

構造化文書を対象とした意味的検索の一手法

森村一雄 吉川正俊 植村俊亮

奈良先端科学技術大学院大学 情報科学研究科

本稿では、文献の表層的な情報を利用するだけでなく、文献の深層的な情報である意味的な概念構造に着目し、文献検索システム側でこれらを抽出し索引として利用するための方法について述べる。さらに、従来の情報検索モデルで行われてきた問合せに比べ、より表現力のある意味的な問合せ手法の例を示し、これらの実現方法について考察する。意味的な概念構造は、その粒度に応じて、構成要素（章や節などの文献の構成単位）内の語句レベルに着目した局所的なものと、構成要素間の関連に着目した大局的なものと二種類に分類できる。本稿では、二種類の概念構造をそれぞれ抽出し、これを文献検索システムの統合的な索引として利用することを提案する。そして、意味的な概念構造を利用者が指定できるような問合せ手法について説明し、その実現方法および実験システムの概要について述べる。

Semantic Information Retrieval for Structured Documents

Kazuo MORIMURA Masatoshi YOSHIKAWA Shunsuke UEMURA

Graduate School of Information Science,
Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara 630-01, Japan

This paper describes a method of extracting conceptual structures from structured full text databases, and shows their application to indices in information retrieval systems. We take notice of semantic structures of documents, as well as information from surface layer of documents and an appearance of them. Furthermore, we represent examples of a way of semantic retrieval using conceptual structures. Conceptual structures are classified into local structures (in documents' component, e.g. chapter, section) and global structures (between documents' component) according to their granularity. We attempt to extract two-level semantic conceptual structures from documents, and to make use of them as indices of information retrieval systems. Then, we present a method of semantic retrieval by using conceptual structures and show an experimental system which realizes semantic retrieval.

1. はじめに

従来の文献検索システムにおいては、索引に用いる情報として、全文情報の表層的な情報(文献名、概要、キーワード、著者名等の書誌情報、出現頻度の高い語句等の重要語情報、章立てなどの構成情報、など)にしか着目していないものが多い。これらの情報のみを利用した場合は、利用者の問合せ時における、検索式の構成に際する表現力(expressivity)は高いとはいえない[5]。

既存の文献検索システムのこのような問題点の原因としては、(1) 検索式の構成における表現力が低いこと、(2) 検索システムの検索対象(索引)として、文献の表層的な情報、すなわち字面のみしか考慮していないこと、などの点が考えられる。

本研究では、これらを鑑み、文献検索において文献の表層情報を利用するだけでなく、文献の深層的な意味構造を利用することを考える。すなわち、(1) 文献を抽象化した意味的な概念構造を文献の深層的な情報として捉え、これを抽出すること、(2) 文献検索の際に概念構造を索引として利用すること、(3) 概念構造を指定することにより、より表現力のある問合せを行なえるような検索式の構成法を用いること、などにより、表層的な情報のみを利用した検索システムに比べてより検索精度を高めることのできるような文献検索システムを提案することを目的とする。

また、本研究では実験材料として、HTML化された論文を公開しているWWWサイト¹より、論文約110本を入手し、実験システムを構築した。

以降、2節では、本研究で文献検索システムの索引として扱う、文献の概念構造について述べる。3節では、概念構造の抽出法の一例について述べる。4節では、抽出した概念構造を用いた問合せ手法について述べる。また、実験システムにの構成についても触れる。最後に、5節で、まとめを述べる。

2. 文献の概念構造

本節では、文献の概念構造について述べる。まず、概念構造の定義を与える。概念構造は、文献内における概念間の意味的な関連をネットワークとして表現するものであり、その粒度により、大局的なものと局所的なものとの二通りを考える。

2.1 概念構造の定義

本節では、概念構造の定義を考える。一般に、文献等は、ある事実や知識の集合として捉えることができるものであるが、個々の事実や知識といったものはそれぞれが独立して存在しているのではなく、何らかの関連や制約を伴って存在していることが多い。さらに、個々の事実や知識は、類似や詳細や抽象といった関係

(制約)の下に、ある種の集合構造を形成することが多い(例:シソーラス)。このような事実や知識の集合を概念と呼ぶことにする。また、ある一つ概念における個々の事実や知識を、概念要素と呼ぶことにする。文献等における概念要素の最小単位は語句である。

