数式データを対象とした複合連想検索システム

岸本　貞弥† 　中西　崇文†† 　村方　　衛††† 　大塚　　透†† 　櫻井　鉄也††
北川　高嗣††

† 筑波大学大学院 理工学研究科
†† 筑波大学大学院 システム情報工学研究科
††† 筑波大学 第三学群 情報学類
E-mail: †{kishimoto,takafumi,murakata,otsuka}@mma.cs.tsukuba.ac.jp,
††{sakurai,takashi}@cs.tsukuba.ac.jp

あらまし　現在，Mathematical Markup Language (MathML) の仕様が公表され，web 上の数式を含む文書における数式が利用できる状況にある．これまで我々は Latent Semantic Indexing(LSI) を用いて MathML で記述された数式を問い合わせとして類似数式検索を実現した．また，特定分野を対象とした連想検索のためのメタデータ空間生成し，意味の数学モデルに適用することで数学用語に対する意味的連想検索を実現した．本稿では，この類似数式検索と，数学用語に対する意味的連想検索を連結した複合連想検索について示す．また，この検索に適した GUI を提案する．
キーワード　MathML, 類似数式検索, 意味の数学モデル, 情報検索

Composite Association Retrieval System for Data of Mathematical Formulas

Sadaya KISHIMOTO†, Takafumi NAKANISHI††, Mamoru MURAKATA†††, Toru OTSUKA††,
Tetsuya SAKURAI††, and Takashi KITAGAWA††

† Master's Program in Science and Engineering, University of Tsukuba
†† Graduate School of Systems and Information Engineering, University of Tsukuba
††† College of Information Sciences, Third Cluster of Colleges, University of Tsukuba
E-mail: †{kishimoto,takafumi,murakata,otsuka}@mma.cs.tsukuba.ac.jp,
††{sakurai,takashi}@cs.tsukuba.ac.jp

Abstract　Mathematical Markup Language (MathML) was released by World Wide Web Consortium (W3C), and We can use mathematical contents on the Web. We have implemented a function of similarity-based retrieval for formulas with Latent Semantic Indexing (LSI), using formulas encoded by MathML as queries. In addition, we have implemented a function of semantic associative search applied to mathematical terms. In this paper, we present composite association retrieval system for data of mathematical formulas and propose a GUI system which is suitable for use with this retrieval system.
Key words　MathML, Similarity-based formulas retrieval, Mathematical Model of Meaning, Information retrieval

1. ま え が き

現在，コンピュータネットワーク上に科学技術分野を対象とした多種多様な情報群が広域に遍在しつつある．また，情報群は増加を続けており，それらのデータ群は，知識・情報の源として重要な存在となっている．このような環境下で，これらのドキュメント群を対象とした，高度な検索方式と知識の発掘方式が重要となっている．

しかしながら，科学論文等，科学技術分野の情報の多くには数式が含まれており，それらの数式の持つ意味が重要となる場合が多い．このような科学論文等の数式を含んだドキュメントについて，意味的な内容を反映した検索を行うためには，数式を対象とした類似検索方式の実現が重要であると考えられる．

これまで，数式や公式を対象とした検索方式として，独自の

インデックス付けを行った数学データベースに対してパターンマッチングによる検索を行う研究 [3] にて実現されている．

数式は，どの演算子が含まれているか，どのような構造になっているか，どのような分野で使われるかなど，見方によって数式の意味合いが変わってくることが多い．例えば，$F = ma$ という式は，構造から見ると単なる m と a の掛け算を表す式である．しかし，応用範囲を考えれば，物理学で言えば「運動の法則」を表す公式であり，もしくは，買い物をしている人にとっては，単なる単価 a のものを m 個購入したときの価格 F という式でもある．つまり，数式を対象とした類似検索方式を実現するためにはこのような見方によって意味合いが変化する，数式の多角性を導入することが重要となると考えられる．

本稿は，数式を対象とした複合連想検索方式の実現について示す．本方式は，ユーザが発行した複数の問い合わせに対して，それぞれの問い合わせに合致した複数の検索方式，つまり複数の計量系で計量し，それらの結果を統合する．このことによって，数式と問い合わせとの関連性を様々な見方から計量を行い，かつ，その結果を統合することにより，ユーザの見方に合致した検索結果を得ることが可能であると考えられる．

