

平成24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25年度
5. 課題番号 2 4 0 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 アツシ 櫻 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で、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、求める情報がどの部分に記述されているのかをユーザ自ら探す労力を軽減する。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、今後重要な技術になることが期待される。

従来のXML情報検索システムでは文書の更新が発生することを想定していないため、本研究では従来のXML情報検索システムを拡張させて文書の更新機能を付与する。その際、文書の更新の発生に応じて高速に差分更新を可能な索引構造を提案し、文書の更新、すなわち文書の追加・削除・書換時における処理を定義した。

また、更新が発生すれば可能な限り高速に更新を反映させるため、検索結果を提示する上で重要なデータのみを索引へ登録するためのフィルタを提案した。更に、検索システム運用初期のような、検索システムに蓄積されている文書数が少量の時点では、文書のスコアリングに用いる統計量のうち文書集合全体から算出される大域的重みは正確な統計量を得ることができない。従って、少数の文書からも正確な大域的重みを推定する手法を提案した。

これらにより、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。

10. キーワード

- (1) XML (2) 情報検索 (3) 文書更新 (4) 差分更新
- (5) 大域的重み (6) INEX (7) (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 おおむね順調に進展している。
 (理由) 当初の研究計画の通り、差分更新システムのためのスキーマの考案と差分更新の高速化のためのフィルタの提案を行い、評価実験においてそれらの有効性を示すことができたため。

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。

今後の研究方針として、当初の研究計画の通り、更新処理の並列化及び分散化を行う。なお、その際、CPU と GPU の両方の演算装置を併用することで高速化を目指す。

13. 研究発表（平成24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 (1) 件 うち査読付論文 計 (1) 件

著者名	論文標題				
Atsushi Keyaki	Fast Incremental Indexing with Effective and Efficient Searching in XML Element Retrieval				
雑誌名	査読の有無	巻	発行年	最初と最後の頁	
International Journal of Web Information Systems (IJWIS)	有	Vol 9, Iss 1	2013	in press	
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)					

〔学会発表〕 計 (3) 件 うち招待講演 計 (0) 件

発表者名	発表標題		
櫻 惇志	更新を考慮した XML 部分文書検索システムの精度の改善		
学会等名	発表年月日	発表場所	
第5回データ工学と情報マネジメントに関するフォーラム 2013 (DEIM 2013)	2013/3/3	ホテル華の湯(福島)	

発表者名	発表標題	
Atsushi Keyaki	Fast and Incremental Indexing in Effective and Efficient XML Element Retrieval Systems	
学会等名	発表年月日	発表場所
14th International Conference on Information Integration and Web-based Applications & Services (iiWAS2012)	2012/12/5	Bali, Indonesia

発表者名	発表標題	
榎 惇志	XML 部分文書検索における索引の高速な差分更新と高精度検索	
学会等名	発表年月日	発表場所
Web とデータベースに関するフォーラム (WebDB Forum 2012)	2012/11/20	秋葉原コンベンションホール(東京)

〔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。

--