

平成25年度科学研究費助成事業（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 特別研究員奨励費 4. 研究期間 平成24年度～平成26年度

5. 課題番号

	2	4	・	7	8	0
--	---	---	---	---	---	---

6. 研究課題名 高精度かつ高速な XML 検索に関する研究

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
ケヤキ 櫻	アツシ 惇志	情報科学研究科	特別研究員 (DC2)

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

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。

本研究で実用的な XML 検索システムを作成するにあたって、情報検索システムの要件である高精度・高速な検索を満たしつつ、Web 検索システムにおいて極めて頻繁に発生する文書の更新に対しても現実的なコストで索引の差分更新を行うことで実現させる。XML 検索システムでは従来の文書検索システムのように検索結果として文書全体を提示するのではなく、文書のうちのユーザが求める部分のみ提示することを目指す。従って、実用的な XML 検索システムを開発することができれば、検索システムを用いるユーザは検索に必要な時間を短縮させ、求める情報が文書中のどの部分に記述されているのかをユーザ自ら探す必要がなくなる。更に、近年モバイル端末やタブレット型 PC からの情報検索の機会が増加しており、必要十分な情報のみを提示する XML 情報検索はディスプレイサイズの小さな環境においても有効な検索技術となるため、今後非常に重要な技術になることが期待される。

前年度の研究成果によって、従来の高精度・高速な XML 情報検索システムと比較して遜色のない性能を維持しつつ、高速に文書の更新を反映させることに成功した。今年度は更に発展させて、検索精度においても従来のシステムを上回ることを目指した。その際、過去に提案した、語の重要度のみに着目するのではなく、各部分、及び、関連する部分の持つ情報も考慮した上でより適切な検索結果を提示することを目指す手法と組み合わせることで、従来のシステム以上の高精度を実現した。

また、XML 部分文書検索技術の Web 検索への拡張を目指し、予め Web 文書の持つ物理構造を論理構造に沿うように再構造化を行い、また、情報要求を満たさない重要でないコンテンツを除去するためのフィルタを適用することで、Web 文書に対しても XML 部分文書検索技術は有効であることを示した。

10. キーワード

- | | | | |
|---------|------------|----------|----------|
| (1) XML | (2) 情報検索 | (3) 文書更新 | (4) 差分更新 |
| (5) 統計量 | (6) Web 文書 | (7) | |

(注)・印刷に当たっては、A4判（縦長）・両面印刷し、左端を糊付けすること。

(8)

11. 現在までの達成度

下欄には、交付申請書に記載した「研究の目的」の達成度について、以下の区分により自己点検による評価を行い、その理由を簡潔に記述すること。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。
<区分>①当初の計画以上に進展している。 ②おおむね順調に進展している。 ③やや遅れている。 ④遅れている。

(区分) ②
(理由) 当初の予定通り、極めて短時間で更新に対応しつつ、高精度かつ高速な XML 検索を実現させた。その際、システムは大規模データに対してもスケールアウトを行うように開発を行った。また、XML 検索技術の実用化の一つの手段である Web 検索システムへの拡張を行うため、XML 検索技術適用のための Web 文書の再構造化手法を提案した。

12. 今後の研究の推進方策

本研究課題の今後の推進方策について簡潔に記述すること。研究計画の変更あるいは研究を遂行する上での問題点があれば、その対応策なども記述すること。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。

開発したシステムが実際にスケールアウトするのかどうか、システム構築を行って評価する。その際、効率的に CPU と GPU のタスクを切り分けることによって、より高性能な並列・分散処理が実現すると考えられる。

13. 研究発表（平成25年度の研究成果）

※ 「13. 研究発表」欄及び「14. 研究成果による産業財産権の出願・取得状況」欄において記入欄が不足する場合には、適宜記入欄を挿入し、それによりページ数が増加した場合は、左端を糊付けすること。

【雑誌論文】 計（ 2 ）件 うち査読付論文 計（1）件

著 者 名		論 文 標 題					
樺惇志, 宮崎純, 波多野賢治, 山本豪志朗, 武富貴史, 加藤博一		文書の更新を考慮した高精度 XML 部分文書検索手法の提案					
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.6, No.4	2	0	1	3	1~16
掲載論文の DOI (デジタルオブジェクト識別子)							
なし							

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
掲載論文の DOI (デジタルオブジェクト識別子)						

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
掲載論文の DOI (デジタルオブジェクト識別子)						

(注)・印刷に当たっては、A4判（縦長）・両面印刷し、左端を糊付けすること。

〔学会発表〕計（ 4 ）件 うち招待講演 計（ 0 ）件

発表者名	発表標題		
Atsushi Keyaki	XML Element Retrieval@1CLICK-2		
学会等名	発表年月日	発表場所	
NTCIR-10	2013年6月18日～ 2013年6月21日	国立情報学研究所（東京都千代田区）	

発表者名	発表標題		
Atsushi Keyaki	Update-Aware Accurate XML Element Retrieval		
学会等名	発表年月日	発表場所	
第5回 データベース、Web、情報マネジメントに関する若手研究者国際ワークショップ (iDB Workshop 2013)	2013年7月21日～ 2013年7月23日	北海道大学（北海道札幌市）	

発表者名	発表標題		
Atsushi Keyaki	A Path expression-Based Smoothing of Query Likelihood Model for XML Element Retrieval		
学会等名	発表年月日	発表場所	
1st ACIS International Symposium on Applied Computing and Information Technology (IIAI-AAI ACIT 2013)	2013年8月31日～ 2013年9月4日	くにびきメッセ（島根県松江市）	

発表者名	発表標題		
樽惇志	XML 部分文書検索技術の Web 文書への適用		
学会等名	発表年月日	発表場所	
第 6 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 2014)	2014年3月3日～ 2014年3月5日	淡路夢舞台（兵庫県淡路市）	

〔図 書〕 計（ 0 ）件

著者名	出版社		
書名	発行年	総ページ数	

14. 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
				出願年月日	

15. 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。

--