科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　　６　日現在

機関番号：　１４４０１
研究種目：　基盤研究(C)
研究期間：　2011～2013
課題番号：　２３５００１２０

研究課題名（和文）問合せ最適化を考慮したＸＭＬデータ交換に関する研究

研究課題名（英文）A study on XML data exchange with query optimization

研究代表者
石原　靖哲（ISHIHARA, YASUNORI）

大阪大学・情報科学研究科・准教授

研究者番号：００２６３４３４

交付決定額（研究期間全体）：（直接経費）　　3,900,000円、（間接経費）　　1,170,000円

研究成果の概要（和文）：あるスキーマに従うXML文書を異なるスキーマに従うXML文書に変換して活用するための基盤技術の開発が目的である．本研究では，変換後の文書への問合せの最適化に応用できるように，変換後の文書候補に共通の性質をXPathという問合せ言語で表現することを目指した．成果として，XPathで記述された式が文書候補の共通構造であるかを効率よく判定できるための条件をいくつか与えた．また，データ交換にまつわる他の理論的問題にも取り組んだ．

研究成果の概要（英文）：Our aim is to develop fundamental technologies for utilizing an XML document conforming to a schema by transforming it to a document conforming to another schema. In this research, we have adopted XPath, a query language to XML documents, to represent a common structure of candidates for transformed XML documents, because such representation will be useful for query optimization. One of our contributions is several conditions where it can be decided efficiently whether a given XPath expression is really a common structure of a given set of candidate documents. We have also tackled other theoretical problems related to data exchange.

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　データベース　問合せ解析　データ統合

１．研究開始当初の背景

近年，さまざまな情報が XML 形式で表現され，やりとりされるようになってきている．XML で表現されたデータの最大の特徴は，半構造の自己説明型データであるという点にある．その特徴を最大限に生かすためには，少々形が異なる XML 文書群であっても自己説明情報を用いてそれらをシームレスに扱えることが望ましい．

本研究で取り組んだデータ交換というテーマは，上でも述べたとおり，あるスキーマに従うデータを異なるスキーマに従うデータに変換して活用するための基盤技術開発が目的である．2 つのスキーマとそれらの間の依存関係，および変換前のデータが与えられたとき，一般に変換後のデータが一意に定まるとは限らない．そのため，変換後のデータ候補に共通の性質を求める問題が精力的に研究されている．この共通の性質は確定解（certain answer）と呼ばれている．確定解の用途としては，それ自身をユーザに提示することはもちろん，変換後のデータ（候補）に対する問合せ処理の際に再利用し，問合せ処理効率を向上させるという用途が挙げられる．

関係データベースにおけるデータ交換の場合，確定解は変換後データ候補（=関係，すなわち組の集合）の共通集合で定義できる．その定義の簡潔さゆえ，これまでに多くの成果が得られている．しかし，XML データベースにおけるデータ交換の場合，そもそも木構造をもつXML 文書の確定解をどのように定義すべきかが自明でないという問題があった．2010 年には，ある文献において，木構造をもつXML 文書の確定解の定義および表現方法と，それを求める問題の計算量についての成果が報告された．この文献では確定解をできるだけ精密に表現することを目的としている．そのため，この文献の方法で表現された確定解を問合せ処理に効率よく再利用できるかは不明であった．

２．研究の目的

本研究では，変換後のデータ（候補）に対する問合せ処理に効率よく再利用できるような確定解の表現方法およびその計算方法を開発することを目指した．具体的には，問合せ処理での再利用を可能とするために，XPath という言語で記述された問合せ式で確定解を表現することを柱とした．XPath は，XML 文書（=木構造）中の頂点集合を指定するための問合せ言語であり，XQuery のような多くのポピュラーなXML 問合せ言語の部分言語でもある．典型的には XPath 問合せは，根頂点からの，一般には分岐をもつ経路として記述される．本研究では，この「根頂点からの分岐あり経路」を記述できる XPath の機能に着目し，変換後のデータ候補に共通の性質（すなわち，木の共通構造）を XPath 問合せで表現することを検討した．

