

研究種目：若手 (B)
研究期間： 2007～2011
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. 研究計画の概要

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一つである。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。現行の 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。型検査とは、プログラムの解析によって、このような誤作動を開発段階ですべて発見する技法である。本研究課題は、XML 変換プログラムの型検査を、「トランスデューサ理論」を用いて、「厳密」かつ「高速」に行うことが目的である。

2. 研究の進捗状況

(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。

3. 現在までの達成度

②おおむね順調に進展している。

上記(1)にあるとおり、当初予定していたトランスデューサ理論に基づいた型検査アルゴリズムを考案し、発表することができた。しかし、対象となるXML変換の枠組みの表現力自体に問題提起があり、(2)～(5)にある通り、従来のXML変換の拡張と、関係するアルゴリズムの研究を行うに至った。これらは当初計画からややずれるものの、(2)(3)にあるとおり有用なマクロツリートランスデューサの拡張にもつながり、研究をより深める方向につながり、全体として順調な達成度と考えている。

4. 今後の研究の推進方策

これまでに、マクロトランスデューサの型検査アルゴリズム、マルチターン・マクロトランスデューサの結合可能性と計算量、XMLのN項問い合わせ問題に関する研究を行った。今後はまず、これらの理論の厳密な証明を完成させ、システム構築して実効率の実験をし、論文誌に発表していく。また型検査アルゴリズムのさらなる改良として順推論やトップダウンアルゴリズム等の手法を追加も検討していく。さらに、当分野であまり使われたことのない機械学習理論を調査し、入力 of 統計的な性質を分析することにより、実際的な効率を向上させることを検討していく。最後に、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、全文英文の著書を発表する予定である。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にまとめていく。(すでにケンブリッジ大学出版社と契約済み。)

5. 代表的な研究成果

(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 1 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.

[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 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：
取得年月日：
国内外の別：

[その他]