これら、概念や概念要素同士の関連や制約、および概念の抽象化などを総称して、概念構造と呼ぶことにする。一般的には、概念構造は、複数の概念要素とそれらの集合である概念、及び概念要素や概念の間の関連とからなる。概念と概念要素は、属性として、その名称(それぞれ、概念名、概念要素名と呼ぶ)と重み(それぞれ、概念の重み、概念要素の重みと呼ぶ)を持つ。関連は、その属性として、関連の種類(関連名と呼ぶ)と関連の度合い(関連の重みと呼ぶ)からなる。概念構造は、直観的には、概念要素(又は概念)を端点、関連を概念要素(または概念)間を結ぶ枝として表した、ネットワーク構造で表現される(図1)。

2.2 文献における概念構造

前節で示した概念構造の定義とその例は一般的なものである。対象を文献に限定して考えると、概念構造は、語と語との関連、文と文との関連、といったように、文献の構造や抽象度に応じた様々な粒度のものが存在する。本研究では、(1) 局所的な概念構造として、文献の構成要素(章や節などの構造的なまとまり)における語句間の関連、(2) 大局的な概念構造として、構成要素同士の関連、に着目し、これらを抽出し、索引として利用することを考える。次に、それぞれの概念構造について述べる。

2.3 局所的な概念構造

文献の構成要素(章や節など)内における語句間の関連からなる概念構造である。概念要素の単位は、語句である。よって、概念は語句の集合となる。

概念と概念要素

概念は、ある条件や制約のもとに概念要素をクラスタリングしたものである。局所的な概念構造においては、概念要素は文献内の語句であるので、概念は、語句をクラスタリングしたものになる。クラスタリングの方法としては従来から様々な手法が提案されている[2]。本研究で用いている手法については、3.1節で説明する。

概念要素間の関連

概念要素の単位を語句とした場合、概念要素間の関連は、端点となる語句がどのように存在しているかによって、様々な分類を行なうことができるが、本研究では、概念要素間の関連の抽出処理²の都合上、同一文脈上にある語句同士で、しかも同一文の中における関連のみを扱う。

概念間の関連

²3.1節にて説明する

¹<http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/WWW2/Proceedings.html> など

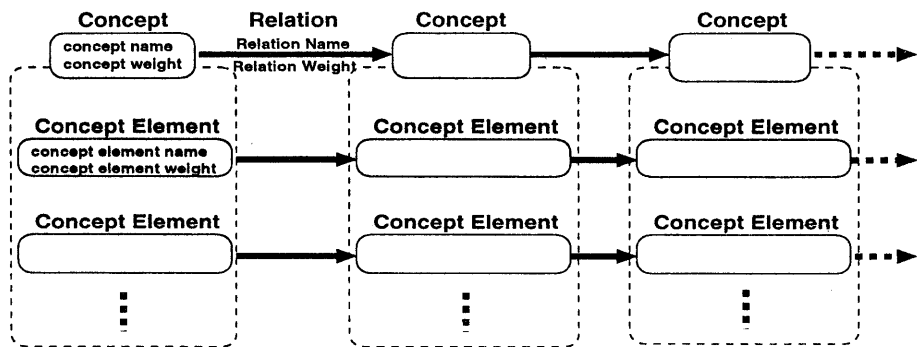


図 1 一般的な概念構造

概念間の関連においては、関連の種類は一概に特定できない。一般に、概念は複数の概念要素（語句）の集合であり、互いに異なる集合間の概念間の関連は特定できない。よって、概念間の関連は、属性として、関連の重みのみを持つ。概念は概念要素を抽象化しているものと捉えることができることから、概念間の関連は、概念要素間の関連よりもより抽象度が高い関連であるといえる。

2.4 大局的な概念構造

大局的な概念構造は、文献の構成要素（章や節など）間の関連による概念構造である。本研究では、検索対象を学術論文に限定している。このことにより、文献の構成面（章や節の構成面）における典型的なスキーマを何通りか想定することができる。

3. 概念構造の抽出

本節では、概念構造の抽出法について、実験により確認した手法について述べる。局所的な概念構造と大局的な概念構造に分けて説明する。

3.1 局所的な概念構造の抽出

基本的な考え方は、浅いレベルの自然言語処理 [6] と、通常の情報検索で良く使われる統計的手法を組み合わせて概念構造の抽出を行なう、といったものである。語句間の関連語の抽出には、形態素解析の結果、動詞句として判定される語句を利用する。

