

数式と索引語による複合連想検索支援システムの構築

本 間 秀 典^{†1} 岡 田 龍 太 郎^{†2} 永 嶋 章 弘^{†2}
杉 山 貴 俊^{†2} 川 口 祐 以^{†2}

現在，教育や研究などの分野において計算機が広く用いられており，専門分野に関する知識や情報が計算機上に蓄積されている．本稿では，数式と索引語による複合検索支援方式を提案する．科学技術の分野において文書の内容を特徴付ける要素として特に重要なものは，その分野における専門用語と数式であると考えられる．専門的な知識に基づいてこれらの提示や入力支援を行うことにより，専門的な知識を持たないユーザに対して効率的な検索支援を行うことが可能なシステムの実現を目指す．

Construction of an Integrated Associative Search Support System Using Mathematical Formulas and Index Words

HIDENORI HOMMA,^{†1} RYOTARO OKADA,^{†2} AKIHIRO NAGASHIMA,^{†2}
TAKATOSHI SUGIYAMA^{†2} and YUI KAWAGUCHI^{†2}

We use computers widely in the field of research and education. This paper propose an integrated associative search support system using mathematical formulas and index words. In the field of science and technology, mathematical expressions and technical terms are often appear and assumed as important features in documents. We realize a search support system for users having no domain knowledge to suggest them based on expert knowledge.

1. 序 論

近年，計算機の高性能化や低価格化，コンピュータネットワークの広域化および高速化に伴い，不特定多数のユーザが自由にメディア情報を発信し，それらを共有することが可能な環境が整備されつつある．膨大なメディア情報が計算機上に蓄積されるようになり，それらを対象とした効率的な検索方式および知識の獲得方式の確立は，重要な研究課題の一つとなっている．

現在，研究や教育など専門性の高い分野においても計算機が広く用いられるようになってきた．e-ラーニングコンテンツや論文データベースなど，多くの情報群が計算機上に蓄積されている．これらを対象とした検索を行う場合，その分野を特徴付ける専門的な用語や表現を用いることが有効であると考えられる．例えば，科学技術の分野では，多くの専門用語に加え，数式をはじめとした数理的な表現が多く出現する．これらの情報を情報検索に活用することで，専門分野を対象

とした情報検索の効率化を図ることが可能となると考えられる．しかしながら，現在の検索システムではユーザ自身が専門用語を想起する必要があるうえ，数式などを表現する統一的な規格が確立されていないことから，専門的な知識を持たないユーザが必要な情報を獲得することは極めて困難である．

このような問題を解決するために，問題解決環境 (problem solving environment; PSE) の重要性が高まっている．PSEとは、『コンピュータ関係の特別の知識やスキルを必要とせず，問題を解決するための計算ハードウェアとソフトウェア環境』を指す．

これまで我々はメタシステムとして提案された問題解決支援システム“MathGUIde”⁴⁾を応用し，以下のシステム群からなる数理分野における問題解決支援システムを構築した³⁾．

- 数学教育における授業支援システム
数式の入力支援と動的なグラフ描画の機能により，数学教育における授業支援を実現した．
- 特定分野を対象とした連想検索システム
特定分野を対象とした書籍の索引に基づく索引語間の意味的連想検索方式¹⁰⁾を応用し，専門用語による数式検索と索引語の想起支援 (サジェスト) による検索支援のプロトタイプを実装した．

本研究では，これらの機能を情報検索および検索支援に特

^{†1} 研究代表者：筑波大学大学院 システム情報工学研究科

Project Leader: Graduate School of Systems and Information Technology, University of Tsukuba

^{†2} 研究メンバー：筑波大学大学院 システム情報工学研究科

Project member: Graduate School of Systems and Information Technology, University of Tsukuba

化することにより、科学技術の分野を対象とした、数式と専門用語のサジェストによる検索支援システム、および数式を検索クエリとして入力および編集するためのインターフェースを実現する。また、数式に関しては、W3C 勧告となっている MathML (Mathematical Markup Language)⁸⁾ という XML ベースの規格による標準化が進んでおり、今後、統一規格による数式の表記および利用が想定される。そこで本研究では、MathML を用いることにより、数式の構造に基づく数式の類似度計量方式の実現を目指す。そして、これらのシステム群を統合することにより、数式と索引語による複合連想検索支援システムを構築する。

本論文の構成は次のようになっている。2 節では、特定分野における専門用語のサジェストの実現方式について述べる。3 節では、木構造に基づく数式の類似検索方式について議論する。4 節では、数理分野における問題解決支援システム “MathGUIDe”⁴⁾ について概説し、その文書検索への応用である “MathGUIDe / web” について述べる。5 節では、複合連想検索支援システムの概要を示す。最後に 6 節で本研究を総括する。

2. 専門家の知識に基づく検索語及び数式のサジェストによる検索支援方式

本節では、専門家の知識に基づく検索語及び数式のサジェストによる検索支援方式について述べる。本方式では、書籍の索引を用いて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を生成し、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。また、ユーザインターフェースを考慮することでより効率の良い検索支援方式の実現を目指す。以下に各方式と実装システムについて述べる。

