

シナリオを用いる論文抄録理解・作成援助システム<sup>†</sup>猪 瀬 博<sup>††</sup> 齊 藤 忠 夫<sup>†††</sup> 堀 浩 一<sup>††</sup>

本論文は、論文抄録の文章を理解して、抄録からキーワードを自動的に抽出するための、シナリオにもとづく手法を提案している。この手法は抄録の文章の意味的深層構造にいくつかの定型があることを利用し、これをシナリオとして、文章を理解するものである。キーワードは構造化された情報として取り出される。本論文はこの方法をインプリメントした結果と、その評価について述べている。シナリオは抄録の作成援助システムとしても利用でき、この原理にもとづく抄録作成援助システムについても述べられている。

## 1. ま え が き

文献検索システムは、主として学術情報や特許情報を対象として、広く普及しつつある。このようなシステムでは検索の手がかりは単純なキーワードによっている。しかし、単純なキーワードの検索では、ユーザが意図した検索条件の設定を十分に行うことが困難であると認識されており、より高度な検索法が望まれている。

従来のキーワードの抽出法は人手によるのが普通である。機械による場合には、単純な統計的手法によっている<sup>1)</sup>。人手による場合には作業品質のばらつきが問題になり、機械による場合には意味理解が行われないうために、キーワード自体の適格性が問題となる。

この問題を解決するために本論文では、論文抄録を自然言語として理解し、そのなかから、構造化キーワードとよぶ、意味を含んだキーワードを自動的に抽出する方法を提案している。

本論文で述べる抄録処理の方法は、従来広く研究されている自然言語処理の研究成果を利用しているものであるが、処理の対象とする文章を科学技術文献の抄録に限定し、抄録の文章のもつ特殊な構造を利用している。科学技術文献の抄録の目的はその論文の概要を読者に手短かに伝えることであり、文献の性質によっていくつかの定型に分類される意味的深層構造をもつ。この構造の定型をフレームによって表現し、フレームに基づいて抄録の文章を理解する。その手続きの記述をシナリオと呼ぶ。これにより、従来の、機械

翻訳などを目標とする汎用の自然言語理解の手法に比べ、単純な手法とすることができる。

本論文では、このような抄録理解の手法を詳述し、その有効性を実証するための実験結果を述べる。実験の結果、抄録に何の前処理や制限を加えなくとも、多くの抄録についてこの手法は有効であることがわかった。しかしながら適切な理解が得られない抄録もあった。

このシステムで理解しやすい抄録は、人間にとっても理解しやすい抄録と考えられ、抄録作成のガイドシステムを作成することが、システムにとっても、人間にとっても理解しやすい抄録を作るために有効であると考えられる。本論文ではこの目的のための抄録作成ガイドシステムとそれによる抄録作成の実験結果についても述べている。

## 2. 構造化キーワード

## 2.1 構造化キーワードの考え

従来から、民話や童話等には話の筋のいくつかの定型が存在することが談話理解の研究で指摘されている<sup>5)</sup>。

論文や事務文書のような、計算機による処理の実用的価値の大きい文章においても、その深層構造や表層構造にいくつかの定型が存在すると考えられる。時不変な構造を見だし、あるいは人為的に規定して利用できれば自然言語の処理は容易になる。

何らかの情報を自然言語の文章として表現するまでの過程とその逆過程である理解の過程には、構文規則をはじめとするいくつかの規則と、構文解析木のような構造をなす情報が存在する。シンタクスとセマンティクスの取扱いについては種々議論、提案がなされているが、ここでは科学技術文献の論文抄録を対象として、自然言語の生成過程を図1のように心理学的に

<sup>†</sup> A Paper Abstract Understanding and Generation Assistance System by Means of Scenario by HIROSHI INOSE, TADAO SAITO and KOICHI HORI (Faculty of Engineering, University of Tokyo).

<sup>††</sup> 東京大学工学部電子工学科

<sup>†††</sup> 東京大学工学部電気工学科

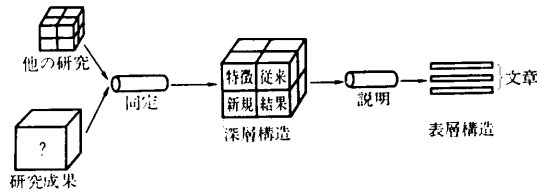


図1 論文抄録の文章生成の過程

Fig.1 Sentence generation process for an abstract of a paper.