概念構造の抽出処理は、概念の抽出と関連の抽出とに大別される。それぞれの処理について説明する。

処理 1：概念の抽出

i) 概念要素の候補のリストアップ

概念要素の候補となる語句は、基本的には語句の出現頻度により決定される。ここで用いている方

法は、「不要語で区切られた複数の語句を連語とみなして、それらの出現頻度を求め、出現頻度がある閾値以上のそれらの語句を概念要素の候補とする」といったヒューリスティック（ルール）に基づく。

ii) 概念要素のグループ化による概念の生成

i) の結果である概念要素の候補リストより、出現頻度がある閾値より大きい語句を、それぞれ概念を代表する概念要素とし、その概念の概念名とする。そして、概念名と部分一致する概念要素を、該当する概念に加える。概念名の重みは、元の概念要素の重みに、その概念に属する他の概念要素の重みを合計した値となる。さらに、概念又は概念要素の重みの値に、次のような正規化処理を施す。

$$\begin{aligned} \text{概念または概念要素の重み} &= \\ &\frac{\text{概念または概念要素の出現頻度}}{\text{概念または概念要素が属する構成要素の語数}} \end{aligned}$$

処理 1 の結果の例を示す（表 1）。

表 1 概念の生成処理の結果の例

概念/概念要素名	重み
<概念>cache	23.1
<概念要素>cache	8.4
<概念要素>field cache analysis	2.1
<概念要素>proxy server cache	2.1
<概念要素>cache tends	2.1
<概念要素>cache scan	2.1
<概念要素>cache fills	2.1
<概念要素>cache analysis program	2.1
<概念要素>cache hit	2.1

処理 2：関連の抽出

i) 概念要素間の関連名の決定

処理概要としては、(1) 先に得られた概念のうち、互いに異なる概念に属する二つの概念要素を含む文を見つける、(2) その文中で、二つの概念要素をそれぞれ主節と目的節として見た場合動詞句に相当する語句を、形態素解析処理を行うことにより抽出する、(3) (2) で得られた動詞句を、二つの概念要素間の関連名とする、といったものである。英文形態素解析ツールとしては、tagger [1] を用いている。

tagger を用いた形態素解析処理を実際のテキストに適用した結果を示す。

When/WRB browsing/VBG the/DT Web,/NNP requests/NNS are/VBP usually/RB sent/VBN directly/RB from/IN the/DT browser/NN to/TO the/DT Web/NNP server/NN of/IN interest,NN wherever/WRB in/IN the/DT world/NN that/WDT may/MD be./CD

ここでは、動詞句に関する形態素タグである /VB,/VBN,/VBG,/VBZ,/VBD に着目する。上記の文で、概念要素 “requests” と概念要素 “browser” との関連名を決める場合は、“requests” と “browser” に挟まれた動詞句である “sent” が、関連名として抽出される。

ii) 関連の重みの決定

関連の重みには、(a) 概念要素間の関連の重み、(b) 概念間の関連の重み、の二種類がある。それぞれの計算方法を示す。

(1) 概念要素間の関連の重み

概念要素間の関連の重みは、リンクの端点となる概念要素名が一致し、かつ、関連名が等しいようなリンクの数を集計することによって求める。

実際には、求めた重みの値に、次のような正規化処理を施す。先に求めた重みを w_{link} とすると、

$$\text{概念要素間の関連の重み} = \frac{\text{概念要素が属する構成要素の文の数}}{\text{概念要素が属する構成要素の文の数}}$$

(2) 概念間の関連の重み

概念間の関連の重みは、概念に属する概念要素間の関連の重みを合計した値となる。

以上により、文献の構成要素（章、節など）内における局所的な概念構造を、各概念を葉、関連名及び関連の重みを枝属性とするネットワークとして表現する（図2）。

実際には、概念構造を文献検索における索引として利用するために、局所的な概念構造に、次のような属性を付加する。

- 構成要素名（題目）
- ロケーション（通し番号、“3.1”といった形式）
- URL（“http://…/…html#sec.1.1”といった形式³⁾）

3.2 構成要素間の類似度の算出

本節では、構成要素間の類似度、すなわち、局所的な概念構造の類似度の算出法について述べる。

この計算法は、ベクトル空間モデル [2] における文献ベクトル間の類似度の計算法を元にした方法で行なう。

ベクトル空間モデルでは、文献ベクトルの類似度は、文献ベクトル同士の内積を求めることで、計算できる。たとえば、文献ベクトル V_{D1}, V_{D2} について、 $V_{D1} = ((network, 40), (cache, 30), (database, 5))$ $V_{D2} = ((query, 20), (network, 10), (database, 40))$ とすると、類似度 $sim(V_{D1}, V_{D2}) = 40 \times 10 + 5 \times 40 = 600$ となる。

