

平成19年9月25日

電子・情報工学専攻	学籍番号	039301
申請者氏名	梅村祥之	

指導教員氏名	増山繁 磯田定宏
--------	-------------

論文要旨(博士)

論文題目	人間の文章閲覧・作成支援のための自然言語処理の研究
------	---------------------------

(要旨 1,200字程度)

知識集約産業を支えるナレッジワーカーは、特に、調査、計画段階で、次のような流れで仕事をすることが多くなっている。

インターネット上の情報を幅広く閲覧→ 調査したいテーマが絞られる  
 →そのテーマに沿った情報の検索→ 検索された文章データベースからの情報抽出  
 → 抽出された情報の分析→ 報告資料作成

これらをコンピュータ支援するシステムの要素技術として、本論文では、

- ・ 文章閲覧における情報推薦実現の要素技術： 興味の推定技術
- ・ 情報抽出を高精度化する要素技術： 交通事故原因表現の抽出技術
- ・ 文章推敲支援の要素技術： 情報不足の判定技術

の3つを取り上げる。

**ユーザーの興味の推定：**  
 文章閲覧を支援する情報推薦・情報フィルタリングを実現するためには、ユーザの意図や興味を推定する必要があり、マウス操作、視線などのユーザの挙動から推定する試みがなされていた。本研究は、文章閲覧時のユーザの挙動データ一画面を送るためのスクロール操作の時間情報一から、ユーザの興味を推定する。従来、興味の程度と閲覧時間が相関することから、閲覧時間によって興味を推定していた。しかし、文の難易度などによっても閲覧時間が影響されるため、速度指標のみでは不十分と考え、それを補完する新たな指標として、文の言語情報から標準的な閲覧時間を推定し、それと実測閲覧時間の乖離の程度を指標化した。被験者6名による約300記事に対する興味を推定する実験により手法の妥当性を確認した。

**交通事故原因表現の抽出：**  
 情報抽出の対象として商品購入の動機、商品の評判（企画部門等で利用）、交通事故の原因（自動車の安全装置開発企画部門等で利用）といったような理由や原因を対象とした情報抽出のニーズが増大している。近年、それに対して、評判分析と呼ばれる手法が提案されてきているが、それらの手法では、人手によりルールを作成する必要があり、多大な労力を要する。本研究は、この中の“交通事故の原因”を題材に自動抽出技術を開発する。具体的には、人手で与えた1つの種表現（交通事故原因を表す表現）から出発し、ブートストラップ的に機械抽出を繰り返すことによって、事故原因表現を取得する。今回、新聞記事データからの抽出実験を行い、良好な評価結果を得た。手法開発上の一例として交通事故原因に関する情報抽出を扱ったが、その他、商品購入の動機や商品の評判など上述の様々な応用例に本手法を適用可能である。

**文章推敲支援のための情報不足の判定：**  
 推敲に関する従来研究では、タイプミスや構文構造の複雑さなどを指摘する手法などが行われていた。本研究では、読みやすさを向上させるために、説明が不足していて論理展開が読み取りにくいと感じられる箇所を検出する技術を扱う。文章としては情報を正確に伝達するための仕事用の文を対象として、文単位での情報不足を対象とする。まず、原文から連体修飾部を欠落させた課題文を生成し、次に、被験者がその箇所に情報不足を感じるかどうかを判定して正解判定データを作成した。その後、正解判定データの一部からSVMによる機械学習を行い、残りのデータを機械判定させる。判定の特徴量は、例えば、修飾部の欠落箇所におけるつながりの滑らかさに関係した語の連鎖に関する統計量である。4名の被験者による約1,000箇所の判定課題の実験により手法の妥当性を確認した。