

令和 6 年 5 月 27 日現在

機関番号：62615

研究種目：基盤研究(C)（一般）

研究期間：2018～2023

課題番号：18K11252

研究課題名（和文）XQueryを用いたデータ相互運用問題におけるデータの信頼性の計算手法の確立

研究課題名（英文）Towards a computational frame work on ranked trust policies in data interoperability problems used in XQuery

研究代表者

加藤 弘之（Kato, Hiroyuki）

国立情報学研究所・アーキテクチャ科学研究系・助教

研究者番号：10321580

交付決定額（研究期間全体）：（直接経費） 2,800,000円

研究成果の概要（和文）：データ相互運用問題におけるスキーママッピングをデータベースビュー定義と捉えることで、ビュー更新問題として問題を再定義した。データベース問合せ言語の数学的基礎になっている、論理に基づく問合せ言語であるDatalogを用いてデータベースに対する有効な更新からビュー定義を導出する手法を提案し、実装によりその有効性を確かめた。
また、ネットワークに閉路が存在する場合の取り扱いを idempotent として取り扱うような数学的枠組みを用意した。これにより各演算子が持つ数学的な性質と来歴情報が持つ数学的な性質の間に準同型が成り立つように定義でき、来歴情報の有用性を確認することができた。

研究成果の学術的意義や社会的意義

同じ意味を持つデータが異なるフォーマットで存在している現状に対して、統一的なデータの利活用を目的とするデータ相互運用問題に取り組んだ。具体的には、データ相互運用問題をデータベース研究分野で古くから取り組まれてきているビュー更新問題に帰着させ、これまでにない言語的観点からこの問題に取り組み、実用上よく使われているビュー定義とビューに対する更新について正しく更新が伝わることを確認した。
これにより、データ相互運用を現実のものとするのが可能となった。

研究成果の概要（英文）：By considering schema mappings in the data interoperability problem as database view definitions, we redefined the problem as a view update problem. We proposed a method to derive view definitions from valid updates to the database using Datalog, a logic-based query language that forms the mathematical foundation of database query languages, and confirmed its effectiveness through implementation. Additionally, we established a mathematical framework to handle cases where the network contains cycles by treating them as idempotent. This allowed us to define a homomorphism between the mathematical properties of each operator and those of provenance information, thereby confirming the usefulness of provenance information.

研究分野：データベース

キーワード：Provenance

1. 研究開始当初の背景

「データ相互運用問題」とは、異なるスキーマのもとに存在している同じ意味を表すデータを統一的に扱う問題として、データベースの研究分野で古くから継続して取り組まれてきた課題であり、新しい技術や応用とともに発展し続けている。実際、AIGMOD/PODS, VLDB, CIDR などのデータベース研究分野のトップの国際会議において、毎回この問題に関するセッションが設けられている。2016年のフォーブス誌によると、データサイエンティストの仕事のおよそ80%は「データを用意すること」に充てられている。ここでいう「データの用意」とは、データ分析のために、データを探し、集め、整理、統合し、管理することである。このような仕事はデータサイエンティストにとって最も煩わしい部分であるとのことである。この「データの用意」において、データ相互運用問題で使われている技術は重要な役割を果たしている。

データ相互運用問題は、近年次の二つの点で新たな方向に拡張されている。1) 対象となるデータモデルは従来の関係データから、データ交換フォーマットの国際標準であるXMLデータへと拡張されている。2) 異なるスキーマの統合手法は、データ全体をカバーするスキーマによる統合から、各ピア間のスキーマの違いを変換で記述するスキーママッピングを用いたP2P手法へと移行している。尚、データ相互運用問題では閉包性を達成するために、使われているデータモデルの問合せ言語を用いてスキーママッピングを記述する。P2Pアプローチは、全体として統合スキーマを用意する必要がないため、現実の問題に適している。問合せは自分がよく知っているピアに対してなされ、そのピアにスキーママッピングで繋がっている全てのピアからデータを検索することが可能となる。もう一つの利点は、ピアの参加離脱に関するモジュール性にある。新たに参加するピアは、最も似ているスキーマを持ち熟知しているピアとの間にスキーママッピングを記述すれば良いし、ピアが離脱する場合、自分と繋がっているスキーママッピングを削除するだけで良い。

