

意味空間を用いた主題分析法†

鶴身 征雄^{††} 打浪 清一^{†††} 手塚 慶一^{†††}

Abstract

In order to analyze the subjects in the documents automatically, the new subject analysis method using the meaning-space is proposed. The meaning-space is the multi-dimensional space where topological relationships among concepts are expressed. The mechanism of various kinds of mapping between the description levels such as surface level, deep level and semantic level are clarified according to our linguistic algebraic model. The algorithm of automatic subject analysis method is described and the I. R. system using this method is constructed. It is confirmed experimentally that this system has a high performance and a short retrieval time.

1. まえがき

言語情報を対象にした検索システムの多くは、検索対象集合を順次細分していき、要求情報を含むと思われる部分集合を抽出する探索方法や、言語情報に離散的なラベルづけを行なう主題分析法などを用いている。しかし、このような離散的な検索システムは、同一概念が異分野に現われたり、意味的に近い概念が離れた位置に配置されたりするために意味的な歪を生じ、概念同志の関連性にしても、有無の2値で表わすために呼出率が低下したりする欠点を有している。

したがって、言語情報を意味的に処理するために、言語情報の持つ位相的な性質の表現や処理に適した連続的なデータ構造と処理手順を持つ情報検索システムが望まれる。連続的なデータ構造を持つシステムとしては、接続行列を用いるものがある¹⁾。この方法は、ロールやリンク情報を文献に付加する離散的な方法にくらべ、言語情報の表現や処理に適しており、シソーラスなどを用いることにより呼出率を高めることも可能である。また、この方法は、キーワードを基礎概念として言語情報を表現している。しかし、その基礎となるキーワード空間が意味的に直交^{†††}していない

め適合率が低くなる恐れがある。検索質問は、重みづけを行なったキーワードを列挙して表わしている。しかし、重みづけは、そのシステムをよく知らないものにとって困難な作業であるので、誰でもが自由に検索を行ないえない欠点を有している。

筆者らは、かつて言語情報の持つ位相的な性質に着目して位相型データ構造なる概念を提案し、論文を構成要素とした意味空間を用いた検索システムを構成し検索実験により、意味空間が言語情報を表現するのに適したデータ構造であること、このシステムが高い検索能率と短い検索時間を持つことを確めた²⁾。しかしこのシステムでは、ファイルの更新、検索質問作成の際の基礎となる標準論文の選択、細かい質問概念の表現が困難である。

そこで本論文では、上記の2つの方法の欠点を補うものとして、意味空間を用いる主題分析法を提案し、主題分析アルゴリズムなどについて検討した。この方法は、基礎概念の間の意味的な遠近関係を表現する空間として、これらの基礎概念を要素とする意味空間を構成し、意味空間上に表現された基礎概念に、それらを結合する接続子によって定まる演算を施すことによ

†††一般にキーワード空間では、キーワードを座標軸として選んでい
る。しかも、それらのキーワードを直交した座標軸として取り扱
っているため、意味的に関連のあるキーワードも、意味的に無関
係であるとして取り扱うことになる。したがって、キーワード空
間を直交座標として取り扱う場合には、意味的に無関係なキー
ワードで座標系を構成するか、意味的に関係あるキーワードを使用
する場合には、それらのキーワードを意味的に無関係な概念に分
解して、それらを座標軸として用いる必要がある。

† Some consideration on an automatic subject analysis method using the meaning space, [by Yukio TSURUMI (University of Osaka Prefecture), Seiichi UCHINAMI and Yoshikazu TEZUKA (Osaka University)].

