Abstract

In the book 'The Theory of Relational Databases' written by David Maier in 1983, the following warning can be found in the introduction to a chapter on incomplete information and database semantics: *The reader is warned that the topics in this chapter are matters of personal taste. Whether the definitions and approaches offered seem right depends on individual intuition.*

Although the literature of incomplete information in relational databases is numerous, no approach in this field proposed since the time of that writing (and before) has been considered a final solution. In particular there are many objections to constructs offered by the de facto standard query language SQL for handling incomplete information.

In this thesis we deal with the problem in a way different from most previous approaches. We take a denotational approach insofar as we start with definitions of different kinds of answers to queries by considering queries as functions. The meaning of answers is determined without references to definitions of operations or to the interpretation of logical query expressions in partial database states. Algorithms for the evaluation of queries formulated in concrete query languages are given later according to these definitions.

We restrict ourselves to two possible meanings of missing values: *a value exists but it is unknown* and *there is no information on a value.* After some introductory definitions we discuss several kinds of answers representing sure and maybe information for the first meaning. Information in answers is sure if it is valid independent of unknown values. Maybe information is valid for at least one possible replacement of unknown values by defined values. Two forms for representing missing values are considered: ω–relations where a single special value is on hand for denoting missing values and V–relations where a set of variables is on hand for this purpose. For both forms we consider queries formulated in relational algebra as well as in tuple relational calculus. Following the discussion of the relevant literature, several evaluation methods for different kinds of answers are presented and an undecidability result is shown for ω–relations. In case of incomplete information, completeness of all interesting kinds of answers is tied up with the tautology problem. In order not to lose too much information, multi–values logics are used by some of our efficient evaluation methods. The classical equivalence theorem concerning relational algebra and relational calculus is proven for some extensions to relational algebra operations and modified interpretations of calculus expressions in case of partial relations.

In the last two chapters we concentrate on the second meaning of missing values mentioned above. We show the problems connected with the definition of sure answers under the 'no information' assumption and present a transformation method which allows to 'adorn' SQL–queries in such a way that answers always represent sure information.

Keywords

relational databases, incomplete information, partial relations, null values, query languages, sure answers, maybe answers, relational algebra, relational calculus, multi–valued logic, SQL semantics