

申請者	学科名	情報システム工学科	職名	教授	氏名	菊井 玄一郎
調査研究課題	地域行政に関するテキスト情報の構造化およびアクセス支援技術の研究					
調査研究組織	氏名	所属・職		専門分野	役割分担	
	代表	菊井玄一郎		情報システム工学科・教授	知識情報処理	研究統括・計画立案, テキストマイニング手法の検討
	分担者	見船 文香	大学院・システム工学専攻(前期), 2年	テキスト話題分析	地方議会会議録に対する話題分析手法の検討	
		福圓 琢真	大学院・システム工学専攻(前期), 1年	方言の形態素解析	地方議会会議録に対する形態素解析手法の検討	
		宇田 一平	同上	地理情報システム, ウェブ応用	ウェブを用いた情報の可視化とアクセスインターフェースの検討	
調査研究実績の概要	<p>地域の課題を解決し, 有効な計画を立案するためには地域の状況や課題に関する知識(データ)を知らなければならない。我々は, 地域情報のうち大量に蓄積されながら活用が進んでいないテキスト情報(文字情報)の活用を支援する技術の研究を進めている。本年度の調査研究では, 地域における課題や解決策などを多く含んでいる地方議会会議録を対象に構造化, および, インデクシングの検討を行った。</p> <p><b>1) データの整備</b></p> <ul style="list-style-type: none"> <li>・地域行政に関するテキストデータとして地方議会会議録のテキストデータを公開ページからダウンロードし計算機で処理可能な形式に変換した(テキスト約2000万文字分)</li> <li>・方言のコーパスを整備した</li> </ul> <p><b>2) 「通告者一覧表」を用いた地方議会会議録の構造化</b></p> <ul style="list-style-type: none"> <li>・地方議会会議録は電子化されてウェブ上に公開されており, 全文検索(文字列検索)は可能なものの, それ以外は会議日と発言者でしか索引づけられておらず内容からの一覧は困難であった。この問題に対して, 本調査研究では「通告者一覧表」が議事録の内容に関するタイトル(概要)になっていることに着目し, 通告者一覧表の最も細かい項目(中項目)と会議録本文の該当箇所とを日本語処理技術によって自動的に対応付けする手法を提案した。なお, この成果は分担者(見船)により電気情報関連学会中国支部大会において発表され, 奨励賞が与えられた(資料目録[1])。</li> </ul> <p>このことにより, 議事録本文に対する「内容による目次」が可能になる。今後はこの手法を組み込んだ地方議会会議録の閲覧システムの作成を行う予定である。</p>					

地域貢献への反映を踏まえて記述のこと

中項目と本文とを自動的に対応付け

発言順位	議席番号	氏名	大項目	中項目・質問要旨	答弁者
4	17	[ ]	1 文化振興行政について	(1) 高木聖顕先生の文化勲章受章を 受け、本市の今後の取組はどうか。 ①市民がこぞって参加できるような 祝賀式長などの計画はあるか。 ②記念館の建設、若しくは増設 示場の設置が必要と考えるがどうか。	市長 教育長
			2 教育行政について	(1) 市中心部の大規模校について ①新設を含め再編成の時期と考 えるがどうか。 ②幼稚園についてはどうか。 ③小学校についてはどうか。 ④中学校についてはどうか。	教育長

通告者一覧表

議長：△△議員  
△△議員：〇〇〇〇〇〇〇〇  
議長：XXXXXXXXXX  
市長：〇〇〇〇〇〇〇〇〇〇〇〇  
.....

会議録本文

調査研究実績の概要

地域貢献への反映を踏まえて記述のこと

2) 岡山弁向けの形態素解析処理の開発

日本語の文章は単語の間に空白をいれずに記述するため、単語の境界が分からない。一方、文章に対して情報検索や索引付け、情報抽出を行うためには日本語文を単語に切り離して品詞や読みなどの属性を付与する「形態素解析処理」は必須の基盤技術である。日本語の標準語に対しては高精度の解析処理ソフトが利用可能であるが、方言には対応していない。現在主流となっている日本語形態素解析手法はあらかじめ正しく形態素解析された大規模なテキストデータから統計的な学習によって処理するものであるが、方言については話し言葉ということもあり、大規模なテキストデータが存在せず、この方法は使えない。そこで、本調査研究では、岡山弁のテキストデータの収集と並行して、機械学習に依らない方法を検討した。その方法は、標準語と岡山弁の表現の言語表現の間にある程度の対応関係があることに着目して、標準語の形態素解析処理のパラメータを岡山弁のそれに自動的に変換するというものである。その結果、解析精度(F値)89%を達成した。

なお、この成果は研究分担者(福圓)によって松山市で開催された情報科学技術フォーラム(FIT2015)で発表され、奨励賞が与えられた(資料目録[2])。

上述の成果は岡山弁の文章に対する情報検索やテキストマイニング等には利用可能な精度と考えられるが、文章の理解に用いるにはまだ不十分である。今後は機械学習の手法を取り入れつつ精度向上を図ることが必要である。

成果資料目録

- [1] 見船文香, 但馬康宏, 菊井玄一郎: “目次との対応付けによる地方議会会議録の話題分割”, 平成27年度電気・情報関連学会中国支部連合大会, 発表番号25-30, (2015) (平成27年度 情報処理学会中国支部 奨励賞)
- [2] 福圓琢真, 但馬康宏, 菊井玄一郎: “標準語用のパラメータを流用した岡山弁向け統計的形態素解析”, 第14回情報科学技術フォーラム(FIT), E-007, Sep., 2015. (FIT奨励賞)