

## 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 6 月 3 日現在

機関番号：14603

研究種目：若手研究（B）

研究期間：2010～2012

課題番号：22700099

研究課題名（和文） 概念モデルと問合せ集合を用いた XML スキーマの導出と更新

研究課題名（英文） Derivation and update of XML schemas using conceptual model and query set

研究代表者

橋本 健二（HASHIMOTO KENJI）

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：90548447

研究成果の概要（和文）：XML スキーマ更新やそれに伴う XML 文書変換を、情報保存の観点において適切に行うための手法を確立することを目指した。本研究では、問合せ保存と呼ばれる情報保存の一形式化に着目し、木変換器や木オートマトンなどを XML 文書に対する変換と問合せの形式モデルとして用いて問合せ保存性の問題定義を行った。そして、XML 文書変換と問合せが与えられたときに、変換前文書に対する問合せ結果を、変換後文書からでも取り出すことが可能な、変換後文書に対する問合せが存在するかを検査する手法と、実際に問合せ保存が満たされる場合に変換後文書に対する問合せを構成する方法を提案した。

研究成果の概要（英文）：This research aims to propose a framework for XML schema update and XML document transformation considering information preservation. We focused on query preservation, which is a formalization of information preservation. We showed that, for some transformation and query classes modeled by tree transducers and tree automata, the problem of deciding the query preservation is decidable and a query that works directly over the transformed data can be constructed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010 年度	1,000,000	300,000	1,300,000
2011 年度	1,100,000	330,000	1,430,000
2012 年度	800,000	240,000	1,040,000
年度			
年度			
総計	2,900,000	870,000	3,770,000

研究分野：総合領域

科研費の分科・細目：情報学 メディア情報学・データベース

キーワード：XML, 文書変換, 情報保存, 問合せ保存

### 1. 研究開始当初の背景

近年、様々な分野で XML の利用が広がっている。XML 文書は、その利用に際し、スキーマによって文書の構造を規定することができる。しかしながら、その柔軟性により、似た情報を扱うが異なる構造を規定したスキーマ

マがしばしば定義されている。また、利用する状況に応じてスキーマが更新される場合も存在する。そのため、あるスキーマに従う XML 文書群を別のスキーマに従うように変換する必要が生ずる。このような文書変換においては、変換前の文書群に含まれる必要な

情報がその変換によって失われないことが望ましい。

XML スキーマ更新に関して、研究代表者はこれまでに、木構造の埋め込み関係に基づく表現能力保存という XML スキーマ間の関係を定義し、その関係に基づき情報が欠落しないスキーマ更新のための有効な手法を提案した。しかし、木構造の埋め込み関係と呼ばれる条件を満たす更新を対象としており、データ利用の状況に応じて情報保存に関する制約の調整ができない。そのため、そのような情報保存の形式化のもとでは、データの変換や更新の自由度が一般に低い。

一方、問合せ保存と呼ばれる、変換の情報保存の一形式化が知られている。問合せ保存は、変換  $T$  と問合せ  $Q$  に対して次のように定義される (図 1) : 問合せ  $Q'$  が存在して、任意の変換前データ  $D$  に対して、 $D$  に対する  $Q$  の結果が  $D$  を  $T$  で変換した結果  $T(D)$  に対する  $Q'$  の結果と一致する。つまり、問合せ可能性の観点において、問合せ  $Q$  で得られる情報がデータ変換  $T$  によって失われないことを意味する。問合せ保存性は、問合せによって保存すべき情報を指定することで、状況に応じて変換が満たすべき制約を設定できるメリットがある。しかしながら、XML のような木構造データに対する変換・問合せに関する問合せ保存性についての研究はほとんど知られていない状況であった。

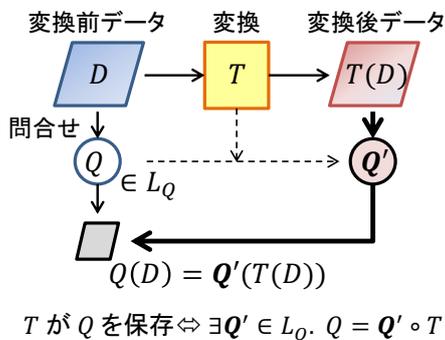


図 1 : 変換の問合せ保存

## 2. 研究の目的

本研究では、XML 文書の変換および問合せを対象に、問合せ保存性に基づく情報保存の形式化を導入して、問合せ保存性の検証および変換後データに対する問合せ (図 1 での  $Q'$ ) の構成を行う手法の開発を目指す。

既存研究により、XSLT や XQuery によって記述される一般の木変換のクラスでは、問合せクラスをラベルによる射影に限定しても、問合せ保存性は決定不能であることが知られている。しかしながら、問合せ保存性が決

定可能なクラスの調査がほとんど行われていない。また、問合せには、XQuery のような XML 文書を返すタイプ (以降、木問合せ) と、XPath のような XML 文書中の要素集合を返すタイプ (以降、頂点問合せ) の 2 種類が存在する。これらの違いが問合せ保存性の定義や決定可能性に与える影響についても言及されていない。そのため、本研究の目的は、木問合せと頂点問合せのそれぞれについて、問合せ保存性が決定可能であるような、なるべく広い変換クラスおよび問合せクラスを明らかにすることである。また、問合せ保存が満たされる場合に、変換後文書に対する問合せを構成する方法を与える。

