

平成22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	ミヤザキ 宮崎 ジュン 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	ウエムラ 植村 シュンスケ 俊亮	奈良産業大学・情報学部	教授
8 0 3 1 4 5 3 2	ハタノ 波多野 ケンジ 賢治	同志社大学・文化情報学部	准教授

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、特に電子文書の中でも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てきた。

既存の手法では、各部分文書ごとのスコア値に基づいて、高いスコア値の部分文書のみをランキングしてユーザに提示していたが、高いスコア値を持つ部分文書の周辺の適合部分文書を取りこぼしてしまい、その結果検索精度を落とす原因となっていた。そこで、高いスコア値を持つ部分文書の周辺を探索し、ある程度適合すると判断される周辺の部分文書を統合するとともに、統合した結果、サイズが変更された部分文書のスコア値を調整する手法を提案した。この提案手法を用いて、国際ワークショップINEX2010のコンペティションに参加したところ、世界二位の検索精度を達成し、提案手法の高い有効性を実証することができた。

引き続きXML部分文書検索に関して、検索精度を落とさず検索処理効率を上げるために、XML文書の経路式に基づく索引語の大域重み付けの既存手法であるIPFを改良し、類似した経路式のある規則によりまとめることにより、識別すべき部分文書のクラス数を減らす手法を提案した。その中で、ISFと呼ぶ提案手法はクラス数を75%削減したにも関わらず検索精度はIPFとほぼ同じであることが判明した。これにより、ISFが効率の良い高精度検索処理に有効であることを明らかにした。

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) XML      (6)
- (7)      (8)      (裏面に続く)

11.研究発表（平成22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			
櫛惇志, 波多野賢治, 宮崎純	有益な検索結果提示のための部分文書再構成手法の提案			
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.4, No.1	2   0   1   1	1-13

著者名	論文標題			
Atsushi Keyaki, Kenji Hatano, Jun Miyazaki	A Query-oriented XML Fragment Search Approach on A Relational Database System			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Journal of Digital Information Management	有	Vol.8, Issue 3	2   0   1   0	175-180

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（ 6 ）件      うち招待講演 計（ 0 ）件

発表者名	発表標題	
Atsushi Keyaki, Kenji Hatano, Jun Miyazaki	Expansion of Result Reconstruction for Effective XML Fragment Search at INEX 2010	
学会等名	発表年月日	発表場所
INEX 2010 Workshop	2010年12月13日	オランダ、ヴグフト

発表者名	発表標題	
Atsushi Keyaki, Kenji Hatano, Jun Miyazaki	A Result Reconstruction Method for Effective XML Search	
学会等名	発表年月日	発表場所
12th International Conference on Information Integration and Web-based Applications & Services	2010年11月9日	フランス、パリ

発表者名	発表標題	
三宅香菜子, 宮崎純, 天野敏之, 藤澤誠, 加藤博一	複数の検索要素を統合する3Dオブジェクト検索システム	
学会等名	発表年月日	発表場所
第3回データ工学と情報マネジメントに関するフォーラム(DEIM2011)	2011年2月28日	修善寺

発表者名	発表標題	
櫛惇志, 波多野賢治, 宮崎純	XML 部分文書検索における緩和大域的重み付け手法の提案	
学会等名	発表年月日	発表場所
第3回データ工学と情報マネジメントに関するフォーラム(DEIM2011)	2011年2月28日	修善寺

発表者名	発表標題	
櫛惇志, 波多野賢治, 宮崎純	再構成された XML 部分文書に対再構成された XML 部分文書に対するランキング手法の提案	
学会等名	発表年月日	発表場所
電子情報通信学会WI2研究会	2010年9月16日	佐渡

発表者名	発表標題		
三宅香菜子, 宮崎純, 藤澤誠, 加藤博一	付加情報を利用した3Dオブジェクト検索システム		
学会等名	発表年月日	発表場所	
平成22年度情報処理学会関西支部大会	2010年9月22日	大阪市	

〔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。

--