モデル化して考える。

論文抄録は、自分の行った研究成果等を人に知らせるための文章である。研究成果は何らかの構造をなした情報として認識できる。最初その構造化された情報は漠然として形のはっきりしないものであるが、いくつかの認識過程を経て、明瞭になり、最後に自然言語の表層構造として表現される。認識のための規則は、自然言語の表層構造に反映される。論文抄録の文章生成の場合、認識の第1段階は、自分の研究成果の他の研究からの区別ないし同定の段階である。第2段階は、研究成果を詳細に説明する段階である。

第1段階は文献抄録の意味的深層構造に対応し、第2段階はそれに基づく表層構造に対応する。ここで、意味的深層構造とは、その抄録によって著者が表現しようとする内容であり、表層構造とは、それを表現した文章である。

文献検索システムにおいては、同定の段階の結果として得られる構造をなした情報を2次情報として蓄えればよいことになる。その実現のための条件は、同定の段階の認識過程が普遍的で、かつ、それが自然言語の表層構造に反映されていることである。論文抄録のような、目的のはっきりした文章においては、この条件が満たされることが十分期待できる。

そこで、意味的深層構造を表した2次情報を構造化キーワードとよぶことにする。構造化キーワードは、図1の同定の結果に対応し、論文の中心的主題が何であり、それが他の論点とどのような関連を有しているかなどのキーワードへの意味づけとそれらのはず構造を表現する。構造化キーワードは、表層構造としての文章から、単語の意味を手がかりに文章作成の逆過程をたどることにより抽出し構成される。

従来の単純なキーワードの羅列は構造化キーワードの、構造の最も単純なものである。また、句構造等に基づくロール付キーワード等<sup>7)</sup>も一種の構造化キーワードと考えることもできる。

また、何をもちいて同定の規則とするか、そしてどの

ように構造化された情報を構造化キーワードとすればよいかは、システム利用の目的と試行錯誤によって定めることができる。その詳細については、次節以降で述べる。

## 2.2 構造化キーワードの表現法

構造化キーワードは一種の知識の表現であるから、フレーム、セマンティックネットワーク、述語論理等の知識の表現法を用いて表現することができる。

構造化キーワードの表現方法に要求される条件は、1) 一般ユーザが文献検索に使用することを想定しているので、表現が計算機の素人にもわかりやすく単純明解であること、2) 文献に述べられた概念を表現するために、概念の階層構造は必要であるが、人工知能におけるような推論、演繹等の高度の知識操作は要求されないこと、などである。

上記の条件から構造化キーワードの表現方法としてフレームを用いることにする。

本論文では概念の枠組を階層的に表現する表現法として、フレームという用語を用いる。高度の知識操作を目的とした Minsky らのフレーム<sup>9)</sup>よりは簡略化された形になっている。フレームはフレーム名とスロット群から成る。フレーム名はそのフレームの表現せんとする概念の名前である。スロット群はその概念を構成する構成要素である。スロットは再帰的にフレームであるか、またはキーワードが代入される。キーワードの代入される前のフレームは一つの概念の内包を与える。すべてのスロットにキーワードを代入することにより一つの概念の外延を得る。

本論文における自然言語の理解は、一つのフレームで表現される概念を話題とする自然言語の文章を読み、すべてのスロットにキーワードを代入して外延を一つ得る過程であると定義する。

たとえば、クーデタを話題とする新聞記事の文章を

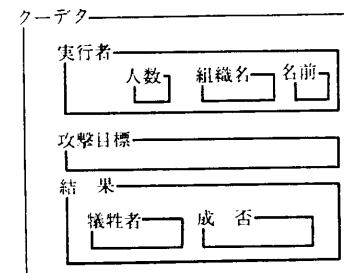


図2 フレームの例：クーデタの概念のフレーム表現  
Fig.2 An example of a frame expression of the concept of coup d'état.