## 3. 研究の方法

木問合せおよび頂点問合せをそれぞれ対象として、形式言語理論を用いた形式手法の知見を利用して、問合せ保存性の検証や変換後文書に対する問合せの構成について検討を行った。

(1) 木問合せを対象 : 木問合せを対象とした問合せ保存は、変換前後の文書に対する問合せ結果の木構造が完全に一致することを要求する。XML 文書をラベル付き順序木、変換と問合せをともに木変換器と呼ばれる形式モデルを用いて表現することとした。そして、問合せ保存性の形式的な定義として、確定性 (determinacy) と包摂性 (subsumption) と呼ばれる 2 種類の定義を与えた (図 2)。確定性は、木変換器  $T$  と  $Q$  に対して、ある木変換関数  $Q'$  が存在して、任意の変換前の木  $D$  に対して、 $D$  に対する  $Q$  の結果が  $D$  を  $T$  で変換した結果  $T(D)$  に対する  $Q'$  の結果と一致することを表す。包摂性は、確定性とほぼ同様であるが、 $Q'$  が変換前の木に対する問合せ  $Q$  と同じ問合せクラス  $L_Q$  に属することをさらに要求する。以上の定義のもとで、変換能力の異なるいくつかの木変換器のクラスにおいて、確定性と包摂性の判定問題の決定可能性について検討した。また、それらの性質が満たされるときに、変換後の木に対する問合せ  $Q'$  を構成する方法についても検討した。

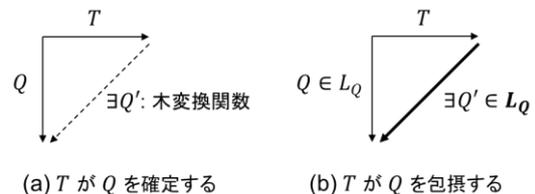


図 2 : 確定性と包摂性

(2) 頂点問合せを対象 : 木問合せを対象とした問合せ保存は、変換前文書の木の親子関係や兄弟関係などの構造情報を、文書変換が保存すべき情報とみなす場合に有効であ

る。一方、頂点問合せを対象とする問合せ保存は、頂点（がもつ値）の組の集合が維持されることを要求する。そのため、木の構造情報よりも、文書内の要素の関連情報が保存すべき情報とみなされる。頂点問合せは、木問合せの場合と比較して、指定される問合せによっては構造情報に関する制約が緩いため、問合せ保存を満たす文書変換の自由度が高いと考えられる。問題の形式化については、(1)と同じく変換は木変換器で表現したが、問合せは木オートマトンで表現される受理実行問合せと呼ばれるモデルを用いた。この問合せクラスは、木上の単項二階述語論理 (MSO) で表現される問合せ能力と等価な能力があることが知られており、XPathの核となる軸や述語などを用いた問合せ能力を包含する。頂点問合せに合わせた問合せ保存性を定義し、以上で挙げたモデルを用いて、(1)と同様に、問合せ保存性の判定を行う手法と変換後の木に対する問合せを構成する方法を検討した。

#### 4. 研究成果

(1) 木問合せを対象とした問合せ保存性：

① (学会発表の④) まず、ラベル付き順序木をFirst-child/Next-sibling符号化によって2分木にして扱うことを前提に、正規先読み付き決定性線形トップダウン木変換器 (DLTR) と呼ばれる木変換器のクラスでXML文書に対する問合せ・変換が表現される場合に、確定性を判定するアルゴリズムを提案し、そのアルゴリズムの正しさを証明した。また、確定性が満たされるならば、確定性を満たす根拠となる変換後の木に対する問合せと等価な、DLTRのクラスの木変換器を構成する方法を与えた。このことから、確定性と包摂性が一致することを示した。一方、変換のクラスを、部分木の複製を許した準同形木変換器とした場合には、問合せを恒等写像のみに限定したとしても、確定性が決定不能であることを示した。また、変換クラスが単項二階述語論理で定義可能な木変換器 (MSOTT) の場合にも、同様に確定性が決定不能であることを示した。

② (学会発表の①, ②) 次に、2分木に限らず一般のランクありラベル付き順序木を対象として、DLTRよりも変換記述能力の高いクラスの木変換器によって変換および問合せが表現される場合についても検討を行った。

変換が一価性線形拡張ボトムアップ木変換器 (s1-XBOT)、問合せが一価性ボトムアップ木変換器 (s-BOT) で表現されるときに、確定性が決定可能であることを証明した。s1-XBOTとs-BOTはいずれもDLTRより真に大きいクラスであり、それぞれ、木構造の階層に関する順序関係を部分的に変更でき、