†† 大阪府立大学計算センター

††† 大阪大学工学部通信工学科

り言語情報を意味空間上に表現し主題分析を行なうものであって、次の特徴を有する。

- 基礎概念を意味的に直交した空間上に表現することにより、適合率の向上が期待できる。
- 接続行列を用いる方法の重みづけに相当する操作が構文構造により自動的に定められる。したがって、表現形式や語句に制限はあるが、自然言語に近い形での質問が可能になる。
- 単語のレベルで質問分析が可能なので、細かい質問概念が表現できる。
- ファイルの構成および更新が簡単であるので、検索者が夫々自分に適した検索ファイルを作ることができる。

2. 主題分析の自動化における言語情報代数系

計算機に自然言語を取り扱わせるためには、その構造を知る必要がある。自然言語に近い形式言語のモデルとして、Chomsky の句構造、変形規則からなるものが代表的である。しかし、このモデルは、文の構成過程のみで処理過程を扱っていない。したがって、言語情報の意味的理解、処理過程を取り扱えるモデルが望まれる。このようなモデルとして、筆者らは、図1に示すモデルを提案した³⁾。

このモデルは、表層、深層、意味の3表現レベルからなり、各レベルは、日常言語、表層表現の生成元、および意味空間に相当している。各表現レベルの要素と内・外算法およびそれらの間の写像 Φ, Ψ を抽出することにより言語情報の処理過程が記述される。

しかし、 Φ, Ψ の抽出は、計量心理学的な手法を要求し、統計的処理が必要となる。したがって、 Φ, Ψ の正確な形を求めるためには、モデルの適用範囲を制限する必要が生ずる。本稿では、以下に示す表題文の特徴から、適用範囲を論文表題に限定した。

(1) 構文的特徴

文献表題のほとんどは、句構造をなしており、通常の文におけるような主部・述部といった構形成態

はとらない。また表題中の語句の間に存在する関連は、助詞や一部の動詞を含むような接続語句によって示される。

(2) 語句的特徴

主要な主題を明示するための指標や説明的事項としての機能を持つ文献表題特有の表現が存在する。

- ……に関する考察。
- ……の……への応用。
- ……に対する一提案。

これらの語句は、表層表現と深層表現の関係を定めるのに有用である。

(3) 意味的特徴

文献表題によって表現される論文内容は、一つの静的な概念を表わしていると考えられる。したがって、動的、情緒的な要素は持たず、語の意味も対象分野が限られると一意性を有しており、意味的な多様性を生じにくい。したがって、語または句を要素とする意味空間上の各語句の位置を表わすベクトルを、それらの語句を結合する接続語句によって定められる線形演算で結合するような操作で、文献概念が表現されると考えられる。

論文表題が与えられると、図2に示す手順で主題分析が行なわれる。各表現レベルの構造を以下に示す。

$$\langle \text{論文の表題} \rangle ::= \langle \text{句} \rangle | \langle \text{論文の表題} \rangle \langle \text{AND} \rangle \langle \text{論文の表題} \rangle \quad (2.1a)$$

$$\langle \text{句} \rangle ::= \langle \text{キーワード} \rangle | \langle \text{空語} \rangle \langle \text{句} \rangle | \langle \text{句} \rangle \langle \text{空語} \rangle | \langle \text{常用語} \rangle \langle \text{句} \rangle | \langle \text{句} \rangle \langle \text{常用語} \rangle | \langle \text{句} \rangle \langle \text{接続子} \rangle \langle \text{句} \rangle | \langle \text{句} \rangle \langle \text{接続子} \rangle \langle \text{常用語} \rangle \quad (2.1b)$$

$$\langle \text{論文の構造} \rangle ::= \langle \text{句の構造} \rangle | \langle \text{論文の構造} \rangle \langle \text{接続子} \rangle \langle \text{論文の構造} \rangle \quad (2.2a)$$

$$\langle \text{句の構造} \rangle ::= \langle \langle \text{キーワード} \rangle \rangle | \langle \langle \text{句の構造} \rangle \rangle ; \langle \text{役割} \rangle \quad (2.2b)$$

$$\langle \text{論文の意味} \rangle ::= \langle \text{句の意味} \rangle | \langle \text{句の意味} \rangle \langle \text{AND 演算} \rangle \langle \text{句の意味} \rangle \quad (2.3a)$$

$$\langle \text{句の意味} \rangle ::= \langle \text{キーワードの意味} \rangle | \langle \text{句の意味} \rangle \langle \text{句} \rangle$$

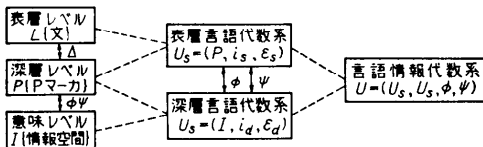


図1 言語情報代数系

Fig. 1 The Linguistic Information Algebraic System.

