

## 審査の結果の要旨

氏名 若木 裕美

本論文は、「Query Refinement based on Comprehensive Representation of Multiple Topics (複数トピックの包括的提示による検索支援に関する研究)」と題し、英文6章から構成されている。WWW 文書などに対するキーワード検索の結果が含む様々なトピックを整理・提示し、さらに検索質問の改善に結びつけることを目的とした研究である。『トピック』について定義を行い、そのようなトピックを形成しやすい単語の抽出方法を提案している。また、提案手法の有効性を多様な種類の文書コレクションを用いて検証し、さらに、提案手法のデモシステムを作成して実験的な環境ではなく実際の検索結果のデータと現実的な検索語を用いた実験を行っている。これにより、提案手法の特性を既存の他システムとの比較を通じて検証したものである。

第1章は、「Introduction」であり、本研究の背景および既存のキーワード型検索エンジンの問題点について指摘している。さらに、本研究の意義について述べると共に、本研究の提案するアイデアを簡単に紹介している。最後に、本論文の構成をまとめている。

第2章は、「Related Work」であり、本研究と関連する研究について「Overview」において広くまとめている。特に、検索結果の内容を整理することに注目した研究の流れと、特徴的な単語の抽出を目的とした研究の2つに大別してまとめている。「Overview」の後には、それぞれの内容に関して、特徴的な研究について細かく調査している。

第3章は、「Tangibility: A New Measure for Topical Term Extraction」と題し、本論文で提案する単語重み付けの方法について説明を行っている。Tangibilityとは本論文で提案する単語重み付け手法のコンセプトであり、『特定の単語とのみ頻繁に共起する単語はトピックを形成しやすい』というものである。Tangibilityの定式化として、まずTNG1・TNG2を予備実験的に提案している。これはTangibilityの有効性を確認するためのものであり、次の第4章の最初の実験でその有効性が確認される。さらに、Tangibilityのコンセプトを追求し最終的な定式化としてTNGを提案している。また、TNGによって抽出された単語のクラスタリングの手法についても述べている。

第4章は、「Experiment for Tangibility」と題し、本論文で提案するTangibilityの定式化に関する実験を包括的に行ったものである。まず、TNG1・TNG2について他の単語重み付け手法との比較実験を行っている。実験データはNTCIR3の検索課題である。また、その後はTNGについて実験を詳細に行っている。他の単語の重み付け手法と比較しながら、抽出された単語がトピックに偏った出現をする単語であるかどうか、また、検索に有効であるかどうかについて実験を行っている。特に、トピックへの偏りについては被験者による人手の判断も行っている。実験データは全7種類で、検索性文書コレク

ション・文書分類用コレクションの両方を利用し、言語も日本語と英語の両方を用いている。

第 5 章は、「Representation System of Multiple Topics with Search Results: Application of Tangibility」と題し、本論文で提案した TNG を用いたシステムを作成している。システムが検索結果として使用するデータは、Google の検索結果の URL 一覧から文書を収集し使用している。検索語としては、実際によく使われた検索語や多義性の問題のある検索語を用いている。比較実験では、検索結果をクラスタリングする公開システム Clusty を比較対象としている。また、提案手法で生成した単語クラスタを被験者に実際に検索に使ってもらい、その有効性や特性について議論している。さらに、検索語の曖昧性の問題について考察・議論を詳細に行っている。

第 6 章は、「Conclusion」と題し、本論文の全体を簡潔にまとめると共に、第 1 章で議論した本研究の意義に対し本研究が解決した点に言及している。

以上これを要するに、本論文は、爆発的に増大する WWW 文書等の情報からの検索に関して、少ない単語で曖昧に表現された質問による検索結果に含まれる多様なトピックを抽出し、そのトピックを表現するのに適切な語の集合を求め、多義性やファセットなどに基づいて分類表示するために有用な統計的な手法を提案し、それを大規模なコーパスにより定量的に評価しまた他システムとの比較を行うことにより、提案手法の有用性を実証したもので、情報検索やテキストマイニングにとって有用な知見が得られており、電子情報学上貢献するところが少なくない。

よって、本論文は博士（情報理工学）の学位請求論文として合格と認められる。