中間出力の複製などが可能なクラスになっている。我々が与えた確定性を判定するアルゴリズムでは、s1-XBOTの逆変換の構成と、その結果とs-BOTとの合成が関数を表現する（入力木に対して出力木が一意に決定される）かの判定を行っている。s1-XBOTの逆変換を表現するために、接ぎ木と呼ぶ指定された正則木言語中の任意の木を非決定的に挿入可能な機能を新たに追加した木変換器のクラス (XBOT+g) を考えた。また、その木変換器のクラスにおける関数性判定が決定可能であることを示した。従来のボトムアップ木変換器の関数性判定が決定可能であることは知られていたが、それらの証明を拡張することでXBOT+gにおいても決定可能であることを新たに証明した。

包摂性についても同様のクラスで決定可能であることを証明した。一般のランクありラベル付き順序木を対象とした場合には、①でのDLTRに対する結果と異なり、一般に確定性と包摂性は一致しない。しかし、確定性の決定可能性の結果から、確定性が満たされるときに、変換後の木に対する問合せ  $Q'$  がs-XBOTで表現可能であるということが示された。そして、s-XBOTが与えられたときにそれと等価なs-BOTが存在するかという判定問題が決定可能であることを証明した。これにより、包摂性についても決定可能であるという結果を与えた。

(2) 頂点問合せを対象とした問合せ保存性 (学会発表の③)：まず、頂点問合せを対象とする問合せ保存性の定義を与えた。その定義においては、XML文書は、データ木と呼ばれる、木の各頂点が自然数（データ値）をもつような木で表現した。一般に、頂点問合せの形式モデルは、木の頂点（の位置）の集合を返すことを想定するため、その定義のもとでは変換前後の異なる2つの木に対する問合せ結果を比較できない。そこで、頂点問合せによって選択した頂点があもつデータ値の集合を問合せ結果とみなすことにした。その上で、問合せ保存を以下のように定義した：ある頂点問合せ  $Q'$  が存在して、任意の変換前の木に対する問合せ結果集合が、変換後の木に対する  $Q'$  の結果集合に包含される。

以上のような定義のもとで、問合せが受理実行問合せ、変換が線形トップダウン木変換器で表現されるときに、問合せ保存を判定するアルゴリズムを与えた。受理実行問合せは、一般に頂点の集合だけでなく  $n$  個の頂点組の集合を返す問合せを記述でき、このような問合せについても対応した。また、問合せ保存が満たされるときに、変換後の木に対する頂点問合せの構成可能であることを証明した。

(3) 今後の課題：

- ① 木問合せを対象とする場合については、確定性が決定可能かそうでないかの、変換クラスの大まかな境界線を与えることができたが、包摂性についてはまだ議論が必要である。包摂性の方が問合せ保存の本来の考え方に近いので、包摂性がより現実的に強力な変換・問合せ能力を有するクラスにおいて決定可能であるかどうかを明らかにするのは重要な課題の1つである。
- ② 頂点問合せを対象とする場合についても、同様により強力な変換クラスにおいて問合せ保存の決定可能性を検討する必要がある。また、今回定義した問合せ保存では変換前後の木に対する問合せ結果に包含関係が成り立つことを要求している。この定義も情報保存の観点では有効と考えられるが、問合せ保存性の本来の概念は変換前後の木に対する問合せ結果集合の一致を要求する。今回は後者の定義に対する判定アルゴリズムを与えることができなかった。後者の定義に基づいた問合せ保存についての決定可能性の調査および問合せ構成アルゴリズムの提案は今後の課題である。

#### 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)

[学会発表] (計5件)

- ① Kenji Hashimoto, Ryuta Sawada, Yasunori Ishihara, and Toru Fujiwara, “Determinacy and Subsumption for Single-valued Bottom-up Tree Transducers”, Proceedings of the 7th International Conference on Language and Automata Theory and Applications (LATA 2013), LNCS 7810, pp.335-346, Bilbao, Spain, 2013.4.2 (accepted).
- ② Kenji Hashimoto, Ryuta Sawada, Yasunori Ishihara, and Toru Fujiwara, “Determinacy and Subsumption for Single-valued Bottom-up Tree Transducers”, 第15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 (PPL2013) カテゴリ 2, 2013.3.4, 福島.
- ③ 宮原一喜, 橋本健二, 関浩之, “決定性線形下降木変換器における頂点問合せ保存”, 電子情報通信学会技術研究報告, SS2012-38, Vol. 112, No. 275, pp. 13-18, 2012. 11. 1, 広島, 平成 24 年度電子情報通信学会ソフトウェアサイエンス研究会研究奨励賞.
- ④ 澤田龍太, 橋本健二, 石原靖哲, 藤原融, “決定性線形トップダウン木変換器における問合せ保存の決定可能性”, 電子情

報通信学会技術研究報告, SS2011-75, Vol. 111, No. 481, pp. 109-114 2012. 3. 14, 沖縄.

#### 6. 研究組織

##### (1) 研究代表者

橋本 健二 (HASHIMOTO KENJI)  
奈良先端科学技術大学院大学・  
情報科学研究科・助教  
研究者番号：90548447

##### (2) 研究分担者

該当なし

##### (3) 連携研究者

該当なし