本研究においては、局所的な概念構造は木構造で表現される。よって、木構造同士のマッチングをとることで類似度を計算できる。ここでは、概念構造をベクトルとみなすことで、次のような計算により類似度を求めることができる。

構成要素 S_1, S_2 の類似度は、以下に示す $\alpha, \beta, \gamma, \delta$ を用いて、次式で求められる。 $(C_1, C_2, C_3, C_4$ はそれぞれ定数を表す。)

$$\text{sim}(S_1, S_2) = C_1 \times \alpha + C_2 \times \beta + C_3 \times \gamma + C_4 \times \delta$$

この式で、 $\alpha, \beta, \gamma, \delta$ は、次のような値である。

 α : 概念の類似度

α は、概念構造における概念名及びその重みを概念ベクトルとみなして、内積計算を行なった値である。

 β : 概念要素の類似度

β は、概念名が一致した概念において、概念要素名及びその重みを概念ベクトルとみなして、内積計算を行なった値である。

 γ : 概念要素間の関連の類似度 γ は、次のような値である。

関連名が一致している概念要素間の関連において、双方の所属する概念名が一致していて、かつ、双方の概念要素名が一致している、概念要素間の関連の重みの合計値

なお、 γ は、完全一致あるいは部分一致する概念が二個以上ある時に算出可能である。

 δ : 概念間の関連の類似度

δ は、双方の概念名が一致している、概念間の関連の重みの合計値である。

なお、 δ も、完全一致あるいは部分一致する概念が二個以上ある時に算出可能。

³データベースに保持される HTML 文書においては、該当する箇所にといったタグを埋め込んであるものとする。

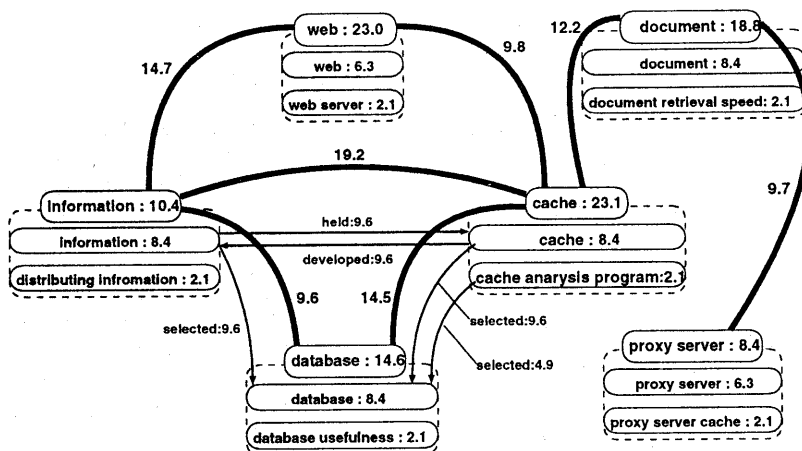


図 2 局所的な概念構造の例

3.3 大局的な概念構造

大局的な概念構造は、文献の構成要素間の関連による概念構造である。本研究では、検索対象を科学技術論文に限定している。このことより、対象となるドメインから得られる知識をトップダウン的に利用することによって、文献の章や節の構成面における一般的なスキーマを何通りか想定することができる。

さらに、文献を次の三つの観点から捉えることにより、大局的な概念構造を三つの独立したビューからなる索引として扱うことを考えていく。

i) 構成要素間の階層構造に着目することによる概念構造

構造化文書を、文献の題名を表わす構成要素を頂点とした木構造として捉える。木構造に基づく隣接している端点（構成要素）間に、何らかの関連があるものとする。枝属性としては、関連名は明示的に抽出することは困難である。よって、関連の重みのみを、3.2節で説明した構成要素間の類似度を求める手法を用いて計算することにより、求める。

ii) 構成要素間の意味的な関連に着目することによる概念構造

「概要」や「序論」に関する構成要素は、それぞれ文献全体を抽象化したものであると考えることができる。よって、「概要」や「序論」に関する構成要素とそれ以外の構成要素には、何らかの関連がある、とみなすことができる。これらの関連の重みを、3.2節で説明した構成要素間の類似度を求

