

目次表題を利用した情報検索手法の検討

5 P-8

和田 光弘

東京大学大学院工学系研究科

片山 紀生 高須 淳宏 安達 淳

学術情報センター研究開発部

1 はじめに

現在までに、全文検索などの情報検索技術が発達したが、蓄積情報の変化により、利用者が検索時に使用できる情報は多様化している。それらの情報を最大限に利用するために検索支援機能の開発が活発になっている。

本稿では目次表題を検索の際に積極的に用いて、検索キーワードの表示や Browsing 検索などを行い、利用者支援に役立てる手法を検討したので、その概念とプロトタイプについて述べる。

2 目次を使用した検索技術の提案

キーワード検索の一般的な問題点として、妥当なキーワードが不明であること、キーワードに対応する文献数が予想できないことが挙げられる。それらを解決するため、キーワードの絞り込み候補や拡張候補を提示する手法¹⁾などがすでに提案されている。

一方、情報生成者と利用者の間での効果的なコミュニケーションを図るために、利用者の知識状態を検索システムが把握し、それに応じて利用者に情報を提示するといった、認知的観点から検索システムを設計する必要も言われている。しかしながら、利用者の知識状態を正確に推し量ることは無理であるし、それを捉えるために情報操作回数を増やすことは逆効果である。

そこで、本システムでは検索システムとそれを使用する利用者の知識とを繋ぐインターフェースとして、目次表題を用いることとした。目次情報の検索利用は、すでに Ariadne²⁾ などで使用されており、将来的に有用な手段と考えられる。

目次表題の利用の利点として、

1. 利用者は目次表題を眺めることで、検索を行なう上での背景知識を得ることができる。
2. その上で、目次表題に即した検索語を表示すれば、検索システムとの意図のずれを防ぎ、検索対象を

A Study of Document Retrieval Method using a Table of Contents

Mitsuhiro WADA¹, Norio KATAYAMA², Atsuhiko TAKASU², Jun ADACHI²

¹Graduate School of Engineering, The University of Tokyo

²Research & Development Department, National Center for Science Information Systems

明確にすることができる。

3. 目次の章や節などの構造を利用することで、ある分野に関して、検索範囲の絞り込みや拡大を行なうことが可能になる。

などが挙げられる。

3 プロトタイプシステムの機能と構成

以上に述べた機能を実現するために、図1のような構成の検索システムを試作した。

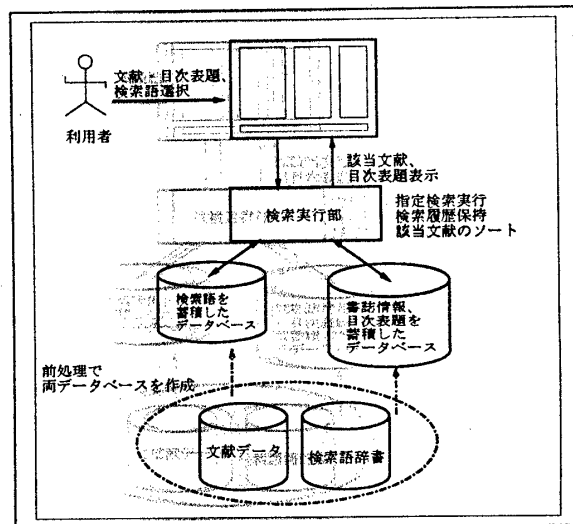


図1: プロトタイプシステムの構成

プロトタイプシステム実現に向けて考慮したのは、

1. 検索語に対しての該当文献数、一文献に対する検索語数などのデータを得るために、前処理してある。
2. 利用者の負担を軽くするため、複雑な検索条件は用いず、マウスによる選択が中心になっている。
3. 利用者が必要な情報に早くたどり着けるよう、章のみや節までなど、目次表題の表示レベルが選択できるように設計している。

の3点である。

なお使用した文献は、電気工学の分野を中心に462冊の文献の表題、及び目次表題である。

3.1 前処理部

目次表題から検索語を抽出するため、EDR³の情報処理分野の専門辞書(約12万語)と学術用語集の電気分野(約1万語)の辞書を使用した。辞書に存在する検索語を利用し、検索語データベースと文献情報データベースを作成しておく。一文献あたりの平均目次表題数は104、一目次表題に付与される検索語数は4であった。