本稿では，MathML を用いた関数や演算子，数学記号の出現による類似数式検索機構と，数式を表す言葉による意味的連想検索機構とを統合した複合数式検索システムを実装する．またこれらのシステムを用いて，本方式の有効性を検証する．ここで複合連想検索とは，様々な計量系から出てくる検索結果を AND や OR などの演算子を用いて結合し，検索結果のリストを得る検索のことである．

まず，2. 章では類似数式検索と，意味的連想検索の各検索機構について述べる．3. 章では数式データを対象とした複合連想検索について述べ，4. 章に実験例を示す．

2. 各検索機構の実現

2.1 類似数式検索機構の実現方式

ここでは，類似数式検索の実現方式について概要を述べる．本方式は MathML で書かれた数式を対象として，与えられた数式とタグの構成が類似した数式を検索するシステムである．本方式の特徴は，数式の演算子に注目して検索を行うことにより，添え字や変数に使う文字の違いなどによる，記述方法が異なる数式においても同様の意味と捉えて検索可能な点にある．

2.1.1 類似数式検索方式の概要

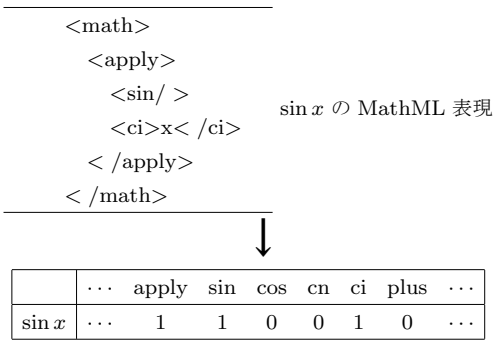
（1） 検索対象の数式群よりデータ行列を自動作成

まず，検索対象の MathML で記述された数式から，その数式の特徴を表すメタデータを抽出する．次にそれらを並べて構成するデータ行列を生成する．この行列により，検索対象となる数式データ群の類似度を計量する空間に表現することができる．メタデータ自動抽出方式については 2.1.2 節で示す．

（2） 問い合わせの数式よりメタデータを抽出

検索対象の数式データと同様に，問い合わせとして与えられた MathML で記述された数式から，その数式の特徴を表すメタデータを抽出する．

（3） 類似度を計量



	···	apply	sin	cos	cn	ci	plus	···
$\sin x$	···	1	1	0	0	1	0	···

MathML 中のタグの種類とその頻度をカウント

図 1 $\sin x$ の例.
Fig. 1 Example of $\sin x$.

上記項目 (1),(2) により抽出されたメタデータから，類似度を計量し，その値の大きい順にソートする．これにより，問い合わせの数式とタグの構成が類似した数式が検索される．本方式では，類似度の尺度としてコサイン尺度を用いている．

2.1.2 MathML で表現された数式を対象としたメタデータ自動抽出方式

本節では，MathML で記述された数式からメタデータを抽出する方式について述べる．本方式は，MathML のタグ情報に注目し，数式の特徴として抽出することにより，数式の演算子に依存した検索を実現するものである．具体的には以下の手順で実現される．

（1） MathML 表現の数式が構成するタグの種類とその出現頻度を導出

対象となる MathML 表現の数式データ $\boldsymbol{d}_i (i = 1, 2, \cdots, n)$ のタグの種類とその出現数をカウントすることで特徴づけする．

$$\boldsymbol{d}_i = (t_{1i}, t_{2i}, \cdots, t_{mi})^{\mathrm{T}}. \tag{1}$$

$t_{1i}, t_{2i}, \cdots, t_{mi}$ は対応する MathML のタグの出現頻度を表す．例として図 1 のように行う．

（2） tf・idf による重み付け

抽出したタグの頻度によってその数式の特徴を表しているが，タグの中には，どの数式にも多く含まれるタグが存在し，各数式の特徴を表す際にノイズとなる可能性がある．本方式では，全文検索においてよく用いられている tf・idf [7], [8] を用いて重み付けを行う．

2.2 数学用語等の言葉を適用した意味的連想検索機構の実現方式

ここでは，数学用語等の言葉を適用した意味的連想検索機構の実現方式について概要を述べる．特定分野を対象とした連想検索のためのメタデータ空間生成し，意味の数学モデルに適用することでこれを実現している．この検索機能によって，問い合わせの語に関連する語を検索することができる．

2.2.1 意味の数学モデルの概要

本節では，人間が様々な印象を表す際に用いられる単語 (以下，印象語) によって表現した問い合わせに対応した情報群を検索することを目的とした意味の数学モデルの概要を示す．詳細は文献 [9]〜[11] に述べられている．

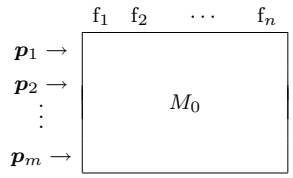


図 2　初期データ行列 M_0 によるメタデータの表現.