める手法を用いて計算することにより、求めることができる。

同様に、「序論」に関する構成要素と「結論」に関する構成要素の間にも関連がある、とみなすことができる。これについても、関連の重みを求めることができる。

iii) 文献の性格に基づく概念構造

個々の文献の持つ性格に基づく概念構造である。文献集合の中には、その章立てを見ることにより、文献の性格を推測することのできるものが存在する。たとえば、「実験」「実験材料」「実験方法」「実験結果」「結果に対する考察」などに関する構成要素がある文献が含まれているとすれば、その文献は、何らかの「実験」に関するものである、と推測することができるだろう。同様に、「開発」に関する文献、「証明」に関する文献、「応用」に関する文献、なども、推測することができる。ここでは、ある文献が「実験」に関するものである、と推測した場合の、概念構造の与え方について述べる。

(1) 以下に相当する構成要素を抽出する。

- 「実験材料」に関する構成要素 (material)
- 「実験方法」に関する構成要素 (method)
- 「実験結果」に関する構成要素 (effect)
- 「実験結果に対する考察」に関する構成要素 (examination)

(2) 枝属性の関連名としては、たとえば構成要素「材料-方法」の間の関連の場合 “material-method”、構成要素「材料-結果」の間の関

連の場合“material-effect”、といった関連名を与える。

関連名は、構成要素間のリンクとして使用できる。たとえば、「実験材料と方法に関する構成要素は？」といった問合せ時、局所的な概念構造を全て検索することなしに、構成要素間のリンク属性をみるだけで結果を得ることができる。

開発主体の論文や、証明主体の論文等も、同様の手法により、関連属性を与える。

なお、構成要素間の関連の重みは、3.2節で説明した方法を用いて計算する。

以上により、文献の構成要素の関連である大局的な概念構造を、ネットワークとして表現する。

3.4 二種類の概念構造の統合

これまで説明してきた二種類の概念構造を統合して表現すると、次のようになる(図3)。

このような統合された概念構造は、文献を、大局的な意味構造と局所的な意味構造といった二種類の視点で捉えるものである。概念構造を用いることによる利点を挙げる。

- i) 様々な形状(大きさなど)を取りうる文献を、(半)構造化されたものとしてみなすことができる
- ii) 通常の全文検索システムに比べ、問合せ時の計算コストが少なくて済む(検索実行時に全文を走査する必要がないため)
- iii) 二種類の粒度をもった概念構造に基づく意味的な検索が可能(二種類の概念構造を、異なった視点からのビューとして利用)

実際の検索は、問合せ式と二種類の概念構造とのマッチングにより行なうことになる。これらについて、4節で説明する。

4. 概念構造を用いた文献検索

本節では、概念構造の利用例として、概念構造を用いた文献検索システムについて述べる。

4.1 概念構造を用いた問合せ手法について

概念構造に基づく問合せ[4, 3]として、次のようなものを考える。ここでは、問合せの種類を、利用者が検索条件を明示するかしないか、に基づいて分類する。

4.1.1 検索条件を明示した問合せ

利用者が条件を明示する問合せである。

- i) 文献の構造に関する問合せ
文献の構成要素の名前と、それに含まれる概念名に関する問合せである。

構文: Include (StructureName = 構成要素名, ConceptName = 概念名)

例: Include (StructureName = “Introduction”, ConceptName = “network”)

検索処理の概要: 局所的な概念構造を索引として使用する。

結果のランキングのためのルール:

- (1) 構成要素 “Introduction” に、概念名 “network” である概念を含む文献
- (2) 構成要素 “Introduction” に、概念要素 “network” が存在するような概念を含む文献

上記の順、かつ、概念/概念要素の重み(この場合、概念/概念要素 “network” の重み)の値の順で結果を返す。

- ii) 概念要素間の関連に基づく問い合わせ
構成要素内の、概念/概念要素間の関連に基づく問合せである。

構文: Related (概念₁, 概念₂, ..., 概念_n)

例: Related (“database”, “network”)

検索処理の概要: 局所的な概念構造を索引として使用する。

結果のランキングのためのルール:

