の情報源を扱う複数の空間を対象に，分野間の共通概念を介して両分野のベクトル要素を統合する．この操作により，共通概念を介して，異分野のその他の概念間をも関係付けることが可能となり，それまで分からなかった異分野の概念間の関連性が分かるようになる．

このように，異分野間に存在する共通概念を介して，各分野の概念間が関連付けられることにより，対象とする複数の分野の検索空間を意味の解釈をともなって統合することを可能としている．

また，異分野の統合を考えた場合，異分野のデータベース群を単純に合成するだけでは十分でない．各分野の情報源を分野別に扱っている限り，両分野の情報間の関連性を統一的に計量することはできない．また，異分野にまたがる情報源や，複数の分野を背景に総合的に述べられている情報源などは，その特徴が分散し

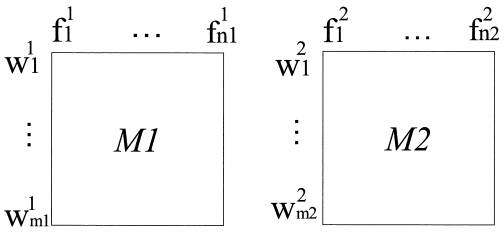

図 3 分野別マトリクス
Fig. 3 Two single-field matrixes ($M1$, $M2$) to be integrated.

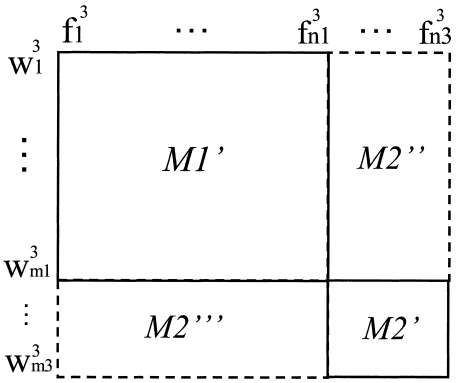

図 4 統合マトリクスの生成
Fig. 4 Integrated matrix $M3$.

てしまうことから，どんなに重要な情報であったとしても，各分野の情報を分野別に検索している限り，どの空間においても上位に検索することができない，という問題がある．

異分野の情報源に対して，分野を統合した視点から統一的に関連性を計量することが重要であると考えるため，各分野を検索空間マトリクスのレベルで統合し，複数の分野に対して統一的に相関量を計算することが可能な，分野統合検索空間を実現する．

分野1のマトリクス（以下，$M1$）および分野2のマトリクス（以下，$M2$）統合による統合分野3のマトリクス（以下，$M3$）生成は，次の3ステップから成る．分野別マトリクスを**図 3** に，統合マトリクス生成を**図 4** に示す．

- Step-1 特徴語（feature）の統合
  $M1$ の feature 群と，$M2$ の feature 群を合成し，その重複を除く．それを $M3$ の feature 群とする．
- Step-2 基本データの統合
  $M1$ の基本データ群と，$M2$ の基本データ群を合成し，その重複を除く．それを $M3$ の基本データ群とする．
- Step-3 特徴付けの設定
  Step-1，2によって決定した feature 群・基本デー

タ群を，それぞれ統合空間 3 のマトリクス（$M3$）の横軸・縦軸に設定し，ベクトル化，すなわち特徴付けを行う．feature・基本データの共通部分におけるベクトル要素に対しては，各マトリクスにおいて設定されているベクトル要素の合成を行うことにより，統合マトリクスのベクトル要素を設定する．これは，分野 1 および 2 の意味的検索空間を対象に，各分野の意味付けを統合する操作にあたる．

$M1$ を対象として，Step-1，2 の操作により抽出した $M1$，$M2$ の共通要素を，$M1$ における特徴付けに反映したものを $M1'$ とする．$M2$ から，$M1$，$M2$ の重複を除いた部分を，$M2'$ とする．基本データが $M1$，$M2$ 共通，かつ，feature が異なる部分を $M2''$ とする．$M2$ のうち，feature が $M1$，$M2$ において共通，かつ，基本データが異なる部分を $M2'''$ とする．

$M3$ は，$M1'$，$M2'$，$M2''$，$M2'''$ を図 4 のように合成することにより，生成される．提案方式では，以上のプロセスにより，分野別マトリクスを対象として，意味的統合をともなったマトリクス統合を実現する．

3. 提案方式の意味的連想検索への適用と実現

ここでは，本稿における提案方式である，分野別に生成されたマトリクスを対象とした意味的検索空間統合方式を，意味の数学モデルによる意味的連想検索[3)~7)] に適用する方式について具体的に示す．

提案方式は，他の意味的検索方式において用いられる意味的検索空間の統合にも適用可能である．

一般的な検索手法であるパターンマッチング検索では，静的かつ明示的に与えられた記述に対する単純なパターン照合でのみ検索を行うが，実際は，データの持つ意味や，データ間の関係性は静的に決めうるものではなく，文脈や状況，あるいはユーザの視点に応じて動的に変化するものである．意味的連想検索では，分野別の専門知識を利用して，その分野の「意味」を形式的に計量することのできるベクトル空間である「メタデータ空間」を生成する．メタデータ空間における文脈解釈，ベクトル計算により，指定した文脈に対して，意味的に近い情報を動的に検索することを可能にしている．

提案方式を，意味的連想検索に適用することにより，分野別に生成されたメタデータ空間が，意味の解釈をともなって統合される．これにより，分野統合的な意味計量が可能な，意味的連想検索を実現する．

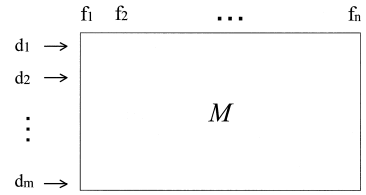

図 5 データ行列 $\mathcal{M}$ によるメタデータ表現
Fig. 5 Metadata represented in data matrix $\mathcal{M}$.

以下に，意味的連想検索に関する概要を述べ，さらに，提案方式を意味的連想検索に適用する際の実現方法について具体的に述べる．

3.1 意味的連想検索の概要

各分野における基本用語によって表現した問合せに対応したメディアデータを検索することを目的とした，意味の数学モデルによるメディアデータ検索方式の概要を示す[3)~7)]．

(1) **メタデータ空間 $\mathcal{MDS}$ 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 $\mathcal{MDS}$）を設定する．**図 5** に示す．

(2) **メディアデータのメタデータをメタデータ空間 $\mathcal{MDS}$ へ写像**

設定されたメタデータ空間 $\mathcal{MDS}$ 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，検索対象データのメタデータが同じメタデータ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 個の基本データ $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_t$ が以下のように付与されていることを前提としている．

$$P = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_t\}. \quad (1)$$

各基本データは，ベクトル表現された特徴を持っている．

$$\mathbf{w}_i = (f_{i1}, f_{i2}, \cdots, f_{in}). \quad (2)$$

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 個の基本データが合成されベクトル表現された後，メタデータ空間 $\mathcal{MDS}$ へ写像される．

(3) **メタデータ空間 $\mathcal{MDS}$ 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 $\mathcal{MDS}$ 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 $\mathcal{MDS}$ 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，閾値を越えた相関値（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．

この操作により，検索者が与えたコンテキストに対して相関の強い軸のみによる部分空間が選択される．与えられたコンテキストによりダイナミックに選択されたこの部分空間上においてメディアデータベクトルのノルムを計量することにより，与えられたコンテキストに対して意味的に相関の強い検索対象データを，ダイナミックに解釈することが可能となる．