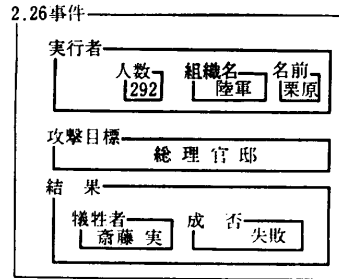


図3 キーワードの代入されたフレームの例  
Fig.3 An example of the frame filled with key words.

理解するとは、図2に示すクータのフレームのスロット群にキーワードを代入してたとえば、図3に示すようなフレーム(2.26事件)を出力することを意味する。

### 2.3 構造化キーワードの構造の決定

構造化キーワードの構造、すなわちフレームの構成は、次の3段階からなる手続きを経て定めることができる。

1° 分析の段階: 取り扱う文章のテーマとする概念のフレーム表現を定める第1段階は、分析の段階である。分析の対象はシステム使用の目的と取り扱う文章の内容である。システム使用の目的から文章を理解して出力すべき情報の構造がトップダウンに定まる。また、実際に取り扱う文章が表現している内容から構造がボトムアップに定まる。

2° 試行錯誤の段階: 第1段階の分析に基づき定めたフレーム表現を用いて、システムを試験運用する。使用経験に基づきフレームを修正する。

3° 規格化の段階: システムの有用性が確認できた後はシステムの理解しやすい文章の書き方の規格を定め、それにそって文章を書いてもらうようにする。

構造化キーワードの手法は事務文章や特許情報の管理にも適用できると考えるが、ここでは科学技術文献の論文抄録を対象にフレームの構成決定の例を以下に示す。

論文抄録のフレームを決定することは、2.1節で述べた論文同定の認識過程の定型を探ることにはほかならない。フレーム決定の第1段階で詳細なフレームを定めるほど、適用できる文章の範囲は狭くなり、また第3段階における規格を厳しくしなければならなくなる。

ドキュメンテーションの専門家の推奨する論文の構成法等も参考に図4に示すような一般的なフレームを定めることも、ある特定の研究分野専用の、たとえば図

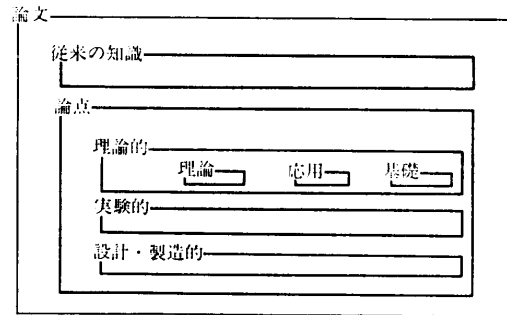


図4 論文のフレーム(1)  
Fig.4 Frame of paper(1).

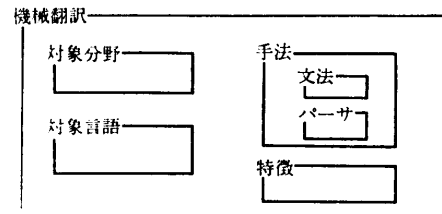


図5 機械翻訳の研究のフレーム  
Fig.5 The frame for study of machine translation.

5に示すような機械翻訳の研究のフレームを定めることも可能である。実際にはシステム使用の形態(文献検索システムにおいては問合せ文の定型)と取り扱う文章の定型を分析して定めなければならない。その例については5章の実験で示す。

## 3. 構造化キーワードの抽出

### 3.1 シナリオを用いるトップダウン抽出

自然言語の文章を理解して構造化キーワードを抽出する手法を考える。

従来の、機械翻訳等を目的とした自然言語理解の研究においては、理解を、形態素分析、構文解析、意味解釈、モデルとの対応等のいくつかの段階に分けてボトムアップに、逐次処理してゆく手法が一般的であった。この手法には、多くの複雑な処理を要する。

これに対し、本論文の目的とする構造化キーワードの抽出においては、理解の目標となるフレームははじめから定まっている。したがって、フレームにそって、トップダウンに直接、単語の意味や局所的な構文を手がかりに文章を理解してゆくことが可能である<sup>3)</sup>。トップダウンに直接処理する方式においては、フレームにあてはまらぬ結果を切り捨てることにより、処理をボトムアップ逐次処理方式に比べて単純化することができる。このような理由から本論文では、後者のトッ