（1） **メタデータ空間 $\mathcal{MDS}$ の設定**

メタデータ空間生成方式については，2.3 節で示す.

（2） **検索対象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，検索対象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.

（3） **メタデータ空間 $\mathcal{MDS}$ の部分空間の選択**

利用者は与える文脈を複数の印象語を用いて表現する．ユーザが与える印象語の集合をコンテクストと呼ぶ．このコンテクストを用いてメタデータ空間 $\mathcal{MDS}$ に各コンテク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, 部分空間が選択される.

（4） **メタデータ空間 $\mathcal{MDS}$ の部分空間における相関の定量化**

選択されたメタデータ空間 $\mathcal{MDS}$ の部分空間において，検索対象データベクトルと検索語列との相関を計量する．メタデータ空間に写像された検索対象データベクトルの部分空間におけるノルムを求めることにより，文脈に対応した検索対象データの探索を行う．部分空間における検索対象データベクトルのノルムの大きさをその文脈と検索対象データとの関連の強さとする.

2.3　メタデータ空間生成方式

本節では，特定分野を対象としたメタデータ空間を，語とページの関係が記述されている書籍の索引を用いて生成する方式を示す．本方式では，検索対象を包含する特定分野について書かれた書籍が存在することを前提としている．本方式は以下の流れで実現する.

（1） 初期行列の設定

まず，対象とする特定分野について書かれた書籍の索引を参照する．索引に出現するキーワードとなる語を特徴語とみなし，索引情報から各ページ数を用いて特徴付ける.

$$\boldsymbol{p}_i = (f_{i1}, f_{i2}, \cdots, f_{in}) \tag{2}$$

ここで i はページ数，f_{ik} は特徴語に対応したページ数について特徴付けた値である．特徴付ける f_{ik} の値は，以下のように決定される.

- 索引中で特徴語がそのページ数を参照している場合 : ”1”
- 索引中で特徴語がそのページ数を参照してない場合 : ”0”

以上から，$\boldsymbol{p}_i$ を用いて，$(\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m)^T$ とすることによって，図 2 のような m 行 n 列の初期データ行列 M_0 を作成する.

（2） 初期データ行列の修正によるデータ行列の生成

(1) で作成した初期データ行列 M_0 には，ページと語の関係を表す行列となっており，ページ同士の関係が反映されていない．初期データ行列 M_0 にページ同士の関係を反映するように修正してデータ行列 M を生成する.

一般的に書籍には目次が付いており，これらの情報を反映することにより，ページ同士の関係を反映したデータ行列 M が生成可能となる．以上により，m 行 $n+\alpha$ 列のデータ行列 M を生成できる．ここで，α は特徴を追加した場合の要素の増加分を表す.

（3） 相関行列 M^TM からメタデータ空間生成

(2) で生成されたデータ行列 M の相関行列 M^TM を計算すると，$n+\alpha$ 行 $n+\alpha$ 列の行列となる．これは特徴語と特徴語の関係を示す行列となる．よって，この相関行列 M^TM を固有値分解し，非ゼロ固有値に対応する固有ベクトルによってメタデータ空間を生成する.

これにより，語と語の関係を計量するメタデータ空間の構成が可能となる.

3.　数式データを対象とした複合連想検索

類似数式検索機能と数学用語等の言葉を適用した意味的連想検索機能を連結して，検索システムを実現することにより，言葉と数式からなる問い合わせに合致した統合された検索結果を得ることを考えた．数式と言葉に対して類似検索機能を用いることで，個々に検索機能を用いる場合よりも優れた結果が得られると考えられる.

3.1　実 現 方 式

数式を対象とした複合連想検索方式の全体概要図を図 3 に示す．本方式は次の流れで実現される.