3.2 目次表題から抽出された検索語による検索

目次表題を指定すると、抽出された検索語とその検索語で検索できる件数が表示される。例えば、「データモデルの拡張—構造表現の強化」という目次表題があった場合、「データモデル」「拡張」のような検索語、そして節以下に含まれる「関係データベース」「オブジェクト指向」などの検索語が抽出され、各々選択可能な形で表示される。利用者はこれらを任意に選択することにより、該当文献を得ることができる(図2参照)。

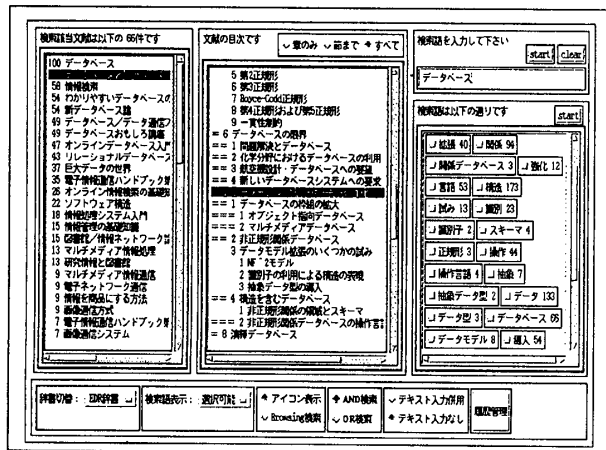


図2: 目次表題からの検索語検索実行画面

上位の目次表題は、下位の目次表題の検索語を含むので、章では表示される検索語が増えて、選択できる分野は広がるし、逆に節では表示される検索語数が少なくなり、狭めることができる。

3.3 目次表題からの Browsing 検索

目次表題からの Browsing 検索とは、指定した目次表題から抽出された検索語から、目次表題の内容を自動的に判断し、それに即した文献を表示させるものである。実行例を示すと、「情報検索の基本概念」のような表題が存在した場合、検索システムはあらかじめ前

処理されていた、「キーワード」、「演算」、「シソーラス」などの検索語を使用し、情報検索やデータベースのような文献を検索結果として返す(図3参照)。

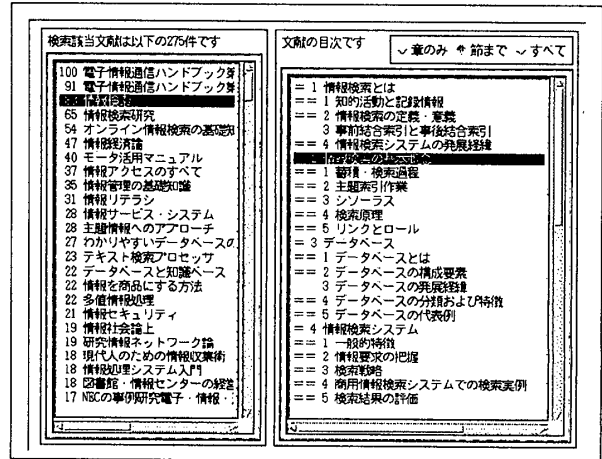


図3: Browsing 検索実行画面

参考文献から文献にリンクを張ることと同意で、目次に動的なリンクを張ったとみなすこともできる。検索分野の指定が容易なるだけでなく、情報発見としても利用できると考えられる。

4 まとめと今後の課題

本稿の手法は、目次表題を検索利用者に表示させることで、検索の際のシステムからの情報提示をめざすもので、検索分野の知識の取得、整理が可能である。特に、目次は普段から接することができるので、利用者は使用の際に特別な知識を必要としない点で優れていると言える。

現在、プロトタイプの評価として被験者による使用実験と、システムの改良を平行して行っている。さらに全文検索などと組合せ、利用者が場合に応じて、検索手法を切替えることができ、さらに積極的に検索支援を行なうようなシステムへの発展などが課題である。

参考文献

- [1] 富士, “キーワード検索における検索ナビゲータ,” 情報処第45回全国大会, 4-231, (1992).
- [2] 長尾他, “電子図書館 Ariadne の開発,” 情報管理, Vol.38, No.3, pp.191-206(1995.6).

³ 日本電子化辞書研究所