様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成　25　年　5　月　31　日現在

機関番号：１２６１２
研究種目：若手研究（B）
研究期間：2010～2012
課題番号：22700008
研究課題名（和文）木トランスデューサに基づく実用的な構造化文書変換の効率化と高信頼化

研究課題名（和文）Efficient implementation and verification of practical structured-document transformation based on tree transducer theory

研究代表者　中野　圭介（ NAKANO KEISUKE ）
電気通信大学・先端領域教育研究センター・准教授
研究者番号：30505839

研究成果の概要（和文）：本研究課題の目標は，木トランスデューサ（TT）の理論を XML 文書などの構造化文書の変換に応用し，TT 理論の実用的側面を検証することである．TT 理論は形式言語理論の研究者らが中心となり数学的興味から研究が進められてきたため，その実用性は疑問視されていたが，本研究の成果により構造化文書などの木構造データやその一般化であるグラフ構造データの変換プログラムの効率化や高信頼化に十分有用であることが確認できた．

研究成果の概要（英文）： The goal of this research project is to demonstrate the practicality of the theory of tree transducers (TTs) by applying it to transformation of structured documents such as the XML format. The TT theory has been intensively studied by researchers in formal language theory from the mathematical point of view. The project representative confirmed the practicality of TT streaming and TT verification by extending existing results on composition and typechecking TTs.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	700,000	210,000	910,000
2011 年度	700,000	210,000	910,000
2012 年度	800,000	240,000	1,040,000
総　計	2,200,000	660,000	2,860,000

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：オートマトン理論・形式言語理論

１．研究開始当初の背景

XML（eXtensible Markup Language）に代表される構造化文書形式は，データベースや文書の木構造の簡潔な表現として広く使われており，異なるソフトウェア間のデータの共有や授受には構造化文書変換が必要不可欠なものであった． 研究代表者のそれまでの研究では，木トランスデューサを利用した構造化文書変換に対し，実行時間の短縮や消費メモリの節減などによる効率化および静的な(実行前の)型検査による高信頼化の基盤理論についていくつか提案してきたが，簡略化された構造化文書変換の形式的モデルを対象としていたために実用性の確証には至っていなかった． これは，形式的モデルでは付加情報や参照関係を含まない純粋な木構造データのみを対象としているためで，実用に合わせた理論の拡張が求められていた．

２．研究の目的

本研究では，構造化文書の中で最も一般的な XML 形式の文書や木構造に参照関係を加えたグラフ構造データを対象とし，従来の研究における理論上の制限を緩和することにより既存の XML 文書変換言語の処理系に応用し，その実用性を確認する．以下では，構造化文書変換の効率化と高信頼化およびグラフ変換の基盤理論のそれぞれについて，背景を踏まえた研究目的を具体的に述べる．

(1) XML 変換の効率化

多くの XML 処理では，XML 文書が表現する木構造を操作する手法（木構造処理）が採用されている． 木構造処理では，入力となる XML 文書に対応する木構造をメモリ上に保存し，その木構造を辿りながら指定された処理を行う形で実装される． 実際，XML 処理に特化した言語である XPath，XQuery，XSLT の処理系は，いずれも木構造処理で実装されているため，扱われる XML 文書と同等以上の大きさのメモリが必要となる． しかしながら，近年顕著になってきた情報爆発やハードディスク価格の下落に伴い，扱われる XML 文書が巨大化してきているため，木構造処理による実装は時間・空間の両面で効率的ではない場合が多い． これは，木構造処理では必要のない木構造までメモリ上に保存されてしまうために起きる問題である．