この部分空間選択機構により，各検索対象データについて，与えられたコンテキストを構成する単語群が共通に持つ要素に対応する部分にのみ着目した相関量を計量することが可能となる．すなわち，コンテキストとして与えられる単語群が共通に持つ要素群（軸群）による部分空間を抽出することにより，検索者の意図をシャープに反映した相関量計算が可能になる．

(4) **メタデータ空間 $\mathcal{MDS}$ 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 $\mathcal{MDS}$ 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している．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．

また，メディアデータを特徴付ける特徴の数が多い場合，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．

3.2 メタデータ空間生成方式

以下に，メタデータ空間の生成プロセスを示す．

(a) 対象とする分野を表現するために必要な特徴語（以下，feature）群を準備する．対象分野の専門辞書などを用いて，各見出し語を説明している説明文中の単語を抽出し，この集合を feature 群とする．これにより，その分野の意味を表現するのに必要な単語群が定義される．

(b) 対象とする分野の基本的な用語である，基本データ群を準備する．(a) と同様に，専門辞書を用いて，見出し用語群を抽出し，この集合を基本データ群と定義する．

(c) feature 群を用いて，各基本データの特徴付けを行う．同様の専門辞書を用いて基本データの説明文を調べ，説明文をもとに，関係のある feature には 1 を，逆の意味で用いられている feature には -1 を，関係のない feature には 0 を，それぞれ設定する．この方法で，すべての基本データに対して，feature による特徴付けを行う．

(d) 以上の feature による基本データの特徴付けマトリクスから，意味的連想検索のためのメタデータ空間を生成する．

以上のプロセスにより，対象分野における意味の形式的な計算を可能とするメタデータ空間を生成する．

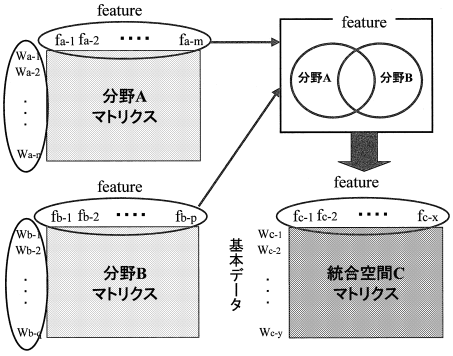

図 6 feature，基本データの統合
Fig. 6 Integration for features and words.

3.3 メタデータ空間統合方式

分野別に生成されたメタデータ空間を対象に，提案方式を適用し，メタデータ空間の意味的統合を実現する．ここでは，分野別に生成されたメタデータ空間 A，B を対象に，メタデータ空間を統合し，統合空間 C を実現する際の，具体的なプロセスを示す．

Step-1 A，B 間における feature 群の統合

A，B 間において，それぞれの feature 群を合成し，feature 語の重複を除く．この集合を，統合空間 C の feature 群と定義する．このプロセスの概要を**図 6** に示す．

Step-2 A，B 間における基本データ群の統合

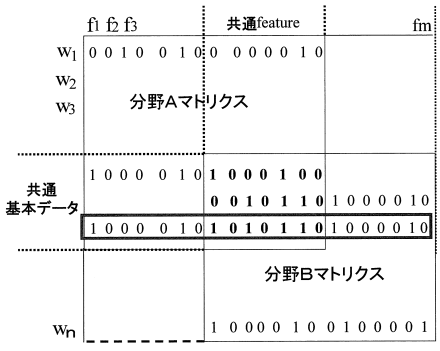

図 7 ベクトル要素の合成
Fig. 7 Integration for vector elements.

同様に，A，B 間において，それぞれの基本データ群を合成し，語の重複を除く．この集合を，統合空間 C の基本データ群と定義する．図 6 に示す．

Step-3 ベクトル要素の統合

A，B それぞれのマトリクスにおいて定義されている 1，−1，0 の各ベクトル要素を合成する．**図 7** にベクトル要素の合成方法を示す．

A，B の合成マトリクスにおける feature および基本データの共通部分を α とし，非共通部分を β とする．このプロセスにおいて重要な点は，共通部分 α において，A，B の共通基本データに対する特徴付け設定のオペレーションである．ここでは，A，B の合成マトリクスにおけるそれぞれの要素について，論理和をとる方法を示す．A，B の合成マトリクスにおける共通基本データ（w）が，A，B それぞれのマトリクスにおいて，feature 語 f1〜f8 によって，以下のように特徴付けされているとする

A における w：f1, f2, f5, f6, f7
B における w：f1, f3, f4, f6, f8

この場合，共通基本データ w に対する特徴付けは，論理和をとって，

C における w：f1, f2, f3, f4, f5, f6, f7, f8

と特徴付け直す．

また，非共通部分 β においては，元のマトリクスにおける特徴付け要素をそのまま用いることとする．

以上のプロセスにより，メタデータ空間 A，B を統合し，新たな統合メタデータ空間 C が生成される．

なお，提案方式は，任意の n 個の空間統合にも同様に適用可能である．

3 つ以上の空間統合は，以下のステップにより実現する．

(1) それぞれ独立に構築したメタデータ空間 A，B に対して空間統合方式を適用し，統合メタデータ空間 C を生成する．

(2) (1) で生成されたメタデータ空間 C と，あらかじめ独立に構築した別のメタデータ空間 D に対して空間統合方式を適用し，統合メタデータ空間 E を生成し，それを新たな C として，さらに (2) を適用する．

提案方式では，3 つ以上の空間の統合を，2 つの空間の統合の組合せにより構成する．すなわち，n 個の空間を 1 ステップで統合するのではなく，2 つの空間を対象に空間統合方式を適用し，そこから生成された空間に対してさらに別の空間を統合する，という方式により 3 つ以上の空間統合を実現する．

本方式を n 個のマトリクスに対して適用する場合，2 つのマトリクスを対象としたマトリクス統合を行う関数を設定する方法と，n 個のマトリクスを対象とした関数を設定する方法が考えられる．

本方式は，まず，2 つのマトリクスを対象として，共通 feature および共通基本データを介して統合し，その結果生成されたマトリクスと，次に対象とするマトリクスを，共通 feature および共通基本データを介して統合するという関数を繰り返し適用することにより，n 個のマトリクスを統合する．

この場合，統合の対象とする全マトリクス間の共通 feature および共通基本データの設定が統一的であるならば，マトリクス間の統合順序は，最終的な統合マトリクスの内容に非依存である．しかし，共通 feature あるいは共通基本データの設定が全マトリクス間で統一的でない場合には，マトリクスの統合順序が，最終的な統合マトリクスの内容に影響を与える．

本方式では，全マトリクス間で統一的な共通 feature および共通基本データが設定できることを前提として関数を適用することにより，この統合の結合律および適用順序の非依存性を実現できる．

3.4 ベクトル要素の統合方式

今回の実現においては，統合の対象となるマトリクス間の共通部分におけるベクトル要素の合成方式として，論理和のオペレーションを適用した．

共通部分 α における用語の合成方法としては，ほかに，論理積，あるいは算術和などいくつかのオペレーションが考えられる．

ここで，論理積を用いるオペレーションは，両分野に共通した要素に相関のない概念については，統合空間においてはその相関を 0，すなわちその概念がないととらえることに対応する．

また，算術和をとるというオペレーションを考えた場合，意味の数学モデルの性質において，1＋1＝2とすると，1と2の間の距離と0と1の間の距離が線形空間（意味空間）において同じになってしまい，意味的な関係を正しく反映することができない．

それに対して，論理和をとるオペレーションは，両分野の一方の要素に相関があれば，統合空間においてはその相関は1となり，その概念が存在することを意味する．分野間統合においては，どちらか一方の分野にその要素が存在すれば統合分野においてその概念が存在する，という仮定がより一般的であると考えられるため，ここでは統合のオペレーションに論理和を用いている．