- (1) 概念名が “database” である概念と、概念名が “network” である概念との関連の重みが大きい文献の構成要素
- (2) 概念名が “database” である概念と、概念要素 “network” が存在する概念との関連の重みが大きい文献の構成要素
- (3) 概念要素 “database” が存在する概念と、概念名が “network” である概念との関連の重みが大きい文献の構成要素
- (4) 概念要素 “database” が存在する概念と、概念要素 “network” が存在する概念との関連の重みが大きい文献の構成要素

の順で結果を返す。

- iii) 文献の構造と、概念要素間の関連とを取り混ぜた問合せ

例: 「構成要素: Experiment に概念: network を含み、かつ、概念: network と概念: cache との関連があるような文献は？」といった問合せ

検索処理の概要: Include 演算と Related 演算の積/和をとることで実現できる。この例の場合、

Include (StructureName = “Experiment”, ConceptName = “network”) and Related (“database”, “network”) となる。

Include 演算と Related 演算の両者とも、結果として URL 集合を返す。よって、Include 演算と Related 演算の結果の積または和を計算することにより、求められる。

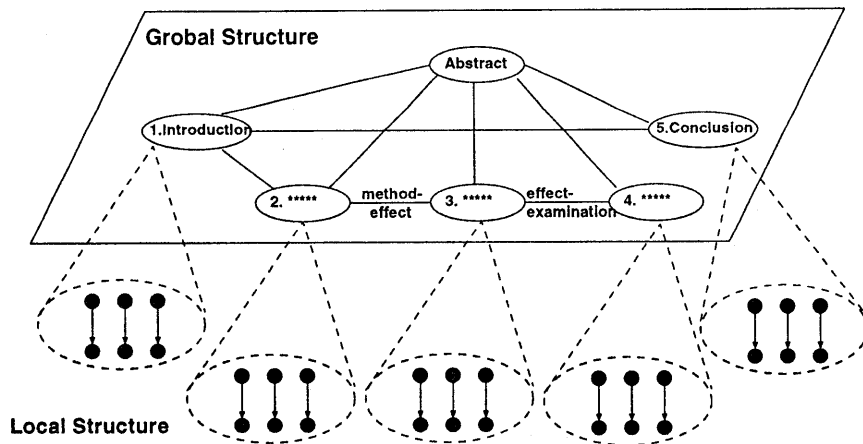


図 3 二種類の概念構造の統合

- iv) 概念要素間の関連名に基づく問い合わせ
局所的な概念構造および大局的な概念構造における、概念要素間の関連名に対する問合せである。

構文： RelationName (関連名 | global 関連名)
例： RelationName (“describe”)

RelationName (global “material-method”)

検索処理の概要： 関連名としては局所的な概念構造における概念要素間の関連、大局的な概念構造における構成要素間の関連、の二種類がある。与えられた関連名に、キーワード “global” が付加されている場合、大局的な関連名として扱う。

- (1) 局所的な概念構造における関連名の指定 (デフォルト)
指定された関連名を含む概念要素間の関連を含む概念構造を持つような、構成要素を返す。
- (2) 大局的な概念構造における関連名の指定 (キーワード “global” で明示)
指定された関連名を含む構成要素間の関連を持つような文献を返す。

結果のランキングのためのルールとしては、関連の重みが大きいものから提示する、といったものである。

- v) 図や表を目印とした問合せ

図や表といったものは、文献の構成要素を抽象化している、と考えられる。これらを目印とした問

合せを考える⁴

構文： figure (概念名), table (概念名)

- vi) 観点を指定する問合せ

「…の立場／文脈／観点 から、…について書かれている文献は？」といった、利用者の観点を指定した問合せである。

より抽象度が高い問合せである。

構文： View (ViewName = 概念名, Concept-Name = 概念名)

検索処理の概要： 大局的な概念構造と、局所的な概念構造を索引として利用する。

- (1) 文献の構成上、親子関係／兄弟関係にあるような構成要素間の関連を探し出す。
- (2) 親／兄にあたる構成要素に、ViewName で示される概念名または概念要素が存在し、かつ、子／弟にあたる構成要素に、ConceptName で示される概念名または概念要素が存在するような、構成要素間の関連を選択する。
- (3) 選択された関連を、ランキング処理を経て、提示する

以上の演算の他に、フィルタとして、次のようなもの考える。

⁴figure, table といった、図や表に関する語が、予め局所的な概念構造として抽出されていることを前提としている。図や表といった文献中の物理的な指標は、本研究で定義している“概念”とは厳密には異なるものであるが、利便性を考慮して、あえて概念構造として扱うこととする。