一方，この問題を回避するために，SAX に代表されるイベント駆動型の XML 文書処理の手法（ストリーム処理）が提案されている．この手法では，入力となる XML 文書の文字列に対し直接指定された処理を行うため，メモリの使用を最小限にとどめることができる．さらに，入力を読み終えるよりも先に部分的な結果を出力するため，実行時間を短縮することも可能である． このようなストリーム処理では時間・空間の両面で効率的に実行可能であるが，XML 文書を文字列のまま扱う必要があり，木構造を自由に辿るような変換が記述しにくいため，プログラムの生産効率が低くなり，バグの修正や仕様変更への対応が容易ではない． すなわち，木構造処理とストリーム処理は実行効率と生産効率において一長一短である．

研究代表者は，木構造処理とストリーム処理の両方の長所を同時に得るために，木構造処理からストリーム処理を自動的に導出する手法をいくつか提案している． これにより，生産効率の高い木構造処理によるプログラムでありながら実行効率の高いストリーム処理の恩恵を得ることが可能になる． 研究代表者は，属性文法の合成を利用した手法を足がかりに，制限された XPath 式をサポートした独自の XML 変換言語 XTiSP を開発した．また，関数型プログラミングによる XML 変換をサポートするために，構造的な再帰関数に対してもストリーム処理が導出可能であることも示している． これらの研究は，木言語理論に関する研究成果の一つである木トランスデューサ（TT）の合成に基づいており，ストリーム処理が導出できるかどうかは元の木言語理論における制限に依存する．この制限により多くの実用的な XML 処理へ応用するには至らなかったが，本研究では，実際に利用されている XML 処理を対象とした上で，これらを扱う上で必要な木言語理論の拡張を行う． 具体的には，以下の 2 点を研究目標とする．

- TT に条件分岐や基本型の値（整数・文字列など)を導入し， 既存の TT 合成理論を拡張し，その制限の緩和を言語 XTiSP にも反映することにより， ストリーム処理を意識せずに木構造処理が可能になる枠組みを目指す．
- XQuery や XSLT などの XML 変換言語によるプログラムから（拡張された）TT への翻訳法を開発し， 既存の XML 変換プログラムを効率的に実行できる枠組みを目指す．

これらの枠組みは最終的に実装に基づいて評価するものとし， 特に，2 つめにおいては既存の XML 変換言語処理系との実行効率の比較を行う．

(2) XML 変換の高信頼化

一般にプログラムの型検査の枠組みは，ユーザの記述したプログラムが意図したものであるかを 実行前に確認する有効な手段であり，プログラムの信頼性を高める上で重要な役割を担っている． しかしながら， XML 変換の型検査は型同士の複雑な包含関係をもつために， 従来の型検査機構をそのまま適用することができない． 本研究では，TT の型検査を応用し，この問題を解決する．XML 文書における「型」とは，対応する木構造がどのような形をしているかを示すもので，一般に DTD や W3C XML Schema や Relax NG などの XML スキーマによって与えられ， 木オートマトンによって受理される正規木言語によって形式化することができる．XML 変換の型検査とは，入力および出力に対応する XML スキーマをそれぞれ S_{in}，S_{out} とすると，与えられた XML 処理が， S_{in} を満たす XML 文書に対して 常に S_{out} を満たす XML 文書を出力することを実行せずに保証することである． 広く使われている XSLT などの XML 変換言語による XML 処理では 型検査がサポートされておらず， 正しい形式に変換されるかどうかは全てユーザの責任となっている．本研究では， これまで研究代表者が行ってきた XML 処理の型検査に関する研究を発展し，実用的な構造化文書変換に対しても適用で

きる枠組みを実装することを目的とする．研究代表者は，オーストラリアの研究所 NICTA（当時）の Sebastian Maneth 博士と共同研究において TT に穴適用(hole application)とよばれる概念を導入することにより 既存の研究より効率的に型検査が行えることを示し， その後の研究において，一般的な TT の合成に対して計算複雑性を改善する手法を示してきた． しかしながら，いずれの型検査も型のサイズに対して指数時間必要とするアルゴリズムを提示しており， 実用上は改善が望まれる． 本研究では，実用・理論の両面からこのアルゴリズムの改善を目指す．