ブダウン直接処理の方式を用いることにする。

トップダウン直接処理方式においては、上述のように、単語の意味から直接フレームを形成するが、その規則をシナリオとよぶことにする。シナリオを実行する機構を、シナリオインタプリタとよぶ。

シナリオは、プログラムで直接実現することも可能ではあるが、記述形式に一般性と汎用性をもたせる目的で、従来から構文解析に広く用いられている ATN (Augmented Transition Network)<sup>4)</sup> を用い記述することにする。ATN は再帰呼び出し可能な有限状態機械の集合であるが、ある状態から別の状態へ遷移の際にレジスタ転送などの任意のアクションを行えるように記述できる点に特徴がある。普通、構文解析用に用いられる場合は ATN は単語を一つずつ読み込んでその品詞により遷移を行いながら解析を行うが、ここでは、任意の条件記述と、アクション記述の機能を利用して、単語の意味や解析の途中結果なども判定して遷移を行いながら、解析結果の構造化された情報を出力するように記述することによりシナリオを実現する。ATN によりシナリオを記述することにした理由は、1) ATN もシナリオも、本質的に動作がトップダウンで、記述に都合がよいこと、2) ATN を用いることにより、フレームにあてはまらない未知の世界の部分については構文解析のみ行うようにすることができること、などである。

### 3.2 シナリオの構成

シナリオは、単語の意味や文の構文を手がかりに、あらかじめ用意したフレームにあてはめて文章を理解するための規則の記述である。シナリオは、あらかじめ用意したフレームから、文章に表される記述内容を次々に期待して処理してゆく動作と、単語の意味や構文から、それに見合ったフレームを呼び出す動作の両方を記述できねばならない。

意味解釈、構文解析、フレームとの対応の三つの機能を同時に結合して実行するために、シナリオを、レキシコン・ドリブン・ルーチン、グラマ・ドリブン・ルーチン、フレーム・ドリブン・ルーチンとよぶ3種類のルーチンに分けて記述することにし、3種類のルーチン間で割込みと情報の受け渡しを行いながら並列実行する機能をシナリオインタプリタの中に作成した。

シナリオインタプリタは基本的には ATN インタプリタであるが、割込みによりボトムアップな制御もできるように拡張した。

辞書としては、単語の品詞分類と単純な意味分類を与えた。

レキシコン・ドリブン・ルーチンは単語の意味と局所的な構文を手がかりに局所的な意味処理を行うルーチン、グラマ・ドリブン・ルーチンは大まかな構文規則に従って文を流し読みしていくためのルーチン、フレーム・ドリブン・ルーチンは、フレームにあてはめて意味の記述を整理するルーチンである。これらの3種類のルーチンの並列実行のための結合の方法と、それぞれの詳細な動作は、以下のとおりである。

① 結合方法：全ルーチンが共有するフラグとレジスタが存在する。これらを用いてルーチン間の同期をとり、情報のやりとりを行う。ただし、レキシコン・ドリブン・ルーチンは、グラマ・ドリブン・ルーチンの実行中に、あらかじめ、シナリオインタプリタに登録しておいた意味の単語に出会ったときに、割込みがわかり、単語の意味に対応したルーチンが起動される。

② レキシコン・ドリブン・ルーチン：フレームを構成する手がかりとなる単語ごとにレキシコン・ドリブン・ルーチンを用意する。たとえば、'応用' という意味の単語に対するレキシコン・ドリブン・ルーチンは、何の何への応用かを探して、処理結果をレジスタに入れ、フラグをたてフレーム・ドリブン・ルーチンに通知する。レキシコン・ドリブン・ルーチンの ATN における状態遷移は単語の品詞分類と意味分類によって定める。したがってレキシコン・ドリブン・ルーチンの機能は局所的な構文解析にすぎない場合もあるが、その機能をグラマ・ドリブン・ルーチンから分離したことによって、グラマ・ドリブン・ルーチンの負担を軽くすることができる。すなわち、注目すべき単語だけを拾って、その周辺を細かく調べるのが、レキシコン・ドリブン・ルーチンの役割である。レキシコン・ドリブン・ルーチンは、文を越えた処理は行わない。

