

論文内容要旨

| | | | |
|--|----------------------|-----|------|
| 報告番号 | 甲 先 第 191 号 | 氏 名 | 北川浩也 |
| 学位論文題目 | 文書要約と文字列検索の効率化に関する研究 | | |
| <p>内容要旨</p> <p>文書要約は、文書を解析することで、重要な内容を簡約した情報に変換することであり、要約の量は、アプリケーションに応じて変更される。近年の情報爆発の時代、また携帯端末の普及により、ユーザが意図する情報を効率よくアクセスするための重要な文書処理技術の一つとして、広く利用されている。文書要約手法は多く存在するが、利便性の高い手法の一つとして、教師なし学習手法である NMF(Non-Negative Matrix Factorization)を利用した手法がある。しかし、NMF の計算量は、文書のサイズに比例するため、文書のサイズにより、要約処理速度が低下する欠点があった。従って、本論文の目的の一つは、この NMF による要約技術を携帯端末などで高速処理するための効率化の提案を行う。また、多種多彩なキーワード抽出を必要とする文書要約の汎用的高速化技術の一つとして、複数文字列検索がある。複数文字列検索とは、対象となる文書から、あらかじめ用意した文字列（キーワード）の集合が出現するすべての位置を検出することである。この技術は、上記 NMF 手法を含めた一般的な文書要約だけでなく、文書処理手法の重要な技術の一つである。従って、本論文では、この複数文字列検索の効率化手法を提案する。</p> <p>以上、本論文は、文書要約と文字列検索の効率化に関してまとめたものであり、まず、NMF 手法による文書要約を高速化する手法を提案し、次に、複数文字列検索の効率化手法を提案する。本論文は、以下に示す 5 章により構成される。</p> <p>第 1 章では、文書要約と文字列検索の背景及び、関連研究についてそれぞれ概説し、本研究の目的と位置づけを明確にするとともに、本研究の成果について述べる。</p> <p>第 2 章では、一般的な文書自動要約手法について概説し、従来研究について概説する。</p> <p>第 3 章では、NMF の教師なし学習を利用した文書要約手法の計算量を削減するために、NMF に用いられる特徴ベクトルの削減手法について提案する。提案手法に対する評価実験では、文書自動要約精度の低下を抑えつつ、高速な文書要約を実行できることを示す。この結果として、従来の NMF 手法に対して、約 8.5 倍の速度で文書要約が高速化できることが分かった。</p> <p>第 4 章では、一般的な文字列検索技法を概説した後、複数文字列に対するマッチング手法の効率化として、トライ法による複数文字列検索による、Aho-Corasick の手法の高速化手法を提案する。実験により、トライ法が Aho-Corasick の手法より、検索速度が 19.8%から 40.1%の間で改善できることが分かった。また、この高速化要因をより深く分析するために、キャッシュ効率についても実験し、トライ法による Aho-Corasick の手法の高速化の根拠を実証した。</p> <p>第 5 章では、本研究の成果をまとめるとともに、今後の研究課題について述べる。</p> | | | |

論文審査の結果の要旨

| | | | |
|---|---|-----|-------|
| 報告番号 | 甲 先 第 191 号 | 氏 名 | 北川 浩也 |
| 審査委員 | 主査 獅々堀 正幹 副査 寺田 賢治 副査 泓田 正雄 副査 青江 順一 | | |
| 学位論文題目 文書要約と文字列検索の効率化に関する研究 | | | |
| 審査結果の要旨 <p>文書要約は、文書を解析することで、重要な内容を簡約した情報に変換することであり、要約の量は、アプリケーションに応じて変更される。近年の情報爆発の時代、また携帯端末の普及により、ユーザが意図する情報を効率よくアクセスするための重要な文書処理技術の一つとして、広く利用されている。文書要約手法は多く存在するが、利便性の高い手法の一つとして、教師なし学習手法である NMF(Non-Negative Matrix Factorization)を利用した手法がある。しかし、NMF の計算量は、文書のサイズに比例するため、文書のサイズにより、要約処理速度が低下する欠点があった。従って、本論文の目的の一つは、この NMF による要約技術を携帯端末などで高速処理するための効率化を提案したものである。また、多種多彩なキーワード抽出を必要とする文書要約の汎用的高速化技術の一つとして、複数文字列検索がある。複数文字列検索とは、対象となる文書から、あらかじめ用意した文字列（キーワード）の集合が出現するすべての位置を検出することである。この技術は、上記 NMF 手法を含めた一般的な文書要約だけでなく、文書処理手法の重要な技術の一つである。従って、本論文のもう一つの目的は、この複数文字列検索の効率化手法を提案したものである。本論文は、文書要約と文字列検索の効率化に的を絞り、NMF 手法による文書要約を高速化する手法や複数文字列検索の効率化手法の提案と実験実証結果をまとめたものである。</p> <p>以上、本研究は、意図する情報へ効率よくアクセスするための文書処理技術である教師なしの文書要約、及び文書要約の汎用的高速化技術である複数文字列検索の効率化に関して新しい手法が提案されており、提案手法の有効性を従来手法との比較実験を通して、実証したものである。従って、本論文は博士(工学)の学位授与に値するものと判定する。</p> <p>なお、本論文の審査には、森田和宏講師の協力を得た。</p> | | | |