Dealing with inconsistencies and incompleteness in database update (position paper)

Giuseppe De Giacomo[1] and Maurizio Lenzerini[1] and Antonella Poggi[1,2] and Riccardo Rosati[1]

[1] Dipartimento di Informatica e Sistemistica "Antonio Ruberti"
Università di Roma "La Sapienza"
Via Salaria 113, I-00198 Roma, Italy
surname@dis.uniroma1.it
[2] INRIA Futurs
Parc Club Orsay-University
4 rue Jean Monod, F-91893 Orsay Cedex, France

Several areas of research and various application domains have been concerned in the last years with the problem of dealing with incomplete databases. Data integration as well as the Semantic Web are notable examples. Surprisingly, while many research efforts have been focusing on several interesting issues related to incomplete databases, as query answering, not much investigation have been done concerning updates. In this position paper we aim at highlighting some of the issues we are dealing with in our work on updates over incomplete databases.

Instance level updates under constraints Our interest in this area stems mainly from the need to deal with updates in Description Logics based ontologies. Description logics (DLs) are logics for expressing the conceptual knowledge about a domain in terms of classes and associations between them [1]. Such logics are currently considered among the most promising formalisms for representing ontologies by the Semantic Web community [2]. DL-based ontologies are often used for accessing data stored in a data layer by means of query answering. Interestingly, the open world assumption is enforced and incomplete information on the data is assumed in this setting. While there is a whole body of research on query answering in such systems, the research on update is very recent [5, 4] . First results show that classical approaches on logical databases update, such as Winslett's [7, 8], are often adequate, as long as one takes into account a clear distinction between intensional (conceptual level) information and extensional (instance level) information typical of this setting. Intensional information, expressed by universal assertions in DLs, is considered immutable, while extensional information, expressed in a form of an incomplete database, is subject to updates. From a database point of view, the setting above requires investigating updates on incomplete databases under a wide variety of constraints, such as keys and foreign keys, more general forms of inclusion dependencies, etc. Thus, such research is revamping interest in updates in databases with incomplete information.

Update expressibility The problem of update expressibility arises [5] as soon as we consider expressive schema and constraints languages. This problem aims at deciding whether given a class of incomplete databases $\mathcal{C}$ satisfying certain constraints and an incomplete database $\mathcal{I}$ belonging to $\mathcal{C}$, the result of an update over $\mathcal{I}$ is expressible through a new incomplete database belonging to $\mathcal{C}$. For many classes $\mathcal{C}$, we have that updates are not expressible. One notable such class is the class $\mathcal{KFK}$ of incomplete databases that are characterized by a relational schema with keys and foreign keys. Notice that, in contrast, query answering in such class is polynomial (actually LOGSPACE) in data complexity [3].

Materialized update vs. virtual update Even if the result of the update is not expressible in a given class $\mathcal{C}$ of incomplete databases, we might be interested in answering queries on the database resulting from the update. Indeed, to do so we do not necessarily need to "materialize" the new state of the database, but actually we could reason on the original database base by taking into account the update in a "virtual" way. In a sense, this is analogous to the distinction between projection via regression vs. progression in reasoning about actions well-known in AI [6]. Along this direction, one interesting problem is understanding whether answering queries after updates over incomplete databases belonging to the previously mentioned class $\mathcal{KFK}$ remains polynomial.

Update and inconsistency Finally, it is worth noticing that updates bring in the general issue of dealing with inconsistency in databases with incomplete information. The standard update semantics addresses the issue of solving inconsistency between the current instance level information (i.e., the data) and what has been asserted by the update, while it does not deal with inconsistencies between the update and intensional level information (i.e., the constraints). It would be interesting to study possible semantics that are tolerant w.r.t. the latter form of inconsistency.

References

1. Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Applications.* Cambridge University Press, 2003.
2. Franz Baader, Ian Horrocks, and Ulrike Sattler. Description logics as ontology languages for the semantic web. In *Mechanizing Mathematical Reasoning: Essays in Honor of Jörg Siekmann on the Occasion of His 60th Birthday*, number 2605 in LNAI, pages 228–248. Springer, 2005.
3. Andrea Calì, Domenico Lembo, and Riccardo Rosati. On the decidability and complexity of query answering over inconsistent and incomplete databases. In *Proc. of PODS 2003*, pages 260–271, 2003.
4. Giuseppe De Giacomo, Maurizio Lenzerini, Antonella Poggi, and Riccardo Rosati. On the update of description logic ontologies at the instance level. Technical report, 2006. Submitted.
5. Hongkai Liu, Carsten Lutz, Maja Milicic, and Frank Wolter. Updating description logic aboxes. In *Proc. of KR 2006*, 2006. To appear.
6. Raymond Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems.* The MIT Press, 2001.
7. M. Winslett. Reasoning about action using a possible models approach. In *Proc. of AAAI'98*, 1988.
8. M. Winslett. *Updating Logical Databases.* Cambridge University Press, 1990.