３．研究の方法

本節では，前節で述べた目的を達成するために当初予定していた方法について述べる．実際には，木構造データだけでなくグラフ構造データも対象にすることで，より広範囲に応用可能な結果が得られたため，やや異なる方法を用いることとなった． しかしながら，基本的なアイデアは一致しているため，ここでは当初の予定していた方法を中心に言及し，実際の方法は次節で述べることとする．

本研究では，研究目的で示した構造化文書変換に関する効率化と高信頼化の達成のため，主に木トランスデューサ（TT）をはじめとする木言語理論の研究とその実用化を行う．構造化文書変換は TT によってモデル化できるが，実用的な XML 処理を考慮する場合には拡張が必要である． TT の拡張は，効率化と高信頼化のそれぞれの実現に対して別々に行う．XML 処理の効率化については，属性文法の核の部分を独立させた属性付き木トランスデューサ（ATT）や，構造的再帰をモデル化したマクロ木トランスデューサ（MTT）の合成に関する既存の理論に対し，特に(効率化が必要となる）大規模な XML 文書を対象とする実用的な処理を想定した拡張を施す．一方，XML 処理の高信頼化に関しては，入出力の XML 文書が XML スキーマに厳格であるような実用例（DocBook 形式から XHTML 形式への変換など）に対して必要となる拡張を施す．初年度は主に XML 処理の効率化の研究を中心に進めるとともに，XML 処理の型検査をはじめとする高信頼化に関連した研究の調査を行う． また，必要に応じて木構造データの一般化であるグラフ構造データについても調査を行い，同様の効率化や高信頼かが可能かについて精査を行う． 2 年目に関しては，初年度の効率化の研究をもとに実装を行い，XML 処理やグラフ変換の型検査の実用化に向けた既存研究の発展に取り組み，必要に応じてその実装と評価を行う．3 年目には，両方の実装を実用的な変換に対して適用し，必要に応じて基盤理論の改良を図る．

(1)　平成 22 年度の取組み

初年度は，TT の理論を整理し，表現力の制限を確認した上で必要な拡張を行う．木言語理論における TT は，木から木への変換を規則の集合として形式化されており，その規則の形に応じて，トップダウン木トランスデューサ（TDTT)，属性付き木トランスデューサ(ATT)，マクロ木トランスデューサ(MTT)などに分類されている． 本研究では，この中から合成に関してよい性質をもつ ATT と，表現力が高く関数型プログラムに似た構造をもつ MTT に対して拡張を試みる．拡張については対象とする実用的な XML 処理に依存するが，TT は木構造の生成や消費の形が限定されているため，少なくとも木構造以外の値やその上の計算による拡張が必要であると考えられる．XML 処理の効率化については，研究代表者の先行研究において，表現力の高い項書換え系という極端な拡張を対象にストリーム処理の導出法を提案しているが，この方法では XML 処理に頻出する木の列の連結に関して非効率であることが分かっているため，上述したような効率化に影響しない別の拡張を考える必要がある． 本研究では，広範囲の XML 処理の効率化を実現するために，MTT を実用的な XML 処理として使うために拡張するアプローチと，先行研究で示した強力すぎる方法に対して規則の形を制限するアプローチの両面から XML 処理の効率化に挑む．

一方，XML 処理の高信頼化に関連しては，TT への拡張は慎重に行う必要がある． TT により XML 処理を表現することで型検査が可能になる理由は，(1) TT による変換は認識可能性を保持する(正規木言語に対する逆写像が正規木言語になる)ことと(2) 正規木言語の包含関係は判定可能であることの 2 点である．TT を拡張するにはこれらを意識する必要があるが，先述の XML 処理の効率化に必要とされる拡張では，一般に(1)の認識可能性の保持を崩してしまい，(2)に必要な包含関係が決定不能問題となってしまう． そこで，本研究では，正規木言語に対して包含関係が決定可能であるようなより広いクラスの木言語（例えば，文脈自由木言語）が存在することなどに着目して TT の拡張を行う．初年度はこのアイデアをはじめとした木言語理論に関する情報収集を中心に行う．また，木構造の一般化であるグラフ構造に対する変換の理論についても調査する．