2.1 意味の数学モデルの概要

本節では、言葉と言葉の関係の計量を実現する意味の数学モデルの概要を示す。詳細は、文献 7), 8) に述べられている。

- (1) メタデータ空間 MDS の設定
メタデータ空間 MDS と呼ばれる、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にマッピングするための正規直交空間 (以下、 MDS) を設定する。
- (2) メディアデータのメタデータを MDS へ写像
設定された 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。
- (3) 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 MDS に各コンテキストに

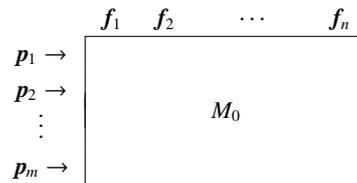


図 1 初期データ行列 M_0 によるメタデータの表現。

対応するベクトルを写像する。これらのベクトルは、 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) 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。

2.2 書籍の索引部を用いたメタデータ空間生成

本節では、書籍の索引を用いたメタデータ空間生成方式について示す。詳細は、文献 10) に述べられている。具体的には、次の流れで実現する。

- (1) 初期データ行列の設定
まず、対象とする特定分野について書かれた書籍の索引に出現する語を特徴語とみなし、索引情報から各ページ番号を用いて特徴づける。各ページを索引語を用いて特徴づけたベクトルを特徴ベクトル p_i とする。
$$p_i = (f_{i1}, f_{i2}, f_{i3}, \dots, f_{in}) \quad (1)$$
ここで i はページ番号、 f_{ik} は特徴語に対応したページ番号について特徴づけた値である。特徴づける 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番号を参照している場合: "1"
 - 索引中で特徴語がそのページ番号を参照していない場合: "0" p_i を用いて $(p_1, p_2, p_3, \dots, p_m)^T$ とすることで、図 1 のような m 行 n 列の初期データ行列 M_0 を作成する。
$$M_0 = (p_1, p_2, p_3, \dots, p_m)^T$$
- (2) 初期データ行列 M_0 の修正によるデータ行列 M の生成
(1) で作成した初期データ行列 M_0 には、ページ番号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。そのため、ある概念が複数ページにわたって書かれている場合、索引に記述されているキーワードとして表される語とページ番号の関係だ

けでは表現しきれず、精度を悪化させる原因となりうる。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、目次には章、節とその題名、そしてページ番号が付与されている。章、節は、ある内容を説明するための論理的な枠であることから、内容に関係のあるページのかたまりとして捉えることができる。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。章、節とページの間を導出したデータ行列 M を生成することにより、章と節は各ページについて内容のかたまりによって、分けられたものであることから、内容に即した語と語の関連を導入することができる。

まず、章、節の番号を特徴語として初期データ行列 M_0 を修正、追加する。章、節番号について該当ページを全て「1」、それ以外のページを「0」と特徴づける。例えば、23 ページが 2 章 3 節に該当する場合、「2」、「2-3」を特徴語として、23 ページの「2」、「2-3」に「1」と特徴づける。

以上より、 m 行 $n+\alpha$ 列のデータ行列 M を生成できる。ここで、 α は章、節番号を特徴として付け加えた分である。

- (3) 相関行列 $M^T M$ 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 $M^T M$ を計算すると、 $n+\alpha$ 行 $n+\alpha$ 列の行列となる。概要図を図 2 に示す。これは特徴語と特徴語の関係を示す行列となる。この相関行列 $M^T M$ 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。本操作の詳細は文献 7)~9) で、示されている。

これにより、語と語の関係を計量する検索のためのメタデータ空間が構成可能となる。

2.3 効率的な検索支援のためのユーザインタフェース

効率的な検索を支援するためには、ユーザにとって使いやすいインタフェースが必要不可欠であると言える。そこで、本システムではユーザインタフェースに以下のような機能を実装する。

2.3.1 専門家の知識に基づいた検索語のサジェスト

2.1 節で述べた意味の数学モデルと 2.2 節で述べた書籍の索引部を用いたメタデータ空間生成方式を用いることで専門家の知識に基づいた検索語のサジェストが実現可能となる。サジェストの手順を以下に示す。

- (1) 検索語の入力
検索語を入力する。ここで入力される語は、索引に含まれる語である必要がある。

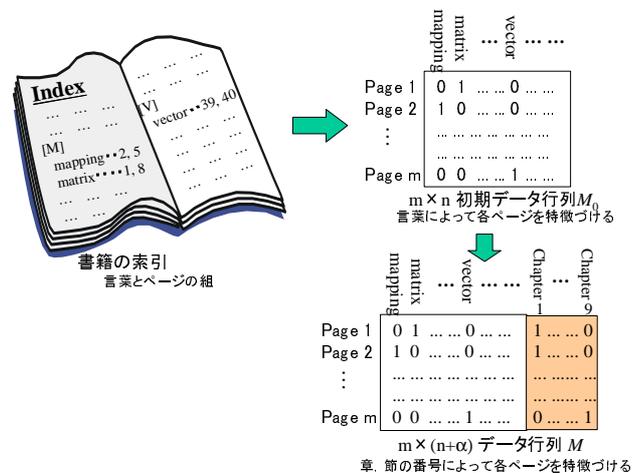


図 2 作成するメタデータの概要図。

- (2) 意味的相関の計量
入力された検索語と各索引語との意味的相関を意味の数学モデルを用いて計量する。
- (3) 語の提示
意味的相関を計量した結果から、相関量の大きい上位数個の語を提示する。