しかし，この統合のオペレーションについては応用依存であり，論理和だけが有効な演算ではなく，ここでは例として論理和を適用しているという位置付けである．

また提案方式は，メタデータ空間A，B間において，A，B間が互いに無関係な分野である場合，すなわち，Step-1における feature 語の重複がなく，また Step-2における基本データの重複もない場合についても，両空間を統合することが可能である．また，A，B間の独立性が高い場合についても同様である．

すなわち，提案方式は，両分野のメタデータ空間の共通性の高さに依存しない空間統合方式である．

関連性の高い分野間では，feature 語の重複，および基本データの重複の割合が高くなり，マトリクスの統合においてベクトル要素間の統合が多く行われる．特に，論理和のオペレーションによる統合で，共通語を多く含む場合は，意味的関係が強く反映される．また，独立性の高い分野間では，それらの重複が少なく，ベクトル要素間の統合は少なくなる．

提案方式では，どちらの場合においても，各々独立に生成された2つのメタデータ空間の両分野に関連するメディア群は，問合せに対して，より高い相関を持つ傾向を示すことを可能にする．この傾向は，統合する対象の各分野間の独立性の高さには依存しない．すなわち，提案方式は，統合対象の分野間の関連性，独立性には依存せず，まったく無関係な分野間での統合においても適用可能な方式である．

なお，今回のシステム実現に適用した意味的連想検索では，横軸に feature 語，縦軸に基本データをおくマトリクスを設定したが，LSIにおいても同様に，横軸に索引語，縦軸にドキュメントを設定するマトリクスが仮定されており，このマトリクスに対して直交化を行うことにより，検索空間を生成している．

LSIにおけるマトリクスに対しても同様に，共通概念を介したマトリクス統合が可能であり，提案方式を適用した空間の統合が可能である．

4. 実　　験

2章で示した提案方式，ならびに3章で示した意味的連想検索に提案方式を適用した実現方式によって，実験システムを実現した．

実験を行い，その有効性を検証する．

4.1 実験目的

実験は，主に次の点を目的として行う．

- 分野別に構築した空間の検索精度を検証する．
- 提案方式を適用した統合空間の検索精度，ならびに有効性を検証する．
- 提案方式を3つ以上の空間統合に適用した際の実現可能性と有効性を検証する．
- 提案方式が統合の対象とする分野間の類似度に依存せず適用可能であることを確認する．

4.2 実験環境

実験では，環境分野および医療分野をテーマに設定した．

環境分野は，地球温暖化問題や森林破壊，食糧問題など，近年，世界的に非常に注目されている対象である．また医療分野も，我々の日常生活に密接にかかわる問題として注目度の高い対象分野であるといえる．社会における環境問題・医療問題への関心は非常に高く，WWW上でも膨大な数のデータやドキュメントが作成，提供されており，それらの効果的な利用に対するニーズは高い．

実験では，あらかじめ分野ごとに個別に構築し，それぞれ独立に検索可能であるメタデータ空間を複数用意し，それらを提案方式により統合することで，新たな分野統合メタデータ空間を生成した．

今回は，以下の2種類の分野間統合を実現した．

実現Aとしては，環境分野の中でも特に，地球温暖化や大気汚染などの大気に関する分野を対象とした「大気メタデータ空間」と，水質汚染や森林破壊などの地上に関する分野を対象とした「地上メタデータ空間」を設定し，これら独立に生成された2分野のメタデータ空間を対象に空間統合プロセスを経て，「大気・地上統合メタデータ空間（＝環境メタデータ空間）」を実現した．

実現Bとしては，環境分野を対象とした「環境メタデータ空間」と，医療分野を対象とした「医療メタデータ空間」の統合を実現した．

実現Aは比較的類似性の高い分野間の統合，実現

表 1 〔実現 A〕実験システムの詳細
Table 1 Details of semantic spaces.

	feature 数	基本データ数	空間次元数
大気メタデータ空間	250	248	243
地上メタデータ空間	318	311	307
統合メタデータ空間	425	469	415
共通（重複）用語数	143	90	—

表 2 〔実現 B〕実験システムの詳細
Table 2 Details of semantic spaces.

	feature 数	基本データ数	空間次元数
環境メタデータ空間	425	469	415
医療メタデータ空間	437	690	436
統合メタデータ空間	806	1127	794
共通（重複）用語数	56	32	—

B は比較的独立性の高い分野間の統合，として位置付けられる．

また，実現 B では，実現 A において大気空間と地上空間を統合して生成された環境空間を，さらに医療空間と統合しており，3 つ以上の空間統合を実現している．

なお，メタデータ空間生成における feature 抽出，および基本データのベクトル特徴付けには，それぞれ，環境分野の専門辞書[9)~11)]，医療分野の専門辞書[12),13)]を利用した．

それぞれの検索空間は，同一のインタフェースから検索可能なアプリケーションとして実現し，WWW からのアクセスを可能とした．ユーザは，ニーズに応じて，分野に特化した検索を望む場合には分野別検索エンジンを，統合的な視点からの検索を望む場合には統合検索エンジンを選択し，検索することが可能である．

なお，分野別マトリクスにおける feature 群および基本データ群を統合し，特徴付けし直して，統合分野マトリクスを生成するプロセスは，Perl 言語によって実現した．

実現 A，実現 B それぞれにおける，分野別に実現したメタデータ空間，ならびに統合プロセスを経て生成された統合メタデータ空間のマトリクス構成は，それぞれ**表 1**，**表 2** のとおりである．

以下，実現 A，実現 B，それぞれについて，提案方式の有効性を検証するための実験 1〜3 を行う．各々の実験方法，実験結果および考察を述べる．

4.3 〔実現 A〕大気空間×地上空間の統合に対する評価実験

まず，実現 A「大気空間×地上空間の統合」に対する実験内容，実験結果および評価について示す．

検索対象データとしては，Web 上で公開されている大気問題・地上問題，それぞれに関連する新聞記事ドキュメントを 30 件ずつ，計 60 件収集し，検索の対象とした．**表 3** に示す．

表 3 〔実現 A〕検索対象ドキュメント
Table 3 Target documents.

地上問題に関するドキュメント	doc01〜doc30
大気問題に関するドキュメント	doc31〜doc60

表 4 〔実現 A〕メタデータ設定例
Table 4 Examples of metadata.

ID	メタデータ
doc01	悪臭 水質 水質汚染 肥料 排出量
doc02	BOD 水質 水質汚染 肥料 河川 ･･
･	･
doc30	下水道 下水処理 悪臭 生活排水 浄化槽
doc31	温暖化 ハロン オゾン層 フロン ･･
doc32	温暖化 気候変動枠組条約 オゾン層 ･･
･	･
doc60	喘息 健康被害 光化学スモッグ ･･

なお，ドキュメントに対するメタデータ設定は，記事の内容から判断し，検索語となりうる用語のうち，記事中に出現した用語をそのドキュメントのメタデータとして設定した．**表 4** にメタデータ設定の一部を示す．

4.3.1 実験 A-1

実験 A-1 では，検索空間統合の前提として，各々分野別に構築した「大気メタデータ空間」ならびに「地上メタデータ空間」それぞれについて，個々の検索精度を検証する．

4.3.1.1 実験方法

実験では，大気・地上それぞれの空間に対して，各々 15 の問合せを発行した．この際，あらかじめ正解とするドキュメントを，各問合せに対して 5 件ずつ設定しておく．実験結果において，上位 10 件中，上位 5 件中に，この正解ドキュメント 5 件が含まれる割合をそれぞれ算出した．この値が高いほど，望ましいドキュメントが上位に検索されていることを示す．大気空間・地上空間の検索結果を，それぞれ**図 8**，**図 9** に示す．