かつては信頼できる専門家によって作成されることでその質が保証されていたデータベースは、近年インターネット上で様々な人たちによって作成、コピー、移動されている。このような状況では、データの質は保証されず、様々な質のデータが混在している。先に述べたP2Pアプローチの利点であるモジュール性により、自分が参加しているP2Pのシステムにどのような質のデータが混ざっているか事前に知ることは困難である。このような状況において、検索結果のデータの信頼性に関する研究は重要な課題である。

2. 研究の目的

本研究の目的は、P2Pに基づくデータ相互運用問題において、スキーママッピングを通じた問合せ結果の信頼性について、来歴情報を使った計算手法の確立である。

3. 研究の方法

上記の目的に対して、以下の方法で研究を行った。

- (1) まず、問合せに対して問合せ言語の演算子の性質に応じて、検索結果の来歴情報を導入する。この時、演算子の数学的性質と来歴情報の数学的性質の間に準同型が成り立つように定義する。
- (2) 上記で定義した来歴情報を更新にも対応できるように拡張する。この時、更新演算と来歴情報の間にも準同型が成り立つように拡張する。
- (3) スキーママッピングが閉路を構成している場合、更新が idempotent となるように来歴情報を定義する。
- (4) スキーママッピングとデータベースビュー定義の関連を調べる。

4. 研究成果

データ相互運用問題におけるスキーママッピングをデータベースビュー定義と捉えることで、ビュー更新問題として問題を再定義した。ビュー更新問題とはデータベース研究分野で古くから取り組まれている問題である。この問題の難しさの本質はビュー定義関数が injective ではないことによる、ビュー更新戦略が複数存在するか場合によっては該当するビュー更新戦略が存在しないことにある。本研究では、これまでにない言語的アプローチでこの問題に取り組み成果を挙げた。具体的には、以下のような成果がある。

- (1) データベース問合せ言語の一つである Datalog を用いて、データベースに対する有効な更新からビュー定義を導出する手法を提案し、実装によりその有効性を確かめた。
- (2) Datalog を用いたビュー更新の記述に関するデバッガを開発した。入力したビュー更新戦略が正しいものかどうかを判定し、正しくない場合に、なぜ正しくないのかを反例を生成することで利用者に通知する枠組となっている。これにより、利用者はビュー更新の記述が容易にできるようになった。
- (3) 非再帰データログを用いたビュー更新問題の解決手法をさらに発展させ、再帰が扱える

- ように対応し、実用的な複数の例を用いてその有効性を確認した。
- (4) 問合せ言語の演算子の数学的な性質と準同型が成り立つよう到来歴情報を定義した。
 - (5) ネットワークに閉路が存在する場合の取り扱いを idempotent として扱うような数学的な枠組みを用意した。これにより、各演算子が持つ数学的な性質と来歴情報が持つ数学的な性質の間に準同型が成り立つように定義でき、来歴情報の有用性を確認することができた。

