Inconsistency and Incompleteness in Data Integration: a Logic-based Approach

Andrea Calì

Dip. di Informatica e Sistemistica
Università di Roma "La Sapienza"
Via Salaria 113, I-00198 Roma, Italy
cali@dis.uniroma1.it

In data integration systems, a global schema provides a common representation of the data stored in a set of heterogeneous sources. Integrity constraints are used to enhance the expressivity of the global schema, yet the underlying data may not satisfy such constraints. In this talk we address the problem of query processing in global-as-view integration systems where key and inclusion dependencies are expressed on the global schema, and the sources are considered sound, but not necessarily complete. In particular, we first show the undecidability of query answering in the presence of key and inclusion dependencies, and then provide a rewriting technique for a maximal class of dependencies for which decidability of query answering holds. We adopt an intensional approach, that consists of rewriting the user query according to the integrity constraints, in such a way that the evaluation the rewritten query over the sources returns the correct answers to the original query.