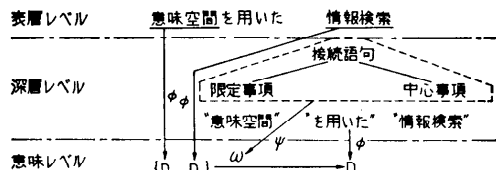


図2 論文表題の主題分析

Fig. 2 The Subject Analysis of the Theses.

の意味)〈接続演算〉 (2.3b)

これらのメタ表現の意味および未定義の構造は、システムの対象とする分野に応じて定められる。

3. 論文表題を用いた主題分析法⁴⁾

2節で述べたモデルに基づき、電子通信学会誌発表論文を対象として主題分析を行なう場合の、各表現レベルの構造、処理系での処理手順について述べる。

3.1 各表現レベルの構造

各表現レベルの構造を定め、意味を明らかにする。

〈キーワード〉 ::= 句構造言語 | 正規表現 | ...

〈AND〉 ::= = | および | ...

〈接続子〉 ::= = | による | のための | ...

接続子としては、助詞だけでなく助詞+助動詞のものも考慮に入れ、表1, 2†を参考にして定めた。また、接続子の機能は、つぎのように仮定した。

- の ; 最も使用頻度が多い。主題とその属性
- または、考察域と主題とを結合する。
- による ; 手段、作用などの限定事項と中心事項

表1 接続語句の出現頻度

Table 1 The Occurrence of the Connector.

接続語句	頻度	接続語句	頻度
の	56.9%	を含む(含んだ)	1.1%
による	8.5	利用する(した)	1.1 *
と	7.7	のための	0.8
〜について	7.4	としての	0.8
における	3.5	を有する	0.8 *
に関する	3.2	に対する	0.5
を用いた	2.9	を伴う(った)	0.5 *
を持つ	2.1	での	0.5 *
および	1.3	のある	0.4 *

* 出現頻度が低く同じ機能を持つ出現頻度の高い接続語句に変換される接続語句。

表2 接続語句の連鎖形態

Table 2 The Patterns of the Connector Sequence in the Title.

連鎖形態	頻度	連鎖形態	頻度
〜の〜	40.0%	〜の〜による〜	1.6%
—	10.9	〜に関する〜	1.6
〜による〜の〜	7.3	〜を含む〜の〜	1.6
〜による〜	4.2	〜を有する〜の〜	1.6
〜の〜に関する〜	3.1	〜を持つ〜の〜	1.0
〜を用いた〜の〜	2.6	〜を持つ〜	1.0
〜の〜の〜	2.6	〜における〜	1.0
〜における〜の〜	2.6	〜のための〜	1.0
〜を用いた〜	2.1	その他	12.6
〜の〜の〜の〜	1.6		

† 電子通信学会誌 238件についての統計。

を結合する。

における; 考察域の指定を行なう。“の”のかわりに用いられることが多い。

に関する; 考察対象を表わす。後に続く語に、常用語がくる場合が多い。

を用いた; 手段、材料などの限定事項を表わす。

を持つ; 機能、性質などの限定事項を表わす。

のための; 目的事項を表わす。

〈常用語〉 ::= 考察 | 応用 | 提案 | ...

論文の意味的内容には影響を与えないが、キーワードに役割を付加する際の手がかりとなる。

〈空語〉 ::= ある | 簡単な | いくつかの | ...

論文の内容に全く影響を与えないような一般的な形容詞のことを指し、意味的处理においては無視する。

〈役割〉 ::= M | P | O | D | C

M ; 論文の中心事項。

P ; 中心事項の属性、特性、説明事項。

O ; 中心事項の目的事項。

D ; 中心事項を限定する属性事項。

C ; 中心事項の考察域に関する事項。

〈キーワードの意味〉 ::= (k_{i1}, k_{i2}, ..., k_{in})

キーワードの意味は、各キーワード間の意味的な遠近関係を保存するようなn次元意味空間上の座標値である。

〈接続演算〉 ::= ω₁ | ω₂ | ...

