

## 自他論旨比較に基づく文書探索システム

中塚 麻奈美 小嶋 弘行

広島工業大学 環境学部 環境情報学科

### 1. はじめに

近年、家庭や業務現場でのコンピュータ利用の普及に伴い、膨大な文書情報が電子文書として取り扱われるようになった。研究開発や文献調査活動においても、協調作業や知識共有支援を目的とした機械的検索による文献利用が増加している。特に、論理的記述が重視される技術論文の検索では、優れた論文を読解参照することによる知識の取得が日常的に行われている。一方、電子文書の増大は、検索結果の膨大化を招き、検索効率を悪化させている。これは、現在の検索エンジンの多くが文書の内容・構成にかかわらず対象文書中において所定のキーワードが存在するか否かのみを基準としてページを拾い上げているためである。

また、検索文書を読解によって得た文書内容そのものや論述、構成などに関する知識は、自己の知識と比較参照でき、業務活動や学習の場面で利用に有効であると予想される。一方、文書への理解・解釈には個人差があり、同一の文書を読んだ場合であっても、読者の経験や知識の差により異なった認識が生じる場合がある。そのため、文書に記述された論旨すなわち議論の展開、筋道が読み取れ、読者の的確な文書理解を助けるツールが必要であると考えられる。

これらを受け、組織における業務活動の効率向上と業務知識の共有の観点から、検索キーワード群に加え、文書の論理構造も利用し得る文書検索、文書作成支援手法が望まれているといえる。

本研究では、比較的文書の論理構造が明確であり、論旨が重要視される学術論文を対象に、自他論旨比較に基づく文書探索システムを提案する。

### 2. 論旨に基づく文書利用支援の検討

従来の文書理解、文書作成支援の方法として

Documents Seeking Systems Based on Comparison  
between Oneself and Others Point of an Argument  
Manami Nakatsuka, Hiroyuki Kojima  
Department of Environmental Information Studies  
Faculty of Environmental Studies

文書中から必要な部分だけを入手するために、自動要約ツールなどによって読み手の文書の量を制御し、短時間で適確に内容を把握する技術や、文書自動生成ツールの研究が進められている。しかし現在、これらには構文解析や意味解析の手法に加え、技術的には対象とするテキストのジャンルを考慮した手法を選ぶ必要性もあり、文章が長く複雑になればなるほどテキストの抽出、要約ともに能力に限界があるのが現状で、文書全文を対象に要約を行った場合、利用目的によっては文書の内容を正しく表現できていない場合も生じる[1]。

一方、人間が文書を検索参照するという行為は、昔から日常的に行われている。このヒューマンプロセスを電子文書利用に活用することが、検索参照の効率化として有効であると考えられる。人間が文書を検索参照するときの特性として、文書の題名から内容を推測、選択し、文書の目次や内容の特徴を表している部分を拾い読みする事で、詳読をすることなく全体の内容を把握し、文書を理解しようとする。この文書論述の構造的な意味を踏まえた人間の文書理解、文書作成行動そのものをコンピュータに自動処理させることは困難である。むしろ上述の人間の文書検索、比較参照による文書生成などの行動プロセスを支援する方式が望ましいと考える。

### 3. 論旨検索、知識共有のための文書構造化

本研究では、論旨を頼りにした論文検索、文書作成支援の方式を提案する。これには、文書中から文章の筋道となる論理構造を抽出する必要がある。そのため、XML記述により章レベルで構造化された論文を対象とし、この文書内の論述構造の抽出を行う。XMLはデータに独自のタグを指定することが可能で、テキスト形式で記述されるため、汎用性を有し様々な場面に適用される。このXMLの特徴を利用し、タグを利用して章ごとに文章を抽出する。その文章を要約ツールを用いてそれぞれ要約することにより重要文を抽出、論旨の生成、文書構造化を行う。これにより、論旨をキーとした文書検索および文書の構造理解から論旨展開知識の共有を図る。