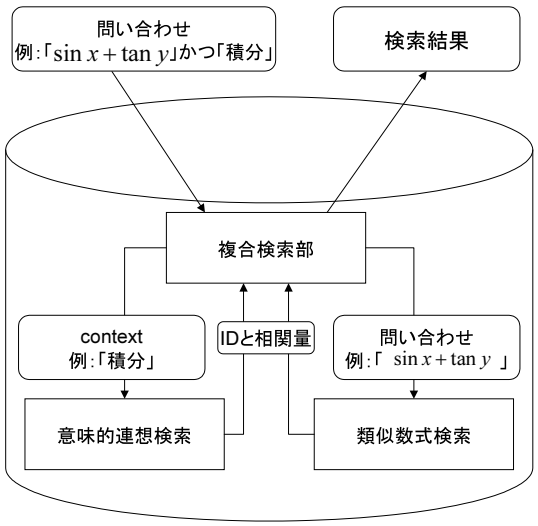


図 3　複合連想検索方式の全体図

Fig. 3　a picture of Composite Association Retrieval

Step1 :　問い合わせ発行

ユーザに検索のための問い合わせを入力してもらう．本方式では，ユーザからの問い合わせは，数式と言葉 (数学用語) から与えられることを想定している.

Step2： 問い合わせの振り分け

ユーザからの問い合わせを数式は類似数式検索機構に，言葉は意味的連想検索機構に振り分ける．

Step3： 各検索機構による結果の統合

各検索機構の結果を基本統合演算子によって統合し，問い合わせに対する検索結果としてユーザに返す．

基本統合演算子「AND」,「OR」について以下に述べる．本システムで対象としている検索機構は，問い合わせに対して，検索対象データの相関量を返すものを想定している．ユーザに出力の際に，この相関量でソートをすることにより，問い合わせに近いものから順に出力することができる．ここでは，独立に実装されている検索機構 A と検索機構 B の検索結果の統合を考える．

検索機構 A で検索した結果を $\boldsymbol{A} = (a_1, a_2, \cdots, a_n)$，検索機構 B で検索した結果を $\boldsymbol{B} = (b_1, b_2, \cdots, b_n)$ とおく．なお，a_i は検索機構 A で検索したそれぞれの検索対象データの相関量の値，b_i は検索機構 B で検索したそれぞれの検索対象データの相関量の値，n は検索対象データの数である．ただし，$0 \leqq a_i \leqq 1$，$0 \leqq b_i \leqq 1$ とする．

このとき,「AND」統合演算子 $\otimes$ を以下のように定義する．

$$\boldsymbol{A} \otimes_{i=1}^{n} \boldsymbol{B} = (\sqrt{a_1 b_1}, \sqrt{a_2 b_2}, \cdots, \sqrt{a_n b_n}) \qquad (3)$$

また,「OR」統合演算子 $\oplus$ を以下のように定義する．

$$\boldsymbol{A} \oplus_{i=1}^{n} \boldsymbol{B} = (\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2}, \cdots, \frac{a_n + b_n}{2}) \qquad (4)$$

3.2 入力用 GUI

本方式では数式の問い合わせに MathML を用いている．しかしながら，数式を MathML で記述するには MathML タグとその文法を知っておく必要があり，検索する際ユーザに入力させるのは現実的ではない．

そこで複合連想検索システムでは，数式の入力をより簡単にするために GUI による入力を考案した．この GUI は「拡張可能な GUI システム “exGUIde”」をもとに作っている．

拡張可能な GUI システムとは，数式の入力メニュー・出力形式をユーザが自由にカスタマイズできるシステムであり，様々な数理ソフトウェアの利用支援が可能である．カスタマイズは XML 定義ファイルと XSLT スタイルシートにより行う．システムの概念図を図 4 に示す．

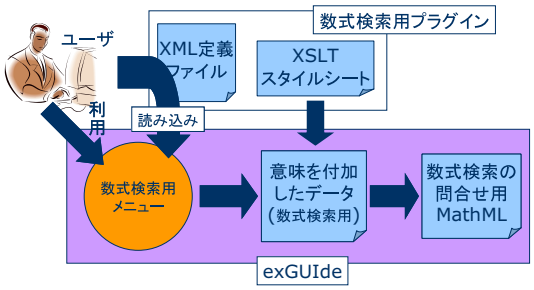


図 4 拡張可能な GUI システムの概念図

拡張可能な GUI システムを実装した Java アプリケーションをここでは “exGUIde” と呼ぶ．図 5 にその概観を示す．この “exGUIde” を用いることで単に GUI を用いて MathML が生成できるだけでなく，検索に必要な意味情報を付加することができる．