4.3.1.2 実験結果

本実験により，大気・地上，両空間の，ほぼすべての問合せにおいて，上位 10 件中の正解ドキュメント再現率が 80%〜100%の高い割合を示した．また，上位 5 件中の正解ドキュメント再現率に関しても，全体の 7 割近くの問合せにおいて，80%以上の高い再現率を示した．

また，たとえば，問合せ「京都議定書」に対する検索結果において，問合せ語である「京都議定書」を直

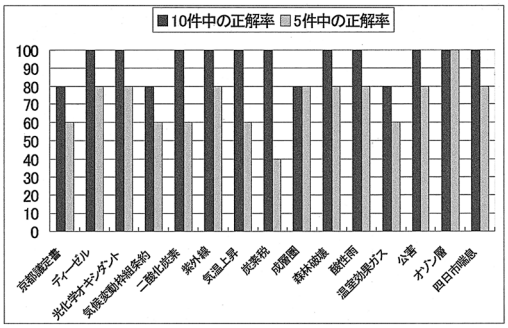

図 8 〔実現 A〕大気空間検索結果
Fig. 8 The result for the air space.

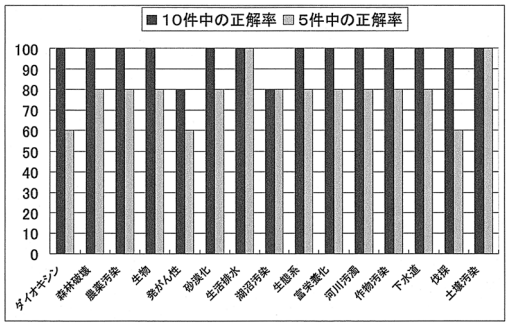

図 9 〔実現 A〕地上空間検索結果
Fig. 9 The result for the ground space.

接メタデータとして持たないドキュメントでも,「フロン 温暖化 オゾン層 オゾン層破壊 二酸化炭素 温室効果 排出量」といった,京都議定書に意味的に近いメタデータを持つドキュメントが上位に検索されており,ダイナミックに,意味的相関量が計量されていることを確認した.

4.3.1.3 考 察

本実験により,大気,地上の各々のメタデータ空間において,高い精度を持って検索が実現されていることを確認した.

統合空間の検索精度検証の前提として,統合の対象となる分野別の検索空間において高い検索精度が得られていることを確認した.

4.3.2 実 験 A-2

実験 A-2 では,大気メタデータ空間,地上メタデータ空間を統合した,統合メタデータ空間の有効性を検証する.分野別空間では上位に検索されないような,分野横断的に言及しているドキュメントが,統合空間においては上位に獲得可能であることを検証する.

4.3.2.1 実 験 方 法

実験データは,大気,地上それぞれに関する新聞記事ドキュメント 30 件ずつ,計 60 件のドキュメントを検索対象とした.

実験では,大気,地上,統合の 3 空間に対して,同一の問合せを発行する.同一の検索語に対する 3 空間の検索結果の順位を比較し,単独空間における分野別検索では上位に得られなかったが,統合空間では上位に検索されているドキュメントのメタデータを対象に,その正当性を検証する.検索結果を**表 5**,**表 6**,**表 7** に示す.なお,doc に続くナンバーはドキュメントの ID であり,隣は相関の強さを表すノルムの値である.

4.3.2.2 実 験 結 果

表 5 の実験結果を見ると,問合せ「二酸化炭素」に対し,ドキュメント 38 は,大気空間における検索結果では 8 位と上位に検索されなかったが,統合空間においては 1 位に検索されている.

ドキュメント 38 のメタデータは,「温暖化 環境庁 京都議定書 温室効果ガス 森林 ガソリン 環境税 炭素税 二酸化炭素 干ばつ」であり,「温暖化 京都議定書 温室効果ガス 温室効果ガス 炭素税」などの,大気の分野において “二酸化炭素” に関連の強いメタデータに加え,「森林 干ばつ」といった,地上の分野において “二酸化炭素” に関連の強いメタデータをあわせ持ったドキュメントであることが分かる.

このように,分野横断的に広く言及しているドキュメント 38 のようなデータが,大気単独空間では上位に検索されなかったが,統合空間では 1 位となったことが確認できる.

また,表 6 は問合せ「環境ホルモン」に対する検索結果である.ここでドキュメント 11 のメタデータに注目すると,「生活排水 BOD 汚水 排水 魚介類 窒素 下水 下水道 浄化 飲料水 リン 湖沼 富栄養化 し尿 赤潮 アオコ プランクトン」といった幅広いメタデータを持っており,ドキュメント 11 は,より広い視点から広範囲にわたって多くのことを述べているドキュメントであることが分かる.そのため,地上単独空間においてはデータの特徴が分散してしまい,30 件中 29 位と低い順位となったが,統合空間では 3 位と上位に検索されていることが確認できる.このように,総合的なメタデータを持つドキュメント 11 のようなデータは,分野別検索空間では上位に検索されないが,統合空間では上位に検索できることを確認した.

また,表 7 を見ると,問合せ「エルニーニョ現象」は,地上空間の検索語として設定されていないため,地上単独空間では検索結果を得ることができなかったが,統合空間では,地上空間のドキュメントに対しても,「エルニーニョ現象」に対する相関量計算が行われ,検索結果として得ることができた.地上空間にお

ける検索では得られなかったドキュメント 15・16・17 が，統合空間では上位 1 位〜3 位に検索されている．これらのドキュメントは，メタデータとして「砂漠化 森林破壊 温暖化 干ばつ 気候変動」などを持つ．エルニーニョ現象は，異常気象や森林へのダメージ，砂漠化などに影響するとされており，その点において，これらのメタデータはエルニーニョ現象に対する相関が強い．また異常気象の 1 つである温暖化も，同じく異常気象を起こすエルニーニョ現象と関連深いメタデータである．したがってこれらのメタデータを持つドキュメントが，統合空間において上位に検索される理由が説明しうる．

4.3.2.3 考　　察

以上の結果から，大気・地上空間における分野別検索では上位に検索することができなかった，分野横断的なドキュメントや，総合的に多くのことに言及しているドキュメントが，統合空間における分野統合的な意味解釈による検索によって，上位に検索されることを確認できた．

本実験により，提案方式によって生成された統合空間において，分野別検索では見つからなかった新しい情報の発見が可能であることを確認した．

4.3.3 実 験 A-3

実験 A-3 では，統合空間において，統合の対象となる分野別空間の性質を保持しつつ，総合的な視点からの検索が行われていることを検証する．統合空間において，検索語に対して望ましいドキュメント群が得られることを確認する．

4.3.3.1 実 験 方 法

実験データは，大気，地上それぞれに関する新聞記事ドキュメント 30 件ずつ，計 60 件のドキュメントを検索対象とした．

実験方法は，統合空間に対して，専門知識をもとにあらかじめ設定した，次の 3 パターンの検索語群を問い合わせる．

1）大気問題に関連の強い検索語群
2）地上問題に関連の強い検索語群
3）大気・地上両分野に関連する検索語群

問合せに対する検索結果のうち，上位 20 件を対象に，その順位に応じて，1 位—20 point / 2 位—19 point / 3 位—18 point / … / 20 位—1 point としてポイントを与える．その後，地上に関するドキュメント（doc01〜doc30），大気に関するドキュメント（doc31〜doc60）ごとに，獲得ポイントを集計することで，問合せに対する各分野のドキュメントの分布を見る．

これにより，たとえば，1）大気に関連の強い検索語，と設定された問合せ語に対する検索結果の分布では，上位 20 件における大気ドキュメントの獲得ポイント数が，総ポイント数に対して占める割合が高いことが望ましい，ということになる．実験結果の集計を，問合せ語の性質別に，**図 10**，**図 11**，**図 12** に示す．

4.3.3.2 実 験 結 果

結果から，大気に関連の強い検索語に対しては，大気に関するドキュメントが上位を占め，地上に関連の

表 5 〔実現 A〕「二酸化炭素」に対する検索結果
Table 5 The result of “Air pollution”.