(2)　平成 23 年度の取組み

XML 処理の効率化については，初年度研究を進めた ATT や MTT の拡張に基づいて実装を行う．この際，これまで開発を続けてきた XML

変換言語 XTiSP の次期リリースを考慮して実装する．特に XPath 式やその問合せに基づく反復処理のサポートは必要な実装の一つであるが，TT への翻訳については研究代表者の先行研究のアイデアをもとに行う．

一方，XML 処理の高信頼化については，初年度に収集した木言語理論のアイデアを踏まえ，TT の拡張による XML 処理の型検査に取り組む．これまで多重指数時間が必要であった型検査アルゴリズムについても，研究代表者の先行研究である穴付き MTT のアイデアなどを用いて，現実的な実行時間で可能かについて検証し，必要に応じて実装も行う．また，XML 処理の型検査が指数時間かかってしまう大きな原因は，認識可能性を維持するような逆写像を利用しているためであるが，TT の規則の形を強く制限した XML 処理に対して（一般に認識可能性を維持しない）順写像を用いて型検査を行う多項式時間アルゴリズムも提案されている． 本研究ではこのアイデアを利用した型検査についても考慮に入れる．また，実用的な XML 処理への適用も考慮して型検査を実装し，その評価を行う．

(3) 平成 24 年度の取組み

XML 処理の効率化については，これまで確立した実装方法と試験実装を参考に本格的な XML 処理を可能にする処理系を実装する．また，その有効性を確認するため，XQuery や XSLT などの既存の XML 処理言語からの翻訳も行う．その際，共同研究で提案した XML 処理の効率化手法によって導出されたストリーム処理が並列計算によって実装しやすい構造をしていることを踏まえ，その実用性も検討する．

一方，XML 処理の高信頼化についても提案した型検査アルゴリズムを実装する．XQuery の効率化についての研究と同様に汎用の XML 処理言語からの翻訳を用いて，実際に使用されている XML 変換プログラムに対して型検査を行い，その妥当性を検証する．

４．研究成果

本研究課題は，木トランスデューサ（TT）の理論を XML 文書などの構造化文書の変換に応用することを目標としている． しかし，XML ではノード間に参照関係を与えることも認めているため，循環構造や共有構造を定義することができ，TT 理論の対象である木構造データよりも一般的なグラフ構造を扱う必要がある． 特に循環構造をもつグラフ構造データを処理する場合，停止しない計算が簡単に記述できないように制御する必要がある． これらの問題が，本研究課題が掲げる高信頼化を達成する上で解決すべきものであった．

平成 22 年度では，トップダウン木トランスデューサ（TDTT）やマクロ木トランスデューサ（MTT）などの TT において基本的な概念となっている構造的再帰形式に着目した．構造的再帰形式は，再帰呼び出しのたびに必ず構造を消費するような形で計算が進められるため，グラフ構造データに対しても必ず停止する計算のみを表現することが可能になる． この結果自体は，既に提案されているグラフ構造変換言語 UnCAL において実現されているものであるが，この言語モデルと TDTT には密接な関連があるため，TT 理論を応用できるのではないかと考え，UnQL プログラムの検証に取り組んだ．ここでいう検証とは，「入力仕様を満たす入力に対する出力が必ず出力仕様を満たす」という TT 理論で確立されている概念のことである． これを UnQL に応用することで，プログラムの高信頼化が実現される． この年度においては TDTT での検証のような決定可能な手続きは発見できなかったが，その部分的な結果として「入力仕様を満たす入力に対する出力集合の見積もり」に成功した．これは，ビュー更新（双方向変換）とよばれる出力の変更を入力に反映する仕組みに直接応用可能である．この他，XML の問合せ言語である XQuery に対する効率化についても研究を行った．