③ グラマ・ドリブン・ルーチン：グラマ・ドリブン・ルーチンは、文中の、句、節の分離と態の判定程度の大まかな構文解析を行うルーチンである。グラマ・ドリブン・ルーチンは、構文的役割によって意味が異なる単語に対処するため、レキシコン・ドリブン・ルーチンの割込みのマスクを制御する機能も果たす。

④ フレーム・ドリブン・ルーチン：フレーム・ドリブン・ルーチンは、レキシコン・ドリブン・ルーチンとグラマ・ドリブン・ルーチンの出力を、フレーム

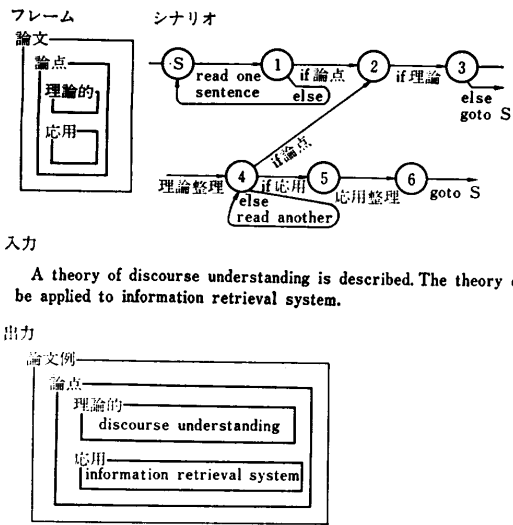


図 6 フレームとシナリオと入出力例

Fig. 6 An example of frame, scenario, input and output.

にそってトップダウンに処理し、構造化キーワードを出力する。文と文のつながりについて、指示代名詞等の複雑な処理は行わず、フレームにあてはまる記述を次々に結びつけていくという方法をとっている。論文抄録等の文章からの構造化キーワードの抽出では、この方法で十分である。

簡単なフレームとシナリオとその入出力の例を図 6 に示す。図 6 の文章が入力されると、まずフレーム・ドリブン・ルーチンからの指示によりグラマ・ドリブン・ルーチンが、最初の文を読み構文解析を行ってゆく。主節の解析を行っている際、theory という単語に出会ったときに、レキシコン・ドリブン・ルーチンがグラマ・ドリブン・ルーチンに割込みをかけ、何の理論かを調べ、(!theory! discourse understanding)\* を主節のレジスタにプッシュし、さらに理論というフラグをたてる。動詞句の解析を行っているときに describe に出会うと、再びレキシコン・ドリブン・ルーチンが割込みをかけ、論点というフラグをたてる。ただし、複文の従属節を解析している場合は describe に出会っても、グラマ・ドリブン・ルーチンからのマスクにより、レキシコン・ドリブン・ルーチンは割込みをかけない。この時点でフレーム・ドリブン・ルーチンは、この文が論点を示した文であること、受動態の文であること、主節に理論的な論点がありそうなことなどを、レキシコン・ドリブン・ルーチンとグラマ・

ドリブン・ルーチンからのフラグで知る。フレーム・ドリブン・ルーチンはフラグとレジスタの中身を調べ、キーワードの discourse understanding を抽出する。さらに応用についての記述がないかを調べるが、この文ではそれを示すフラグはたっていないので、応用の記述を期待するところで処理を中断しリセットすべきフラグとレジスタをリセットして次の文を読むことをグラマ・ドリブン・ルーチンに指示する。第 2 文も同様に処理されるが、第 2 文の先頭にある theory については、やはりレキシコン・ドリブン・ルーチンで処理され (!theory! nil) がレジスタに入れられるが、フレーム・ドリブン・ルーチンは、応用の記述もしくは新たな論点しか期待していないので、これは棄却される。このあと、応用のスロットをうめるキーワードが抽出されて、処理が終了する。

#### 4. 構造化キーワードからの文章の生成

自由に書かれたどのような文章からも構造化キーワードを抽出できることが理想である。しかし文章理解のシナリオが 100% 正しく機能するためには、あらかじめ、取り扱う文章に規格を定めることが望ましい。また、システムと対話しながらキーワードを代入することにより規格に沿った文章を生成する文章生成システムも有用であると考えられる。規格化は計算機ばかりでなく人間が読む時の理解のしやすさにもつながる。