キーワードの意味や句の意味を結合して新しい句の意味を作り出す線形演算である。

3.2 主題分析アルゴリズム

論文表題を利用した主題分析アルゴリズムについて述べる。表層言語代数系に基づいて、論文表題を構文解析し、図3に示す導出木を抽出する。この導出木の各節は、〔〈接続子〉, 〈役割〉〕, または、〔〈キーワード〉, 〈役割〉〕と記され、前者は非終端語、後者は終端語である。根は論文表題を表わし、Sと記す。終端語を含む部分導出木は、〈句の意味〉に対応している。

図4, 5に、論文表題を、意味空間のメタ構造に変

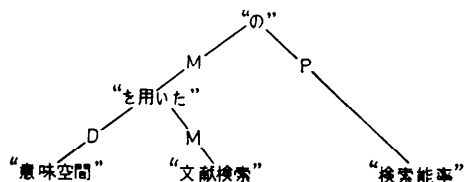


図3 論文表題の構文解析

Fig. 3 The Syntactic Analysis of the Theses.

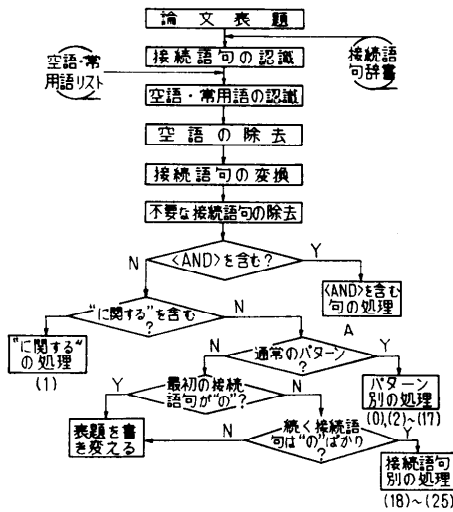


図 4 表題の主題分析手順

Fig. 4 The Flow of the Subject Analysis.

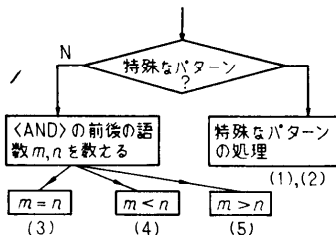


図 5 <AND> を含む句の処理

Fig. 5 The Management of the Phrase Containing <AND>.

換する規則を適用するための前処理の手順を示す。接続語句の変換では、表 1 の *印の低出現語句を同じ機能を持つ高出現語句に変換する。

- を用いた。 → を有する → を持つ。
- を伴う → を含む。 → のある → を持つ。
- の → における。

AND 機能を持つ接続語句を含む場合は、図 5 の手順で AND を含まない句に分解した後図 4 の処理を行なう。付録 1 に示した変換規則からも明らかのようにこの手順は再帰的な手続きを含んでおり、図 4 の④の地点を入口として何回も繰り返される。

3.3 論文表題の意味づけ

論文表題の意味づけは、意味レベルにおいて行なわれる。論文表題の意味は、句の意味、キーワードの意味、接続演算から構成される。したがって、論文表題

の意味づけは、これらのメタ表現の意味づけにより行なわれる。すなわち、図 2 の ϕ , ψ を明らかにすることが必要である。

3.3.1 <キーワードの意味>の意味づけ

キーワードの意味は、従来、樹枝状のリストを用いて表現されることが多かった。しかし、このような表現方法は、キーワード間の意味的な関係を十分に表現しているとは言い難い。そこで筆者らは、意味空間なる概念を提案し、その上でキーワードの意味を表現した⁵⁾。その手順を図 6 に示す。キーワードの意味的な遠近関係を接続行列を用いて求め、それらと n 次元超球面上に配置したキーワード間の距離との大小関係の単調性を保持するように類似度解析法を用いてキーワードを配置し、その球面上の位置をキーワードの意味と考える。この処理は深層言語代数系に基づいて行なわれる。

