

平成 23 年 6 月 15 日現在

研究種目：若手研究 (B)
研究期間： 2007～2010
課題番号：19700007
研究課題名 (和文) トランスデューサ理論による厳密かつ高速な XML 型検査アルゴリズム
研究課題名 (英文) Precise and Fast XML typechecking algorithm based on transducer theory
研究代表者 細谷 晴夫 (Haruo Hosoya)
東京大学・大学院情報理工学系研究科・講師
研究者番号：50335296

研究成果の概要 (和文)：

国際標準データ形式 XML を専門に扱うデータ変換言語のために、実行前に誤りを発見できる型検査とそれに関連する技術を研究した。スキーマという XML の特殊な機能を厳密に扱うため、ツリートランスデューサ理論を用い、また積オートマトンを使って高速な型検査アルゴリズムを開発した。さらに、トランスデューサの複数の返り値による拡張、二階述語論理による XML 問い合わせ、XML 多相型の研究も行った。成果の集大成を英文著書にまとめた。

研究成果の概要 (英文)：

Research has been conducted on typechecking and related techniques for languages specialized to processing of XML, an international standard data format. In order to treat the special functionality of XML called schemas, theory of tree transducers has been exploited and an efficient typechecking algorithm has been developed by using alternating tree automata. Furthermore, transducers extended with multiple return values, XML query based on monadic second-order logic, and XML polymorphism have also been studied. All these results have been summarized in a monograph written in English.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007 年度	800,000	0	800,000
2008 年度	800,000	240,000	1,040,000
2009 年度	900,000	270,000	1,170,000
2010 年度	700,000	210,000	910,000
年度			
総計	3,200,000	720,000	3,920,000

研究分野： 総合領域

科研費の分科・細目：情報学・情報学基礎

キーワード：計算機言語、XML、オートマトン

1. 研究開始当初の背景

XML とは、国際産業標準規格機構 W3C が 1998 年に制定した汎用データ形式であり、現在最も多くのソフトウェアで採用されているフォーマットのひとつである。XML が急速に広まった理由に、XML にはスキーマというデータの制約条件を指定する機構が提供されており、それによりユーザが自由に独自のフォーマットを定義することができるということがある。そのことは同時に XML から XML へのデータ変換をする必要性が高まった。しかしこのとき、変換を行うプログラムが誤作動を起こすと、とくにそれが運用段階であると、修復のために多大なコストがかかる。既存の XML 技術では、これはできるだけ多くの入力を用いてテストすることによって検査するが、この手法ではテストの漏れがないことは保証不可能である。そこで注目されたのが静的解析技術、すなわちプログラムの実行前のある種の間違いを検出する技術である。通常のプログラミング言語ではそのような技術は盛んに研究されているが、XML の変換プログラムは、スキーマという特殊な機構が使われるためそのままでは使えない。従って、XML 専用の静的解析の技術が望まれていた。

2. 研究の目的

従来のプログラミング言語の静的解析の方法論に、型検査というものがある。これは、プログラムのデータ処理の流れをデータ形式である「型」を用いて解析する方法である。本研究では、

- (1) XML のスキーマを型と見なしたような型検査の方式を構築すること、および
- (2) そのような型検査を備えた XML 変換言語をより使いやすくするような言語機構とその実装手法を開発すること、が目的である。

3. 研究の方法

- (1) まず、XML 変換プログラムの型検査を構築するため、本研究では、「トランスデューサ理論」を用いることにした。XML 型検査の方式としては、関数型言語の型検査に基づいた XDuce などの方式があったが、厳密な型検査は原理的に不可能であった。それに対し、トランスデューサ理論を用いることにより、「厳密」な型検査が可能であることは理論的に知られているため、それを積ツリーオートマ

トンという枠組みを用いて「高速」なアルゴリズムを構築する。

- (2) 型検査機能を活かすような言語機能として、以下のようなものについて研究する。
 - ① トランスデューサという枠組みは、戻り値を一つしか返せないという制約があり実用上使いにくいという問題点が合った。そのため、複数の値を返せるような拡張を考え、その理論的な性質を探る。
 - ② ツリー変換言語において重要な機能として、「ツリー問い合わせ」がある。これをツリーオートマトンと密接な関係のある二階述語論理(MSO)を用いて構築し、その理論的性質と実装法を探る。
 - ③ プログラムにより柔軟性を持たせるように多相型という、型をパラメタ化できる機能を考案し、理論的性質と実装方式を探る。

