

電子・情報 工学専攻	学籍番号	983725	指導教員氏名	増山 繁
申請者氏名	酒井 浩之			磯田 定弘

論 文 要 旨(博士)

論文題目	Studies on Automatic Document Summarization Using Statistical Methods (統計的手法を用いた文書自動要約に関する研究)
------	--

(要旨 1,200字程度)

本論文の目的は、文書自動要約に必要な各種の要素技術の開発を行い、その有効性を検証することである。本論文では、まず、複数文書要約においてユーザが知りたい情報を“要約要求”と定義し、ユーザの要約要求を考慮しそれに適合した要約を生成できる複数文書要約手法を提案する。具体的には、要約対象となる、ある事柄に関連した複数文書からその事柄に関連のあるキーワードを抽出しユーザに提示する。ユーザは提示されたキーワードから要約要求に適したキーワードを選択する。その選択されたキーワードによって生成される要約が変化する。提案した要約手法の評価のために、国立情報学研究所主催の、検索と要約の評価のためのワークショップNTCIR4における要約タスク(TSC3)に参加した。その結果、生成された要約の内容評価において良好な成績を得ることができた。また、ユーザのインタラクションによる効果を評価し提案手法の有効性を確認した。

複数文書要約には抽出した重要文を更に圧縮する文圧縮技術が必要になる。本研究では文圧縮を文の不要箇所を削除することで実現する。本論文では削除する不要箇所を対象を、連用修飾節、動詞連体修飾節、多重連体修飾節とした、不要箇所削除による文圧縮手法をそれぞれ提案する。従来の不要箇所の認定においては、人間が作成したヒューリスティックな規則に基づく手法が提案されていたが、本論文で提案する手法はコーパスの統計情報に基づく手法であり、従来手法より高い網羅率および精度を達成できることが特徴である。連用修飾節では、省略できる可能性のある連用修飾節を含む節に対して同一の動詞をもち、かつ、格助詞出現の差異が認められる節をコーパスから検索し、検索された節対から省略可能な連用修飾節を認定する。動詞連体修飾節では、動詞連体修飾節が修飾している名詞に対して“修飾されやすさ”、“修飾多様性”をコーパスから調べ、修飾される頻度が低い、もしくは、修飾する動詞の種類が限定されている名詞を修飾する動詞連体修飾節を省略可能と認定する。同時に、重要な内容を持つ名詞を修飾し、かつ、重要な内容を持たない動詞連体修飾節を削除可能と認定する。多重連体修飾節では、ある1つの名詞を複数の連体修飾節が修飾している場合(多重修飾)に限定し、連体修飾節と修飾している名詞との結合度を、相互情報量に基づいた計算式で重みとして計算する。そして、複数の連体修飾表現の重みの相対的な差で省略可能な修飾表現を認定する。評価の結果、従来手法に比べて良好な精度および再現率を達成でき、各手法の有効性を確認した。

我々の複数文書要約システムにおいて、ユーザに提示されたキーワードはある語の略語である場合がある。略語とその元の語は同一の意味を持っているにもかかわらず別の語として認識されてしまう。略語とその元の語との対応関係を取得できれば、情報検索や文書要約において有用である。そのため、略語とそれに対応した元の語との対応関係をコーパスから自動的に獲得する手法を提案する。提案手法を評価した結果、既提案手法よりも良好な結果を得ることができ手法の有効性を確認した。