3.3.2 <演算係数>の意味づけ

<演算係数>の意味づけは、深層レベルにおける内、外算法と、意味レベルにおけるそれらとの対照表を作ることを意味する。すなわち、導出木の分解、合成などの演算に対応して情報空間内で行なわれる演算を定式化した形で抽出する必要がある。本稿では、林氏の数量化理論第 IV 類⁶⁾を用いて演算を抽出する。この方法は、 m 個の個体または属性が与えられた時、それらに相互の類似性を表わす何らかの量 e_{ij} の行列が与えられていると仮定し、この行列をもとにして、 m 個の個体または属性に 1 次元的数値を与えるものである。この方法を採用した理由を以下に示す。

- (1) 意味空間は、線形演算を保存しているので合成概念と被合成概念との関係を 1 次元尺度上で述べることができれば十分である。
- (2) 意味空間から、統計的に演算を抽出する方法にくらべ少ない被験者でよい。
- (3) 一人の被験者に与えられるアンケートの量が少ない。

林氏の方法を用いるには、語句の間の類似度が必要であるが、それはアンケートにより求める。2 つの語句、それらを接続子で結んだ句、および、“と”で結ばれた句の 4 つを与え、それら間の意味的な遠近関係を求めた。このようにして求めた類似度を用い、数量化理論を適用して 1 次元尺度上に割り当てられた数値より、接続子の前後の語 (A, B なるベクトルで表わされているとする。) にかかる係数を A_0, B_0 。また“の”の場合を、 A_0, B_0 とした時、演算を表わす指

標は次のようになる†.

$$\omega = A/B \text{ if } A/B \leq 1, \\ = A/B \text{ if } A/B > 1.$$

ここで, $A = A_0/2A_d$, $B = B_0/2B_d$.

3.3.3 <論文表題>の意味づけ

論文表題の意味づけは、図2の D_1 , D_2 から D_3 を求める操作を意味する。すなわち、論文表題を図5, 6に示した手順で処理し、付録に示した書換規則を適用して、<キーワードの意味>を被演算項、<接続演算>を二項演算子とする逆ポーランド記法で表わす。つぎに、これらの各項に値を代入することにより、論文表題が意味空間上に表現される。論文表題が<AND>を含む場合は、それぞれの句の意味空間上の領域の和集合で表わされる。

以上の操作により、論文表題の主題分析が行なわれる。この主題分析法を用いた検索システムとその実験例をつぎに示す。

4. 検索実験

4.1 実験の概要

論文表題を利用した主題分析法を用いた検索実験の概要を述べる。ここで想定している検索システムは、粗い検索を行なった後の出力文献に対して、より詳しい検索を行なうものである。したがって、対象文献の表わす概念の範囲も限られており、その数も数千個以内と仮定している。この制限により、語句の多義性を減少している。図6, 7に実験のフローチャートを示す。語空間、空語、常用語のリスト、書換規則の作成には、電子通信学会誌分冊Cの238件の論文を用いた。論文空間には、オートマトン、パターン認識、多値論理の分野より選んだ28件の論文を取めた。キーワードとしては、論文の表題およびあらましから80語を選んだ。

○ 語空間の構成 キーワード・文献の出現頻度行列を用いて、キーワード間の類似度を求め、類似度解析法を用いて語空間を構成する。この際、制限条件として、単に語の間の距離が類似度と単調な関係を有

† A, B に接続子を施して得られる語を C で表わし、各ベクトルの終点を a, b, c とすると、

$$A_0 : B_0 = \vec{ac} : \vec{cb},$$

$$C = \frac{B_0 A + A_0 B}{A_0 + B_0}$$

となる。接続子が“と”の場合には、 A_d, B_d を A_0, B_0 の代わりに用いる。これらの係数を演算指標として用いることも可能であるが、ここでは、 A, B の概念領域の広さを考慮するという意味で、 A_0, B_0 を A_d, B_d で補正した係数を演算指標として用いている。

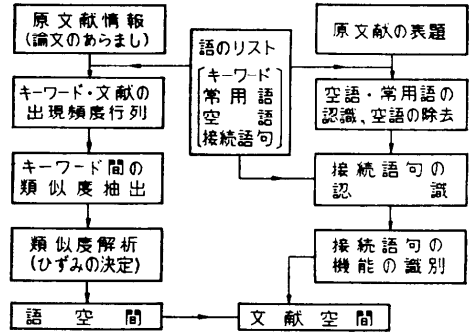


図6 実験フロー (ファイリング)

