

平成 29 年 6 月 14 日現在

機関番号：12612

研究種目：若手研究(B)

研究期間：2013～2016

課題番号：25730002

研究課題名(和文) 定理証明支援系による形式木言語理論の定式化

研究課題名(英文) Mechanized formalization of formal tree languages in proof assistants

研究代表者

中野 圭介(Keisuke, Nakano)

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：30505839

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：木構造データから木構造データへ変換するモデルである木トランスデューサ理論は、プログラミング言語理論やデータベース理論において重要な役割を果たしており、1960年代から盛んに研究が行われている。本研究の目的は、この理論を計算機の上で定式化し、その中の定理に対して機械的に検査可能な形の厳密な証明を与えることである。この成果により、今後の同分野の着実な発展が期待され、さらに堅牢なソフトウェア開発への応用の基盤を提供することが可能となる。

研究成果の概要(英文)：The theory of tree transducers which models tree-to-tree transformations has been explored since 1960's and played an important role in theory of programming languages and databases. The goal of this project is to formalize the theory in computer where all proofs of theorems for tree transducers are mechanically certified. It enables us to easily and steadily expand the theory and apply it to development of robust softwares.

研究分野：情報学基礎

キーワード：形式言語理論 定理証明支援系 木トランスデューサ

## 1. 研究開始当初の背景

本研究においては、「定理証明支援系」と「形式木言語理論」の2つが重要な背景となるため、それぞれについて簡単に説明する。

### [定理証明支援系]

定理証明支援系は、数学的な定理を計算機上で対話的に証明するためのシステムのことで、論理の飛躍や誤った定理の適用などがないように計算機が証明を検査するため、ヒューマンエラーのない厳密な証明や定理の再利用性の向上に貢献している。定理証明支援系については、アイデア自体は古くから存在していたが、Coq や Agda や Isabelle/HOL をはじめとするツールの発展と普及により、ここ数年で大きな注目を集めている技術の一つとなっている。最近では、定理証明支援系の理論や応用に特化した複数の国際会議が新たに開催されるなど非常に注目を集めている。特筆すべきは、マイクロソフトの研究グループによる群論の奇数位数定理の証明で、Coq を用い6年半かけて定理の厳密な証明を与えたことが2012年に報告されている。この過程で、さまざまな数学の定理に有用な Coq のライブラリや機能拡張が開発され、群論のみならず他の数学分野の定式化も盛んに行われるようになった。

### [形式木言語理論]

形式木言語理論は、形式言語理論の一分野で、木構造をもつデータの集合（「木言語」という）のクラスを扱うものであり、木オートマトンや木トランスデューサなどを対象とした理論である。木オートマトンは、よい性質をもつ木言語を定義するための枠組みであり、木トランスデューサは木構造データから木構造データへの関数を定義するための枠組みである。これらを対象とした形式木言語理論は、数学的な興味にとどまらず、データベース研究者の間では XML などの木構造をもつデータを処理するプログラムの正当性を検査するために利用され、プログラミング研究者の間ではソフトウェアのモデル検査や効率的なプログラムの自動導出をするための道具として用いられるなど、様々な場面で応用されている。

## 2. 研究の目的

形式木言語理論は上述の通り堅牢なソフトウェアを開発する上で重要な基盤となる枠組みであるが、誤った定理に基づいて新たな定理を証明してしまったり、定理に必要な条件を見落としとして利用してしまったりすれば、その堅牢性や信頼性が大きく揺らぐことになる。本研究では、定理証明支援系により形式木言語理論の枠組みを定式化し、研究者やソフトウェア開発者にとって有益な基盤として利用できるようにすることを目的とする。

定理証明支援系における定式化は、たとえ

その理論の正当性が揺るぎないものであっても、全ての定理に証明を付加する必要があるため、それほど単純ではない。実際、形式言語理論の重要な定理の一つである Myhill-Nerode の定理は、Wu らが定理証明支援系 Isabelle/HOL により定式化を行っているが、証明の手法によっては1000行近くの証明スクリプトで記述されており、簡潔に記述された1つの定理を証明するだけでも多大な労力が必要であることが分かる。また、Coq などの構成的証明が求められる定理証明支援系では、原則的には背理法による証明ができないため、たとえば、「述語  $P$  を満たす値が必ず存在する」というような形をした存在命題に対する証明として、「 $P$  を満たす値が存在しないと仮定すると矛盾が起きるのでこの命題は真である」という間接的証明ではなく、「値  $v$  が  $P$  を満たすのでこの命題は真である」という具体例を用いた証明を記述することが必要である。この制約は特別な場合を除いて避けることができない。

本研究においては、これらの点に注意を払いながら定理の証明やそれに必要な補題を設計し、形式木言語理論において重要な定義や定理に対し、できるだけ多くのものを定式化することを目標とする。

