

様 式 F - 7 - 1

科学研究費助成事業（学術研究助成基金助成金）実施状況報告書（研究実施状況報告書）（平成 23 年度）

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 基盤研究(C) 4. 補助事業期間 平成 23 年度～平成 25 年度

5. 課題番号

2	3	5	0	0	1	2	1
---	---	---	---	---	---	---	---

6. 研究課題 メニーコアプロセッサ時代における構造化文書の高精度かつ高速検索の実現

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 ジュン	情報科学研究科	准教授
	宮崎 純		

8. 研究分担者

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

XML文書検索の研究は、これまで検索の高精度化、すなわち検索結果中に含まれる適合文書の割合をいかに高めるか、にのみ重点がおかれ、従来のマクロな文書検索と比べて、部分文書検索は根本的に計算コストが大きいにも関わらず、処理時間についてはあまり考慮されてこなかった。更に、高精度検索の手法が、単語の高度な重み付けによる方法や、確率的言語モデルを利用しており、文書の更新を全く考慮していないあるいは更新に弱いという問題があった。

この問題に対して、文書の動的な更新を念頭に置き、高速な更新操作、高速な問合せを可能としつつ検索精度を保つために、索引のスキーマと更新コストを低減するための二つのフィルタを提案した。67万XML文書からなるテストコレクションを利用して評価実験を行った結果、提案索引スキーマを利用することで新たに追加された文書へも適切に重み付け可能であり、提案するフィルタを利用することで検索精度を維持しつつ索引構築時間を4割程度、索引サイズを1割程度削減可能であることを明らかにした。

一方で、マルチコアプロセッサ向けの効率のよい汎用データ構造の維持管理手法の提案も行った。このデータ構造はマルチコアプロセッサを有効に動作させるために必要な要件の一つであるキャッシュのヒット率を高めるために、繰り返し使用されるリスト構造について、計算の進行に従い動的にデータ構造を再構築してキャッシュヒット率を落とさないようにするものである。この提案データ構造を、試験的にあるデータマイニング手法に適用したところ、処理速度が25%以上向上することを明らかにした。

10. キーワード

- | | | | |
|----------------|-----------|------------|-----------|
| (1) 情報検索 | (2) XML | (3) 部分文書検索 | (4) 高精度検索 |
| (5) マルチコアプロセッサ | (6) 高性能計算 | (7) | (8) |

11. 現在までの達成度

(区分)(2) おおむね順調に進展している。

(理由)

本研究の目標である、文書の更新を考慮しつつ、高速かつ高精度のXML部分文書検索を可能とする課題に関して、検索精度を維持しつつ、top-k問合せを可能とするためtag-term索引とpath索引の二つを利用する。しかし、更新処理に関係する索引の維持コストを下げるために、検索結果となりうる確率の極めて低い索引語や複雑なパスの部分文書を予めフィルタで取り除いている。従来更新を考慮しつつ、高精度、高速性を同時に満たす検索の研究はほとんど行われていなかったにも関わらず、本研究結果は、先駆的アイデアとその良好な結果が得られたことを鑑みると、本研究の初年度の達成度としては十分であると考えられる。

一方、マルチコアあるいはメニーコアプロセッサを使用する際に問題となる、共有データのデータ構造の効率の良い維持管理について、キャッシュヒット率を維持し処理速度を上げるために、本来の計算とは関係のない、データ構造の再構築処理にコストをかけても結果的に処理速度を向上させることが可能であることを示した点で、今後のXML部分文書検索の処理性能向上に利用できる可能性を示した点において、研究期間を考慮すると順調な進捗といえる。

12. 今後の研究の推進方策 等

(今後の推進方策)

文書の更新を考慮しつつ、高速かつ高精度のXML部分文書検索を可能とするための二つの索引と二種類のフィルタに関して、文書の削除や文書中の一部の更新処理の評価について今後行っていく予定である。また、この提案手法は基本的なアルゴリズムとデータ構造の提案にとどまっており、マルチコア、メニーコアプロセッサを利用する並列アルゴリズムとはなっていないため、処理手順や全体に占める計算コストの割合を考慮して、適切な並列アルゴリズムの設計を進めていく予定である。しかしながら、XML部分文書検索のための索引処理は、多くの部分で大きなデータ並列性を持つ問題であり、並列アルゴリズムの設計は今後の研究課題であるものの、提案したマルチコアプロセッサ向けの効率の良い維持管理手法を組み合わせ、より良い処理方式を提案していく予定である。その際にGPUの利用を考慮することで、より効率の高いアルゴリズムの設計を行っていく方針である。