Fig. 6 The Flow of I.R. Experiment. (Filing)

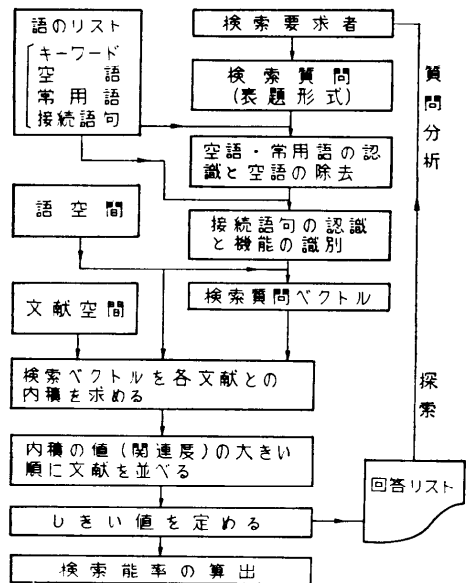


図7 実験フロー (検索)

Fig. 7 The Flow of I.R. Experiment. (Query Analysis & Searching)

しているというだけでなく、擬似的な比例関係を満たすことをつけ加えて空間を構成する。

○ 文献の抄録 文献の表題を接続語句と接続されている語句に分け、空語、常用語などの処理を行ない、各語句に役割を付加する。AND 機能を持つ語があれば、いくつかの句に分けそれぞれが一つの文献を表わすと考えて処理する。つぎに、各語句に与えられた役割とそれらを接続する語句とから演算を定め、各語句を表わすベクトルを合成して文献を表わすベクトルを作成する。

○ 質問分析 文献の抄録と同様に、質問概

念を表わすベクトルを作成する。

○ 探索 論文空間を質問ベクトルに垂直な平面で切り出し、そこに含まれる文献を抽出する。この平面の位置を変えることにより、呼出率を調節できる。

4.2 実験結果

図6, 7の処理手順にしたがい、阪大大型計算機センター TSS システム (NEAC 2200-500) の端末 (50ボー) を用いて行なった検索実験の結果を示す。

図8に、検索ファイルの作成時および探索時間と検索対象文献数との関係を示す。これらの関係を1次式で近似すると次のようになる。

検索時間

$$t_r = 0.24n + 0.26 \text{ (秒)}$$

ファイル作成時間

$$t_c = 0.54n + 0.14 \text{ (秒)}$$

ここで、 n は文献数、ファイル作成に用いたキーワ

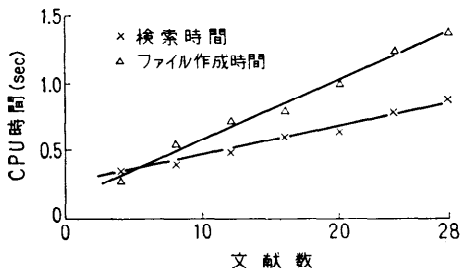


図8 検索時間

Fig. 8 The Calculation Time for the Theses Retrieval.

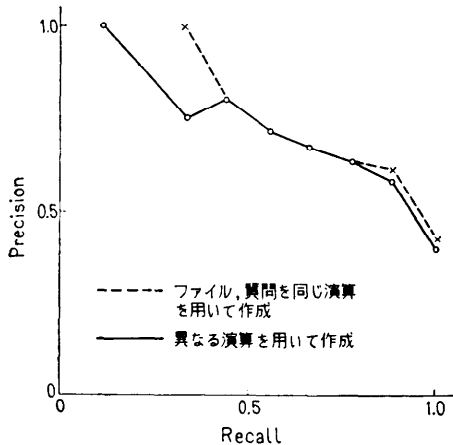


図9 検索能率

Fig. 9 Precision and Recall Factor.

