

# 文書の意味構造を用いたブラウジング機能の開発

5F-8

野上謙一\* 中本幸夫\* 矢島真人\*\* 田野崎康雄\*\*  
 \*東芝コンピュータエンジニアリング(株) \*\* (株) 東芝

## 1. はじめに

ワープロ等の普及により電子化された文書が大量に作成できるようになった。これらの大量の文書の中からユーザの要求に合った文書を検索し、効果的に提示できるシステムが望まれている。しかし、従来文書は物理的なページ単位あるいは巻物のように表示されているため、文書の内容を把握するには、文書の先頭から読んでいかなければならないのが現状である。これらの問題を解決するために、文書要約支援システム等の研究がされているが、現状の技術では十分な要約は実現できていない。

我々は文書の内容を表現するために用いられる語句、例えば『目的』のような語句の構文的な使われ方を解析し、文単位に意味属性を付与することで文書を構造化する手法を提案してきた<sup>[1][2]</sup>。今回は文書の意味構造解析によって抽出した構造化文書を用いて文書を階層的に一覧表示できるブラウジング機能を開発したので報告する。

## 2. 文書の意味的構造化

### 2-1. 意味的構造

我々の提案している意味的な文書の構造化とは、文書の章、節などの論理構造を見出し番号等の形態的な情報を用いて構造化するのではなく、文書の内容を表現するために用いられる語句の構文的な使われ方を解析し、文単位に『目的』『課題』などの属性(以下意味属性と呼ぶ)を付与することで、文書の意味的構造化を実現している。

### 2-2. 意味属性の付与

今回、我々が行った意味的構造化は、文書中より『目的』『課題』などの内容を述べた文(内容明示文)を抽出し、抽出した内容明示文に対して、『目的』などの意味属性を付与することにより行う。意味属性の付与には、意味属性に対応して複数の同一内容を表現する構文パ

ターンを記憶した意味辞書を用いる。この意味辞書とパターンマッチングにより用いて意味属性を付与する方法を図1で示す例を用いて説明する。

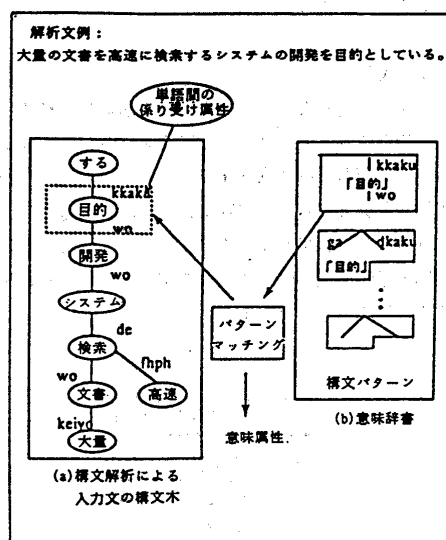


図1 意味属性付与方法

### (1) 構文解析による構文木の生成

構造化対象文書の各文を構文解析し構文木を生成する。例えば『大量の文書を高速に検索するシステムの開発を目的としている。』という文の構文解析結果の構文木を図1(a)に示す。各語句と語句は係り受け関係により木構造で関係づけられる。

### (2) 意味辞書の構文パターンとのマッチング

図1(b)には『目的』を表す複数の構文パターンが登録されていることを示す。(1)で生成された構造化対象文書の構文木が意味辞書に登録されている構文パターンに含まれているかを部分的なパターンマッチングにより行う。構文木に一致した構文パターンを含んでいれば、この文に対し意味属性を付与する。図1においては、意味辞書

中の構文パターンの1つが入力文の構文木の中に含まれているので、この文に対し意味属性『目的』を付与する。

また図1の例では入力文と意味辞書の構文パターンに「目的」という同一注目語句を含んでいる。しかし、意味属性『目的』を表す語句として、「目標」「狙い」「方針」などの類義語的な表現で表される同一内容の文に対しても同一意味属性が付与できる。

### 3. ブラウジング機能

#### 3-1. 実験

技術文書を対象として文書の意味的構造を木構造で一覧的に表示できるブラウジング機能を作成した。図2は文書の意味構造解析で意味属性が付与された文を階層的に表示した例である。本実験では『目的』『方法』『結果』『背景』『意見』『特徴』『課題』『内容紹介』の8種類の意味属性について、意味属性とそれに対応する原文を一覧的に表示できる。

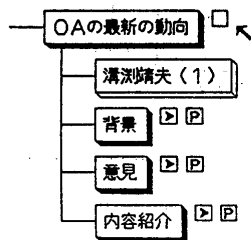


図2

#### 3-2. ユーザインターフェース

実験では、ユーザの指示によりトップダウンで文書の内容を表示できるインターフェースを実現した。まず、意味構造解析によって抽出された意味属性が木構造で表示される。意味属性の付与のされ方は、文書の書かれ方により異なっている。1つの文書に8種類全ての意味属性が付与されるわけではなく、出現順序も文書によって異なる(図2)。木構造のトップレベルには「表題」を表示し、横にあるボタンを押すことにより、その文書の「著者名」およびその意味解析で付与された意味属性名を階層的に列挙表示する。これにより、この文書がどのような内容を述べようとしているかの骨格を直感的に把握することができる。更に意味属性名の横のボタンを

押すことにより具体的な文を表示する。表示された文中には、その意味属性を特定する注目語句と言い回し(図3において意味属性『意見』の「～と考えられる」など)が使用されている。

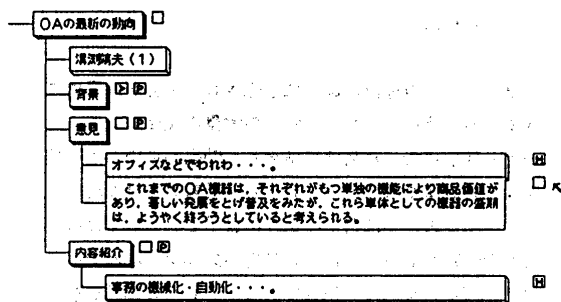


図3

#### 3-3. 検討

意味解析処理により構造化された文書を意味属性単位に階層表示を行うことにより文書の内容を直感的に把握できるブラウジング機能を実現した。しかし、今回行った意味属性の付与は、一文に含まれる語句や言い回しのみを対象にしているため、「以下に～についての特徴を述べる。」のように複数文に関連する場合や1文に複数の意味属性が付与されている場合には意味属性付与が行えない。そのためには今後このようなことを考慮した研究が必要であると考えられる。

### 4. あとがき

文書の内容を簡潔に把握できるように、意味的に文書を構造化表示するブラウジング機能について述べた。

文書中より内容明示文を抽出する方法および意味辞書については、本全国大会の関連発表である「文書への意味属性付与のための意味辞書の拡張」で詳細に報告する。

#### 参考文献

- [1] 岩井他：意味解析を用いた文書構造化手法、情報処理学会第43回全国大会論文集、1991
- [2] 矢島他：文書への意味属性付与のための意味辞書の開発、情報処理学会第43回全国大会論文集、1991