



ルールベースプログラミングのXMLの検証と処理への応用

著者	マリン ミルチア
発行年	2011
その他のタイトル	Applications of rule-based programming to verification and transformation of XML
URL	http://hdl.handle.net/2241/115253

機関番号：12102

研究種目：基盤研究（C）

研究期間：2008～2010

課題番号：20500025

研究課題名（和文） ルールベースプログラミングのXMLの検証と処理への応用

研究課題名（英文） Applications of rule-based programming to verification and transformation of XML

研究代表者

マリン ミルチア (Marin Mircea)

筑波大学・大学院システム情報工学研究科・講師

研究者番号：60396603

研究成果の概要（和文）：

妥当なXML文書の仕様モデルとしての正則生垣言語（RHL: Regular Hedge Language）に関する研究を行い、非線形正則生垣パターンをXML問い合わせ言語として用いる可能性を探った。RHLの新しい表現方法および積集合、商集合、product derivative、factorを計算する新しいアルゴリズムを提案した。非線形正規表現パターン内のすべての変数の組の型に対する健全かつ完全な型推論アルゴリズムの構成に、RHL factorizationをどのように適用するかを示した。

研究成果の概要（英文）：

We studied regular hedge languages (RHLs) as specification models of valid XML documents, and the possibility to use nonlinear regular hedge expression patterns to define a pattern-based query language for XML. We proposed new representations of RHLs, and algorithms for the computation of intersection, quotient, product derivative, and factors of RHLs. We have shown how to use RHL factorization to define a sound and complete type inference algorithm for the tuple of all variables in a nonlinear regular hedge expression pattern.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	1,300,000	390,000	1,690,000
2009年度	1,100,000	330,000	1,430,000
2010年度	1,000,000	300,000	1,300,000
総計	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・ソフトウェア

キーワード：XML、ルールベースプログラミング

1. 研究開始当初の背景

XMLは、Web上でデータを表現するための標準であり、多くのプログラミング言語やツールがXMLデータモデルで動作するように設計されている。

ルールベースのプログラミングスタイルは、Webアプリケーションの設計に必要な検証、変換、クエリなどの機能を強力にサポートする。XML文書の変換は、以下の要素から構成

される書換規則によって与えられる。(1) 木構造のパターン、(2) パターンマッチングによって抽出されたXMLデータから結果となる値を計算する式、(3) ルールが適用される場合に満たされなければならない条件。ルールの適用は戦略言語によって制御される。

2. 研究の目的

このプロジェクトの目的は、XMLデータモデ

ルで動作するように特化したルールベースのプログラミング言語を設計、実装することである。この言語は、ウェブアプリケーションを実装する上で必要となる妥当性検証、クエリ、変換などの機能をサポートする必要がある。

3. 研究の方法

我々の研究の出発点は我々が開発を行ってきた ρ Log である。 ρ Log は、戦略的書き換え及びいくつかの高階パターンマッチングの機能をサポートしている。サポートしている高階機能には列変数、関数、変数、およびコンテキスト変数などが含まれている。これらの機能は、特に XML 文書などの任意の大きさの木構造データからデータを抽出するために有用である。

XML 文書の妥当性を保証するために、我々は、正規表現に基づく型とその上の型検査を検討した。また、オブジェクト指向プログラミングと正規表現型を統合するために、正規表現型を部分型によって拡張することも試みた。

我々は、 ρ Log のマッチングのメカニズムを拡張し、以下をサポートする健全かつ完全なマッチング手続きを設計した：正規表現パターン、シーケンス及びコンテキスト変数に対する所属制約。この手続きは、XML クエリとドキュメントの妥当性検証に有用である。次の型推論問題は、XML 処理の検証において重要な役割を果たす。

- ルールベースのプログラム P 、入力型 T_{in} の XML 文書 D 、および出力ドキュメントの型 T_{out} が与えられる。
- 文書 D をプログラム D で変換した結果が出力ドキュメントの型 T_{out} を持つか？

我々は、始めに密接に関連する問題を検討した。XDuce のような局所的型推論アルゴリズムは、個別にすべてのパターン変数の正確な型を計算する。局所的型推論は、低い計算量を持ち簡単に実装することができるが、マッチングの結果として得られるパターン変数の値は、相互に依存していることを無視している。この欠点を克服するために、我々はすべてのパターン変数のタプルの正確な型を計算する大域的な型推論アルゴリズムを検討した。これらの大域的な型推論アルゴリズムははるかに複雑である。それらを理論的に分析するために、生垣言語に関する因子化 (Factorization) の新しい理論を展開した。

我々が提案した因子化アルゴリズムを実

装するために、Mathematica を使用した。Mathematica を使用する理由は、Mathematica がシーケンス変数及び関数変数に関する機能を備えているからである。また、バックトラックやコンテキスト変数をサポートすることも容易である。

ρ Log は、当初、Mathematica で実装されていたが、利用性を高めるために Java と Prolog による再実装を行った。Java による実装は、このプロジェクトの最終段階で行われた。主な課題は、 ρ Log の計算体系の効率的な Java による実装を明らかにすることである。

$P\rho$ Log は、SWI-Prolog による ρ Log のプロトタイプの実装である。これは、オーストリアの研究所 RISC-Linz の研究者との緊密な協力で、このプロジェクトの外で行われた。

