

氏 名 田中 順平

学位(専攻分野) 博士(情報学)

学位記番号 総研大甲第 2325 号

学位授与の日付 2022 年 3 月 24 日

学位授与の要件 複合科学研究科 情報学専攻
学位規則第6条第1項該当

学位論文題目 Programmable Strategies for Co-Existence of Relational
Database Schemas

論文審査委員 主 査 蓮尾 一郎
情報学専攻 准教授
加藤 弘之
情報学専攻 助教
関山 太朗
情報学専攻 助教
胡 振江
北京大学 計算機科学技術系 教授
武市 正人
東京大学 名誉教授

博士論文の要旨

氏名 田中 順平

論文題目 Programmable Strategies for Co-Existence of Relational Database Schemas

The co-existence of relational database schemas is an important feature of a database. Today, most information systems consist of applications and relational databases. They continuously evolve to meet the ever-changing real world. Applications are well supported by the technology to maintain and operate multiple versions for continuous and iterative development. However, it is a challenging task in databases to make one relational database schema (source schema) evolve to a new relational database schema (target schema) and make them co-exist by sharing data for concurrently running multiple application versions or applications.

The existing work proposes the view-based approach to realize the co-existence of relational database schemas in a database. It gives logically computed view instances of both schemas transformed from a shared database. Update sharing is realized between view instances through updates of the database. A data structure of the database is compatible between a database schema suited for source schema (source-side database schema) and a database schema suited for target schema (target-side database schema). When updates against view instances of target schema increase, the source-side database becomes unsuited for view instances of target schema. The existing work makes data migration into the target-side database based on the target-side database schema available.

However, there exist three problems. First, a co-existence strategy to specify how to evolve a schema and which case data is shared between view instances of schemas is limited to the predefined one, which lacks the flexibility. Second, there is no systematic methodology to realize co-existence strategies based on the view-based approach. This problem causes difficulty in designing user intended strategies and realizing them. Especially it is challenging to design the source-side and target-side database schemas with auxiliary relation names to control update sharing between view instances of schemas. Third, due to the second problem, there is no systematic methodology for data migration between the source-side database and the target-side database for an arbitrary co-existence strategy. This problem makes the co-existence of relational database schemas with flexible data migration difficult for any co-existence strategies.

In order to make the co-existence of relational database schemas more practical, this thesis aims to make co-existence strategies for the co-existence of relational database schemas programmable by solving the problems mentioned above.

To solve the first problem, we propose a Datalog-based DSL (domain-specific language) to specify various co-existence strategies. DSL is designed for the following three items: describing definitions of source schema and target schema, describing schema evolution to specify how to evolve relations of source schema to a relation of target schema, and describing backward

update sharing to specify how to share updates against a relation of target schema to relations of source schema. We assume a relationship between source schema and target schema is kept after schema evolution. Thus update sharing from relations of source schema to a relation of target schema follows the relationship specified by schema evolution. Furthermore, we introduce a property, the consistency of updates, and its verification method. We show that the proposed DSL can describe co-existence strategies of the existing work and other co-existence strategies.

To solve the second problem, we propose methods to systematically derive two types of bidirectional transformations (BXs for short) from a given co-existence strategy: BX between source-side or target-side database and view instances of source schema, and BX between source-side or target-side database and a view instance of target schema.

We define the source-side database schema to consist of base relation names having the same data structure with relation names of source schema and auxiliary relation names. A key idea of deriving BXs on the source-side database is that BX between the source-side database and the view instance of source schema is basically identity mapping between the base relation and the view instance. BX between the source-side database and the view instances of target schema is derived so that a co-existence strategy is realized between the base relations and the view instance of target schema. Auxiliary relations are systematically specified to compute the updated view instances from the updated base relations and the auxiliary relations without loss or gain.

To have a database suited for target schema, we define the target-side database schema to consist of a base relation name having the same data structure with a relation name of target schema and auxiliary relation names that are different from auxiliary relation names of the source-side database schema. A key idea of deriving BXs on the target-side database is that schema evolution is replaced to BXs between target-side database and view instances of source schema, and backward update sharing is replaced to BX between target-side database and the view instance of target schema. Auxiliary relations are systematically specified to compute the updated view instances from the updated base relations and the auxiliary relations without loss or gain.

We evaluate the usefulness of the proposed methods by showing results of derived BXs and execution time of writing and reading against view instances of source schema and target schema. Co-existence strategies of the existing work and other co-existence strategies are evaluated.

To solve the third problem, we propose a method of data migration based on the derived BXs. The source-side database is migrated to the target-side database by the following two steps: the BXs on the source-side database compute view instances of source schema and target schema, then such view instances are transformed into the target-side database by the BXs on the target-side database.