2.3.2 Ajax による画面遷移のないサジェスト

Ajax を用いた画面遷移のない語のサジェスト機能を実装する。これにより、画面遷移によるストレスを軽減できると考えられる。

2.3.3 索引語との前方一致による検索語の補完

入力された語と前方一致する索引語を補完候補として表示する。これにより入力の手間の減少することできる。また、ユーザにない知識の提示が可能となる。

2.3.4 専門書籍の章立てを用いた検索語の提示

本検索支援システムでは、検索を行おうとしている専門分野についての知識をほとんど持っていないユーザを想定しており、Google サジェストのように、ユーザの入力と前方一致する語を提示する方法では、以下のような問題に陥る可能性が想定される。

- (1) 知識が乏しいため目的の情報と（意味的な）相関の強いキーワードが分からず、曖昧なキーワードで検索を開始する
- (2) 検索結果を見てもどれがより目的に合っているか分からない
- (3) 幾つかのページを閲覧するが、適切な修正キーワードが見つからない
- (4) (1)~(3) を繰り返し、望んだ検索結果が得られないまま検索を終了してしまう

この問題を解決するために、専門書籍の章立てに基づき索

引語を分類して提示する方式を用いる。専門書籍の章立てでは、ある内容を説明するための論理的な枠であることから、それらに基づいて語を提示することは意味的に近い語を提示するという点において効果的であると考えられる。これにより、その分野の関する知識がほとんどないユーザにも効果的な検索支援ができると考えられる。

2.3.5 数式のサジェスト

実装するシステムでは、検索を行う際に数式を入力する必要があるが、数式入力是非常に手間のかかる作業である。その手間を軽減し、検索効率の低下を防ぐため数式のサジェストを行う。Web上や文書中の数式は、数式のみで書かれていることはまれであり、周辺にテキストを持つことがほとんどである。そこで、周辺のテキストをメタデータとして付与した数式データを作成し、2.1節で述べた意味の数学モデルと2.2節で述べた書籍の索引部を用いたメタデータ空間生成方式を組み合わせることで、数式のサジェストを行う。

2.4 検索支援システムの構成

本方式に基づく検索システムの実装を行った。また、ユーザにとってより良いインタフェースを模索するため、2通りのユーザインタフェースによる実装を行った。一方は、2.3.3節で述べた索引語との前方一致による検索語の補完と数式のサジェストによる方式、もう一方は、2.3.4節で述べた専門書籍の章立てを用いた検索語の提示による方式である。

以下に実装システムについて示す。

2.4.1 実装環境

WebブラウザFirefox3上で動作するシステムを構築した。ページ遷移をなくし、ユーザのストレスを減らすため、GUI部にはAjaxを採用した。Ajaxライブラリとして、prototype.jsを利用した。サーバ側の処理にはPHPを用いた。また、2.2節で述べたメタデータ空間の生成に用いる書籍には、高校数学の参考書である“高校数学公式活用辞典²⁾”及び線形計算の教科書である“線形計算¹⁾”を用いた。

2.4.2 システム概要

システム概要図を図3に示す。索引語との前方一致による検索語の補完と数式のサジェストによる方式では、検索支援は以下の流れで行われる。スクリーンショットを図4に示す。

- (1) 検索語の入力
任意の検索語を入力する。
- (2) 検索語と前方一致する索引語の提示
入力された語と前方一致する索引語を索引語群から見つけ出し、それらを提示する。
- (3) 専門家の知識に基づいた検索語と数式の提示
2.1節、2.2節、2.3節で述べた、意味の数学モデル及び書籍の索引部を用いたメタデータ空間生成方式を用いることで、専門家の知識に基づいて提示された索引語と意味的に関連のある検索語及び数式を提示する。

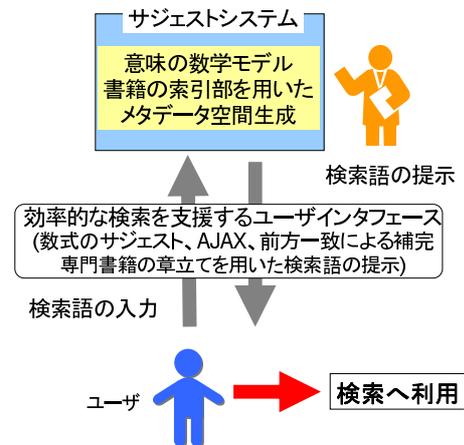


図3 システム概要図。

(4) 検索語及び数式の選択

提示された検索語と数式の中から任意のものを選択する。

専門書籍の章立てを用いた検索語の提示による方式では、検索は以下の流れで行われる。スクリーンショットを図5に示す。

- (1) 章立ての提示
専門書籍に基づいた章立てを提示する。
- (2) 章の選択
提示された章から任意のものを選択する。
- (3) 章に現れる索引語の提示
選択された章に現れる索引語を提示する。
- (4) 専門家の知識に基づいた検索語の提示
2.1節、2.2節、2.3節で述べた、意味の数学モデル及び書籍の索引部を用いたメタデータ空間生成方式を用いることで、専門家の知識に基づいて提示された索引語と意味的に関連のある検索語を提示する。
- (5) 検索語及び数式の選択
提示された検索語から任意のものを選択し検索を行う。