- i) Document ()
文献全体を表す URL のみを返す。
- ii) Chapter ()
章を表す URL のみを返す。
- iii) Section ()
節を表す URL のみを返す。
- iv) Subsection ()
副節を表す URL のみを返す。
- v) SelectLocationName(ロケーション名)
指定したロケーション名 (例: “1.1”, “1-3” など) を選択する。

4.1.2 検索条件を明示しない問合せ (類似検索)

4.1.1節で示した問合せを利用者が繰り返す過程において、ある問合せ結果に類似している文献を検索することができれば、利用者にとってより便利である、と考えられる。本節では、このような、利用者が検索条件を明示せずに問合せを行なうことを考える。一例としては、「…という既知の文献の…節に類似している文献 (の構成要素) は？」といったものである。(類似検索と呼ぶ)

i) 類似検索

構文: similar_to (既知文献の URL)

検索処理の概要: 既知文献の URL より該当する構成要素を探し、構成要素同士の類似度を計算することにより、類似文献の構成要素を検索する。検索結果としては、類似度の値の高い順にランキングを行なうことにより、提示する。

4.2 実験システムの構成

4.1節で示した問合せを実現するための構文を利用者が直接記述するのではなく、上位層に利用者にとって扱いやすいユーザインタフェース (UI) 層を設けることを考える。UI としては、[7]と同様に、WWW ブラウザを用いることを考える。それぞれの問合せは、独立したモジュールとして実装することを考える。利用者の検索要求は、httpd の CGI 機能を利用して、該当するモジュールへ引数として渡され、検索が実行される。検索結果としては、概念構造の属性の一つである URL 集合が返される。

検索システムの処理概要を示す。

- i) 利用者は、WWW ブラウザの検索フォームよりインタラクティブに条件を入力すること等により、問合せを表現する。
- ii) 問合せ解析 CGI スクリプトは、検索条件を引数として、適切なモジュール (CGI スクリプト) を呼び出す。
- iii) 呼び出されたモジュールは、索引 (概念構造ファイル) を検索し、検索結果を URL のリストとして返す。
- iv) 問合せ解析 CGI スクリプトは、検索結果を表示する。
表示方法としては、次のようなものがある。
 - 返された URL リストを持った HTML 文書を合成する
 - 返された URL リストにアクセスして HTML 文書を表示する

実際には、これらの問合せ解析処理の前処理として、概念構造の抽出処理が入る。概念構造の抽出処理は、一回実行しておけば、文献データベースに文献が追加されない限り、行なう必要はない。

5. まとめ

本稿では、文献の表層的な意味情報のみだけでなく、文献の深層的な概念構造情報に着目し、これを抽出し索引として用いることで、より表現力のある文献検索を行なうことのできるようなシステムを提案した。

さらに、粒度の異なる二種類の概念構造を示し、それらの抽出法の一例を示した。さらに、抽出された概念構造を用いた意味的な問合せを行なうことのできるような文献検索システムについて、その問合せ手法の詳細と、システムの全体構成について示した。

参考文献

- [1] Eric Brill. Some Advances in Transformation-Based Part of Speech Tagging. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, 1994.
- [2] William B. Frakes and Ricardo Baeza-Yates, editors. *Information Retrieval - Data Structures & Algorithms*. Prentice-Hall, 1992.
- [3] Michael Fuller, Eric Mackie, Ron Sacks-Davis, and Ross Wilkinson. Structured Answers for a Large Structured Document Collection. In *SIGIR '93 16th Annual International ACM SIGIR Conference*, pp. 204-213, 1993.
- [4] Pekka Kilpeläinen and Heikki Mannila. Retrieval from hierarchical texts by partial patterns. In *Proc. of ACM SIGIR '93*, pp. 214-222, 1993.
- [5] Gonzalo Navarro and Ricardo Baeza-Yates. A Language for Queries on Structure and Contents of Textual Databases. In *ACM SIGIR '95*, pp. 93-101, July 1995.
- [6] 鬼塚真, 大沼守一. 単文と意味ネットワークによるオブジェクトモデル設計法. ソフトウェア科学会 第10回オブジェクト指向計算ワークショップ, pp. 1-6, 1994.
- [7] 大山敬三. インターネットに適応した全文データベース検索システムの構成. 情報処理学会 情報学基礎研究会 研究報告, 37-3, pp. 15-22, May 1995.