The target-side database is migrated to the source-side database by the opposite operations. We evaluate the usefulness of the proposed method by showing the execution time of data migration. Co-existence strategies of the existing work and other co-existence strategies are evaluated.

博士論文審査結果

Name in Full
氏名 田中 順平

Title
論文題目 Programmable Strategies for Co-Existence of Relational Database Schemas

本博士論文は、関係データベーススキーマの共存戦略を柔軟に記述する言語とその実現に関するものである。関係データベースでは、スキーマを継続的に進化させ、データを共有しつつ複数のスキーマを共存させる技術が必要とされているものの、その実現にはいくつかの困難がある。先行研究では、更新されたデータをスキーマ間で共有し合う規則や条件が予め固定されていて、これを共存戦略として柔軟に定めることが難しい。また、ビューを用いて共存戦略を実現するための体系的な手法が存在していない。本研究の主な貢献は (1) 共存戦略を記述するためのドメイン特化言語(DSL)の提案、(2) 2種類の双方向変換により共存戦略を実現する手法の実装と評価、(3) 新しいスキーマに適したデータベースインスタンスを得るデータ移行により新しいスキーマの効率を改善する手法の実現、である。

本論文は、英語で記述されており、全 6 章から構成されている。

第 1 章は序論である。研究の背景、研究目的、既存研究の問題点、本研究の主要な貢献、論文全体の構成を述べている。

第 2 章では基礎知識の導入を行っている。スキーマの共存に関わる Datalog の基礎知識、スキーマ進化の基礎知識、ビューを利用したスキーマ共存と更新データ共有の実現手法、双方向変換の基礎知識、およびそれぞれの関連研究について議論している。

第 3 章では、スキーマの共存を記述するためのドメイン特化言語(DSL)の文法、意味論、およびその言語による記述手法の詳細、更新に対する一貫性の静的な検証手法を示している。また、提案した DSL による多数の共存戦略を例示している。

第 4 章では、第 3 章で提案している DSL により記述された共存戦略を、スキーマ進化のもとであるソーススキーマの表と同じ構造を持つ実表をもつデータベースを元に、ソーススキーマとスキーマ進化後のターゲットスキーマの表をビューインスタンスとして実現する手法を提案している。そこでは 2 種類の双方向変換を導出し、スキーマ間で共有される更新データをデータベースの実表に格納するとともに、実表からだけでは得ることのできないターゲットスキーマのデータを補助表に格納することで共存戦略を実現する手法を示している。さらに、この手法を実装し、評価実験により、先行研究で議論されている共存戦略およびそれ以外の共存戦略を記述してそれらが実現可能であることに加えてデータ更新の実行時間を計測している。

第 5 章では、ターゲットスキーマの表と同じ構造のデータベースを元に、より一般的に

共有戦略を実現する手法を議論している。ここでは第4章の手法を発展させて、2種類の双方向変換を導出し、スキーマ間で共有される更新データをデータベースの実表に格納し、スキーマ進化の前後のスキーマのいずれかにおいて失われる属性のデータと、実表からだけでは得ることのできない双方のスキーマのデータを補助表に格納することによって共存戦略を実現する手法を示している。さらに第4章と第5章で導出された双方向変換により、それぞれのデータベース間でのデータ移行が可能であることを示している。これらの手法を実装するとともに評価実験により、先行研究で議論されている共存戦略とそれ以外の共存戦略を記述し、それらが実現可能であることとおよびデータ移行が可能であることを示し、それらの実行時間を計測して評価している。

第6章では論文のまとめと今後の課題を述べている。

審査会において、出願者は関係データベーススキーマの共存戦略を記述する言語とその実現法について、45分のスライド発表と75分の質疑応答によって行われた。スライド発表では、研究の背景、先行研究の課題、研究目的および主要な貢献が要領よく説明された。その後、審査委員による質疑に対して的確な回答がなされた。

以上のように、本論文では、関係データベーススキーマの共存戦略を記述する言語とその実現に関して新しい手法を与えた。この貢献により、関係データベーススキーマの共存戦略を柔軟に定義するとともに正しく実現し効率的に運用できるようになり、理論だけでなく実践的な観点からもこの課題に対する本研究の貢献が大きいと認められる。なお、本研究の成果は1件の雑誌論文、2件の査読付き国際会議論文、2件の国内研究会論文という形で対外的に公開され評価を得ている。また、開発したシステムはインターネットで公開され関係者が自由にダウンロードできるようにされている。以上の理由により、審査委員会は本論文が博士の学位の授与に値すると判断した。