3. MathMLを用いた木構造に基づく数式の類似検索

数式は科学分野を初めとする幅広い分野で用いられている表現方式であり、方式やアイデア等を直接的に表現するものとして文書中に登場する。そのため、数式を含む文書において、数式はその文書の特徴付ける要素の中でも大きな位置を占めていると考えられる。従って数式が示す内容の類似性が計量できれば、類似する方式、アイデア等を提示している文書の検索が可能になると考えられる。このような方式が実現すれば、論文を書く際に関連研究を探す、特許を申請する際に自分の提案したものが既にあるものかどうかを探すなど、様々な利用が可能となると考えられる。また、数学と物理な

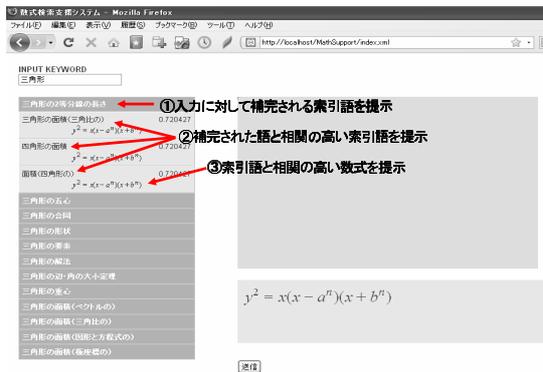


図 4 実装システムのスクリーンショット.

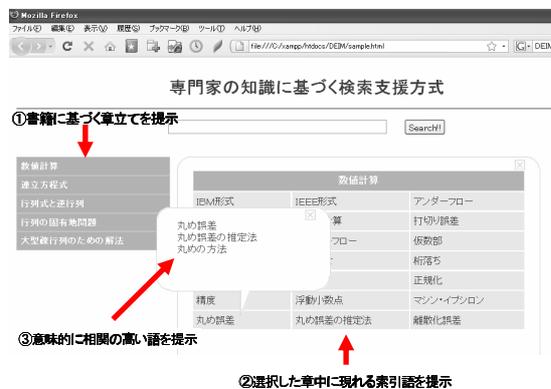


図 5 実装システムのスクリーンショット.

ど分野が違う文書においては、使われている専門用語が異なるためキーワードによる一致検索が効果を発揮しない事が考えられる。ここで、数式の構造を利用した検索を用いる事で、分野をまたいだ検索が有効に行えると考えられる。

近年、数式は W3C 勧告となっている MathML (Mathematical Markup Language)¹⁸⁾ という XML ベースの規格での標準化が始まっている。MathML には、Web などでの実際の表記を重視した Presentation Markup と、内容的な構造情報を厳密に表すことを重視した Content Markup があり、数式の内容的な類似を測るには Content Markup が適していると考えられる。ただし、現在は Presentation Markup の方が普及しており、また $\text{T}_\text{E}\text{X}$ による数式表現からの自動変換も不完全ではあるが行われている。これはつまり、Presentation Markup は、図形としての数式における、各記号の位置情報を表現する事が出来るということである。

著者らは文献 6) において、MathML(Content Markup) を用いた数式検索を提案している。これは TF-IDF を用いて MathML におけるタグを要素として重みを付けたベクトルについての計量を行うものである。しかしながら、数式の構造情報を扱っていない事が課題として挙がっていた。また、

TF-IDF は「多くの検索対象に出現する語は重要度が下がる」方式であるが、数式の類似性を計量する手法としては適さないのではないかと考えた。例えば等号は出現頻度は高いが、式の構造を決める大きな手がかりであると考えられるからである。

本稿ではより内容的に類似性の高い数式の検索を実現するために、MathML の構造情報を利用した類似数式検索方式の提案を行う。

3.1 関連研究

文献 1) において橋本らは、Presentation Markup について、数式の各ノードへの XPath が一致するものの有無を調べるという手法を用い、かつ XPath についての転置インデックスを作成する事により、高速な数式検索を実現している。この方式を用いる場合、ある数式の構造のまとまりが、木構造の高さが一段違うだけの場合にも全く見つからなくなるため、本研究とは方向性が異なる。

小田切らは、Content Markup について、数式そのものをクエリとするのではなく、「sin を含む積分」といったような条件を用いて、部分的な構造を指定することができる検索方式を提案している¹²⁾。これは数式の内容的な検索という観点からは、数式の構造を指定でき、必ず要求に合ったものが検索されるという点で優れている。また、ここでも TF-IDF を数式にそのまま適用するのは必ずしも適切でないという考察が加えられていた。

3.2 数式の構造に関する考察

XML は木構造データを表現できるデータ記述言語であるので、MathML はそのまま数式を木構造で表現した形式で表されている。しかし、任意の数式が与えられた時、その表記は一意ではない。例えば、 $a - b + c$ という式は、 $a + (-b) + c$ 、 $(a - b) + c$ 等に対応する複数の MathML 表記が可能である。このような場合、数式の表現する内容と MathML での表記を適切に対応付ける事が重要である。そこで、MathML の表記と数式の構造の関係についての考察を行った。主要なものをいくつか以下に示す。

3.2.1 木構造構築の基本的な指針

MathML は、 $\langle math \rangle$ タグを root とし、その後は $\langle apply \rangle$ タグを前置記法における括弧として用いる形で表現される。よって階層構造としては、 $\langle apply \rangle$ タグの直後に出現する要素を演算子とし、その後続く要素をオペランドとして、演算子の下に置くこととする。このようにして、数式の内容に基づく木構造を構築する。 $a = (b + 2)c$ という式の MathML 表記と、それから構築される木構造を図 6 に示す。