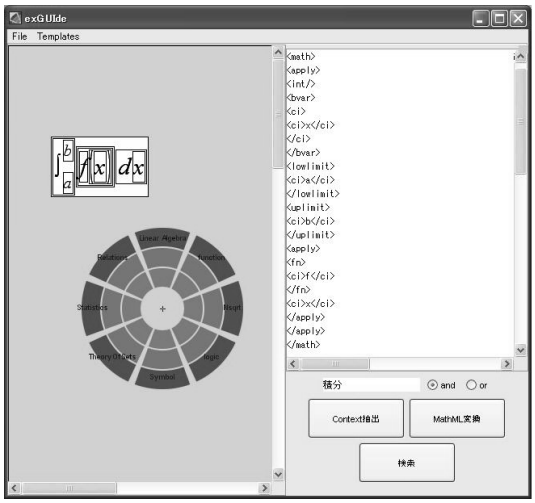


図 5 複合連想検索システムの入力用 GUI

4. 実　　験

本方式に基づくシステムを構築し実験を行った．実験例 1 では，意味的連想検索機能を実現するためのデータに,「マグロウヒル大学演習　線形代数」[14] の索引を用いて作成したデータ行列を使用した．数式データは「線形代数学の基礎」[15] より作成した．実験例 2 では，意味的連想検索機能を実現するためのデータに,「基礎物理学 第 2 版」[16] の索引を用いて作成したデータ行列を使用した．数式データは「Essential 物理学」[17] より作成した．

4.1 実 験 環 境

実験環境を表 1 に示す．また，使用言語は Perl と Java である．

表 1 実 験 環 境.

(サーバ)		
OS：	Solaris8	
HTTP サーバ：	Apache	version1.3.17
言語：	Perl	version5.6.1
	Java	version1.4.1

(クライアント)		
OS：	Windows XP	Home Edition
Web ブラウザ：	Internet Explorer	version6.0
プラグイン：	MathPlayer	version2.0b

4.2 実 験 例 1

4.2.1 実 験 方 法

実験例 1 では，意味的連想検索機能を実現するためのデータに,「マグロウヒル大学演習　線形代数」[14] の索引を用いて作成したデータ行列を使った．具体的には，索引を用いて各ページを索引に出現する 376 語で特徴づけを行った．ただし，索引で参照されないページについては省略した．この操作により，

149 行 376 列の初期行列となった．

検索対象の数式データとして，MathML で書かれた 36 個の数式とそれぞれの数式に対して付与された言葉を用いた．数式と言葉は「線形代数学の基礎」[15] より選んだ．数式データは，ID と数式と言葉のデータを 1 セットにしている．数式データの例を表 2 に示す．

表 2　実験用の数式データ例．

ID	式	言葉
1	$y = f(x)$	1 対 1 の写像
2	$\mathbb{R}^n$	n 次元，R^n
3	$\|\|a\|\|$	ノルム
4	$\cos\theta = \frac{a \cdot b}{\|\|a\|\|\|\|b\|\|}$	角，内積，ノルム
5	$\mathrm{Ker}(f) \equiv \{x \in \mathbb{R}^n \| f(x) = 0\}$	核，Ker_f
⋮	⋮	⋮

4.2.2　実 験 結 果

類似数式検索機構と意味的連想検索機構のそれぞれの検索結果として問合せ「$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$」の場合，問合せ「単位行列」の場合をそれぞれ表 3，表 4 に示す．そして，複合連想検索の検索結果として問合せ“「$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$」and「単位行列」”の場合をそれぞれ表 5 に示す．これらは，検索結果の上位 5 件を示している．

また，複合連想検索の検索結果として問合せ“「$\mathrm{rank}(A) = r$」and「行階数」”の場合，問合せ“「$\mathrm{rank}(A) = r$」or「行階数」”の場合をそれぞれ表 6，表 7 に示す．これらは，検索結果の上位 5 件を示している．

表 3　実験結果 1-1(類似数式検索機構)．

問合せ：「$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$」

順位	ID	式	相関量
1	(35)	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1.000
2	(27)	$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$	1.000
3	(26)	$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$	0.975
4	(34)	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	0.975
5	(28)	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$	0.948

実験結果 1-1 において，類似している式が上位に上がっていることがわかる．これは数式のみの検索でも比較的よい結果を示している．しかしながら，類似した数式が多数ある場合は，値の差が大きく現れないため，適合率の低下を招く恐れがある．

実験結果 1-2 において，上位の「単位行列」の次に「上三角行列」があがっている．これは意味的連想検索において，「上三角行列」という言葉そのものを入れなくても「単位行列」という言葉によって「上三角行列」が検索されたことを意味している．

