

氏名(本籍)	と だ ひろ ゆき (香川県)		
学位の種類	博 士 (工 学)		
学位記番号	博 甲 第 4476 号		
学位授与年月日	平成 19 年 7 月 25 日		
学位授与の要件	学位規則第 4 条第 1 項該当		
審査研究科	システム情報工学研究科		
学位論文題目	トピックに注目した文書検索結果の構造化に関する研究		
主 査	筑波大学教授	理学博士	北 川 博 之
副 査	筑波大学教授	Ph.D.	田 中 二 郎
副 査	筑波大学教授	理学博士	大 保 信 夫
副 査	筑波大学教授	工学博士	西 原 清 一
副 査	筑波大学准教授	博士(工学)	山 本 幹 雄

論 文 の 内 容 の 要 旨

デジタル化された大量の文書情報から利用者が望む情報を抽出するためのシステムとして、キーワードベースの検索システムが幅広く利用されている。しかし、このようなシステムを実際に利用する場合、膨大な検索結果に直面し、真に欲しい情報を獲得できないことがしばしば生じる。特に、検索条件のあいまいさにより膨大な検索結果中に所望の情報が埋もれてしまう場合や、検索の目的が検索結果全体から特徴的な情報や概要を抽出したいという場合などに、このような問題が顕著になる。

本研究では、このような背景のもと、検索結果中に含まれるトピックに注目し、検索結果を構造化することで利用者の情報検索を支援するアプローチについて考察している。具体的には、以下の2つの手法を提案している。

- (1) 固有名詞を利用した検索結果クラスタリング
- (2) グラフ分析を利用したトピック構造マイニング

(1)の手法では、キーワードベースの検索システムにおいて、検索条件のあいまいさにより、所望の情報が検索結果中に埋もれてしまうという問題に対して、検索結果に含まれる固有名詞に注目した検索結果のクラスタリングを行うことで、絞り込み検索を支援する手法を提案している。また、IREXのテストコレクションを用いた評価により、絞り込み検索の精度を従来手法と比較し、より少ない検索結果の文書を処理対象とすることで高い精度を得ることを示している。また、本手法は、現在Web上のポータルサイトgooにて、ニュース記事検索、ブログ記事検索および話題提示システムに導入され、多くの一般利用者に利用されている。

(2)の手法は、検索結果全体から特徴的な情報や概要を抽出したいという要求に対応するものである。具体的には、文書をノード、文書間の関係をエッジで示す「文書集合グラフ」で検索結果を表現し、Interested Reader Modelに基づく分析を行うことで、検索結果中に含まれる主要なトピックやトピック間の関係、トピックと個々の文書の関係を明らかにし、ユーザの情報取得を支援する手法を提案している。新聞記事を用いた評価およびNTCIR-4の「トピック分類タスク」に基づく評価実験により、クラスタリングやトピック抽出に適用した場合、従来手法と比較して高い精度をもつことを示している。また、新聞記事を用いた評価実験

により、特定のトピックの主要な内容や派生的な内容を述べた文書を効率的に選択できる事を示している。

さらに、ニュース記事やブログ記事のようにタイムスタンプが付いた文書を処理するために、「文書集合グラフ」構築時に、文書間の時間的な近さを考慮する事で、タイムスタンプを意識したトピック構造マイニングを行う手法を提案している。また、評価実験により、タイムスタンプを利用しない場合と比較し、トピック抽出、クラスタリングの精度が向上することを示している。

審 査 の 結 果 の 要 旨

デジタル化された文書情報の増大に伴い、文書検索システムの与える検索結果から利用者が必要とする情報の獲得を支援する技術はますます重要となっている。本研究は、トピックに注目した文書検索結果の構造化に関する二つの手法を提案したものであり、手法としての新規性と評価実験を通じた有効性が認められる。また、手法(1)に関しては、実システムにおけるサービスとして導入され、現在も一般利用者に利用されている点も評価できる。以上により、本研究は情報工学上の貢献が大きいと判断される。今後、二つの手法の融合や効率化などの残された研究課題に関して研究が進められることが期待される。

よって、著者は博士(工学)の学位を受けるに十分な資格を有するものと認める。