

**科学研究費助成事業 研究成果報告書**

平成 27 年 5 月 15 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24700019

研究課題名(和文) プロセッサ数に比例する並列速度向上を保証するXML処理言語の研究

研究課題名(英文) Study for Parallel XML Processing Language that Guarantees Parallel Speedups

## 研究代表者

森畑 明昌 (Moriyata, Akimasa)

東京大学・総合文化研究科・講師

研究者番号：10582257

交付決定額(研究期間全体)：(直接経費) 2,400,000円

研究成果の概要(和文)：XMLを並列に処理するプログラミング言語に向けての研究を行った。特に木構造を並列処理する際の困難を解決するための基礎理論について成果が得られた。また、これに加え、関数型プログラムを分割統治並列処理する場合の一般的な理論や、本研究でのアプローチがコンパイラにおけるデータフロー解析にも有用であるとの洞察が得られた。以上の結果は、XML処理のみならず、並列プログラミングを考える上での基礎技術となり得る。

研究成果の概要(英文)：We studied for developing a parallel XML processing language. We solved some difficulties in parallel processing of tree structures including XMLs. Moreover, we found a theory for divide-and-conquer parallelization of functional programs and showed a preliminary result that our approach is also useful for parallel data-flow analyses in compilers. These results will be useful not only for parallel XML processing but also for further studying parallel programming in general.

研究分野：プログラミング言語

キーワード：並列処理 XML 関数プログラミング

### 1. 研究開始当初の背景

近年の計算機は、集密度の物理的な限界や消費電力の低減のため、家庭用 PC でさえ複数のコアを持つのが当たり前になっている。そのため、並列計算は、科学技術計算等の専門的な用途に限らず、一般のユーザにとっても必須となりつつある。

しかし、効率の良い並列計算、特に良い台数効果（プロセッサ数に対する並列速度向上）の達成は、現状ほとんどのユーザの手に余る。逐次プログラムの多少の改良では、アムダールの法則としても知られているように、数台以上のプロセッサではほとんど速度向上が得られない。一方、台数効果に優れたアルゴリズムの設計は、逐次計算の場合とは全く異なる着想を要し、専門家以外にはきわめて難しい。

PC が多数のコアを持ち並列計算が一般的になった暁には、台数効果が効率の近似となる。よって台数効果に優れた並列プログラムの構成をサポートする手法はきわめて重要な研究課題である。

### 2. 研究の目的

本研究の目的は、最適な台数効果が保証された XML 変換言語を設計することである。この言語を利用すれば、ユーザは、行いたい変換を記述するだけで、最適の、すなわち漸近的にプロセッサ数に比例する速度向上を得ることができる。

本研究では具体的には以下の 2 点を明らかにする。

- XML 変換のコアをなす計算パターン（例えば XPath クエリ、木の構造変換など）について、台数効果に優れた並列評価方法を与える。さらにその評価方法が最適な台数効果を示す範囲を明らかにする。
- 実用的な XML 変換を台数効果の保証できる計算パターンにコンパイルする手法を与える。さらにその手法で扱える範囲のみが自然に記述できる変換記述言語を設計する。

以上によって、台数効果の保証された XML 変換言語を与える。

本研究は XML 処理の高速化という実用上重要な課題を扱っている。また、計算量の保証という理論的な興味深さを併せ持つ。加えて、各 PC が多数のコアを持つ時代に向けた次世代並列計算言語の基礎としての意義も持つ。

さらに、本研究は次世代汎用並列言語の基礎に繋がる可能性がある。汎用言語が台数効果を保証できる部分集合を持つなら、ユーザはできる限りその範囲で処理を記述することにより、効率良い並列計算を簡単に行うことができる。

### 3. 研究の方法

変換記述言語は XQuery の部分集合、特にそのコアをなす FLWOR 式を考える。XQuery は

広く使われている宣言的な XML 変換記述言語であり、本研究の目的に良く合致する。

並列評価には並列木縮約を用いる。並列木縮約は、入力木の形状によらず最適な台数効果を得ることができるため、本研究には必須である。

以上の前提の元で、以下の方針で研究を進めた

1. まず、コアとなる計算パターンの並列評価法を与える。具体的には、XPath クエリ、木構造間の構造変換、木上の簡単な数値などの計算、の 3 種について、台数効果に優れた並列評価法を与え、さらに台数効果の最適性を保証できる範囲を特定する。
2. 次に、XQuery から、先に議論した計算パターンへのコンパイル技法を与える。主な研究課題としては、大きな変換を各計算パターンへ分割する手法、および各計算パターンの合成を適切に扱う技法が挙げられる。さらに、このコンパイル技法が適切に扱える範囲を特定し、XML 変換言語を与える。

なお、本研究の目標は並列 XML 処理のための基礎理論を与えることである。実装はアイデアの確認のための小規模なものにとどめる。実用レベルのシステムの構築は本研究の範囲外である。

### 4. 研究成果

雑誌論文 では、XML のような平坦な木構造に対する並列木縮約を行うと、各頂点の子の数がそれほど大きくない木の場合に比べ、効率が出にくいという問題に取り組んだ。当該論文では、この困難を解決するアルゴリズムを与え、さらにそのアルゴリズムと既存の並列木縮約アルゴリズムの関係を明らかにした。これにより、並列木縮約の理論が自然に XML のような木構造へと拡張できる。