表 4　実験結果 1-2(意味的連想検索機構)．

問合せ：「単位行列」

順位	ID	言葉	相関量
1	(35)	単位行列	0.938
2	(34)	単位行列	0.938
3	(13)	上三角行列	0.938
4	(14)	上三角行列，行列式	0.518
5	(6)	逆行列，単位行列	0.483

表 5　実験結果 1-3(複合連想検索)．

問合せ：「$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$」and「単位行列」

順位	ID	式	言葉	相関量
1	(35)	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	単位行列	0.969
2	(34)	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	単位行列	0.956
3	(27)	$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$	行列	0.506
4	(26)	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	行列	0.5
5	(28)	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$	行列	0.493

表 6　実験結果 1-4(複合連想検索)．

問合せ：「$\mathrm{rank}(A) = r$」and「行階数」

順位	ID	式	言葉	相関量
1	(7)	$\mathrm{rank}(A) = r$	行階数	0.884
2	(20)	$B = P^{(-1)}AP$	相似	0.138
3	(5)	$\mathrm{Ker}(f) \equiv \{x \in^n \| f(x) = 0\}$	核，Ker_f	0.083
4	(8)	$\|A\| = \det A$	行列式	0.071
5	(33)	$\mathbf{ab} = \|\mathbf{a}\| \, \|\mathbf{b}\| \cos\theta$	内積	0.063

表 7　実験結果 1-5(複合連想検索)．

問合せ：「$\mathrm{rank}(A) = r$」or「行階数」

順位	ID	式	言葉	相関量
1	(7)	$\mathrm{rank}(A) = r$	行階数	0.891
2	(11)	$B = \begin{pmatrix} A & b \end{pmatrix} \equiv \cdots \quad (a)^*$	拡大行列	0.321
3	(5)	$\mathrm{Ker}(f) \equiv \{x \in^n \| f(x) = 0\}$	核，Ker_f	0.176
4	(20)	$B = P^{(-1)}AP$	相似	0.140
5	(32)	$A_{3,2}$	行列	0.120

*表中に収まらないため，数式 (a) は以下に別記した．

$$B = \begin{pmatrix} A & b \end{pmatrix} \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ & & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

実験結果 1-3 において，実験結果 1 において 2 番目にあった ID(27) のデータは,「AND」の統合演算によって 3 番目に順位が下がり，相関量も小さくなっている．これは「AND」の統合演算によって，言葉と数式の両方が適合している数式データが上位にあがることを表している．

したがって，実験結果 1-1，1-2，1-3 から本方式による検索結果は，数式データに対して類似数式検索と意味的連想検索を個別に適用した場合よりも適合率のよい結果が得られると考えられる．

実験結果 1-4 において，最上位にある ID(7) のデータ以外は相関量が低く，有意な結果ではない．実験結果 3 と同様,「AND」の統合演算は，言葉と数式の両方が適合している数式データが上位にあがりやすいことを示している．

実験結果 1-5 において，実験結果 4 においては現れていなかった ID(11) のデータが,「OR」の統合演算によって 2 番目に順位があがっている．これは「OR」の統合演算によって，言葉と数式のどちらかが適合している数式データが上位にあがることを表している．

したがって，実験結果 1-5，1-6 から本方式における「AND」と「OR」の統合演算を使い分けることで，より柔軟な検索を行うことができると考えられる．

4.3 実 験 例 2

4.3.1 実 験 方 法

実験例 2 では，意味的連想検索機能を実現するためのデータに,「基礎物理学 第 2 版」[16] の索引を用いて作成したデータ行列を使った．具体的には，索引を用いて各ページを索引に出現する 778 語で特徴づけを行った．ただし，索引で参照されないページについては省略した．この操作により，191 行 778 列の初期行列となった．

検索対象の数式データとして，MathML で書かれた 325 個の数式とそれぞれの数式に対して付与された言葉を用いた．数式と言葉は「Essential 物理学」[17] より選んだ．数式データは，ID と数式と言葉のデータを 1 セットにしている．

4.3.2 実 験 結 果

類似数式検索機構と意味的連想検索機構のそれぞれの検索結果として問合せ「$F = mg$」の場合，問合せ「運動方程式」の場合をそれぞれ表 8，表 9 に示す．ただし，表 8 では 1 番目の順位のデータが多いので，5 件を超えて示した．そして，複合連想検索の検索結果として問合せ “「$F = mg$」and「運動方程式」” の場合をそれぞれ表 10 に示す．これらは，検索結果の上位 5 件を示している．

表 8　実験結果 2-1(類似数式検索機構).

