# 修 士 論 文 の 和 文 要 旨

| 研究科・専攻 | 大学院　情報理工学研究科　情報・通信工学専攻　博士前期課程 | | |
|---|---|---|---|
| 氏　　名 | 田中 英人 | 学籍番号 | 1031062 |
| 論 文 題 目 | 1 文字タグづけ手法実現のための重要語抽出アルゴリズムの提案 | | |

要　旨

長大な文書情報が多く存在する近年の Web 上においては、適切な文書管理能力が必要とされている。特に、効率的な情報検索を行なうためには、あらかじめ文書の特徴を表す付加情報を持たせておく方法が考えられる。そこで、文書から特徴的な単語を抜き出す手段としてキーワード抽出の技術に注目した。さらに、もう一つの文書整理の手段として，タグを付加する手法が挙げられる。内容が関連した複数の文書に同一のタグをつけることで文書間に関係性を持たせることが可能であり、管理および検索への効果が期待される。本研究では、以上に挙げたキーワード抽出技術およびタグづけの技術を組み合わせることで、文書に対するタグを自動抽出する技術に着目した。しかしながら、タグづけの技術に関しては「表記揺れ」と呼ばれる問題が存在し、文書からキーワード抽出したものをそのままタグとして付加した場合、文書自体の書式や著者などの相違によってタグ同士の関連性が損なわれてしまうと考えた。上記の問題点を解決するために、本研究では Web 上のコンテンツに対して 1 文字でタグづけを行なう手法を提案する。これは、タグづけを行なう際に「1 文字で付加」という制約を与えるものであり、1 文字で表現するがゆえの直感性や連想性を活かせるのではないかと期待している。この手法の概念に基づき、対象となる文章から 1 文字単位で重要語を抽出するアルゴリズムを構築することとした。計算指標としては、キーワード抽出技術である tf-idf 法の概念や、文章中の出現位置、含まれる文の長さ等の情報を利用している。作成したアルゴリズムについて、実験により機械学習を実施することで最適な計算パラメータを決定した。得られたパラメータを適用したアルゴリズムの性能について評価を行ない、その有用性が示された。