統合空間	地上空間	大気空間
doc038 - 0.542610	doc015 - 0.670507	doc035 - 0.699915
doc049 - 0.514518	doc012 - 0.590775	doc049 - 0.655442
doc055 - 0.512407	doc029 - 0.561615	doc031 - 0.652075
doc015 - 0.495991	doc018 - 0.553930	doc052 - 0.643005
doc036 - 0.478958	doc016 - 0.553720	doc047 - 0.641100
doc047 - 0.476990	doc019 - 0.462415	doc037 - 0.625432
doc037 - 0.471426	doc004 - 0.423458	doc032 - 0.579965
doc052 - 0.448463	doc020 - 0.402324	doc038 - 0.551605

表 7 〔実現 A〕「エルニーニョ現象」に対する検索結果
Table 7 The result of “El Niño”.

統合空間	地上空間	大気空間
doc015 - 0.462327		doc035 - 0.663350
doc017 - 0.437274		doc031 - 0.645682
doc016 - 0.433494		doc052 - 0.612261
doc047 - 0.427136		doc049 - 0.605606
doc052 - 0.425687		doc047 - 0.585391
doc038 - 0.412043		doc032 - 0.578213
doc018 - 0.409883		doc037 - 0.548321
doc049 - 0.408066		doc038 - 0.526348

表 6 〔実現 A〕「環境ホルモン」に対する検索結果
Table 6 The result of “Environmental hormones”.

統合空間	地上空間	大気空間
doc026 - 0.526868	doc007 - 0.480487	doc051 - 0.491995
doc025 - 0.494758	doc006 - 0.442158	doc057 - 0.414567
doc011 - 0.460521	doc025 - 0.440210	doc058 - 0.384171
doc008 - 0.423884	doc005 - 0.371614	doc054 - 0.367133
doc007 - 0.415094	doc026 - 0.370191	doc034 - 0.332517
doc027 - 0.410646	doc008 - 0.359828	doc048 - 0.316946
doc003 - 0.407271	～	doc041 - 0.300517
doc021 - 0.402031	doc011 - 0.141531（29 位）	doc050 - 0.294542

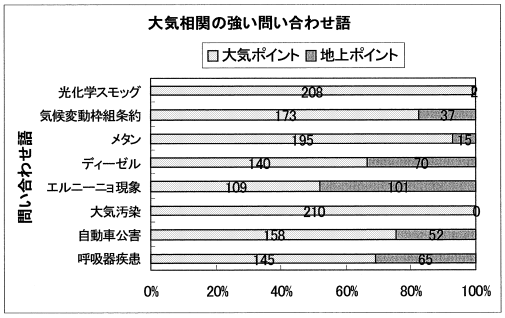

図 10 〔実現 A〕大気相関の強い問合せ結果
Fig. 10 The result of queries related to the air field.

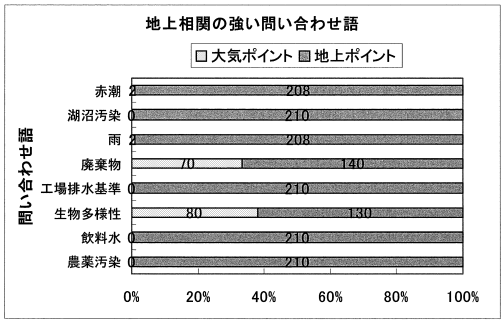

図 11 〔実現 A〕地上相関の強い問合せ結果
Fig. 11 The result of queries related to the ground field.

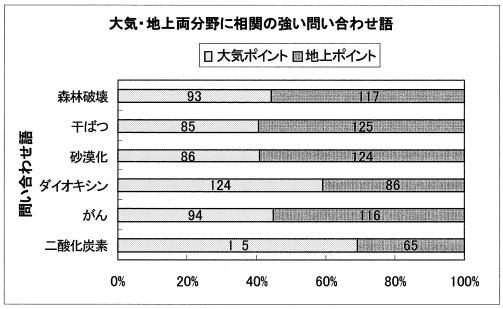

図 12 〔実現 A〕両分野に相関の強い問合せ結果
Fig. 12 The result of queries related to both air and ground fields.

強い検索語に対しては，地上に関連の強いドキュメントが高い割合で上位に検索されていることが確認できる．また，大気・地上の両分野に関連の強い検索語群に対しては，統合的な視点から，両分野のドキュメントを総合的に獲得できていることが確認できる．

4.3.3.3 考　　察

本実験により，各分野に特徴的な検索語を問い合わせた場合には，各分野のドキュメントが上位に得られており，統合空間においても，各分野の性質を保持できていることを確認した．

また，両分野に関係する検索語を問い合わせた場合には，分野統合的な視点から，両分野のドキュメント群を総合的に獲得できることを確認した．

以上，〔実現 A〕大気空間×地上空間の統合に対する実験により，提案方式によって実現した統合空間においては，各分野を意味的に統合し，新たな統合的視点から意味的計算を可能とするメタデータ空間を実現できたこと，および，その検索精度を検証した．

表 8 〔実現 B〕検索対象ドキュメント
Table 8 Target documents.

環境分野に関するドキュメント	500 件
医療分野に関するドキュメント	500 件

表 9 〔実現 B〕メタデータ設定例
Table 9 Examples of metadata.

記事 ID	メタデータ
980122259	インフルエンザ ウイルス 感染 結核 風邪
980210202	がん ウイルス 遺伝 遺伝子 炎症 感染 肝炎 ‥
971023101	ダイオキシン ホルモン 環境 尿 農薬
980623148	かいよう がん 胃炎 抗生物質 腸
980723299	ストレス ホルモン 糖尿病 尿 脳 肥満 老化 ‥
001224107	温室効果ガス 温暖化 気候変動 京都議定書 森林 ‥
001117222	ダイオキシン 汚染 環境基準 環境庁 土壌汚染 ‥
000528158	森林 生態 生態系 緑化 林業
980422027	オゾン層 フロン 温室効果 温暖化 代替フロン ‥
981207252	汚染 下水 河川 環境ホルモン 環境基準 水質汚染 ‥

4.4 〔実現 B〕環境空間×医療空間の統合に対する評価実験

次に，実現 B「環境空間×医療空間の統合」に対する実験内容，実験結果および評価について示す．

なお，ここで統合の対象とする環境空間は，実現 A において，大気空間と地上空間を統合して生成された大気・地上統合空間（＝環境空間）である．

検索対象データとしては，毎日新聞記事 CD-ROM[14] データベースより抽出した新聞記事ドキュメントを対象とした．環境分野に関する記事ドキュメントを 500 件，医療分野に関連する記事ドキュメントを 500 件抽出し，各分野のメタデータ空間の検索対象とした．また，これらを合計した 1,000 件のドキュメントを統合空間の検索対象とした．**表 8** に示す．

検索においては，各検索対象データには，その記事の内容を表すメタデータが，複数の単語の並びにより設定されていることを前提としている．

本実験におけるドキュメントに対するメタデータ設定は，検索語としてエントリされている用語のうち，記事中に出現した用語の集合を，そのドキュメントのメタデータとして設定した．メタデータ設定の例を**表 9** に示す．

4.4.1 実 験 B-1

実験 B-1 では，「環境メタデータ空間」と「医療メ

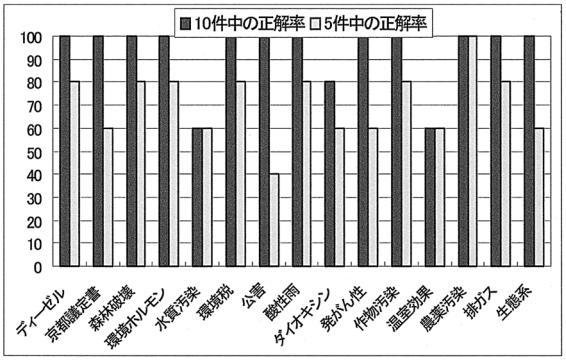

図 13 〔実現 B〕環境空間（大気×地上統合空間）検索結果
Fig. 13 The result for the environmental space.