3.2.2 加減算の順序

並列な関係にある項が階層的に上下関係になることは不自然であるため、加算と減算が多項演算子としても単項演算子としても使える事を利用して、階層的に上下の少ない形で表

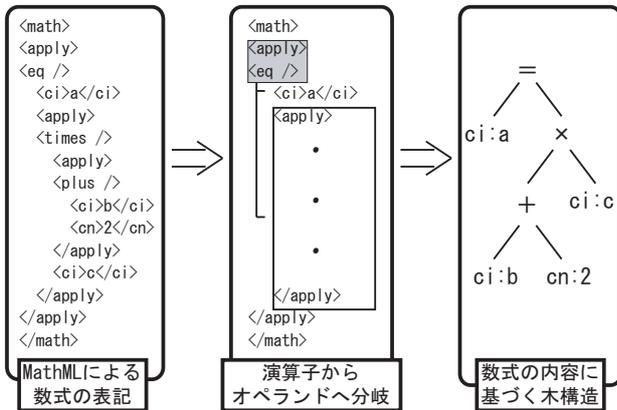


図 6 $a = (b + 2)c$ の MathML および木構造での表現 .

加算のみ $a + b + c$	減算のみ $a - b - c$	混在 $a - b + c$	括弧あり $a + (b + c)$

図 7 加減算と木構造の関係 .

現する . 具体的には以下のように行う .

- (1) 加算のみ 加算でまとめる
- (2) 減算のみ 減算でまとめる
- (3) 加算と減算が混在 加算でまとめ, 減算は単項演算子として用いる
- (4) 括弧あり 表記者の示す優先順位に従う

以上を図 7 にまとめて例を示す .

3.2.3 乗除算の順序

乗除に関しては, 加減算のように統一する方法も考えられるが, 数式の表記に従う方が自然であると考えられる . 例えば万有引力を表す式である以下の式 (2) は, 式 (3) のように変形すれば乗算を多項演算子として用いて表記する事も出来るが, G, Mm, r^2 , それぞれが一つのまとまりとして意味を持っており, それを構造に反映させるものとする .

$$F = G \frac{Mm}{r^2} . \quad (2)$$

$$F = GMmr^{-2} . \quad (3)$$

3.2.4 定数の扱い

π, e などの定数は, 葉ノードに用いられる $\langle ci \rangle$ の中に登場するが, それ自体が数式において意味を持つと考えられるので, 独立した葉ノードとして扱うべきであると考えられる .

3.3 提案する検索方式

数式において, 各演算子は特徴を決めるものとしては対等

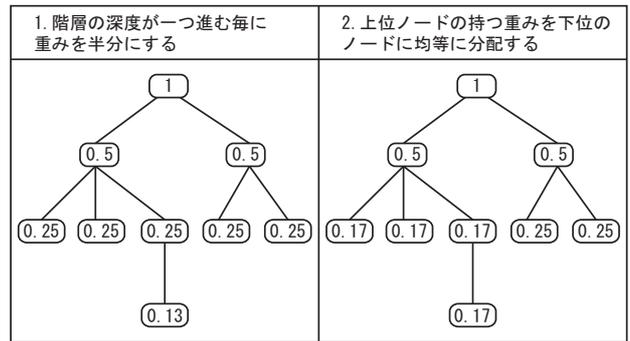


図 8 木構造の各ノードへの重み付け .

ではない . 例えば, 数式を特徴付ける優先度としては, 関係演算子 (等号や不等号), 作用素 (微分や積分), 関数 (任意の $f(x)$ や, \sin 等), 算術演算子 (加減乗除), 分野ごとの定数 (π, e 等), 変数, 数字の順に高い, などの指標が考えられる .

本研究では「木構造の階層が浅いほど重要度が高い」という指標を定めた . これは演算子の出現する木の階層の上下関係と, 式の構造を決める要素の重要度が一致しているという仮定によるものである . この方式は数式の各要素の重要性を, 簡易的にではあるが包括的に表現できると考えている .

3.3.1 数式の内容に基づく木構造を利用した MathML タグへの重み付け

3.2 節による指標を用いて構築した木構造を利用して, 各ノードに対応する MathML タグについての重み付けを行う . 重み付けの方式として以下の 2 種を提案する .

- (1) 階層の深度が一つ進む毎に重みを半分にする
- (2) 上位ノードが持つ重みを下位のノードに均等に分配する

木構造への重み付け手法についての例を図 8 に示す .

3.3.2 同じ MathML タグを持つ式における重みの計量

検索対象となるベクトルとしては, 各 MathML タグの持つ重みを一つとするため, 各ノードに対応するタグに重みをつけた後, 同じタグが一つの式中で出てきた場合, その重みを一つの値とするための計量を行う . その計量方式として以下の 2 種を提案する .

- (1) 同じタグに対応する重みをすべて足し合わせる
- (2) 同じタグに対応する重みのうち最大のものを採用する

3.3.3 類似度の計量

以上に示した 2 つの段階においてそれぞれどちらかを選択し, 数式を MathML タグとその重みからなるベクトルで表現し, コサイン尺度を用いて類似度を計量する .