より具体的には，以下の 3 点を達成することを本研究の目標とした．

(1) さまざまな XPath 問合せクラスについて，そのクラスが確定解をどの程度精密に表現できるのかを，先行研究との比較も含めて明らかにする．
(2) 以下の 2 つの問題について網羅的に検討する．
   - 確定解判定問題: 与えられた XPath 問合せが与えられた XML 文書集合の確定解を表しているかを判定する問題．
   - 確定解導出問題: 与えられた XML 文書集合の確定解である XPath 問合せを求める問題．

   具体的には，さまざまなXPath 問合せクラスや XML 文書集合クラスの組み合わせに対して，確定解判定/導出問題の計算複雑さを明らかにする．さらに，決定可能となる組合せについては判定/導出アルゴリズムを実装し，現実的にどれくらいの計算時間を要するのかを明らかにする．
(3) 確定解判定/導出問題の解を用いて問合せ最適化を行うシステムを試作する．そして，スキーマ統合やスキーマ進化など，データ交換の実際的な場面にこのシステムを適用し，効果を確認する．

３．研究の方法

本研究では，上述の 3 つの目標の達成を目指したことはもちろん，XML データ交換や XPath 問合せの静的解析にまつわる問題のうち，本研究に関連が深いものや知見を活かせそうな問題にも幅広く着手した．具体的には，ある限定された XPath クラスに対する確定解判定問題や確定解導出問題に取り組むのと並行して，以下で述べる XPath 充足可能性問題や XML スキーママッピングの絶対整合性判定問題などに取り組んだ．

XPath 充足可能性問題とは，与えられた XPath 問合せと XML スキーマに対し，その XML スキーマに従っていてかつその XPath 問合せで指定された構造をもつ XML 文書が存在するかを判定する問題である．これは確定解判定問題と双対な問題であるため，充足可能性問題を検討して得られる知見は，確定解判定問題を検討する上でも有用な知見となり得る．また，充足可能性問題の判定結果それ自身が，問合せの最適化に有用であることも知られている．

一方，XML スキーママッピングの絶対整合性判定問題とは，2 つの XML スキーマ（ソーススキーマとターゲットスキーマの対）とそれらの間の依存関係が与えられたときに，ソーススキーマに従う任意の XML 文書について，その依存関係を満たしていてかつターゲットスキーマに従う XML 文書が存在するかを判定する問題である．したがって，XML スキーママッピングが絶対整合性をもたない場合，それによって指定される XML データ交換はデータの変換を行えない（ソースの

文書に対応するターゲットの文書が存在しない）可能性がある．このため，絶対整合性はデータ交換がもつべき基本的性質のひとつであるといえる．

４．研究成果

(1) 確定解判定問題

まず，ワイルドカードを含む XPath 問合せクラスを対象として，与えられた DTD のもとでの XPath 式確定解判定問題の計算複雑さを網羅的に調査した．ワイルドカードはパス和演算の制限された形であり，実用上頻繁に用いられる演算子である．成果としては，パス和演算を含む XPath 問合せクラスに対しては確定解判定問題が coNP 困難であることを示した．さらに，ワイルドカード，子軸，子孫軸，述語のうちのどれか 3 つの組み合わせであれば確定解判定問題が多項式時間可解であること，および 4 つすべてを含む問合せクラスに対しては coNP 困難であることを示した．これにより，実用上意味のある問題設定で，確定解判定問題が効率よく解けるための境界を一部明らかにできた．

その後，さらに兄弟軸をも含む XPath 問合せクラスを対象として検討を行った．その結果，子軸，兄弟軸，述語に加えて子孫軸かワイルドカードを含むクラスに対しては多項式時間可解であること，子軸，子孫軸，兄弟軸，ワイルドカードを含むクラスに対しては coNP 困難であること，などの結果を得た．