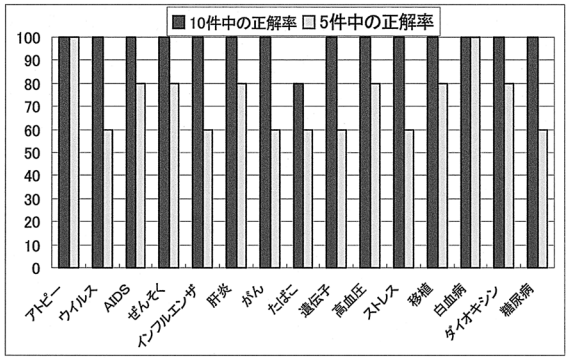

図 14 〔実現 B〕医療空間検索結果
Fig. 14 The result for the medical space.

タデータ空間」それぞれについて，その検索精度を検証する．なお，環境メタデータ空間は，実現 A において，大気空間と地上空間を統合して生成された統合空間である．

環境メタデータ空間に対する検索実験により，提案方式を適用して生成された統合空間の検索精度についての検証を行う．

4.4.1.1 実 験 方 法

検索は，表 8，表 9 に示した新聞記事ドキュメントを検索対象データとした．

実験では，環境・医療それぞれの空間に対して，各々 15 の問合せを発行した．この際あらかじめ，各問合せに対して正解とするドキュメントを 5 件ずつ設定しておく．

実験結果において，上位 10 件中，上位 5 件中に，正解ドキュメント 5 件が含まれる割合をそれぞれ算出した．この値が高いほど，望ましいドキュメントが上位に検索されていることを示す．

環境空間（大気×地上統合空間），医療空間の検索結果を，それぞれ**図 13**，**図 14** に示す．

4.4.1.2 実 験 結 果

図 13，図 14 を見ると，環境・医療の両空間において，ほぼすべての問合せに対して，上位 10 件中の正解ドキュメント再現率は 80%〜100%の高い割合を示していることが確認できる．また，上位 5 件中の正解ドキュメント再現率に関しても，半数以上の問合せにおいて，80%以上の高い再現率を示している．

4.4.1.3 考 察

実験により，環境，医療の各メタデータ空間において，高い検索精度を持って検索が実現されていることを確認した．

特に，大気空間と地上空間を統合して生成された環境空間においても高い再現率が示されており，提案方式を適用した統合空間においても，高い検索精度の検索が実現されることを確認した．

本実験により，提案方式による空間統合の前提として，統合の対象となる検索空間，および提案方式を適用した統合空間において高い検索精度が得られていることを確認した．

4.4.2 実 験 B-2

実験 B-2 では，提案方式により環境，医療の各メタデータ空間を統合した，統合メタデータ空間に対して有効性の検証を行う．

具体的には，分野別空間では上位に検索されないような，分野横断的に言及しているドキュメントや，広範囲に多くのことについて述べているドキュメントなどが，統合空間においては上位に検索されることを検証する．

4.4.2.1 実 験 方 法

検索は，各メタデータ空間においては，表 8，表 9 に示した，環境，医療それぞれに関する新聞記事ドキュメント 500 件ずつ，統合空間においては，それらを合計した計 1,000 件を検索対象データとした．

実験では，環境，医療，統合の 3 つの空間に対して，同一の問合せを発行する．同一の検索語に対する 3 空間の検索結果の順位を比較し，分野別空間では上位に得られなかったが，統合空間では上位に検索されているドキュメントのメタデータを対象に，その正当性を検証する．

検索結果を**表 10**，**表 11** に示す．なお，左側の 9 桁のナンバーは，新聞記事の日付から構成されるドキュメントの ID であり，隣は相関の強さを表すノルムの値である．

4.4.2.2 実 験 結 果

表 10 は，検索語「大気汚染」に対する問合せ結果である．結果を見ると，問合せ「大気汚染」に対して，

表 10 〔実現 B〕「大気汚染」に対する検索結果
Table 10 The result of "Air pollution".

順位	統合空間	環境空間	医療空間
1	001217119 - 0.757872	000625162 - 0.395851	971004314 - 0.532862
2	000826025 - 0.746731	000622083 - 0.365665	980514260 - 0.495517
3	990506132 - 0.746605	000120132 - 0.353257	980508179 - 0.483583
4	001128042 - 0.732719	000119220 - 0.349681	980606209 - 0.461891
5	980617296 - 0.729570	980903276 - 0.341292	971225110 - 0.455851
6	001106140 - 0.722226	980417025 - 0.324443	980417168 - 0.454193
7	980822158 - 0.721012	971015081 - 0.318102	980417134 - 0.428897
8	001015134 - 0.719486	001220024 - 0.313622	981223001 - 0.427919
9	000828030 - 0.718638	001201344 - 0.312631	981207252 - 0.407963
10	990630163 - 0.715068	000201030 - 0.312409	000503002 - 0.367019
～	～	～	～
155	990128019 - 0.550391	000630141 - 0.166195	980731226 - 0.202048
156	990705164 - 0.549977	990523208 - 0.164348	000603039 - 0.201249
157	000317080 - 0.549225	001128042 - 0.163946	980318249 - 0.201079
158	990817003 - 0.548865	000914166 - 0.161578	980609034 - 0.200232
159	990430215 - 0.548798	990409037 - 0.160712	980226112 - 0.199906
160	000916016 - 0.548206	000430155 - 0.160669	980108022 - 0.199830
～	～	～	～

表 11 〔実現 B〕「環境ホルモン」に対する検索結果
Table 11 The result of "Environmental hormones".