構造化キーワードのフレームが与えられると、文章の規格は次のようにして定まる。

規格 I (文と文のつながりの規定)：フレームを構成するスロットを埋める内容の記述をしなければならない。

規格 II (一つの文の文法)：スロットの埋め方が一意に定まる記述をしなければならない。

これらの規格を満たす文章を自動的に生成する文章生成システムの実現が望まれる。そこで、われわれは実現が比較的容易な、穴うめ式の文章生成シナリオを作成した。

穴うめ式の文章生成シナリオは表層構造のレベルでユーザが穴うめと選択肢の選択を行いながら文章を生成するためのシナリオである。穴うめは、句や節のレベルで行うので、表現の自由度が残される。また、追加の説明を行うための文の挿入も許している。

図 6 に示したフレームにあてはまる文章を生成するための生成シナリオのユーザへの要求の例を図 7 に示す。ユーザと対話をしながら、より自動的にアブスト

\* !でかこむことによってスロット名を表し、そのあとにそのスロットに適合するキーワードを記すことにする。

A theory of (専門概念を代入せよ) is given in this paper.

The theory (選択せよ { can be applied to  
explains  
is based on })

(専門概念を代入せよ). (その他の説明を行え).

図 7 生成シナリオの要求例

Fig. 7 An example of a requirement from the generating scenario.

ラクトを生成するシステムも作成し実験中であるが、これについては続稿としたい。

## 5. 実 験

### 5.1 実験対象

前章までに提案した手法の有効性を実証するための小規模な実験を行った。

論文抄録理解の実験対象として、ACM Transactions on Database Systems の1978年 Vol.3, No.4 から1979年 Vol.4, No.4 までの5冊の全論文31件の抄録を採用した。

本論文の手法は日本語にも適用できると考えるが、今回は、入出力の容易な英語に限って実験を行った。また、実験の対象とする文章は、人間が読んでわかりやすい論旨の明解な文章が望ましい。今回採用した文章は、この条件を比較的よく満たしていると思われる。

また、本論文の手法の有効性を実証するためには、30前後の対象件数は、適切な規模であると考えられる。これより1桁ないし2桁規模を大きくすると、本論文の手法の問題以外の用語の時代的变化などの問題が支配的となると思われるからである。大規模化のためには2章で述べたように、人為的な規格を設けることを期待している。

### 5.2 実験方法

シナリオインタプリタを Pascal を用いて作成した。使用した計算機は DEC LSI 11/23-Unix オペレーティングシステムおよびアイ電子 AICOM C5 である。作成した辞書の語彙は約1,000単語である。

実験の第1段階として、抄録文章の深層構造の定型を探るための分析を前節で述べた実験対象に対して行った。分析の方法としては、グラマ・ドリブン・ルーチンのみからなるシナリオにより、構文解析を行いその結果に対して、人手により述語や主要な名詞に対する意味付け、文と文のつながりに対する意味付けを行い、分類し、構造化キーワードのフレームを定めた。

実験の第2段階として、第1段階の結果に基づく構

造化キーワードを抽出するためのシナリオを作成し、抽出の実験を行った。レキシコン・ドリブン・ルーチンは10種類ほど用意した。

実験の第3段階として、規格に沿った文章を生成するシナリオを作成し、簡単な実験を行った。作成した生成シナリオは4章で述べたように、穴うめをユーザに要求するシナリオである。

### 5.3 実験結果

実験の第1段階から、次のような分析結果を得た。

① 抄録の文章は、従来の結果(専門概念の説明、問題点の把握や提起を含む)を述べる部分と、著者の研究結果を述べる部分からなる。著者の研究の主題は、describe, show, give などの述語を用いて明示される。主題が明示されるより、前の文は従来の結果に関する文、後の文は著者の研究成果を説明する文と考えてよい。