平成 23 年度においては，前年度に取り組んだグラフ構造データの変換言語 UnCAL に対し，高信頼化に関する研究を行った． 具体的には，入力となるグラフ構造データの集合 S と UnCAL プログラム f に対して，与えられたグラフ構造データ G が出力されるか否か，すなわち $\exists x \in S: f(x) = G$ が真となるかを判定する問題（ビュー更新可能性問題）に取り組んだ． 入力の集合 S はスキーマとよばれる満たすべき仕様として与えられており，無限個の要素を含むことも考えられるため，一般にこの判定を行うことは困難である． そこで，スキーマの記述能力をうまく制限することでこの判定問題を有限の計算で解くことができることに着目し，グラフ構造データ間の模倣関係に基づくスキーマを提案し，この問題を解くことに成功した． この成果は双方向変換とよばれるデータベースの更新を容易にする枠組みと密接に関連しており，今後の実用性の検証が期待される．また，このグラフ構造データに対するスキーマのアイデアは，TT の理論におけるスキーマに相当する木オートマトンの概念からヒントを得たものであるが，具体的な対応関係は明らかになっておらず，その解明は今後の課題の一つである． この他，TT の理論による XML 変換の検証と似た手法によるグラフ構造データ変換の検証やその応用にも取り組んでいる．

最終年度である平成 24 年度においては，一般のグラフ構造データ変換プログラムだけでなく，当初の目的である XML 変換プログラムへの TT 理論の応用にも取り組んだ． 特に，最も広く使われている構造化文書変換言語の一つである XQuery 言語のプログラムに対して，先述の MTT へ翻訳するアルゴリズムについて研究は有用な成果の一つである．MTT に対する最適化や検証は多くの研究者によって提案されてきたが，MTT は極度に洗練されたモデルであったため，その実用性は疑問視されていた． しかしながら，今回の翻訳アルゴリズムの提案により，その実用性を証明されることとなった． XQuery 言語で記述されたプログラムを MTT に翻訳することにより，このプログラムに対する最適化や検証することが可能となる． ただし，最適化や検証の対象となっていた MTT の表現力の限界により，XQuery 言語のプログラム全体をカバーすることはできなかったが，XPath とよばれる XQuery プログラム記述の核となる機構については広く対応しており，十分実用的であることを確認した． また，TT 理論における最近の成果として，変換プログラムの出力を圧縮する方法が提案されており，これについても本研究で確立した手法が直接応用可能であることが期待され，最終年度を終えた後も Oxford 大学の Sebastian Maneth 博士と共同研究を進めている（平成 25 年 5 月現在）．

また，平成 24 年度においても引き続きグラフ変換理論の応用についても研究を進めた．UnCAL 言語で記述されたグラフ変換プログラムに対しても，構造化文書変換と同様に最適化と検証の両面に取り組み，この成果はモデル変換などのソフトウェア開発技術へ直結するものである．

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 10 件）いずれも査読有

1) Soichiro Hidaka, Zhenjiang Hu, Kazuhiro Inaba, Hiroyuki Kato, Keisuke Nakano: GRoundTram: An Integrated Framework for Developing Well-Behaved Bidirectional Model Transformations. Progress in Informatics, No.10, pp.131-148, 2013.
DOI: 10.2201/NiiPi.2013.10.7

2) Keisuke Nakano: Metamorphism in Jigsaw. Journal of Functional Programming, 23(2), pp.161-173, 2013.
DOI: 10.1017/S0956796812000391

3) Keisuke Nakano: Shall We Juggle, Co-inductively? In Proceedings of 2nd International Conference on Certified Programs and Proofs, pp.160-172 2012.
DOI: 10.1007/978-3-642-35308-6_14