(次年度の研究費の使用計画)

未使用額が生じた要因は、研究の進捗状況に合わせ、予算執行計画を変更したことに伴うものである。また、次年度の請求額と合わせた執行計画は以下の通りである。

マルチコア、メニーコアプロセッサでのXML部分文書検索の処理を考慮しているため、最新の8コアプロセッサ搭載、さらに最新のGPUが搭載されている実験用の計算機をそれぞれ購入し、並列アルゴリズムの設計、検証、評価を行う予定である。さらに、昨年度中に有効性を明らかにした未発表の研究成果ならびに次年度に得られた研究成果に関する対外発表と、論文出版を中心に研究費を利用する予定である。

13.研究発表(平成23年度の研究成果)

〔雑誌論文〕計(3)件 うち査読付論文 計(3)件

著者名	論文標題			
Atsushi Keyaki, Kenji Hatano, Jun Miyazaki	Result Reconstruction Approach for More Effective XML Element Search			
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	有	Vol.7	2 0 1 1	360-380
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			
松原裕貴, 宮崎純, 藤澤誠, 天野敏之, 加藤博一	CC-PAID: CPUキャッシュを有効利用した並列時系列パターンマイニングアルゴリズム			
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.4	2 0 1 1	88-100
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			
Atsushi Keyaki, Kenji Hatano, Jun Miyazaki	Relaxed Global Term Weights for XML Element Search			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Comparative Evaluation of Focused Retrieval, LNCS	有	Vol.6932	2 0 1 1	71-81
掲載論文のDOI(デジタルオブジェクト識別子)				

〔学会発表〕計(8)件 うち招待講演 計(1)件

発表者名		発表標題	
Yuki Matsubara, Jun Miyazaki, Goshiro Yamamoto, Yuki Uranishi, Sei Ikeda, Hirokazu Kato		CCDR-PAID: More Efficient Cache-Conscious PAID Algorithm by Data Reconstruction	
学会等名		発表年月日	発表場所
27th ACM Symposium On Applied Computing (SAC2012)		2012.3.27	Riva del Garda, Italy

発表者名		発表標題	
宮崎純, 鬼塚真		データクラウドを支える技術と研究動向	
学会等名		発表年月日	発表場所
第10回 先端的データベースとWeb技術動向講演会(招待講演)		2011.6.11	東京工業大学

発表者名		発表標題	
櫻 惇志, 宮崎 純, 波多野 賢治, 山本 豪志朗, 加藤 博一		XML 索引の更新コスト削減のための部分文書の統計量に基づくフィルタの評価とその最適化	
学会等名		発表年月日	発表場所
第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012)		2012.3.3	シーサイドホテル舞子ピラ神戸

発表者名		発表標題	
吉武 亮, 宮崎 純, 山本 豪志朗, 加藤 博一		スカイラインの近傍探索を可能とする拡張スカイライン演算の実装と評価	
学会等名		発表年月日	発表場所
第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012)		2012.3.3	シーサイドホテル舞子ピラ神戸

発表者名	発表標題	
福澤優, 宮崎純, 山本豪志朗, 加藤博一	カラムストアとローストアを利用したOLAP問合せ処理における消費電力と処理速度の関係について	
学会等名	発表年月日	発表場所
第4回データ工学と情報マネジメントに関するフォーラム (DEIM2012)	2012.3.3	シーサイドホテル舞子ピラ神戸

発表者名	発表標題	
櫻 惇志, 宮崎 純, 波多野 賢治, 山本 豪志朗, 加藤 博一	XML情報検索のための動的な索引管理手法の一提案	
学会等名	発表年月日	発表場所
情報処理学会研究報告データベースシステム	2011.11.3	工学院大学

発表者名	発表標題	
松原 裕貴, 宮崎 純, 山本 豪志朗, 浦西 友樹, 池田 聖, 加藤 博一	データアクセスの改良による時系列パターンマイニングアルゴリズムの高速化	
学会等名	発表年月日	発表場所
情報処理学会研究報告データベースシステム	2011.11.3	工学院大学

発表者名	発表標題	
吉武亮, 宮崎純, 藤澤誠, 天野敏之, 加藤博一	情報推薦のためのSkyline演算の拡張	
学会等名	発表年月日	発表場所
情報処理学会関西支部支部大会	2011.9.22	大阪大学中之島センター

〔図書〕計(0)件

著者名	出版社			
書名			発行年	総ページ数
			年 月 日	

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

--