② 研究の主題は、理論的主题、実験的主题、設計製造的主题の3種類に大別できる。ただし、これらが組み合わさった研究主题もある。

③ 研究の中身を説明するための述語は、理論的主题、実験的主题、設計製造的主题のそれぞれに対して1~数種類ずつ定まっている。

以上が分析結果であるが、システム使用者の要望に応じて、手段、道具、性能等、必要なフレームを追加することは、可能である。

図8に示す構造化キーワードのフレームを定めた。また、単語の意味を表1に示すように分類した。

次に、実験の第2段階として、図8のフレームにあてはまる構造化キーワードを抽出した結果について述べる。図9に入力した文章と出力された構造化キーワードの例を示す。

主観的評価が混じることは、禁じえないが、作成したシナリオは31論文中、27(約87%)の論文抄録に対

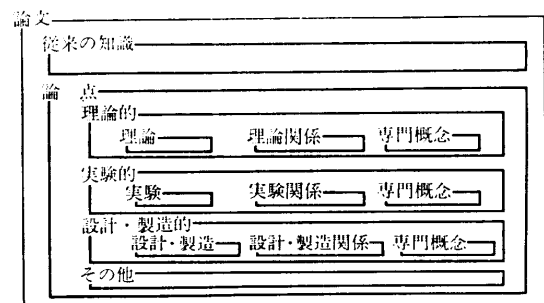


図 8 論文のフレーム(2)

Fig. 8 Frame of paper (2).

入 力

A classified, or clustered file is one where related, or similar records are grouped into classes, or clusters of items in such a way that all items within a cluster are jointly retrievable. Clustered files are easily adapted to broad and narrow search strategies, and simple file updating methods are available. An inexpensive file clustering method applicable to large files is given together with appropriate file search methods. An abstract model is then introduced to predict the retrieval effectiveness of various search methods in a clustered file environment. Experimental evidence is included to test the versatility of the model and to demonstrate the role of various parameters in the cluster search process.

出 力

```
ronten=
  riron
    method-ruj
      FILE CLUSTERING
    applied-to-ruj
      LARGE FILES
  riron
    theory-ruj
    explain-ruj
      RETRIEVAL SEARCH CLUSTERED FILE
  jikken
    jikken
    test
```

図 9 入出力例

Fig. 9 An example of input and output.

しては、人間により抽出した構造化キーワードとほぼ等しい構造化キーワードを抽出することができた。人間は、すべての抄録から図8のフレームにあてはまる構造化キーワードを抽出できた。

最後に実験第3段階の生成シナリオによる文章生成の結果を示す。

図10に出力例を示す。( )はユーザが穴うめした部分, (.....)は選択肢を選んだ部分, (.....)は、ユーザの入力した語句をシナリオが引用した部分である。

穴うめと選択肢と固定部分だけから成るシナリオでも一応満足できる文章を作成できることが示された。

5.4 評 価

① 構造化キーワードの有効性：図9に示した論文と類似の論文を検索するために、東京大学大型計算機

センタ Tool-IR システムを用いて INSPEC-1C データベース (1980 年) で、file, clustering をキーワードとして検索すると、4 文献が出力される。これから、大きなファイルに適用できるファイルクラスタリングの手法と実験について述べた論文に絞りこむことを考える。この操作は従来のキーワードだけではむずかしいが、図9の構造化キーワードを用いると可能となり、1980年に同様の文献はなく、0文献に絞られる。「大きなファイルに適用できる」という条件を除くと、1文献残る。同様に他の30種類の構造化キーワードについても INSPEC-1C (1980年) について検討を行い、ユーザの使い方にもよるが、1/50~1/4に正しく絞ることができることを確認した。この結果から、大きなデータベースで、ある程度検索した後の細かい絞り込みに構造化キーワードは有用であると考えられる。

② シナリオの有効性と限界：実験結果で示したように、31 文献中 27 文献については図8のフレームに

表 1 単語の意味の分類

Table 1 Classification of the meaning of words.

分 類	単 語 の 例	
理 論	theory 類	theory, model, concept, notion
	method 類	method, technique, strategy
	capability 類	capability, characteristic, property
理論関係	applied 類	applied, used, extended
	explain 類	explain, predict, represent
	based on 類	based, extension, use
実 験	experiment	
実 験 関 係	test, explore	
設 計・製 造	design, implementation	
設計・製造関係	execute, improve	
論 点 記 述	describe, give, show, discuss	
専 門	file, storage, query	
そ の 他 有 意	large, small, good	
無 意	a, the, in, on	

(METHOD\_OF) (NATURAL LANGUAGE UNDERSTANDING) IS GIVEN. THIS (METHOD) (IS BASED\_ON) (CONCEPT\_OF) (FRAME). THIS (METHOD) (IS APPLIED\_TO) (METHOD\_OF) (AUTOMATIC INDEXING). (CONCEPT\_OF) (TOP-DOWN UNDERSTANDING) IS GIVEN. THIS (CONCEPT) (OVERCOMES) (PROBLEM\_OF) (AMBIGUITY OF NATURAL LANGUAGE).