4) Soichiro Hidaka, Zhenjiang Hu, Kazuhiro Inaba, Hiroyuki Kato, Kazutaka Matsuda, Keisuke Nakano, Isao Sasano: Marker- Directed Optimization of UnCAL Graph Transformations. In Selected/revised papers from 21st International Symposium Logic-Based Program Synthesis and Transformation, pp.123-138, 2012.
DOI: 10.1007/978-3-642-32211-2_9

5) 中野圭介，日高宗一郎，胡振江，稲葉一浩，加藤弘之：模倣に基づくグラフスキーマを利用したビュー更新可能性判定．コンピュータソフトウェア，29(2)，pp.174-192，2012.
DOI: 10.11309/jssst.29.2_174

6) Kazutaka Matsuda, Kazuhiro Inaba and Keisuke Nakano: Polynomial-Time Inverse Computation for Accumulative Functions with Multiple Data Traversals. In Proceedings of ACM SIGPLAN Partial Evaluation and Program Manipulation, pp.5-14, 2012.
DOI: 10.1145/2103746.2103752

7) Kazuhiro Inaba, Soichiro Hidaka, Zhenjiang Hu, Hiroyuki Kato, Keisuke Nakano: Graph-Transformation Verification using Monadic Second-Order Logic. In Proceedings of 13th international ACM SIGPLAN symposium on Principles and practices of declarative programming, pp.17-28, 2011.
DOI: 10.1145/2003476.2003482

8) Isao Sasano, Zhenjiang Hu, Soichiro Hidaka, Kazuhiro Inaba, Hiroyuki Kato, Keisuke Nakano: Toward Bidirectionalization of ATL with GRoundTram. In Proceedings of 4th International Conference on Theory and Practice of Model Transformations, pp. 138-151, 2011.
DOI: 10.1007/978-3-642-21732-6_10

9) H iroyuki Kato, Soichiro Hidaka, Zhenjiang Hu, Keisuke Nakano, Yasunori Ishihara: Context-preserving XQuery fusion. In Proceedings of 8th Asian Symposium on Programming Languages and Systems, pp. 255-270, 2010.
DOI: 10.1007/978-3-642-17164-2_18

10) Soichiro Hidaka, Zhenjiang Hu, Kazuhiro Inaba, Hiroyuki Kato, Kazutaka Matsuda, Keisuke Nakano: Bidirectionalizing Graph

Transformations. In Proceedings of 15th ACM SIGPLAN International Conference on Functional Programming, pp. 205-216, 2010.
DOI:10.1145/1863543.1863573

〔学会発表〕(計 6 件)

1) Keisuke Nakano: Progress Report on the Rho Property of B Combinators, 37th TRS meeting, 仙台, 2012 年 11 月.
2) Keisuke Nakano: Towards Certified Model Transformation. Atlanmod-BiG Joint workshop on Bidirectionality in Model Transformations, パリ フランス, 2012 年 9 月.
3) Keisuke Nakano: Metamorphism, Jigsaw and String Rewriting. 36th TRS Meeting, 松江, 2012 年 2 月.
4) Keisuke Nakano: View updatability checking for graph queries. Shonan seminar on Automated techniques for higher-order program verification, 葉山, 2011 年 9 月.
5) Keisuke Nakano: View updatability checking with a graph schema. 5th International Workshop on Bidirectional Transformation in Architecture-Based Component Composition, 西安 中国, 2011 年 5 月.
6) 中野圭介, 日高宗一郎, 胡振江, 稲葉一浩, 加藤弘之: 模倣に基づくグラフスキーマを利用したビュー更新可能性. 第 13 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, pp. 146-160, 札幌, 2011 年 3 月.

６．研究組織

(1)研究代表者

中野　圭介 (NAKANO KEISUKE)
電気通信大学・先端領域教育研究センター・准教授
研究者番号：30505839