順位	統合空間	環境空間	医療空間
1	001217119 - 0.608303	990622146 - 0.490062	971004314 - 0.833046
2	990506132 - 0.602118	991209025 - 0.488231	990207146 - 0.805216
3	000826025 - 0.597233	001008152 - 0.482937	001203184 - 0.798642
4	001128042 - 0.588013	001219087 - 0.479456	980226374 - 0.796882
5	980617296 - 0.588009	990319268 - 0.471976	000228033 - 0.796095
6	001015134 - 0.583953	000917206 - 0.462762	980315082 - 0.789393
7	980822158 - 0.580817	000420149 - 0.459485	001018015 - 0.788370
8	001106140 - 0.579942	000724149 - 0.458495	980827211 - 0.784964
9	000828030 - 0.579025	000308112 - 0.455430	000208028 - 0.783548
10	990630163 - 0.575484	991008090 - 0.450773	000111020 - 0.783444
～	～	～	～
345	991002024 - 0.397573	990529311 - 0.165141	001023062 - 0.676683
346	990206130 - 0.396505	000828030 - 0.164548	981216292 - 0.676617
347	971023101 - 0.396302	001217119 - 0.164532	990403229 - 0.676071
348	980302004 - 0.396166	001112159 - 0.164246	971120001 - 0.676011
349	980904187 - 0.395591	001213222 - 0.164041	980518076 - 0.675793
350	990708027 - 0.395425	001015140 - 0.163888	001128142 - 0.675600
～	～	～	～

ドキュメント〔001128042〕は，環境空間における検索結果では 500 件中 157 位と上位に検索されなかったが，統合空間においては 4 位と上位に検索されている．

ドキュメント〔001128042〕のメタデータは「SPM ディーゼル 煙 汚染 汚染物質 温暖化 海 環境基準 環境庁 基準 軽油 健康 健康被害 公害 酸化物 自動車 NOx 法 大気汚染 窒素 窒素酸化物 二酸化窒素 排ガス 排ガス規制 排気 浮遊粒子状物質 硫黄」であり，「SPM ディーゼル 煙 汚染 汚染物質 温暖化 大気汚染 窒素酸化物 排ガス ···」などの，環境分野において "大気汚染" に関連の強いメタデータだけでなく，「健康 健康被害 公害」といった，医療分野において "大気汚染" に関連の強いメタデータを両方あわせ持ったドキュメントであることが分かる．このように，両分野に横断的に広く言及しているドキュメント〔001128042〕のようなデータが，環境単独空間では 157 位と上位に検索されなかったが，統合空間では 4 位と上位に検索できたことが確認できる．

また，表 11 は，検索語「環境ホルモン」に対する問合せ結果である．結果を見ると，問合せ「環境ホルモン」に対して，ドキュメント〔000828030〕は，環境空間における検索結果では 500 件中 346 位であったのに対して，統合空間では 9 位に検索されている．

ドキュメント〔000828030〕のメタデータを見ると，

「SPM がん ディーゼル 汚染 温室効果 温室効果ガス 海 環境基準 環境庁 基準 軽油 健康 健康被害 公害 酸化物 自動車 NOx 法 自動車排ガス対策 石油 大気汚染 炭素 窒素 窒素酸化物 銅 二酸化炭素 二酸化窒素 排ガス 排ガス規制 排出量 発がん性 浮遊粒子状物質 霧 硫黄」となっており，環境分野において“環境ホルモン”に相関の強いメタデータ群だけでなく，「がん 健康 健康被害 公害 発がん性」など，医療分野において“環境ホルモン”に相関の強いメタデータを持ち合わせており，環境分野，医療分野にまたがって総合的に“環境ホルモン”に相関の強い内容を持つドキュメントであることが分かる．このように“環境ホルモン”に相関が強いドキュメントであるにもかかわらず，環境分野における分野別検索では特徴が分散してしまうことから，500 件中 346 位と上位に検索されなかったが，環境，医療の統合空間においては，これらのメタデータが総合的に計算され，ドキュメント〔000828030〕のような両分野に総合的に言及しているドキュメントが上位に検索されたことが分かる．

4.4.2.3 考　　察

実験 B-2 より，両分野に横断的な内容を持つドキュメントや，総合的に多くのことについて言及しているドキュメントが，両分野を統合した統合空間においては，分野統合的な意味解釈により上位に検索可能であることを確認した．これらのドキュメントは，環境・医療空間における分野別検索においては，その特徴が分散してしまうことから上位に検索することができなかった．

本実験により，提案方式によって生成された分野統合空間において，分野別検索では見つからなかった新しい情報の発見が可能であることを確認した．

また，両分野を統合した空間において，たとえば，環境分野の概念としてエントリされている検索語によって，その概念を持たない医療分野のドキュメントに対しても相関量を計算し，検索結果を得ることが可能である．

本実験により，比較的類似性の高い分野間の統合である実現 A に対して，より独立性の高い分野間の統合である実現 B においても提案方式の有効性が示されており，提案方式が統合の対象とする分野間の類似性の高さに依存せず有効な方式であることを確認した．

4.4.3 実 験 B-3

実験 B-3 では，環境・医療の統合空間において，統合の対象とした各分野の性質を保持しつつ，総合的な視点からの検索が行われていることを検証する．

具体的には，統合空間の検索において，統合の対象とした環境分野，医療分野それぞれの検索空間の性質を失うことなく保持しつつ，なおかつ分野統合的な視点からの検索も実現していることを確認する．

4.4.3.1 実 験 方 法

検索は，表 8，表 9 に示した，環境，医療それぞれに関する新聞記事ドキュメント 500 件ずつの合計 1,000 件を検索対象データとした．

実験は，統合空間に対して，次の 3 つのパターンの検索語群を問い合わせる．これらは，辞書などの専門知識をもとに，検索語の性質をあらかじめ設定する．

(1) 環境分野に関連の強い検索語群
(2) 医療分野に関連の強い検索語群
(3) 環境・医療の両分野に関連する検索語群

統合空間に対してこれら 3 パターンの検索語群を問い合わせ，各検索語に対する検索結果において，ドキュメントの分布から，検索語の性質を反映した結果が得られているかどうかを確認する．

方法としては，各問合せに対する検索結果のうち，上位 100 件を対象に，その順位に応じて，1 位—100 point / 2 位—99 point / 3 位—97 point / … / 100 位—1 point としてポイントを与える．その後，環境分野のドキュメント，医療分野のドキュメントごとに獲得ポイントを集計し，問合せに対する検索結果における各分野のドキュメントの分布を見る．

これにより，たとえば，「環境分野に関連の強い検索語」と設定された検索語に対する検索結果の分布では，上位 100 件における環境分野ドキュメントの獲得ポイント数が，総ポイント数に対して占める割合が高いことが望ましい，ということになる．実験結果の集計を，問合せ語の性質別に，**図 15**，**図 16**，**図 17** に示す．

4.4.3.2 実 験 結 果

図 15，図 16 から，環境分野に関連の強い検索語群に対しては環境に関するドキュメントが上位を占め，医療分野に関連の強い検索語に対しては医療に関するドキュメントが高い割合で上位に検索されていることが確認できる．また，図 17 から，環境・医療の両分野に関連の強い検索語群に対しては，統合的な視点から，両分野のドキュメントを総合的に獲得できていることが確認できる．

4.4.3.3 考　　察

本実験により，統合空間において各分野に特徴的な検索語を問い合わせた場合，各分野のドキュメントが上位に得られており，このことから，統合空間においても，統合の対象とした各分野の性質を保持できていることが確認できる．

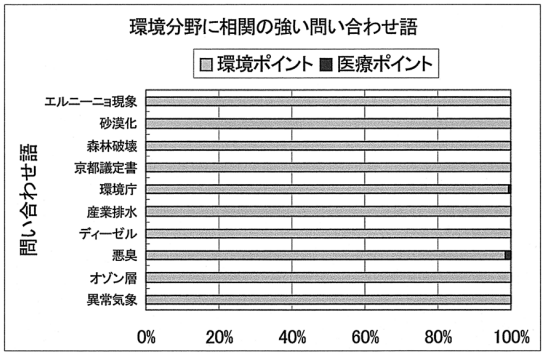

図 15 〔実現 B〕環境分野に相関の強い問合せ結果
Fig. 15 The result of queries related to the environmental field.

