

平成20年度図書館情報メディア研究科プロジェクト研究 研究成果報告書

種 目	プロジェクト研究	萌芽研究	研究代表者 氏 名	藤井 敦
研究課題	種々のコンテキストを考慮した事典コンテンツの検索システム			
研究組織（研究代表者及び研究分担者）				
氏 名	所属研究機関・部 局・職	現在の専門	役割分担	
藤井 敦	筑波大学・大学院図 書館情報メディア研 究科・准教授	自然言語処理, 情報検索	研究代表者・研究統括	
研究目的				
<p>研究代表者は、大量のオンライン情報をあたかも「巨大な百科事典」のように検索して種々の調査を支援するシステムの実現を目的とし、10年近く研究と開発を進めてきた。当該萌芽研究は、前述の目的に対して、過去の蓄積に基づいて手法やシステムをさらに高度化することを目的とした。具体的には、World Wide Web、放送メディア、特許情報として配信される情報を再構成し、多種多様な用語に関する解説型のコンテンツを構築することと、当該コンテンツを検索するシステムを開発している。当該萌芽研究は、a)コンテンツ構築の高度化と、b)ユーザの検索意図に応じた検索手法を提案した。</p>				
研究成果				
<p>上記「a)コンテンツ構築の高度化」について、Web上のフリー百科事典である「Wikipedia」を解析することで、「人々が用語説明を編集するモデル」を自動的に学習し、当該モデルを用いてWebや特許の不均一な情報を組織化する手法を提案した。具体的には、「動物」や「病気」といった用語の種類ごとに、説明に必要な観点の構造（動物なら「生態」や「形態」、病気なら「診断」や「治療」など）および説明に使われる単語の頻度分布をWikipediaから自動的に学習し、用語説明に関するモデルを構築した。その結果、Wikipediaに収録されていない用語も含めて、「Webに存在する一般的な用語」から「特許情報にしか含まれない高度な専門用語」まで多種多様な用語を対象として、多面的な説明情報を生成することが可能になった。</p> <p>上記「b)ユーザの検索意図に応じた検索手法」について、ユーザの検索質問を自動分類し、分類結果に基づいて検索手法を動的に変更する情報検索モデルを提案した。その結果、単一の検索手法しか用いない場合に比べて、検索精度を有意に向上させることに成功した。</p>				
代表的な研究発表・特許等の成果一覧、特記事項等				
<ul style="list-style-type: none"> <li>・ <a href="#">Atsushi Fujii</a>. Modeling Anchor Text and Classifying Queries to Enhance Web Document Retrieval. Proceedings of the 17th International World Wide Web Conference, pp.337-346, 2008.</li> <li>・ <a href="#">Atsushi Fujii</a>. Producing an Encyclopedic Dictionary Using Patent Documents. Proceedings of the 6th International Conference on Language Resources and Evaluation. 2008.</li> <li>・ <a href="#">藤井 敦</a>, 三條場 旭彦. Wikipedia を用いた用語説明のモデル化と事典的検索への応用. 人工知能学会 第20回セマンティックウェブとオントロジー研究会資料, SIG-SWO-A803-01, 2009.</li> </ul>				