3.4 総括

本稿では, 数式的内容的な類似を反映した木構造の構築のための考察と, その指針に合った数式の類似検索方式の提案を行った .

今後の課題として、提案方式を実装し、検証を行う。また、数式の構造的な類似に関する考察を継続して行っていく。例えば積分の演算子は、オペランドとして「下限」「上限」「積分変数」「被積分関数」を取り、提案手法ではそれらを同等に扱っているが、構造的な重要度としては差異があるのではないか等の議論があると考えられる。

その他、一般的な木構造の類似検索などの手法についても調査し、数式に適用する際に必要となる事項について考察を行う。

4. 問題解決支援システム “MathGUIDe”

先行研究において、問題解決支援システム “MathGUIDe” を提案した⁴⁾。“MathGUIDe” はマウスやキーボードを用いて、GUI による数式作成が可能なソフトウェアであり、複合連想検索システム⁵⁾への問い合わせ機能を持ち、検索システムの結果出力ウィンドウを持つ。また、作成した数式の計算、グラフ描画を行うことができる。“MathGUIDe”では、Box 型構造文法¹⁵⁾によって数式の作成・変換を可能としている。この Box 型構造文法によって、他の数理ソフトウェアの利用を可能としている。“MathGUIDe”では、計算・グラフ描画にフリーの数値計算ソフトウェアである “Scilab”¹⁶⁾を用いている。Scilab を用いることによって、高度な数値計算やグラフ描画が可能となる。

本プロジェクトでは、検索支援として数式を用いている。そのため数式の作成や編集機能が必要であると考えられる。そこで、“MathGUIDe”の数式作成・編集機能を Web システム化して利用した。

本節では、“MathGUIDe”及び“MathGUIDe”の Web システム化について述べる。

4.1 Box 型構造文法

Box¹⁵⁾は、数式の表記に関するデータ構造を表すものであり、XML 形式を用いて表現する。Box を用いることによって、2 次元的な数式表記が可能となる。また、Box を入れ子構造にすることで、複雑な数式構造を表現することができる。図 9 に Box の例を、図 10 に複雑な数式の例として $y = \sum_{k=0}^n a_k x^k$ を示す。

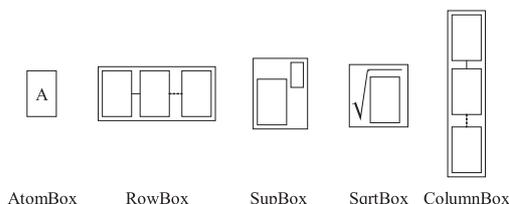


図 9 Box の例。

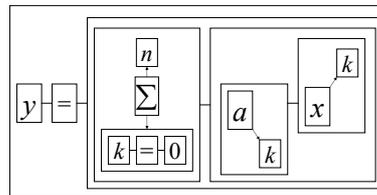


図 10 複雑な Box 構造の例。

Box は 2 次元的な数式表記が可能となるが、意味に関する情報を持っていないため、その Box が何を意味しているのかわからないことがある。例えば図 11 の Box を表 1 のような Box 構造で表すと、A の右上に T があるという表記に関する情報のみを持っており、この構造は転置行列を表すのか、A の T 乗を表すのかこれだけでは判断できない。



図 11 意味を確定できない Box の例。

そこで Box データに意味に関する情報も持たせることができるよう、XML 形式で表した Box データに対して属性として意味を付加する方法が提案された¹⁴⁾。

例として、表 2 に、表 1 の Box 構造に transpose(転置)であるという意味を持たせた Box 構造を示す。このように、自然な数式表現が可能な Box データを用いて、意味を持った数理ソフトウェアの数式データを表すことができるようになる。

表 1 意味を持たない SupBox。

```
<SupBox>
  <AtomBox char="A"/>
  <AtomBox char="T"/>
</SupBox>
```

表 2 意味を持つ SupBox。

```
<SupBox content="transpose">
  <AtomBox char="A"/>
  <AtomBox char="T"/>
</SupBox>
```

4.2 Rings Menu

“MathGUIDe”では、Rings Menu¹³⁾を用いて数式を作成する。Rings Menu には数式や命令の Box テンプレートを登録でき、外側の Ring 中のテンプレートを選択すると、そのサブメニューが内側の Ring 中出现する。このようにすることで、複数階層の構造を表現することができる。Rings Menu においては、内側から外側でなく外側から内側にメニューを展開

してゆく。内側から外側にメニューを展開していくと、外側のメニューを選択する際に Ring 中の反対側にあるメニューを選択したい場合、内側に入らないように回り込んで選択をしなくてはならない。これを逆にすることで、内側に入り込んでも親メニューが変わらずにすむので、高速な入力が可能となる。Rings Menu の欠点として、あるテンプレートの内側には次の階層のテンプレートが存在しているため、階層を多くすればするほど、一つ一つのテンプレートの持つ面積が狭くなり、選択するのが困難になってしまうという点が挙げられる。

図 12 に Rings Menu の概観を示す。



図 12 Rings Menu の概観。

4.3 トランスレータ

“MathGUIde” ではトランスレータを用いて Box データの変換を行っている。トランスレータの関連図を図 13 に示す。

- MathML トランスレータ
Box データから MathML への変換を行う
- Box データトランスレータ
MathML から Box データへの変換を行う
- Scilab ソースコードトランスレータ
MathML から Scilab ソースコードへの変換を行う