学会発表 では、既存の並列 XPath 問合せ手法の計算量の誤りを指摘し、よりよい計算量を持つアルゴリズムを提案した。XPath 問合せは XML 処理で最も重要な要素の一つであり、これに効率の良いアルゴリズムを与えることは重要な課題である。

学会発表 および では、マクロ木変換器と呼ばれる木構造変換のモデルに対する最適な並列評価方法を与えた。XML に対する問合せのような、木構造から木構造を得る計算、すなわち木変換については、その効率の良い並列処理の方法が知られていなかった。本研究では、マクロ木変換器というよく知られたクラスに対しては、並列木縮約により効率よく実現できるということを示した。この結果は並列 XML 処理に対する理論的な基盤たり得る。

以上の結果は、主に研究の方法(1)に関する成果である。研究の方法(2)については、本研究の最中に Hakuta ら（引用文献[1]）によって行われた研究が利用できる。この研究

は、XQuery をマクロ木変換器に翻訳する手法を示している。彼らの目的は、XQuery をストリーム処理することであるが、彼らの変換そのものはこの目的に特化したものではない。この結果を学会発表 および と組み合わせることで、広い範囲の XQuery プログラムを並列評価することができる。

以上をふまえ、結果的には本研究の目標はおおむね達成できたと言える。しかし、Hakuta らの手法は XQuery の全てを扱うことはできないため、以上の手法で得られたものが実用上十分であるか否かはさらなる検討が必要である。

上記のコアな結果に加え、雑誌論文 では、本研究のアプローチがコンパイラにおけるデータフロー解析に対し単純で実用的な並列実装を与えることを指摘し、また、雑誌論文 では、関数型言語に対する一般的な分割統治計算の導出法を示した。これら結果は、本研究のアプローチが、XML 処理に特化したものではなく、一般性を持ったものであることを示している。

#### <引用文献>

[1] S. Hakuta, S. Maneth, K. Nakano H. Iwasaki: XQuery streaming by Forest Transducers, IEEE 30th International Conference on Data Engineering, ICDE 2014, 952-963 2014.

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

#### [雑誌論文](計 3 件)

A. Morihata, K. Matsuzaki: Parallel Tree Contraction with Fewer Types of Primitive Contraction Operations and Its Application to Trees of Unbounded Degree. IPSJ Transaction on Programming, Vol. 7, No. 5, pp. 1-9, 2014. (査読有)

S. Sato, A. Morihata: Syntax-Directed Divide-and-Conquer Data-Flow Analysis. In: Programming Languages and Systems - 12th Asian Symposium, APLAS 2014, Proceedings, Lecture Notes in Computer Science, Vol. 8858, pp. 392-407, Springer, 2014. (査読有)

A. Morihata: A Short Cut to Parallelization Theorems. In: ICFP'13: Proceedings of the 18th ACM SIGPLAN International Conference on Functional Programming, Boston, Massachusetts, September 25 - 27, 2013, pp. 245-256, ACM, 2013. (査読有)

#### [学会発表](計 7 件)

A. Morihata: A Datatype- and Editing-Operation-Generic Incremental Computation. 情報処理学会第 103 回ブ

ログラミング研究会, 産業技術総合研究所(東京都・江東区), 2015年3月9日.  
A. Morihata: A Short Cut to Parallelization Theorems (特別講演). 日本ソフトウェア科学会第 28 回大会, 名古屋大学(愛知県・名古屋市), 2014年9月8日.

森畑明昌: 漸増計算法の第三リスト準同型定理による系統的導出. 第 16 回プログラミングおよびプログラミング言語ワークショップ, 阿蘇の司 ビラパークホテル(熊本県・阿蘇市), 2014年3月7日.

森畑明昌: XPath問合せの Work Efficient な分散並列評価. 日本ソフトウェア科学会第 30 回大会, 東京大学(東京都・文京区), 2013年9月13日.

森畑明昌: 分割統治計算可能性の多相型を用いた保証. In: 第 15 回プログラミングおよびプログラミング言語ワークショップ, 御宿東鳳(福島県・会津若松市), 2013年3月4日.

森畑明昌: 並列木縮約を用いたマクロ木変換器の並列評価. 日本ソフトウェア科学会第 29 回大会, 法政大学(東京都・小金井市), 2012年8月23日.

A. Morihata: Macro Tree Transformations of Linear Size Increase Achieve Cost-Optimal Parallelism (Extended Abstract). 1st International Workshop on Trends in Tree Automata and Tree Transducers, TTATT 2012, Nagoya Univ. (Nagoya, Aichi), Japan, June 2, 2012.

#### [図書](計 0 件)

[産業財産権]  
出願状況(計 0 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

取得状況(計 0 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
取得年月日:  
国内外の別:

[その他]  
ホームページ等

特になし

6. 研究組織

(1) 研究代表者

森畑 明昌 (MORIHATA, Akimasa)  
東京大学・総合文化研究科・講師  
研究者番号：10582257

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：