ード数は80である。

図9に適合率と呼出率の関係を示す。破線は、ファイル作成と質問分析に同じ接続演算を用いた場合、実線は、異なる接続演算を用いた場合の能率を表わす。適合率と呼出率が等しくなる値は、両方法とも67%、適合率、呼出率が100%となる所では、同じ演算を用いる方法が、33%、42%と異なる演算を用いる方法にくらべ高い能率を示している。呼出率が40%以上の範囲では、両者ともほぼ同じ能率を示している。

この実験より、つぎのことが結論される。

○ 検索時間が短いので、会話型オンライン・システムで用いるのに都合がよい。

○ ファイル作成時間は、図8に示すように、文献数に比例して増加する。しかし、対象文献数が千個程度なら、10分以内でファイルの作成が可能であるから、検索要求者が、各自に適したファイルを作ることができる。これにより適合率の向上が期待できる。

○ 語空間を構成する際の類似度を、共出現頻度行列を用いて定めた。このため、多義性を有する語句を含む論文または検索質問に対して適合率が悪くなる傾向がある。

○ いくつかの関連の少ない語を用いて検索を行なうと、全く関係のない論文が検索される。したがってAND機能を持つ接続語を含む質問を行なう場合は、いくつかの質問に分けて行なうのが能率的である。

○ 接続演算のかわりに重みづけを用いても、接続行列を用いる方法より高い能率が得られた。したがって、語レベルで意味空間を構成することにより、適合率が良くなることが明らかになった。しかし、論文レベルで意味空間を構成する方法にくらべ、語から句を合成する過程で意味的なひずみ入るため、適合率はいく分低くなる。

5. むすび

語レベルの意味空間を用い、句構造を持つ論文表題を自動的に主題分析する方法を提案した。論文レベルの意味空間を用いる方法にくらべ、検索能率の点で劣るが、ファイルの作成が容易、検索質問が自然語で行なえ、質問者にシステムについての余計な知識を要求しないなどの利点を持つ。また、接続行列を用いる方法にくらべ、検索能率、検索時間などの点で秀れている。

今後の問題としては、語から句を合成する際の意味の損失を少なくすることがある。そのためには、書換

規則を充実させる必要がある。すなわち、形容詞の処理、連語の分離と処理、接続語句の機能の詳しい研究が必要である。しかし、このシステムは、オープンなシステムだから、書換規則の変更や付加はシステムの他の要素に影響を与えない。したがって、上記の研究の結果をシステムに組み込むことは容易である。

これらの問題が解決されれば、このシステムは、ファイルの作成、変更が容易、自然語で検索質問が可能、検索時間が短いなどの利点があるので、今後の T. S. S. 用の検索システムとして有用である。

付録 書換規則

1. AND を含まない場合の処理

- (0) $\langle N \rangle$, n
- (1) $\langle NP_1$ に関する $NP_2 \rangle$, $[NP_1][NP_2]f_{01}$
- (2) $\langle N_1$ の $N_2 \rangle$
- i) $S(N_2) \in U \vee Fr(N_1) \leq Fr(N_2)$, $n_1n_2f_{13}$
- ii) else , $n_1n_2f_{11}$
- (3) $\langle N_1$ における $NP_2 \rangle$, $n_1[NP_2]f_{31}$
- (4) $\langle N_1$ による $NP_2 \rangle$, $n_1[NP_2]f_{22}$
- (5) $\langle N_1$ を用いた $NP_2 \rangle$, $n_1[NP_2]f_{42}$
- (6) $\langle N_1$ を持つ $NP_2 \rangle$, $n_1[NP_2]f_{52}$
- (7) $\langle N_1$ を含む $NP_2 \rangle$, $n_1[NP_2]f_{62}$
- (8) $\langle N_1$ のための $NP_2 \rangle$, $n_1[NP_2]f_{72}$
- (9) $\langle N_1$ としての $NP_2 \rangle$, $n_1[NP_2]f_{84}$
- (10) $\langle N_1$ に対する $NP_2 \rangle$, $n_2[NP_2]f_{94}$
- (11) $\langle N_1$ を用いた N_2 の $N_3 \rangle$
- i) $S(N_3) \in U$, $n_1n_2f_{42n_3f_{13}}$
- ii) else , $n_1n_2n_3f_{11f_{42}}$
- (12) $\langle N_1$ の N_2 の $N_3 \rangle$, $[N_1N_2]n_3f_{13}$
- (13) $\langle N_1$ の N_2 による $N_3 \rangle$
- i) $N_3 \in U$, $n_1n_2f_{13}$
- ii) $S(N_1) \in U$, $n_1n_2f_{13n_3f_{22}}$
- iii) else , $n_1n_2n_3f_{22f_{11}}$
- (14) $\langle N_1$ による N_2 の $N_3 \rangle$
- i) $N_3 \in U$, $n_1n_2f_{22}$
- ii) $S(N_1) \in U$, $n_1n_2f_{22n_3f_{13}}$
- iii) else , $n_1[N_2N_3]f_{22}$
- (15) $\langle N_1$ の N_2 の N_3 の $N_4 \rangle$, $[N_1N_2]n_3f_{13n_4f_{13}}$
- (16) $\langle N_1$ N_2 $N_3 \rangle$, $n_1[N_2N_3]f_{31}$
- (17) $\langle N_1$ における N_2 の $NP_3 \rangle$
- i) $NP_3 = N_3, S(NP_3) \in U$
 $\vee Fr(N_2) \leq Fr(N_3)$, $n_1n_2f_{52n_3f_{13}}$
- ii) else , $n_1n_2f_{52}[NP_3]f_{11}$