両方の実装は、unranked tree の有限列である生垣に対して動作する。 $P\rho$ Log は、列変数とコンテキスト変数に対する正規言語への所属制約をサポートしている。Java による実装は、Java の型システムと正規表現型システムを統合している。両方の実装は XML の検証と変換に適したルールベースのプログラミング環境を提供している。

4. 研究成果

ρ Log を XML に関する検証と変換に適するものにするために、非線形正規表現パターンによってその計算体系を拡張した。このクラスのパターンは、XDuce の正規表現パターンに似ているが、パターン変数に対する型推論問題ははるかに複雑である。

我々は、我々が提案したパターンのクラスの型推論問題の広範な研究を行った。パターン変数に対する局所的な型推論の代わりに、我々は、パターン内のすべての変数の組の正確な型を推論するアルゴリズムを構築した。我々が提案したアルゴリズムは、生垣言語の因子化の計算に基づいている。

我々は、Conway による正規言語の因子化の理論を正規生垣言語に一般化し、正規言語におけるいくつかの結果が正規生垣言語においても保存されることを証明した。正規生垣言語が有限の因子化を持つことを証明し、その計算のためのアルゴリズムを同定した。このアルゴリズムは高い計算量をもっており、その効率は生垣言語とその因子の表現方法に大きく依存している。論文 [7] は生垣言語に関する線形方程式による表現を提案し、論

文[2]は簡約された決定的かつ完全な線形生垣オートマトンによる表現を提案し、論文[4]は因子オートマトンによる表現を提案している。

これらの理論的結果に加えて、我々はJavaによる ρ Logの再実装に取り組んだ。Javaによる実装は、 ρ Logの計算体系をXDuceのような正規表現型と統合している。文字列型、int, bool, long, float, doubleなどの基本型もサポートしている。この実装は、3種類のソースファイルを読み込むように設計されている：入力ドキュメント、スキーマ、およびルールベースのプログラム。ルールベースの変換の結果は生垣として得られ、シリアル化することでテキストファイルに保存できる。

ρ Logの最初の計算体系は、all-match戦力で動作するよう設計されていた。Javaによる実装では、我々は、XML文書変換のための有用なsingle-match戦略（最長一致と最短一致）によってそれを拡張した。また、以下の評価戦略のための演算子を追加した：leftmost innermost, leftmost outermost, parallel innermost, parallel outermost。

XMLドキュメントの検証は、XMLスキーマに対応する正規表現パターンに対して、XML文書の生垣表現を照合することによって達成される。

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7件)

1. M. Marin, T. Kutsia. On the Computation of Quotients and Factors of Regular Languages. *Frontiers of Computer Science in China*. 4(2):173-184, 2010. 査読有
2. M. Marin, T. Kutsia. Regular Hedge Language Factorization Revisited. In Sheng Yu, editor, *Proceedings of the 14th Intl. Conference on Developments in Language Theory (DLT 2010)*. August 17-19 2010, London, Ontario, Canada. LNCS 6224, pp. 328-339. Springer. 査読有
3. T. Kutsia, M. Marin. Order-sorted Unification with Regular Expression Sorts. In Ch. Lynch, editor, *Proceedings of the 21st International*

Conference on Rewriting Techniques and Applications (RTA 2010), July 11-13, 2010, Edinburgh UK. Vol. 6 of the Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl, Germany, pp 193-208. 査読有

4. M. Marin, A. Craciun. Type inference for regular expression pattern matching. In T. Ida, V. Negru, T. Jebelean, D. Petcu, S. Watt, D. Zaharie, editors, *Procs. of SYNASC' 2010*, pp. 366-276. Timisoara, Romania, 2010. IEEE Computer Society. 査読有
5. M. Marin, A. Craciun. Factorizations of regular hedge languages. In S. Watt, V. Negru, T. Ida, T. Jebelean, D. Petcu, D. Zaharie, editors, *Procs. of SYNASC' 2009*. Timisoara, Romania, September 2010. IEEE Computer Society. 査読有
6. B. Dundua, T. Kutsia, M. Marin. Strategies in PpLog. *PNTCS 15:32-43*. 2010. 査読有
7. M. Marin, T. Kutsia. Linear Systems for Regular Hedge Languages. In J. Grundspenkis, M. Kirikova, Y. Manolopoulos, L. Novickis, editors, *ADBIS 2009. Revised Selected Papers*. September 7 2009, Riga, Latvia. LNCS 5968, pp. 104-112. 2010. 査読有

[研究発表] (計2件)

1. T. Kutsia, M. Marin. Order-sorted unification with regular expression sorts. In Ch. Lynch, P. Narendran, editors, *Procs. of UNIF 2009*. August 2, 2009, Montreal, Canada.
2. M. Marin, T. Kutsia. Matching with Membership Constraints for Hedge and Context Variables. In M. Marin, editor, *Procs. of UNIF 2008*, pp. 55-68. July 18, 2008, Castle of Hagenberg, Austria.

[その他] (計1件)

1. ホームページ等

<http://www2.score.cs.tsukuba.ac.jp/projects/RBPforXML>

6. 研究組織

(1) 研究代表者

マリシ ミルチア (MARIN MIRCEA)
筑波大学・大学院システム情報工学
研究科・講師
研究者番号：60396603

(2) 研究分担者

井田 哲雄 (Ida Tetsuo)
筑波大学・大学院システム情報工学
研究科・教授
研究者番号：70100047

研究分担者

南出 靖彦 (Minamide Yasuhiko)
筑波大学・大学院システム情報工学
研究科・准教授
研究者番号：50252531