問合せ :「$F = mg$」

順位	ID	式	相関量
1	(24)	$F = mg$	1.000
1	(30)	$v = gt$	1.000
1	(120)	$f = ce$	1.000
1	(123)	$f = n\theta$	1.000
1	(303)	$E = \hbar\omega$	1.000
1	(305)	$p = \hbar k$	1.000

表 9　実験結果 2-2(意味的連想検索機構).

問合せ :「運動方程式」

順位	ID	言葉	相関量
1	(48)	運動方程式	0.866
1	(293)	運動方程式	0.866
3	(24)	質量，重力加速度	0.646
3	(25)	万有引力定数，質量	0.646
5	(60)	運動方程式，運動量	0.629
5	(118)	減衰運動，運動方程式	0.629
5	(119)	減衰運動，運動方程式	0.629

表 10　実験結果 2-3(複合連想検索).

問合せ :「$F = mg$」and「運動方程式」

順位	ID	式	言葉	相関量
1	(24)	$F = mg$	質量，重力加速度	0.804
2	(25)	$F = G\frac{Mm}{r^2}$	万有引力定数，質量	0.563
3	(118)	$\ddot{x} + 2\gamma\dot{x} + \omega^2 x = 0$	減衰運動，運動方程式	0.441
4	(293)	$p = \frac{m}{\sqrt{1-\beta^2}}v$	運動方程式	0.410
5	(119)	$x = Ae^{-\gamma t}\cdots$ (b)*	減衰運動，運動方程式	0.399

*表中に収まらないため，数式 (b) は以下に別記した.

$$x = Ae^{-\gamma t}\sin(\sqrt{\omega^2 - \gamma^2}t + a)$$

実験結果 2-1 において，類似している式が上位に上がっていることがわかる．いずれも積の形をした数式である．上位 6 件は類似した数式であり値の差が全くない．実験結果 2-2 において，上位 5 件中 3 件には問合せにある「運動方程式」という言葉が入っているが，2 番目と 3 番目のデータには入っていない．しかしながら,「質量」「重力加速度」「万有引力定数」は「運動方程式」と関わりの深い言葉である．実験結果 2-3 において，最上位に現れている ID(24) のデータは実験結果 2-1 と実験結果 2-2 の上位にも現れていた．しかしながら，実験結果 2-1 に現れていた式は ID(24) を除いてどれも上位 5 件には入っていない．代わりに，実験結果 2-2 の上位に現れていた ID(25) の式が 2 番目にきている．したがって，実験結果 2-1，2-2，2-3 から一方の検索機構からの出力結果により，他方の検索結果がフィルタリングをかけられたような結果が得られたことがわかる．

同様に，類似数式検索機構と意味的連想検索機構のそれぞれの検索結果として問合せ「$E = h\nu$」の場合，問合せ「光量子」の場合をそれぞれ表 11，表 12 に示す．ただし，表 11 では 1 番目の順位のデータが多いので，5 件を超えて示した．そして，複合連想検索の検索結果として問合せ “「$E = h\nu$」and「光量子」” の場合をそれぞれ表 13 に示す．これらは，検索結果の上位 5 件を示している．

実験結果 2-1 において，検索を行う際に検索をしたい式そのものを入力できなくても，その数式に類似している式が上位に上がってくることがわかる．数式に使用している文字によるパターンマッチングではこのような検索結果は出ない．しかしながら，上位 6 件は類似した数式であり値の差が全くない．実験結果 2-2 において，上位 5 件中どのデータにも「光量子」という言葉が入っていない．しかしながら,「光子」「アインシュタインの関係式」は「光量子」の関わりの深い言葉である．実験

表 11　実験結果 3-1(類似数式検索機構).

問合せ :「$E = h\nu$」			
順位	ID	式	相関量
1	(24)	$F = mg$	1.000
1	(30)	$v = gt$	1.000
1	(120)	$f = ce$	1.000
1	(123)	$f = n\theta$	1.000
1	(303)	$E = \hbar\omega$	1.000
1	(305)	$p = \hbar k$	1.000

表 12　実験結果 3-2(意味的連想検索機構).

