

論文内容の要旨

博士論文題目 デジタル図書館のための文書主題の抽出とその検索への応用

氏 名 堀井 千夏

本研究の目的は、単語の概念を求め、文書の主題や利用者の検索意図を推測して検索精度を高める新しい文献検索手法を提案し、21世紀型図書館として話題になっているデジタル図書館を実現することにある。

まず、本論文では、文書中の出現単語と検索質問を単なる文字列として処理するのではなく、単語が深層的にもつ意味をその単語の概念として求め、文書の主題や利用者の検索意図を推測して検索精度を高める手法を提案する。また、提案手法を近年の飛躍的な情報活用場として注目されている「デジタル図書館」の文献検索に適用し、本手法の実用化への問題点を考察している。本論文は以下の6章からなる。

第1章では、研究のモチベーションを記述し、概念情報に基づく文書主題の抽出法の必要性と有用性について記述している。第2章では、デジタル図書館の特徴と問題点を明らかにし、文書主題の必要性とその関連研究を概観している。第3章では、文章の深層的な主題を抽出するために、概念空間の出現頻度から算出した主題の表現度と、単語の出現頻度から算出した主題の主張度を特徴パラメータとする新しい文章主題抽出手法を提案している。また、これらの特徴パラメータの有用性を検証するために、技術論文を対象とした評価実験を行っている。第4章では、デジタル図書館への具体的な応用ケースとして、上記の特徴パラメータを利用することにより、意味検索を可能とする文献検索システムを試作し、評価実験を行う。また、ユーザの多様な検索目的を可能にする文書構造に基づいた主題抽出実験を行い、提案手法が司書エージェントして機能することを実証している。

第5章の結論では、本研究の成果を総括し、提案手法が急速に多様化かつ大規模化しているネットワーク上の情報源を検索するためにも有効な手法となるために残された課題を整理している。

| | |
|----|------|
| 氏名 | 堀井千夏 |
|----|------|

(論文審査結果の要旨)

本論文は、デジタル図書館における意味検索に直結する文献検索手法に関する研究であり、概念パスの表現度と単語の主張度という特徴パラメータで文書主題を特徴付けることにより、検索者の意図を反映した検索が可能となる手法を提案した先端的論文である。従来、来館者の希望を聞き取ることにより的確に文献を探してくれる司書が、図書館の優劣を左右していたが、21世紀にはデジタル図書館の普及が予測されることから、情報検索機能をもった司書エージェント機能の付加が不可欠である。この社会的要請に子と得るために、検索者の意図を的確に判断し、不要情報を削除し必要情報のみを検索してくれるシステム開発の可能性を追求した本論文の成果は、以下の2点に要約される。

1. 文書主題を的確に抽出する単語の意味の曖昧性を解消するために、多義語の意味の分離と同義語の統合を概念情報によって実現する手法を考案し、単語の深層的な意味表現技術を開発した。
2. 文書に含まれる多様な概念情報を絞り込んで文書主題を抽出するために、概念情報の出現頻度から算出した内容の表現度と単語の出現頻度から算出した内容の主張度を特徴パラメータとした検索システムを構築し、技術論文を対象にした評価実験を実施し、検索者の意図が的確に反映した検索が可能であることを実証した。

以上述べたように、本論文は、文書主題を単なる単語の出現頻度ではなく、概念空間を導入することにより検索者の意図を的確に反映した特徴パラメータを用いた新しい文書主題抽出法を提案し、その有効性をデジタル図書館システムに適用して検証した実証的研究である。

これらの研究成果は、学会論文誌1件、査読付国際学会3件として公表されていること、ならびに国際学会論文(連名者)3件などを鑑みると、意味検索という新しい情報検索研究分野の発展に学術面での貢献は大きいと認めることができる。また、21世紀の図書館として社会的にも注目されているデジタル図書館に関する研究であり、社会的ニーズに応える真の実用研究としての貢献も大きいことを示している。

よって、本論文は博士(工学)の学位論文として価値あるものと認める。