## 3. 研究の方法

本研究では、形式木言語理論のうち、特に研究代表者の専門分野の一つである木トランスデューサ理論について定式化を行う。ただし、先述のマイクロソフトの研究グループによる奇数位数定理のプロジェクトの経過からもわかるように、1960年頃より積み上げられた全ての形式木言語理論を個人の力で定式化することは現実的ではない。そこで、木トランスデューサ理論が重要な役割を果たす具体的な応用を精査し、その応用で用いられる定義や定理を対象を絞って定式化を進める。この成果より、形式木言語理論の着実な発展や理論に基づく堅牢なソフトウェア開発が可能となることが期待される。

なお、本研究では定理証明支援系 Coq を用いて形式木言語理論の定式化を行う。Coq は、依存型・自動証明・プログラム抽出・文書化の全ての機能を兼ね備えた数少ない定理証明支援系であり、2002年にフィールズ賞を受賞した数学者 Vladimir Voevodsky 博士による自らの理論の定式化においても利用されている。本研究では、定理証明支援系の中でも数学の定式化が活発に行われていて、かつ、ソフトウェア開発にも利用されている Coq を利用することで、理論・実用の両面で有益な基盤の開発を行う。

## 4. 研究成果

本研究の目的は、木トランスデューサをはじめとする形式木言語理論を定式化することであったが、どの理論を定式化するかを精査する必要があった。このため、過程での副産

物として、木トランスデューサ理論に関する多くの成果が得られたので、こちらについても報告を行う。

なお、平成 26 年度から平成 27 年度にかけて木トランスデューサ理論における未解決問題の 1 つであった「木から文字列へのトップダウントランスデューサの等価性判定の決定可能性」について解決がミュンヘン工科大学の Helmut Seidl らによって発表された (①)。等価性判定とは、与えられた 2 つの木トランスデューサがどんな入力に対しても同じ結果を出力することを判定することで、一般的には入力の種類が無限に考えられるために判定することはできない。この問題に対し、Seidl らは木トランスデューサの形を制限することで、仮に入力の種類が無限にあったとしても有限時間で判定する方法を発見した。この成果や手法が非常に興味深いことから、定式化する対象としてこれも含めることとした。また、当初定式化する予定だった文字列に対する形式言語理論についても平成 25 年度に Christian Doczkal らによって実現されたため (②)、本研究における形式木言語理論の定式化はこれを参考にすることとした。

以下、各年度についての成果を報告する。

#### [平成 25 年度の成果]

初年度においては木言語理論より深く研究が行われている文字列言語理論 (通常の形式言語理論) の定式化に関する調査と、定式化すべき木言語理論の応用として木トランスデューサの逆変換について調査・研究を進めた。また、5 月に開催された木トランスデューサの専門家のみが招待されるダグストゥールセミナーに参加し、スタック機構を導入した木トランスデューサの拡張を紹介し、構造化文書に適用できることも報告した。本セミナーでは、同様の応用を考えている研究グループも参加していたため、意見交換を行い、次年度での論文発表につながった。

#### [平成 26 年度の成果]

2 年目には木トランスデューサの等価性判定に関する新たな成果が発表されたため、その定式化に向けた定理証明支援系ライブラリの整備を行った。また、これと並行して、木トランスデューサ理論のデータベース分野への応用である構造化文書処理の効率化についても研究を進め、国際会議で発表を行った。本研究では、XQuery と呼ばれる実用的な問合せ言語によるプログラムを木トランスデューサに翻訳することで、型検査やストリーム処理などの木トランスデューサのさまざまな利点が見られることを報告した。型検査とは、与えられた木トランスデューサが入出力の仕様を満たすかを自動的に判定するものであり、モデル検査にも応用されている。また、ストリーム処理とは、木構造データをメモリに展開せずに処理を進める手法

で、特定の木トランスデューサについてはストリーム処理が可能であることが研究代表者によってすでに示されている。今回は XQuery による問合せからこのトランスデューサに翻訳できることを示すことで、XQuery のストリーム処理系を実現することに成功した。

#### [平成 27 年度の成果]

3 年目は、Doczkal らの文字列言語理論の定式化を参考に形式木言語理論の定式化を試みるとともに、等価性判定にも応用可能な語方程式の解法の 1 つであるマカニン法の定式化にも着手した。Doczkal らによる文字列言語理論の定式化は、マイクロソフトの研究グループによる Coq の MathComp 拡張を利用しており、木言語理論の定式化においてもそれに従った。また、木トランスデューサの等価性判定については多項式環イデアルを利用したアルゴリズムが Seidl らによって提示され、これに対する調査を進めた。このアルゴリズムは理論上の可能性のみを論じていたため実用的ではないものと予想されるが、今後の拡張や効率化について取り組む価値のある興味深い問題であることがわかった。

#### [平成 28 年度の成果]

4 年目は定式化すべき応用として木トランスデューサの等価性判定や型検査について研究を進め、実装についての効果を確認した。等価性判定については、昨年度までに調査を行った Seidl らのアルゴリズムを実装したが、実用的な木変換に対しては少しの工夫だけでは現実的な時間で等価性を判定できないことがわかった。この定式化については、多項式環に関する定式化も伴うため、部分的なものにとどまっている。また、木トランスデューサの型検査については、実用的な構造化文書変換に適用可能なアルゴリズムを開発し、実装を行い、定式化も進めた。

