Relaxed Notions of Equivalence

T

Decidability of Equivalence

which solves this problem

This is called the Halting problem, and is undecidable. There is no algorithm which solves this problem, nor can anyone ever develop such an algorithm.

Optimizing Schema Mappings with Relaxed Notions of Equivalence

Emanuel Sallinger

Motivation

Schema mappings play an important role in several areas of database research – above all in data integration and data exchange. The foundation of optimizing schema mappings has been laid in a recent paper by Fagin et al., where new relaxed notions of “equivalence” between two schema mappings have been introduced. Fagin et al. clearly demonstrated the potential of these notions in optimizing various kinds of schema mappings.

Goal

We investigate the potential of the relaxed notions of equivalence in optimizing schema mappings that allow to express both data exchange as well as integrity constraints.

On the one hand, we want to clarify if the relaxed notions of equivalence allow for additional optimization potential. On the other hand, we analyze several decidability questions.

SETTING

Tuple-Generating Dependencies

Tuple-generating dependencies (tgds) generalize foreign-key constraints on databases.

Source-to-target tgds can express data exchange.

Formally, they are described as

\[ \forall x (\psi(x) \rightarrow \exists y \psi(x,y)) \]

Schema Mappings

A schema mapping is given by

- source and target database schemas \( S, T \)
- dependencies \( \Sigma \), describing their relationship

Dependencies can describe source-to-target (s-t) data exchange or constraints on the target (t).

Equality-Generating Dependencies

Equality-generating dependencies (egds) generalize key constraints on databases.

Target egds can express integrity constraints.

Formally, they are described as

\[ \forall x (\psi(x) \rightarrow x_i \neq x_j) \]

CONCEPTS

Optimization

Optimization means replacing a schema mapping by a simpler, but still “equivalent” one.

Simpler can mean

- Removing redundant dependencies (subset-minimality)
- Finding a smaller schema mapping (cardinality-minimality)
- Using less complicated formulas as dependencies (antecedent-minimality)

Relaxed Notions of Equivalence

Logical equivalence, is quite strict and does not give rise to much optimization potential. Therefore relaxed notions of equivalence were introduced.

Data-exchange (DE) equivalence does not distinguish mappings which behave in the same way for data exchange.

Conjunctive-query (CQ) equivalence does not distinguish mappings which behave similarly for answering conjunctive queries.

They form a hierarchy, successively giving rise to added optimization potential.

MAIN RESULTS

Optimizing the tgds

For optimizing the s-t tgds,

- DE- and CQ-equivalence have the same power as logical equivalence
- for optimality criteria based on arbitrary real-valued functions

This includes optimality criteria like subset-, cardinality- and antecedent-minimality.

Optimizing the egds

For optimizing the target egds,

- while we gain optimization potential,
- common optimization tasks are undecidable for relaxed notions of equivalence

This holds for the optimality criteria like subset-, cardinality- and antecedent-minimality.

Decidability of Equivalence

DE- and CQ-equivalence are undecidable for schema mappings based on s-t tgds and target egds.

METHODS

Halting Problem

Does a computer, doing an arbitrary computation, eventually halt?

This is called the Halting problem, and is undecidable. There is no algorithm which solves this problem, nor can anyone ever develop such an algorithm.

Reduction

For showing that a number of tasks of our interest are undecidable, we use a principle called reduction.

If we can solve the Halting problem by solving our tasks, since we know that the Halting problem is undecidable, our task must also be undecidable.