問合せ :「光量子」			
順位	ID	言葉	相関量
1	(300)	アインシュタインの関係式	0.820
2	(278)	光子	0.679
3	(291)	静止質量	0.304
4	(194)	点電荷，電界	0.286
5	(195)	点電荷，電界	0.286

表 13　実験結果 3-3(複合連想検索).

問合せ :「$E = h\nu$」and「光量子」				
順位	ID	式	言葉	相関量
1	(278)	$h\nu$	光子	0.822
2	(192)	$F = \delta qE$	クーロンの法則	0.409
3	(118)	$P = \varepsilon_0(1 + \chi_e)E$	電束密度	0400
4	(293)	$D = \varepsilon E$	電束密度，電解，誘電率	0.388
5	(119)	$D = \varepsilon_0 E + P$	電束密度	0.399

結果 2-3 において，最上位に現れている ID(278) のデータは実験結果 2-2 の上位に現れていた．しかしながら，実験結果 2-1 には現れていない．したがって，実験結果 2-1，2-2，2-3 からそれぞれ検索機構からの出力結果により，両方の検索結果を考慮したような結果が得られたことがわかる．

4.3.3 考　　察

実験結果から，本方式は複数の検索機構を用いることで検索結果をより問合せに適合させることができた．また，複数の検索機構を用いることで単一の検索機構では得られない検索結果を得ることができた．

5. む　す　び

本稿では，数式データを対象とした複合連想検索について示し，この検索に適した GUI を提案した．また，実験例を示し考察を行った．GUI を用いることで数式の問い合わせが容易に作成でき，本システムの有用性が高まると期待できる．また，本方式を適用することにより，ユーザは言葉と数式との組み合わせにより，対象とする数式からなるコンテンツの検索が可能となり，ユーザの意図と合致した検索が可能となると考えられる．

今後の課題として，より大きな数式データに対する本システムの適用，数式を含んだ文書を対象とした統合的なデータベースシステムの実現，数式の構造を考慮した検索手法の検討が挙げられる．

文　　献

[1] M.W. Berry, and S.T. Dumais, and G.W. O'Brien, "Using linear algebra for intelligent information retrieval," SIAM Review, vol.37, no.4, pp. 573–595, December 1995.

[2] "W3C Math Home," W3C. http://www.w3.org/Math/

[3] 三枝義典，阿部昭博，佐々木建昭，増永良文，佐々木睦子 "数式処理システム GAL における数学公式データベースのインデキシング手法，" 信学論 (D–I)，vol.J74-D-I，pp.577–585，Aug 1991.

[4] "World Wide Web Consortium," W3C. http://www.w3.org/

[5] 中西 崇文, 岸本 貞弥, 櫻井 鉄也, 北川 高嗣，"複数の書籍の索引部を用いたメタデータ空間拡張統合方式，" 日本データベース学会 Letters(DBSJ Letters), Vol.3, No.1, pp141-144, 2004.

[6] "TtM, a TeX to MathML translator," Ian Hutchinson. http://hutchinson.belmont.ma.us/tth/mml/

[7] G. Salton, and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. and Management, vol.24, no.5, pp.513–523, 1988.

[8] G. Salton, and C. Buckley, "Improving retrieval performance by relevance feedback," J. Am. Soc. Inf. Sci., vol.41, no.4, pp.288–297, June 1990.

[9] T.Kitagawa, Y.Kiyoki, "The Mathematical Model of Meaning and its Application to Multiodatabase Systems," *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering, Interoperability in Multidatabase Systems*, pp.130-135, April 1993.

[10] 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, " 信学論，D-II，vol.J79-D-II，no.4，pp.509–519，1996.

[11] Y.Kiyoki, T.Kitagawa, and T.Hayama, "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning," Multimedia Data Management – using metadata to integrate and apply digital media –, McGrawHill, A. Sheth and W. Klas(editors), Chapter 7, 1998.

[12] 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 ," 情処学論 : データベース, vol.40, no.SIG5(TOD2), pp.15-27，1999.

[13] 河本穣，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.

[14] Seymour Lipschutz, 加藤明史訳，マグロウヒル大学演習　線形代数 (上)(下)，オーム社，東京，1995.

[15] 水本久夫，線形代数学の基礎，培風館，東京，2000.

[16] 後藤憲一，小野廣明，小島彬，土井勝，基礎物理学 第 2 版，共立出版，東京，2004.

[17] 阿部龍蔵，Essential 物理学，サイエンス社，東京，2003.