これらの成果は後述の学会発表⑤他 1 件にて対外発表した．

(2) 確定解導出問題

導出される確定解は，なるべく情報量が大きいことが望ましい．より形式的には，極大な確定解（すなわち，それが確定解であることが他の確定解から導出できないような確定解）を導出できることが望ましい．本研究では，それを実現するための基礎技術として，確定解の極大性判定問題に取り組んだ．具体的には，子軸，子孫軸，述語，ワイルドカードから成る XPath 問合せクラスを対象として，与えられた DTD のもとで与えられた XPath 式が極大な確定解かどうかを多項式時間で判定できるための条件について検討し，その見通しを得た．

この成果は，研究指導学生 1 名の卒業論文としてまとめており，現在，対外発表の準備中である．

(3) 充足可能性判定問題

研究代表者が先行研究において提案した，disjunction-capsuled DTD と呼ばれる DTD の実用的な部分クラスおよびその派生クラスを対象として，XPath 充足可能性問題の計算複雑さについて検討し，これらの問題が効率よく解けるための新たな条件をいくつか得た．具体的には，現実世界の DTD のほとんどすべてをカバーしている十分に広いクラスである RW-DTD という DTD クラスや，それよりわずかに小さいクラスである MRW-DTD というクラスを提案した（図１）．そして，それらのもとで，いくつかのクラスの XPath 充足可能性問題が効率よく解けることを示した（表１）．これらの成果は学会発表①，④，他 1 件にて発表している．

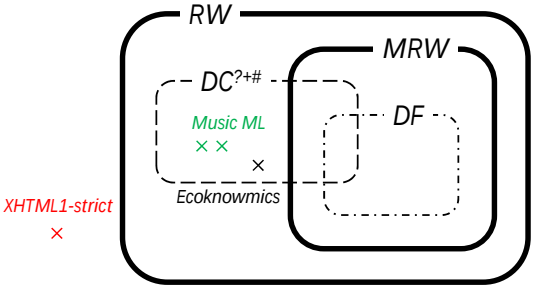

図 1　DTD クラス間の包含関係

表 1　XPath 充足可能性の計算複雑さ

↓	↓*	↑	↑*	→+	←+	∪	[ ]∧	[ ]	any	RW	MRW	DF	DC?+#
+	+			+	+				NPC	P	P	P	P
+		+		+	+				NPC	NPC	P	P	P
+				+	+		+		NPC	NPC	P	P	P
+						+	+	+	NPC	NPC	NPC	NPC	P
	+						+		NPC	NPC	NPC	NPC	P
+	+	+							NPC	NPC	NPC	NPC	P

[ ]∧: 論理積のみを含む述語　　P: 多項式時間可解, NPC: NP完全

(4) XML スキーママッピングの絶対整合性判定問題

XML スキーママッピングの絶対整合性が効率よく解けるような，実用的な DTD クラスおよび依存関係のクラスを提案した．具体的には，上述の MRW-DTD のサブクラスである MDC?+#-DTD という DTD クラスを提案した．選ばれた 27 個の現実世界のスキーマのうち，効率よく絶対整合性判定可能な既知のスキーマクラスでは 5 個しかカバーできていなかったところを，16 個まで拡大することができた．

この成果の一部は後述の学会発表②にて対外発表しており，同発表は Best Student Paper Award を受賞した．また，海外の論文誌への掲載が決定している（掲載巻号は未定）．

(5) システムの試作と評価

XPath 充足可能性の多項式時間判定アルゴリズムを計算機上に実装し，充足可能性に要する実時間を評価した．その結果，XML データの標準的なベンチマーク対し，数十ミリ秒程度で判定が可能であることを確認した．これにより，研究期間内には達成できなかったが，極大な確定解の導出ならびにそれに基づいた最適化についても，十分に実用的な範囲内の時間で行えると期待できることがわかった．

この成果は，研究指導学生 1 名の卒業論文としてまとめており，現在，対外発表の準備

中である.

