

自動抄録機能をもつ対話的文書検索システム — システムの機能と構成 —

2V-7

住田一男 酒井哲也 小野顕司 伊藤悦雄 三池誠司 武田公人
(株) 東芝 研究開発センター

1. はじめに

近年、ワークステーションの計算機パワーの増大にともない、全文文書を検索対象とした全文検索システムの実用化が進みつつある。しかし、現在実用化されている全文検索システムでは、検索してきた文書を表示する場合、検索文書のタイトルの一覧を表示するか、原文をそのまま表示するにすぎない [1]。

検索は、結果から統計的な情報を作成すること、あるいは検索した文書を読み、理解すること、内容を参考にし、再利用すること等を目的として行われる。このため、検索システムにおいては、検索速度や精度の点だけではなく、検索結果の提示方法も配慮し、効率的な検索を可能にする必要がある [2]。

効率的な検索インタフェースの構築を目的として、ディレクトリ構造のような情報の階層構造を利用し、大量情報を可視化する試みがなされている [3]。しかし、情報伝達の中心である言語情報についての扱いがこれまで未検討であった。我々は、効率的な検索を目的として、検索結果の文書から自動的に抄録を生成し提示することを特長とする文書検索システム BREVIDOC (Broadcatching system with an essence viewer for retrieved documents) を試作した。本稿では、試作したシステムの構成ならびに機能を述べる。

2. 対話的文書検索システム

システムの開発にあたり以下の2点を重視した。

- 一覧性

情報全体を把握し、目標箇所へすばやく到達するため、情報の一覧性を配慮する必要がある。

また、たくさんの文書から希望の文書を選択する際、特に文書が長くスクロールを頻繁に行わなければならない場合、目の疲れを経験する。このような問題を解決するためにも、様々な段階での一覧性を確保しなければならない。

- ユーザ主導性

利用者にとって、詳しく読みたい場所や量は、す

べての文書に対して一率というわけではない。概要を知るだけで十分な文書もあるし、内容を詳しく理解したい文書もある。画一的な表示ではなく、利用者の必要度に応じた表示が望ましい。

上記の2点に加え、文書検索では、希望の検索結果が得られない場合、検索された文書を読んだり、検索条件を代えたりなどの試行錯誤を行いながら検索する。このため、ユーザインタフェースは対話性を重視し、効率的な操作を行えるようにすることが必須である。

3. 機能と構成

前節の点を考慮して、自動抄録機能をもつ対話的文書検索システムを、エンジニアリングワークステーション (東芝 AS4000 シリーズ) 上に実現した。図1にシステムの構成を図示する。システムは、検索に先立って起

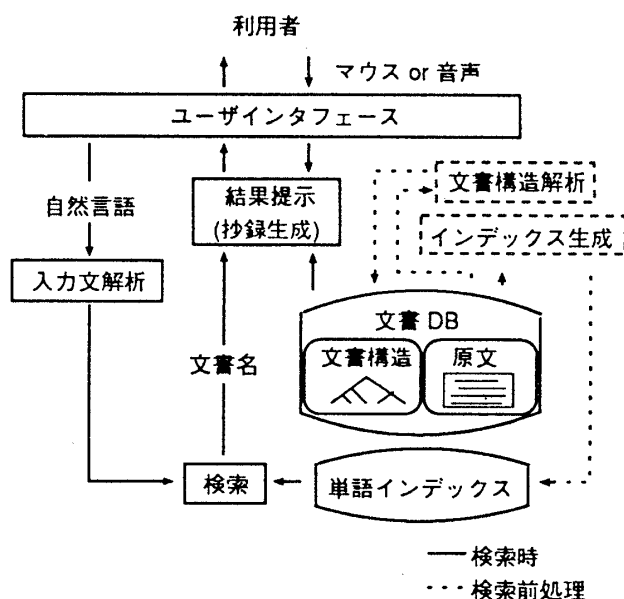


図1: システムの構成

動されるモジュール(破線で図示)と、検索時に起動されるモジュール(実線で図示)の2種類に大分できる。システムは次の6つのモジュールから構成される。

- 文書構造解析 [4]

章や節の構造、各章や節を構成する文章中の文間の修辭的な関係を抽出し、文書構造として表現する。抽出された文書構造は、原文書と対応づけて文書DB中に格納する。

● インデックス生成

高速な全文検索のために文書中に現れるすべての単語を抽出し、単語インデックスを作成する。抽出された単語は、その単語が含まれる文の“意味役割”¹と対応づけて単語インデックスに格納している [5]。

● 入力文解析

利用者が入力する自然言語文を解析し、検索コマンドに変換する。検索コマンドは意味役割と単語列とのペアで表現している。例えば、「被曝の低減を目指したもの」というような入力文に対しては、「目的:被曝/低減/」というように検索コマンドに変換される。