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 2件/うちオープンアクセス 3件）

1. 著者名 Asano Yasuhito, Cao Yang, Hidaka Soichiro, Hu Zhenjiang, Ishihara Yasunori, Kato Hiroyuki, Nakano Keisuke, Onizuka Makoto, Sasaki Yuya, Shimizu Toshiyuki, Takeichi Masato, Xiao Chuan, Yoshikawa Masatoshi	4. 巻 1457
2. 論文標題 Bidirectional Collaborative Frameworks for Decentralized Data Management	5. 発行年 2022年
3. 雑誌名 Communications in Computer and Information Science	6. 最初と最後の頁 13～51
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-93849-9_2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Asano Yasuhito, Cao Yang, Hidaka Soichiro, Hu Zhenjiang, Ishihara Yasunori, Kato Hiroyuki, Nakano Keisuke, Onizuka Makoto, Sasaki Yuya, Shimizu Toshiyuki, Takeichi Masato, Xiao Chuan, Yoshikawa Masatoshi	4. 巻 1457
2. 論文標題 Bidirectional Collaborative Frameworks for Decentralized Data Management	5. 発行年 2022年
3. 雑誌名 Communications in Computer and Information Science	6. 最初と最後の頁 13～51
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-93849-9_2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 田中順平、Vandang Tran、加藤弘之、胡振江	4. 巻 14
2. 論文標題 プログラム可能なスキーマ共存戦略の実現手法	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌プログラミング(PRO)	6. 最初と最後の頁 15～33
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Tran Van-Dang, Kato Hiroyuki, Hu Zhenjiang	4. 巻 13
2. 論文標題 BIRDS	5. 発行年 2020年
3. 雑誌名 Proceedings of the VLDB Endowment	6. 最初と最後の頁 2897～2900
掲載論文のDOI（デジタルオブジェクト識別子） 10.14778/3415478.3415503	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Tran Van-Dang, Kato Hiroyuki, Hu Zhenjiang	4. 巻 13
2. 論文標題 A Counterexample-Guided Debugger for Non-recursive Datalog	5. 発行年 2020年
3. 雑誌名 Programming Languages and Systems (LNCS, volume 12470)	6. 最初と最後の頁 323 ~ 342
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-64437-6_17	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Van-Dang Tran, Hiroyuki Kato, Zhenjiang Hu	4. 巻 15
2. 論文標題 Programmable View Update Strategies on Relations	5. 発行年 2020年
3. 雑誌名 PVLDB	6. 最初と最後の頁 726-739
掲載論文のDOI (デジタルオブジェクト識別子) 10.14778/3377369.3377380	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

[学会発表] 計6件 (うち招待講演 0件 / うち国際学会 2件)

1. 発表者名 Vandang Tran, Hiroyuki Kato, Zhenjiang Hu
2. 発表標題 Toward Recursive View Update Strategies on Relations
3. 学会等名 Ninth International Workshop on Bidirectional Transformations (Bx 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Jumpei Tanaka, Van-Dang, Tran, Hiroyuki Kato, Zhenjiang Hu
2. 発表標題 Toward Co-existing Database Schemas based on Bidirectional Transformation
3. 学会等名 Third Workshop on Software Foundations for Data Interoperability (SFDI 2019+) (国際学会)
4. 発表年 2019年

1. 発表者名 Van-Dang Tran, 加藤弘之, 胡振江
2. 発表標題 A Declarative Framework for Updatable Views in Relational Databases
3. 学会等名 WebDBフォーラム
4. 発表年 2019年

1. 発表者名 Yasuhito Asano, Zhenjiang Hu, Yasunori Ishihara, Hiroyuki Kato, Makoto Onizuka, and Masatoshi Yoshikawa
2. 発表標題 Controlling and Sharing Distributed Data for Implementing Service Alliance
3. 学会等名 Second Workshop on Software Foundations for Data Interoperability
4. 発表年 2019年

1. 発表者名 Yasunori Ishihara, Hiroyuki Kato, Keisuke Nakano, Makoto Onizuka and Yuya Sasaki
2. 発表標題 Toward BX-based Architecture for Controlling and Sharing Distributed Data
3. 学会等名 Second Workshop on Software Foundations for Data Interoperability
4. 発表年 2019年

1. 発表者名 Yasuhito Asano, Dennis-Florian Herr, Yasunori Ishihara, Hiroyuki Kato, Keisuke Nakano, Makoto Onizuka and Yuya Sasaki
2. 発表標題 Flexible framework for data integration and update propagation: system aspect
3. 学会等名 Second Workshop on Software Foundations for Data Interoperability
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 Lu, Q., Zhang, W., Zhang, Y., Peng, Y., Kato, H., Wang, W., Xiao, C. (Eds.)	4. 発行年 2020年
2. 出版社 Springer	5. 総ページ数 193
3. 書名 Software Foundations for Data Interoperability and Large Scale Graph Data Analytics	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------