

平成21年度科学研究費補助金実績報告書（研究実績報告書）

1. 機関番号 1 4 6 0 3 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 特定領域研究 4. 研究期間 平成21年度～平成22年度
5. 課題番号 2 1 0 1 3 0 3 5
6. 研究課題名 情報爆発時代の構造化文書の超高精度検索とその実現
7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
4 0 2 9 3 3 9 4	フガナ ミヤザキ ジュン 宮崎 純	情報科学研究科	准教授

8. 研究分担者（所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。）

研究者番号	研究分担者名	所属研究機関名・部局名	職名
0 0 2 0 3 4 8 0	フガナ ウエムラ シュンスケ 植村 俊亮	奈良産業大学・情報学部	教授
8 0 3 1 4 5 3 2	フガナ ハタノ ケンジ 波多野 賢治	同志社大学・文化情報学部	准教授
	フガナ		
	フガナ		
	フガナ		

9. 研究実績の概要

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

Web文書に代表される電子文書は近年急激に増加しており、その中から有用な情報を取得することが重要となってきた。特に文書のサイズが大きくなると、ユーザの情報要求に適合する箇所を文書中から発見することも困難となっている。

本研究では、電子文書の中でもXML文書に焦点を絞り、ユーザの情報要求に対して文書中で最も適合する一部分、すなわち部分文書の検索手法の高精度化について研究を行った。また、XML文書进行处理するための問合せ言語XQueryの高性能化についても研究を行った。より具体的には、文書中における適合箇所と文書構造の関係を利用して、より良い検索結果を出力するための文書スコアリング方式を提案し評価を行った。また、複数の適合箇所が存在する場合には、それらを再構成してより適合度の高い検索結果を出力するための手法の提案を行った。前者に関しては、部分文書に含まれる検索キーワード数の割合を考慮し、従来の文書スコアリングと組み合わせることにより、検索精度が向上することが分かった。後者に関しては、隣接する複数の適合部分文書に関して、検索結果のサイズが大きくなり過ぎないように制限しつつ、同レベルの位置にある適合部分文書どうしは結合、包含関係にある適合部分文書はそれらを上位の部分文書に併合することにより、既存の手法と比較して40%以上の検索精度の改善ができることを明らかにした。さらに、文書サイズと検索結果となり得る部分文書の最大値の関係を実験的に明らかにした。

10. キーワード

(1) 情報抽出 (2) 情報システム (3) ディレクトリ・情報検索

(4) 構造化文書 (5) (6)

(7) (8) (裏面に続く)

11.研究発表（平成21年度の研究成果）

〔雑誌論文〕 計（ 2 ）件 うち査読付論文 計（ 2 ）件

著者名	論文標題			
油井誠,宮崎純,植村俊亮,加藤博一,山名早人	ロックフリーGCLOCKページ置換アルゴリズム			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
情報処理学会論文誌データベース	有	Vol.2, No.4	2 0 0 9	32-48

著者名	論文標題			
Fuminori Kimura, Akira Maeda, Kenji Hatano, Jun Miyazaki, Shunsuke Uemura	Utilizing Web Directories For Translation Disambiguation In Cross-Language Information Retrieval			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Trends in Communication Technologies and Engineering Science Series (LNEE)	有	Vol.33	2 0 0 9	95-107

著者名	論文標題			
雑誌名	査読の有無	巻	発行年	最初と最後の頁

〔学会発表〕 計（ 7 ）件 うち招待講演 計（ 0 ）件

発表者名	発表標題	
Makoto Yui, Jun Miyazaki, Shunsuke Uemura, Hayato Yamana	Nb-GCLOCK: A Non-blocking Buffer Management based on the Generalized CLOCK	
学会等名	発表年月日	発表場所
IEEE International Conference on Data Engineering (ICDE 2010)	2010年3月4日	アメリカ・ロングビーチ

発表者名	発表標題	
Atsushi Keyaki, Jun Miyazaki, Kenji Hatano	A Method of Generating Answer XML Fragment from Ranked Results	
学会等名	発表年月日	発表場所
INEX 2009 Workshop	2009年12月7日	オーストラリア・ブリズベン

発表者名	発表標題	
Atsushi Keyaki, Kenji Hatano, Jun Miyazaki	A Scoring Method of XML Fragments Considering Query-Oriented Statistics	
学会等名	発表年月日	発表場所
International Conference on the Applications of Digital Information and Web Technologies	2009年8月5日	イギリス・ロンドン

発表者名	発表標題	
樺惇志, 波多野賢治, 宮崎純	XML 検索技術を利用した検索結果の構成手法	
学会等名	発表年月日	発表場所
情報処理学会第72回全国大会	2010年3月11日	東京

発表者名	発表標題	
樺惇志, 波多野賢治, 宮崎純	XML 部分文書の再構成に基づく検索結果の提示手法	
学会等名	発表年月日	発表場所
第2回データ工学と情報マネジメントに関するフォーラム(DEIM 2010)	2010年3月1日	淡路島

発表者名	発表標題	
宮崎純	多様なコンテンツを活用するための高性能XMLデータ処理	
学会等名	発表年月日	発表場所
第2回ユニバーサルコミュニケーション研究会	2009年11月13日	奈良

発表者名	発表標題	
樺惇志, 波多野賢治, 宮崎純	索引語の統計量を用いたXML部分文書検索法の組合せ利用とその効果	
学会等名	発表年月日	発表場所
情報処理学会データベースシステム研究会	2009年7月28日	神戸

〔図書〕 計 (0) 件

著者名	出版社		
書名	発行年	総ページ数	
	■ ■ ■		

12. 研究成果による産業財産権の出願・取得状況

〔出願〕 計 (0) 件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

〔取得〕 計 (0) 件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別

13. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関するwebページがある場合は、URLを記載すること。

--