

Database Reformulation Under Strong Storage Space Constraints (research summary)

Rada Chirkova and Michael R. Genesereth

Stanford University, Stanford CA 94305, USA
{rada,genesereth}@cs.stanford.edu

Abstract. *Database reformulation* is the process of rewriting the data and rules of a deductive database in a functionally equivalent manner. We focus on the problem of automatically reformulating a database in a way that reduces query processing time while satisfying strong storage space constraints.

Our research objective is to explore how database reformulation can be automated for databases whose rules are written using successively more expressive query languages, i.e., various extensions of datalog. In addition, we plan to study reformulation of databases with integrity constraints. We have already obtained decidability results for unary databases whose rules can be expressed in nonrecursive datalog with negation [2], as well as for databases of arbitrary arity whose rules can be expressed as conjunctive queries [1].

In the life cycle of a database system, there are recurring problems whose solution involves transformations of the database schema and/or queries defined on the schema. Prominent examples are database design, data model translation, schema (de)composition, view materialization, and multidatabase integration. Interestingly, nearly all these problems can be regarded as aspects of the same problem in a theoretical framework that we proceed to describe.

Consider an abstract database transformation problem. Suppose the input to the problem includes the schema and rules of a deductive database and also a set of elementary queries which, together with some algebra, forms a query language on the database. Suppose the goal of the transformation is to build an “optimal” structure of the database with respect to the requirements and constraints that are also provided in the input.

Generally, we require that transformations of database schema and rules be performed in such a way that the resulting database satisfies three conditions. First, it should be possible to extract from the transformed database, by means of the input query language, exactly the same information as from the original database. Second, the result should satisfy the input requirements, such as minimizing query processing costs. Finally, the result should satisfy the input constraints; one pervasive constraint is a guarantee of a (low) upper bound on the disk space for storing the transformed database. Notice that since the input

does not include a specific database instance, all three conditions must hold for all instances of the input database and of the transformed database.

We have described and formally specified this problem, which we call *database reformulation*; we consider logic-based approaches to its solution. Database reformulation is the process of rewriting the schema and rules of a deductive database in a functionally equivalent manner: it outputs a new schema and new rules, as well as a rewriting of (all elementary queries of) the input query language in terms of the new schema and rules. By specifying various input requirements and constraints, the database reformulation problem translates into any of the database schema/query transformation problems mentioned above.

We focus on database reformulations whose input requirement is to minimize the computational costs of processing the given elementary queries, under strong storage space constraints that guarantee no more than linear increase in database size. In this formulation, the database reformulation framework is most suitable for dealing with the problems of view materialization and multidatabase integration.

We have obtained decidability results for reformulating unary databases whose rules can be expressed in nonrecursive datalog with negation [2], as well as for databases of arbitrary arity whose rules can be expressed as conjunctive queries [1]. Our long-term objective is to extend research in database reformulation to deductive databases whose queries and views are formulated in progressively more general standard query languages (datalog and its extensions), as well as to databases with integrity constraints.

References

1. Rada Chirkova and Michael R. Genesereth. Linearly bounded reformulations of conjunctive databases. *In Proceedings of the Sixth International Conference on Rules and Objects in Databases (DOOD)*, July 2000.
2. Rada Chirkova and Michael R. Genesereth. Linearly bounded reformulations of unary databases. *In Proceedings of the Symposium on Abstraction, Reformulation and Approximation (SARA)*, July 2000.