4. 研究成果

(1) 積ツリーオートマトンという枠組みを用いたマクロツリートランスデューサの型検査アルゴリズムを提案した。これは、同じ型検査を、よりコンパクトに結果を計算できるだけでなく、様々な最適化をすることが可能になり、従来の決定ツリーオートマトンを用いたものよりも圧倒的に高速であることが分かった。

(2) ツリートランスデューサを拡張したマルチリターン・ツリートランスデューサを提案し、表現力が真に向上することを数学的に証明した。この証明では、マルチリターン・ツリートランスデューサの一つの例が、通常のツリートランスデューサでは表現できないことを示す方法を用いた。

(3) 上記のマルチリターン・ツリートランスデューサが、決定性トップダウンツリートランスデューサというクラスの XML 変換を、前側に連結した場合も、後ろ側に連結した場合も、ひとつのマルチリターン・ツリートランスデューサで表現できることを証明した。マクロツリートランスデューサでは、連結性に関して良い性質が成り立たないが、本研究の結果はマルチリターンによって拡張すれば良い性質が得られることを示すことができた。

(4) XML 変換に多相型を導入し、型の一部を

パラメタ化することによってプログラムの汎用性を高める研究を行った。多相型は関数型言語の研究で標準的に使われる技術であるが、XML 型検査においては、正規表現という従来にない概念を扱うため、マークつきツリーオートマトンという概念を導入し、XML のための多相型を実現した。

(5) XML 変換言語のコアの機能である、ツリー問い合わせに関する研究を行った。特に XML で重要な MSO 論理による、 n 項問い合わせという問題を解くアルゴリズムを提案した。これは、入力ツリーのサイズと、出力の集合のサイズの両者に関して線形時間で計算でき、コンパクトな表現で実行可能でき、従来の手法と比較して大きな優位性があることを示した。

(6) これまでの研究成果を、全世界の研究者や企業がより効率よく学びやすいように、全文英文の著書を執筆し、出版した。この著書には、本研究の成果はもとより、その背景的な知識を含め、この分野に関する網羅的でわかりやすいものとしてまとめることができた。

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 2 件)

Haruo Hosoya, Alain Frisch, and Giuseppe Castagna. Parametric polymorphism for XML. ACM Transactions on Programming Languages and Systems, 32(1):2:1-2:56, 2009.

Kazuhiro Inaba, Haruo Hosoya. Compact representation for answer sets of n -ary regular queries. Theoretical Computer Science. 411(38-39), 3481-3492, 2010.

[学会発表] (計 5 件)

Alain Frisch and Haruo Hosoya. Towards practical typechecking for macro tree transducers. In DBPL, pages 246-260, 2007.

Kazuhiro Inaba and Haruo Hosoya. XML transformation language based on monadic second-order logic. In Programming Language Technologies for XML (PLAN-X), pages 49-60, 2007.

Kazuhiro Inaba, Haruo Hosoya, and Sebastian Maneth. Multi-return macro tree

transducers. In Proceedings of Conference on Implementation and Applications of Automata (CIAA), pages 102-111, 2008.

Kazuhiro Inaba and Haruo Hosoya. Multi-return macro tree transducers. In Programming Language Technologies for XML (PLAN-X), 2008.

Kazuhiro Inaba and Haruo Hosoya. Compact representation for answer sets of n -ary regular queries. In Proceedings of Conference on Implementation and Applications of Automata (CIAA), pages 94-104, 2009.

[図書] (計 1 件)

Haruo Hosoya, Foundations of XML Processing: The Tree-automata Approach. Cambridge University Press, 2010. (全 238 ページ)

[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 :
取得年月日 :
国内外の別 :

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

細谷 晴夫 (Haruo Hosoya)
東京大学 大学院情報理工学系研究科・講師
研究者番号 : 50335296

(2) 研究分担者 ()

研究者番号：

(3) 連携研究者 ()

研究者番号：