- (18) $\langle N_1$ を含む N_2 の $NP_3 \rangle$
- i) $NP_3 = N_3, S(NP_3) \in U$
 $\vee Fr(N_2) \leq Fr(N_3)$, $n_1n_2f_{52n_3f_{13}}$
- ii) else , $n_1n_2f_{52}[NP_3]f_{11}$

2. AND の処理

- (1) $\langle NP_i$ AND (その)
 NP_i への応用 \rangle , $[NP_i]; [NP_i][NP_i]f_{42}$
- (2) $\langle NP_i$ AND その $NP_i \rangle$
- i) $NP_i \in U$, $[NP_i]$
- ii) else , $[NP_i]; [NP_i][NP_i]f_2$
- (3) $\langle NP_i$ AND $NP_i \rangle$
- i) $m = n \uparrow$, $[NP_i]; [NP_i][NP_i]f_2$
- ii) $m < n$, $[NP_i \cdot C_{1m} \cdot N_{1(m+1)} \dots N_{1n}]; [NP_i]$
- iii) $m > n$, $[NP_i]; [N_{f_{1,1}} \cdot C_{f_{1,1}} \dots C_{f_{(m-n)}} NP_i] \uparrow \uparrow$

注 $\uparrow m, n$ は AND の前後の句を構成する語数。

$\uparrow \uparrow NP_i, NP_i$ は AND の前後の句を表わす。 N_{fi}, C_{fi} は、 i 番目の句および接続子を示す。

参考文献

- 1) G. Salton: Associative document retrieval techniques using bibliographic information, J. of ACM (1963).
- 2) 鶴身征雄, 打浪清一, 手塚慶一: 意味空間を用いた情報検索, 信学会論文誌 D, 昭 47-5.
- 3) S. Uchinami, Y. Tezuka and K. Kasahara: A model of linguistic information algebra, p. 711, Technology Report of the Osaka University, Osaka University (1970).
- 4) 鶴身, 打浪, 手塚: 意味空間を用いた主題分析法, 信学会 A 研資 (1972-4).
- 5) 鶴身, 打浪, 手塚, 笠原: 多次元配列法による情報空間を媒介にした情報検索に関する一考察 信学全大, 昭 43.
- 6) 安田三郎: 社会統計学, 丸善, pp. 211~215, (1969).
- 7) 西村勉彦, 岩坪秀一: 計算意味論の実験, 情報処理, (1970-3).
- 8) 安部憲弘, 豊田順一, 田中幸吉: 文献表題を利用した索引作成の自動化に関する考察, 情報処理, (1970-12).
- 9) R. N. Shepard: The analysis of proximities; multidimensional scaling with an unknown distance function I, Psychometrika, 27 (June 1962).

(昭和 47 年 8 月 19 日 受付)

(昭和 47 年 11 月 13 日 再受付)