最終目標であった形式木言語理論の定式化については部分的には終えることができたが、想定していた全ての応用を考慮した定式化については完成していない。これについては、次年度以降の後続プロジェクトである「定式化された形式木言語理論に基づくソフトウェア基盤技術の開発」において、引き続き進める予定である。特に、Seidl らによる等価性判定アルゴリズムについては、正当性の証明の定式化を行うことで、そのアイデアの本質を追究する。

#### <引用文献>

- ① Helmut Seidl · Sebastian Maneth · Gregor Kemper, Equivalence of Deterministic Top-Down Tree-to-String Transducers is Decidable, *Proceedings of 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 943-962, 2015,

doi:10.1109/FOCS.2015.62

- ② Christian Doczkal · Jan-Oliver Kaiser · Gert Smolka, A Constructive Theory of Regular Languages in Coq, *Proceedings of 3rd International Conference on Certified Programs and Proofs (CPP)*, pp. 82-97, 2013, doi:10.1007/978-3-642-25379-9\_3

## 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 8 件)

- ① Hiroyuki Kato · Soichiro Hidaka · Zhenjiang Hu · Keisuke Nakano · Yasunori Ishihara, Context-Preserving XQuery Fusion, *Mathematical Structures in Computer Science*, 25(4), pp. 916-941, 2015, doi:10.1017/S096012951300008X, 査読有.
- ② Shizuya Hakuta · Sebastian Maneth · Keisuke Nakano · Hideya Iwasaki, XQuery Streaming by Forest Transducers, *Proceedings of IEEE 30th International Conference on Data Engineering (ICDE)*, pp. 952-963, 2014, doi: 10.1109/ICDE.2014.6816714, 査読有.
- ③ Makoto Onizuka · Hiroyuki Kato · Soichiro Hidaka · Keisuke Nakano · Zhenjiang Hu, Optimization for Iterative Queries on MapReduce, *Proceedings of the Very Large Database Endowment (PVLDB)*, Vol. 7, No. 4, pp. 241-252, 2013, doi: 10.14778/2732240.2732243, 査読有.
- ④ Kazutaka Matsuda · Kazuhiro Inaba · Keisuke Nakano, Polynomial-Time Inverse Computation for Accumulative Functions with Multiple Data Traversals, *Higher-Order and Symbolic Computation*, 25(1), pp. 3-38, 2013, doi: 10.1007/s10990-013-9097-8, 査読有.
- ⑤ Soichiro Hidaka · Kazuyuki Asada · Zhenjiang Hu · Hiroyuki Kato · Keisuke Nakano, Structural Recursion for Querying Ordered Graphs, *ACM SIGPLAN Notices ICFP'13*, 48(9), pp. 305-318, 2013, doi: 10.1145/2544174.2500608, 査読有.
- ⑥ Kazuyuki Asada · Soichiro Hidaka · Hiroyuki Kato · Zhenjiang Hu · Keisuke Nakano, A Parameterized Graph Transformation Calculus for Finite Graphs with Monadic Branches, *Proceedings of 15th International Symposium on Principles and Practice of Declarative Program-*

*ming* (PPDP), pp. 73-84, 2013, doi: 10.1145/2505879.2505903, 査読有.

- ⑦ 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, 査読有.
- ⑧ Keisuke Nakano, Metamorphism in Jigsaw (Functional Pearl), *Journal of Functional Programming*, 23(2), pp. 161-173, doi: 10.1017/S0956796812000391, 2013, 査読有.

[学会発表] (計 6 件)

- ① 中川涼太 · 中野圭介, 属性文法合成による関数融合の実装, 第 19 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 (PPL2017), 2017 年 3 月 10 日, 華やぎの章慶山 (山梨県笛吹市).
- ② 高橋祐多 · 中野圭介, 木から文字列への決定性トップダウン変換の等価性判定の実装, 第 19 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 (PPL2017), 2017 年 3 月 9 日, 華やぎの章慶山 (山梨県笛吹市).
- ③ 阿部和敬 · 中野圭介, マクロ森トランスデューサの実用的な型検査に向けて, 第 113 回プログラミング研究発表会, 2017 年 3 月 3 日, 東京大学 (東京都文京区).
- ④ Keisuke Nakano, On properties of B-terms, *NJ Programming Languages and Systems Seminar*, 2017 年 5 月 19 日, フィラデルフィア (アメリカ).
- ⑤ 渡邊秀隆 · 中野圭介, Grammatical Framework における語彙データの自動生成, 言語処理学会第 21 回年次大会, 2015 年 3 月 19 日, 京都大学 (京都府京都市).
- ⑥ Keisuke Nakano, XML Stream Processing based on Tree-Transducer Composition, *Dagstuhl Seminar on Tree Transducers and Formal Methods*, 2013 年 5 月 7 日, ダグストゥール (ドイツ).

## 6. 研究組織

### (1) 研究代表者

中野 圭介 (NAKANO, Keisuke)

電気通信大学・大学院情報理工学研究所・准教授

研究者番号: 30505839