(---- は選択肢の選択  
 ——— は自由な入力  
 ..... はシステムによる引用 を示す。)

図 10 生成シナリオ出力例

Fig. 10 An example of the output of the generating scenario.

Record structures are generally efficient, familiar, and easy to use for most current data processing applications. But they are not complete in their ability to represent information, nor are they fully self-describing.

図 11 キーワードが抽出されなかった抄録の例

Fig. 11 An example of abstracts from which keywords could not be extracted.

適合する構造化キーワードが抽出された。失敗した抄録の例(一つの抄録の全文)を図 11 に示す。この中に "Incompleteness of record structures is described." という表現があればシナリオによってキーワードを抽出できる。しかし、図 11 の文章のように、何が中心の論点かを推論しなければならないような間接的な表現になっているとシナリオはうまく機能しない。フレームを複雑な形にしたときもシナリオの手法が有効かどうかは問題になるが、スロットの種類を直接表すような表現で書かれている限りはシナリオは有効である。文章のあちこちの記述から推論してスロットを埋めることはシナリオだけでは困難である。

## 6. むすび

論文抄録の文章の意味を理解し、構造化キーワードを抽出する手法を提案した。また、構造化キーワードに基づく文章の生成にも触れた。本手法は、文章の意味的深層構造に定型が存在することを前提としている。実験により、その前提の妥当性と手法の有効性を示した。

本論文の手法は、個人用あるいは一研究室用の小規模な文献データベースへの適用は比較的容易であると考えられるが、大規模なシステムに拡張するためには、用語や表現法あるいは内容の統制等、膨大な地道な統計的研究が必要であろう。将来は、統計学的成果を生かして規格化された文章が学会等では流通することを期待する。

本論文で提案した手法を用いて抽出された構造化キーワードを蓄積したり、マッチングをとって検索したりするためには、フレームを取り扱うための専用の機構が必要であるが、それには、最近の知識工学の成果が適用できよう。さらに、構造化キーワードに、さまざまな情報を付加することにより、高度の知識ベースへと発展させる可能性もある。

## 参 考 文 献

- 1) 中井: 機械補助索引(MAI)について, 情報管理, Vol. 19, No. 4, pp. 247-259(1976).
- 2) Schank, R. C. and Abelson, R. P.: *Scripts, Plans, Goals and Understanding*, John Wiley and Sons, New York(1977).
- 3) Schank, R. C. et al.: *Parsing Directly into Knowledge Structures*, Proc. 6th IJCAI, pp. 772-777(1979).
- 4) Woods, W. A.: *Transition Network Grammars for Natural Language Analysis*, *Comm. ACM*, Vol. 13, No. 10, pp. 591-606(1970).
- 5) 田中: 談話理解の構造, 情報処理, Vol. 20, No. 10, pp. 889-895(1979).
- 6) 長尾: 自然言語の理解, 情報処理, Vol. 19, No. 10, pp. 952-961(1978).
- 7) 絹川, 木村: 日本語文構造解析による自動インデクシング方式, 情報処理論文誌, Vol. 21, No. 3, pp. 200-207(1980).
- 8) Nishida, T. and Doshita, S.: *The Framework of Knowledge Representation and Its Retrieval in LGS—The Literature Guide System*, Proc. 6th IJCAI, pp. 662-664(1979).
- 9) Winston, P.H.: *Artificial Intelligence*, Addison-Wesley, Reading (Mass.) (1977).
- 10) 堀, 齊藤, 猪瀬: シナリオを用いて構造化されたキーワードをアブストラクトから抽出する一手法, 情報処理学会, 計算言語学研究会, 25-2(1981).  
(昭和 57 年 3 月 2 日受付)  
(昭和 57 年 6 月 15 日採録)