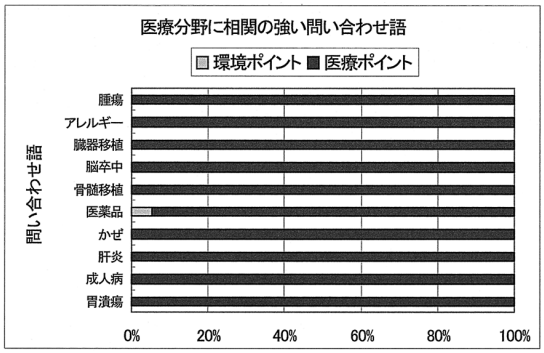

図 16 〔実現 B〕医療分野に相関の強い問合せ結果
Fig. 16 The result of queries related to the medical field.

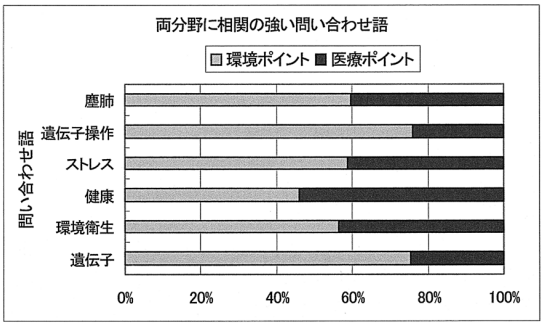

図 17 〔実現 B〕両分野に相関の強い問合せ結果
Fig. 17 The result of queries related to both environmental and medical fields.

また，両分野に関係する検索語を問い合わせた場合には，分野統合的な視点から，両分野のドキュメント群を総合的に獲得できることを確認した．

本実験により，提案方式によって実現した統合空間においては，統合の対象とした各分野を性質を失うことなく保持しており，一方の分野に特徴的な概念に対しては，分野別空間における分野に特化した検索時の精度を落とすことなく，意味的に関連の強いドキュメント群の獲得を実現していることを確認した．

同時に，両分野に関連する概念や両分野に横断するような概念に対しては，新たな分野統合的視点からの意味解釈による計量が行われ，分野をまたいだ総合的な情報獲得が実現できていることが分かった．

4.5 実験全体のまとめ

以上，実現 A，B に対する実験 1〜3 により，以下のことを確認した．

提案方式を適用した統合空間においても，高い検索精度を実現していることを確認した．

統合空間における分野統合的な視点からの検索により，異分野のドキュメント間に内在する関連性を統一的に計算することが可能となった．

分野横断的な内容や広く総合的に言及しているような，分野別検索では埋もれていて見つからなかったドキュメントを，提案方式による分野統合検索空間においては上位に検索することができ，新しい情報の発見が可能であることを確認した．

また，統合空間においては，統合の対象とした各分野の性質を保持しつつ，分野統合的な視点からの総合的な情報獲得を実現していることを確認した．

以上，実現 A，B に対する実験から，比較的類似性の高い分野間の統合である実現 A，より独立性の高い分野間の統合である実現 B，どちらの場合においても提案方式の有効性が示されており，提案方式が統合の対象とする分野間の類似性の高さに依存せず有効な方式であることを確認した．

さらに，実現 A，B に対する実験から，提案方式は，3 つ以上の空間統合にも適用可能であり，その場合にも，高い検索精度，分野横断的な情報の発見など，2 つの空間を統合した場合と同様の有効性を示すことを確認した．

5. 結　　論

本稿では，複数の分野を対象とした，意味的検索空間統合方式を提案した．

提案方式により，分野別に構築された既存のベクトル空間のマトリクスを対象に，分野間の共通概念（共通語）を用いて，複数の意味的検索空間を，意味の解釈をともなって統合することを可能とした．提案方式を，意味的連想検索へ適用し，環境問題および医療分野を対象としたシステム実現ならびに実験により，その実現可能性と有効性について検証した．

提案方式による分野別検索空間の統合により，分野別検索では上位に検索されなかった新たな情報の発見や，分野を越えた新たな関連性の発見が可能であるこ

とを確認した．

現在の検索エンジン，特に，ベクトル空間を用いて，専門的で高度な情報検索を目指す検索エンジンは，その分野の専門知識をエンジンに組み込む必要があるため，必然的に分野別に構築されることになる．検索エンジンの世界では，このように，静的に定められた分野に限定して情報検索が行われてきたが，実際には，様々な事象や概念は分野を超越して複雑に関連しあっているため，分野に限定されない，統合的な視点からの情報検索を実現することは，非常に重要である．

本稿において提案した，意味的検索空間を対象とした分野統合方式は，分野という概念を柔軟かつダイナミックにとらえ直し，そこに新たな発想や知識を生み出すための基礎となる方式である．これまで分野という限られたドメインで行われていた検索空間の構築に対して，提案方式を用いることにより，検索ドメインを拡張していくことが可能となる．

提案方式は，統合対象の分野間における関連性の強さ，独立性の高さには依存せず適用できる点が特徴である．関連性の高い分野間では，feature 語の重複，および基本データの重複の割合が高くなり，ベクトル要素間の統合が多く行われる．また，独立性の高い分野間では，それらの重複が少なく，ベクトル要素間の統合は少なくなる．提案方式は，どちらの場合においても，両分野に関連するドキュメント群は，問合せに対してより高い相関を持つ傾向を示すことを可能にする．

また，本稿では，提案方式を適用した 3 つ以上の検索空間統合を実現した．実験により，提案方式は 3 つ以上の空間統合にも適用可能であり，その場合においても有効性を示すことを確認した．

今回は，提案方式を意味的連想検索に適用した際の実現を述べたが，今後はそれに加え，LSI など他のベクトル空間モデルに提案方式を適用する際の実現方式を研究し，より汎用な計算モデルに適用可能な方式として，分野統合環境の整備・発展を目指す．

謝辞 本研究の提案システムの実現を担当いただき，また，貴重なご助言をいただいた慶應義塾大学政策・メディア研究科吉田尚史博士に感謝いたします．

参 考 文 献

1) 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).
2) Dumais, S.T., Furnas, G.W., Landauer, T.K. and Deerwester, S.: Using latent semantic analysis to improve information retrieval, *Proc. CHI'88: Conference on Human Factors in Computing*, New York, pp.281–285, ACM (1988).
3) Kitagawa, T. and Kiyoki, Y.: The 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 (Apr. 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).
5) 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 (1995).
6) 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).
7) 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).
8) Berry, M.W., Drmac, Z. and Jessup, E.R.: Matrices,Vector Spaces, and Information Retrieval, *Society for Industrial and Applied Mathematics*, Vol.41, No.2, pp.335–362 (1999).
9) 日外アソシエーツ：環境問題情報事典第 2 版，p.477, 日外アソシエーツ株式会社，東京 (2001).
10) 上田豊甫ほか：ハンディー版環境用語辞典，共立出版株式会社，東京，p.332 (2000).
11) 荒木 峻ほか：環境科学辞典，p.1015, 東京科学同人，東京 (1985).
12) 後藤 稠ほか：CD-ROM 最新医学大辞典 第 2 版 画像増補版，医歯薬出版，東京 (1999).
13) 日野原重明ほか：看護・医学事典第 5 版，p.1107, 医学書院，東京 (1992).
14) 毎日新聞社〔著作権者〕：CD-5yrs，毎日新聞 1996–2000（ハイブリッド版），日外アソシエーツ，東京 (2001).

(平成 13 年 12 月 21 日受付)
(平成 14 年 3 月 25 日採録)

（担当編集委員 掛下 哲郎）

石原 冴子（正会員）

2000 年慶應義塾大学文学部図書館・情報学科卒業．現在，慶應義塾大学政策・メディア研究科修士課程在学中．異分野データベース統合システム，感性データベースシステムの研究に従事．

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，助教授を経て，1996 年，慶應義塾大学環境情報学部助教授，1998 年同学部教授．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，電子情報通信学会各会員．