また、本方式の検討には論文の序論と結論部に着目した。学術論文などの論理的構造特性を有する文書の場合、序論と結論の部分に最も重要なことが書かれているといわれている。そのため、序論と結論の部分を利用し論文検索を行うことでキーワードのみの検索に比べ関連文書を見つけやすいのではないかと考えた。したがって、検索のための文書構造化として序論と結論の部分を要約し、重要キーワードの抽出とキーワードの出現順序を決める方式を採った。

#### 4. 論旨比較による文書作成高度化支援

論旨とは、「論文・議論の主旨・議論の筋道」と定義される。そのため論旨は個人の文章の書き方に大きく依存しており、コンピュータで自動的に自他論旨を比較することや、自己の文書に反映させることは現時点では難しいと予想される。そのため、人間の文書理解や文書比較、参照を支援する方式について述べる。

人間の文書参照という行動を支援するために、自他論旨を並べて提示した。優れた文書と自己の文書を比較することは、自己の論述の矛盾摘出や自己論旨の発展、論理構造の表現知識取得などに効果があると考えられ、文書理解の迅速化、個人の能力育成および業務の改善、効率化が期待できる。また、読者の経験や知識の差により同一の文書を読んだ場合であっても、文書への理解、解釈には個人差がある。論旨を提示、読解することにより、こうした認識のずれを明示できる確かな文書理解を図る効果も期待される。

#### 5. 論旨比較に基づく文書探索システム

本研究の文書探索方式として、論旨に着目した検索方式を提案する。方式検討には学術論文を対象とした。学術論文は論理的構造特性を有する。その構造に着目し、章ごとの要約文と抽出専門用語の提示によって論旨を表現した。また本システムの試作実験のため、あらかじめ研究室メンバーの論文をデータベースに保存した。

具体的な処理手順は、XMLで構造化記述した論文の定義タグ<introduction>、<conclusion>を頼りに序論と結論とその他の章に分け、重要文抽出、要約生成して表示する[2]。そして、序論と結論の部分の要約文を利用して、重要語抽出ツールを用いて接続語と頻度による重要キーワードの抽出を行う[3]。このキーワードの出現順序を頼りにデータベースから全文検索エンジン Namazu を利用して論文検索を行う。上記機能を、Javaのサーバサイドプログラミング技術である JSP/サーブレットを用いて実装した。

次に本システムの論旨生成および実行画面を図1に示す。はじめにユーザは比較したい、例えば自己の論文を表示させ、論文の論旨として章の要約と重要語を提示させる。重要語とその順序を踏まえた検索を実行、関連のある論文が検索表示される。ユーザは表示された論文と論旨を参照することにより、論旨展開の差異、自己と他者の論旨の矛盾を把握する。



図1 論旨生成および検索実行画面

#### 6. おわりに

文書検索の効率向上と文書作成高度化支援の観点から、論旨による文書検索、比較表示の方式について検討した。以下に検討結果をまとめる。

- (1) 文書の論理構造化として、章、段落ごとの要約と重要語提示により論旨を表示、文書の内容理解や論旨構成知識の習得などに供する文書探索支援方式を提示した。
- (2) XML タグと要約、重要語抽出ツールを用いた文書論理構造の抽出機能を構築した。
- (3) 情報検索の検索エンジンとして、XMLで構造化された論文を対象に文書論理構造抽出機能により得られた重要キーワードの出現順序を用いた全文検索エンジン Namazu による検索機能を実装した。

#### 参考文献

- [1] 奥村学：テキスト自動要約，情報処理，Vol. 45, No. 6 (2004)。
- [2] 簡易要約器 Posum： <http://www-cl.tufts.ac.jp/pub/tools/posum/>
- [3] ” 専門用語自動抽出システム ” のページ：  
<http://gensen.dl.itc.u-tokyo.ac.jp/>