4.4 Web システム化

本プロジェクトでは、検索支援システムにおいて数式の作成・編集を行えるようにするため、“MathGUIde” を web 化

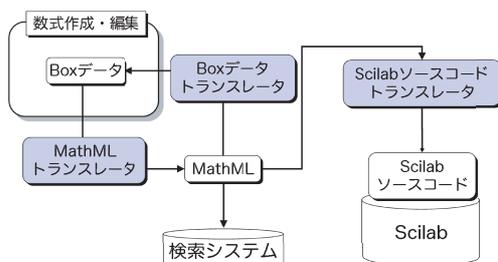


図 13 トランスレータの関連図。

したシステム “MathGUIde/web” の提案・実装を行った。今回は数式作成・編集部分について Web システム化を行った。

実装にあたって Ajax(Asynchronous JavaScript + XML)¹⁷⁾ を用いることで、ページ遷移なしで数式操作を行えるようにした。Ajax とは、JavaScript や XML, DOM(Document Object Model), XMLHttpRequest などの技術を組み合わせることで非同期通信処理を実現する方法である。

実装を行うにあたり、以下のような切り分けで “MathGUIde/web” の構築を行った。

- GUI に関する処理 (Event 検知, 数式表示等)
クライアント側 javascript
- Event 処理, 数式の作成・編集処理, トランスレータ等
サーバ側 java

プロトタイプシステム “MathGUIde” は Java で書かれており、サーバプログラムに Java のソースを使いまわすことによって実装の簡略化を行った。新たに GUI に関する処理の実装を行うにあたり、JavaScript のメジャーライブラリである prototype.js を使い、RingsMenu や数式の描画に Dojo ToolKit を用いている。また、通信に関しては JavaScript-Java 間に DWR フレームワークを利用している。

システムの概要図を図 14 に示す。検索システムでは、数式は MathML を用いて表している。MathML を Box データトランスレータを介することで Box に変換し、“MathGUIde” によって編集が行えるようになる。また、Box は MathML トランスレータによって MathML に変換し、数式検索システムで編集・作成した数式が利用可能になる。

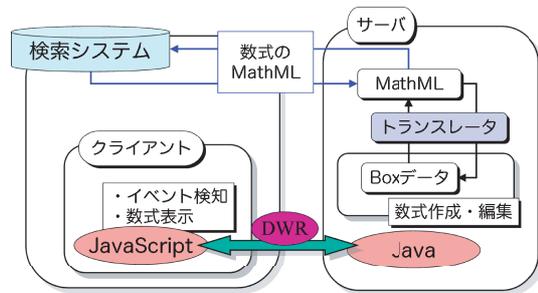


図 14 “MathGUIde/web” の概要図。

5. 数式と索引語による複合連想検索支援システムの構築

前章までのモジュール群を統合し、専門性の高い知識群を対象とした専門用語と数式による複合連想検索支援システムを実装する。システムは以下のように構成される。

- (1) 語の提示方式の選択
2.3.3. 節及び 2.3.4. で述べた、索引との前方一致による補完及び専門書籍章立てに基づいた検索語の提示が

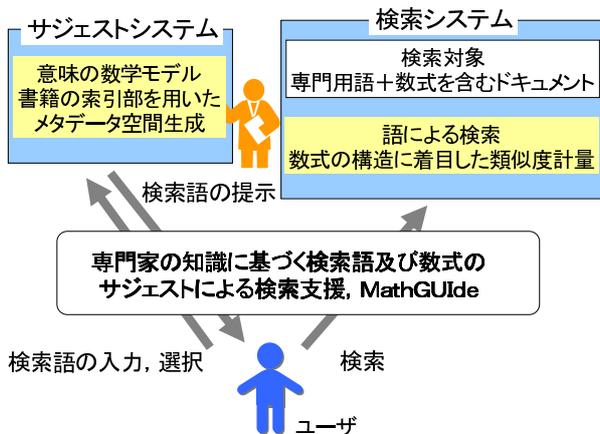


図 15 複合連想検索支援システムの概要図。

ら語の提示方式を選択する。

- (2) 検索語の入力または選択
索引との前方一致による補完を選択した場合は、任意の語を入力する。専門書籍章立てに基づいた検索語の提示を選択した場合は、任意の章と索引語を選択する。
- (3) 専門家の知識に基づいた検索語と数式の提示
2節で述べた方式を用いて、専門家の知識に基づいて選択された索引語と意味的に関連のある検索語及び数式を提示する。
- (4) 検索語及び数式の選択
提示された検索語と数式の中から任意のものを選択する。
- (5) 数式の編集
必要に応じて、選択した数式を4節で述べた MathGUIDe を用いて編集する。
- (6) 検索語及び数式による検索
選択された検索語を用いて文書検索を行う。さらに、3節で述べた方針に基づいて数式による検索を行う。これらを組み合わせることにより、数式と専門用語による複合的な連想検索支援方式を実現する。

前方一致による検索語入力支援機能を利用した場合のシステムの実行例を図 16 に示す。この例では、検索語を途中まで入力するだけで候補となる用語が補完されて提示され、さらに補完された各語と意味的に関連のある用語と数式がサジェストされる。ユーザは、これらの中から目的にあったものを選択し、検索クエリとして用いることが可能となる。さらに、選択した数式を必要に応じて編集して利用することも可能である。