(6) その他

木変換器間の包摂判定や問合せ間の型振舞い等価性判定といった，データ交換において有用となる新しい基礎技術の開発にも取り組んだ（後述の学会発表③他 1 件にて対外発表済み）．さらに本研究では，XML データベース問合せ解析技術の別の応用として，推論攻撃に対する安全性検証法についてもいくつかの成果を得た（後述の雑誌論文①他 2 件にて対外発表済み）.

(7) まとめ

当初の目標であった，確定解に基づく問合せ最適化に関しては，残念ながら十分な成果が出たとはいえず，引き続きブラッシュアップを続けていく予定である．一方で，XPath 充足可能性判定や XML スキーママッピングの絶対整合性判定など，関連した問題で多くの重要な成果を得ることができた．そのため，全体としては，データベース理論分野の発展に十分貢献できたと考えている.

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 1 件）

① Kenji Hashimoto, Hiroto Kawai, Yasunori Ishihara and Toru Fujiwara, Decidability of the Security against Inference Attacks using a Functional Dependency on XML Databases, IEICE Transactions on Information and Systems, Vol. E95-D, No. 5, pp. 1365-1374, 10.1587/transinf.E95.D.1365 (2012-05) 査読有.

〔学会発表〕（計 17 件）

① Yasunori Ishihara, Nobutaka Suzuki, Kenji Hashimoto, Shogo Shimizu and Toru Fujiwara, XPath Satisfiability with Parent Axes or Qualifiers Is Tractable under Many of Real-World DTDs, Proceedings of the 14th International Symposium on Database Programming Languages, http://arxiv.org/abs/1308.0769, 2013 年 8 月 30 日, Conference Center of Riva del Garda (Trento, Italy).

② Hayato Kuwada, Kenji Hashimoto, Yasunori Ishihara and Toru Fujiwara, The Consistency and Absolute Consistency Problems of XML Schema Mappings between Restricted DTDs, Proceedings of the 15th International Asia-Pacific Web Conference, Lecture Notes in Computer Science 7808, pp. 228-239, 2013 年 4 月 5 日, Novotel Sydney on Darling Harbour (Sydney, Australia).

③ Yasunori Ishihara, Kenji Hashimoto, Atsushi Ohno, Takuji Morimoto and Toru Fujiwara, Typing XPath Subexpressions With Respect to an XML Schema, Proceedings of the 5th International Conference on Advances in Databases, Knowledge, and Data Applications, pp. 128-133, 2013 年 1 月 30 日, Novotel Marques del Nervion (Seville, Spain).

④ Yasunori Ishihara, Kenji Hashimoto, Shogo Shimizu and Toru Fujiwara, XPath Satisfiability with Downward and Sibling Axes Is Tractable under Most of Real-world DTDs, Proceedings of the 12th International Workshop on Web Information and Data Management, pp. 11-18, 2012 年 11 月 2 日, Sheraton Maui Resort & Spa (Maui, Hawaii).

⑤ Kenji Hashimoto, Yohei Kusunoki, Yasunori Ishihara and Toru Fujiwara, Validity of Positive XPath Queries with Wildcard in the Presence of DTDs, The 13th International Symposium on Database Programming Languages, http://www.cs.cornell.edu/conferences/dbpl2011/papers/dbpl11-hashimoto.pdf, 2011 年 8 月 29 日, The Westin (Seattle, Washington).

〔その他〕

ホームページ等

http://www-infosec.ist.osaka-u.ac.jp/~ishihara/research/

６．研究組織

(1)研究代表者

石原 靖哲（ISHIHARA, Yasunori）

大阪大学・大学院情報科学研究科・准教授

研究者番号：00263434