平成 13 年 2 月22 日

電子・情報工学専攻	学籍番号	943708
申請者氏名	大竹 清敬	

指導教官氏名	増山 繁　教授 磯田 定宏 教授

論 文 要 旨 (博士)

論文題目	Studies on Relevant Document Retrieval and Summarization (関連文書検索とその自動要約に関する研究)

(要旨 1,200 字程度)

本論文の目的は，人間の知的活動を支援し，増強するためのいくつかの技術を発展させることである．そのために，ある文書に関連する文書群を検索するための手法を提案する．次に関連する複数の新聞記事をまとめて要約する手法を提案する．さらに，より高品質な要約のために必要な格フレームを語順を考慮して自動獲得する手法について検討する．そして，格フレーム獲得において問題となる名詞の多義性を解消する手法を提案する．

計算機およびネットワークの発展によって，膨大な量の情報を容易に得ることができるようになった．その一方で，生物としての人間の情報処理能力はほとんど変化していない．そのため，膨大な情報の中から高速・高精度に必要とする情報を集め（検索），素早くその内容を理解（要約）することが高度情報化社会では要求される．さまざまな情報のなかでも文書は人間の知的活動の基礎となるため極めて重要な要素である．

これまでの多くの情報検索研究は検索質問に基づくシステムを想定している．一方で，検索質問としてキーワードではなく，文書そのものを示し，その文書に関連する文書を検索する手法が考えられる．このような手法はいくつか提案されてきたが，どのような索引語単位を用いるべきかについての研究はほとんどない．そこで，本研究では，索引語の単位として形態素の連接を用いた関連文書検索法を提案し，形態素のみを索引語の単位として用いる手法と比較した．その結果，本研究で提案する関連文書検索の有効性を確認した．

初期の自動要約に関する研究は，単一の文書においていかに重要な文を選択するかに主眼が置かれたが，1995 年以降になって，複数の文書をまとめて要約する場合の問題点も検討されるようになった．本研究では，対象を新聞記事としたとき，その表現上の特殊な構造から，ヒューリスティックスによって十分な要約手法が構築できると考え，手法を考案した．この複数記事要約手法をアンケートによって評価した結果，文章が自然で，適切な要約であることが明らかになり，本手法の有効性を確認した．

また，自動要約に関する研究が進む一方で，より自然で読みやすい要約の生成は依然として困難である．その原因のひとつとして構文解析誤りがある．日本語の構文解析においては，格構造解析が必要であり，そのために格フレームは重要な役割を果たす．しかしながら，既存の格フレーム辞書は人手により収集・整備されているため，量的に不十分である．そのため，格フレームの自動獲得が望まれる．そこで，本研究では，従来重要視されてこなかった語順に着目し，単一言語コーパスから格フレームを自動獲得する手法を検討した．また，コーパスからの格フレーム獲得において問題となる名詞の多義性を解消する手法を検討し，大規模なコーパスを用いた実験により提案手法の有効性を確認した．