● 検索

単語インデックスを参照し、検索コマンドにしたがって文書を検索する [5]。検索コマンドで意味役割が指定された場合、その意味役割に対応づけて単語インデックスに格納されている単語情報に基づいて文書を検索する。

● 結果提示

文書群、文書一覧、抄録、原文の4つの詳細度で検索結果を提示する。文書群の提示では検索条件と近い文書をグループ化して表示する。一方、文書一覧の提示ではタイトル、著者名などを一覧表示する。抄録提示で提示する抄録は、文書DB中に格納されている文書構造を参照し、その構造に基づいて生成する [6]。

● ユーザインタフェース

検索の実行、表示する文書や節の指定など各種コマンドは、マウス入力とともに、音声入力も可能である。マウス入力の場合、離れた箇所にあるボタンを押す際、マウスの移動などの操作が繁雑となる場合がある。コマンドの音声入力を可能にすることで、このようなマウス操作を軽減することができる。なお、音声認識機能は、ソフトウェアのみで動作し外付けのハードウェアは一切使用していない [7]。

4. 動作

利用者が検索条件として、自然言語文を入力すると、まず始めに、入力文解析部がその文を解析し検索コ

¹“意味役割”とは、文書の種類に対応して含まれているだろうと予測される情報を指し示すために、指標として設けたものである。例えば、科学技術論文などの場合、「話題」、「背景」、「目的」、「特徴」、「結論」、「課題」などがこれに相当する。

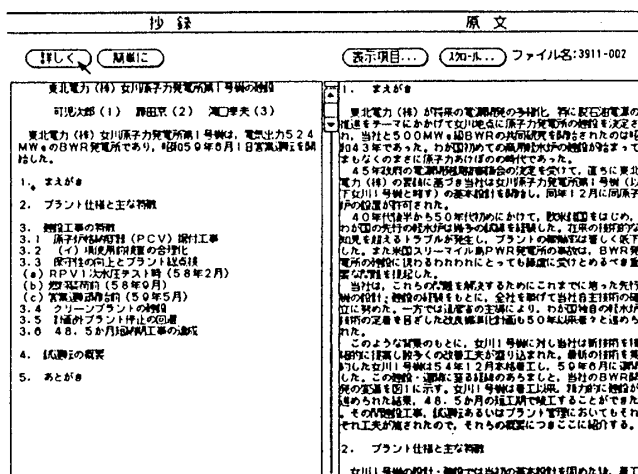


図 2: 原文と抄録提示

マンドに変換する。次に、検索部が、単語インデックスにしたがって文書検索を行い、検索条件に対応する文書を取り出す。

結果提示では、検索条件と類似性の高い順に検索文書をグループ化して表示する(文書群表示)。また同時に、文書群表示で表示している文書群のうち最も類似度の高いものの文書一覧、さらにその一覧の最初に位置する文書の抄録と原文を同時に表示する。抄録と原文の提示状況を図2に示す。

提示されている抄録は、利用者の指示に応じて詳しくを変えることが可能である。抄録提示において節が指定されるとともに、詳しくの指示ボタン(「詳しく」、「簡単に」)が押されることにより、その指示にしたがって節ごとに表示する抄録の詳しくを変更する。

5. おわりに

試作した自動抄録機能をもつ全文検索システムの機能と構成について述べた。システムは、検索結果の抄録を提示し、提示された抄録の任意の箇所を詳しく表示するという文書検索のための新しい文書提示インタフェースを有している。今後、試作したシステムを用い検索行為の効率という観点からの評価を行う予定である。

参考文献

[1] 例えば寺地元他:「フルテキストサーチシステム Bibliotheca/TS の開発(3)」, 第45回情全大, 第3分冊, 2C-5.
 [2] 三輪眞木子:「データベースサーチャーの視点」, 情報解説, Vol.33, No.10, pp.1162-1170.
 [3] Card, S.K., et al.: "The Information Visualizer, an Information Workspace," *ACM Proc. SIGCHI'91*, pp.181-188.
 [4] Sumita, K., et al.: "Document Structure Extraction for Interactive Document Retrieval Systems," *ACM Proc. SIGDOC'93*, pp.301-310.
 [5] 三池誠司他:「自動抄録機能をもつ対話的文書検索システム - 検索機能 -」, 第48回情全大, 第3分冊, 2V-8.
 [6] 小野顕司他:「自動抄録機能をもつ対話的文書検索システム - 自動抄録機能 -」, 第48回情全大, 第3分冊, 2V-9.
 [7] 永田仁史他:「ワークステーションにおける音声認識機能の開発」, 信学技報, HC 91-19, pp.63-70.