6. 結 論

本稿では、科学技術の分野に現れる数式と専門用語を対象

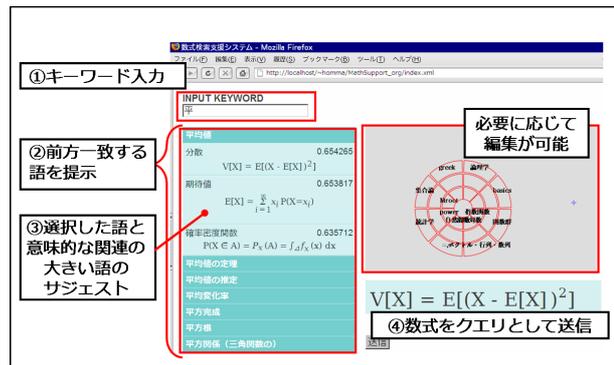


図 16 システム実行例。

とした想起支援および検索クエリとしての利用支援を実現するため、専門用語間の意味的な関連に基づくサジェスト機能、専門的な知識を持たないユーザへの専門用語選択の支援を行うためのインターフェース、および MathML に基づく木構造に基づく数式の類似度計量方式について議論し、これらのプロトタイプを実装した。そして、これらのシステム群を統合することにより、数式と索引語による複合連想検索支援システムの構築を行った。

今後の課題として、より適切な数式の類似度評価指標の確立、および類似度計量における数式データ構造の自動取得などがある。また、複数の分野にまたがる検索対象を想定したシステムの拡張や提案方式の性能評価が挙げられる。

謝辞 本研究の一部は、魅力ある大学院教育イニシアティブ「実践 IT 力を備えた高度情報学人材育成プログラム」による。

参 考 文 献

- 1) 橋本英樹, 土方嘉徳, 西田正吾, MathML を対象とした数式検索のためのインデックスに関する調査, 情報処理学会研究報告, Vol.2007-DBS-142, pp.55-59 (2007).
- 2) 岩瀬重雄, 高校数学公式活用辞典 第三版, 旺文社 (2002).
- 3) 岸本貞弥, 本間秀典, 芳村 亮, 杉山貴俊, 成嶋 諒, 数理分野における問題解決支援システムの構築, システム開発型研究プロジェクト 2007 年度研究成果報告, 筑波大学 (2008).
- 4) Kishimoto, S., Murakata, M., Sakurai, T. and Kitagawa, T., Problem-Solving Support System for Mathematical Sciences, *Proc. the Third IEEE International Workshop on Databases for Next-Generation Researchers*, pp. 79-84 (2007).
- 5) Kishimoto, S., Nakanishi, T., Murakata, M., Otsuka, T., Sakurai, T. and Kitagawa, T., An Implementation Method of an Integrated Associative Search for Mathematical Expressions, *Proc. the IASTED International Conference on Databases and Applications*, Innsbruck, pp. 160-167 (2006).
- 6) 岸本貞弥, 中西崇文, 櫻井鉄也, 北川高嗣, 栃木敏子,

MathML を用いた類似数式検索方式の実現, 第 14 回データ工学ワークショップ (DEWS2003) 論文集 (2003).

- 7) 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 (1993).
 - 8) Kiyoki, Y., Kitagawa, T. and Hayama, T., A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning, *Multimedia Data Management – using metadata to integrate and apply digital media –*, *Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media* (Sheth, A. and Klas, W., eds.), McGrawHill, chapter7 (1998).
 - 9) 清木 康, 金子昌史, 北川高嗣, 意味の数学モデルによる画像データベース探索方式とその学習機構, 電子情報通信学会論文誌 (D-II), Vol.J79-D-II, No.4, pp.509–519 (1996).
 - 10) 中西崇文, 岸本貞弥, 櫻井鉄也, 北川高嗣, 特定分野を対象とした連想検索のための書籍の索引部を用いたメタデータ空間生成方式, 電子情報通信学会論文誌, Vol.J88–D–I, No.4, pp.840–851 (2005).
 - 11) 名取 亮, 線形計算, 朝倉書店 (1993).
 - 12) 小田切健一, 村田剛志, MathML を用いた数式検索, 第 22 回人工知能学会全国大会 (JSAI2008) 論文集 (2008).
 - 13) 大塚 透, 数理ソフトウェアのための入力支援システム生成系に関する研究, 修士論文, 筑波大学 (2005).
 - 14) 大塚 透, 櫻井鉄也, 拡張可能な GUI システム “exGUIDe” を用いた数理ソフトウェア利用支援, 日本応用数理学会 2004 年度年会後援予稿集, 日本応用数理学会 (2004).
 - 15) 櫻井鉄也, 趙 燕結, 杉浦 洋, 鳥居達生, 自然な数式表記のためのユーザインタフェース, 日本応用数理学会論文誌, Vol.6, No.1, pp.147–157 (1996).
 - 16) Inria., Scilab Home Page. <http://www.scilab.org/> .
 - 17) Garrett, J. J., Ajax : A New Approach to Web Applications. <http://www.adaptivepath.com/ideas/essays/archives/000385.php> .
 - 18) W3C Math Home <http://www.w3.org/Math/> .
-