Effective Ontology-Based Data Integration

Riccardo Rosati

Dipartimento di Informatica e Sistemistica
Sapienza Università di Roma
Via Ariosto 25, 00185 Roma, Italy
`rosati@dis.uniroma.it`

The goal of data integration is to provide a uniform access to a set of heterogeneous data sources, freeing the user from the knowledge about where the data are, how they are stored, and how they can be accessed. One of the outcomes of the research work carried out on data integration in the last years is a clear conceptual architecture, comprising a global schema, the source schema, and the mapping between the source and the global schema. In this talk, we present a comprehensive approach to ontology-based data integration. We consider global schemas that are ontologies expressed in OWL, the W3C standard ontology specification language, whereas sources are relations, managed through a data federation tool that wraps the actual data. The mapping language has specific mechanisms for relating values stored at the sources to objects that are instances of concepts in the ontology. By virtue of the careful design that we propose for the various components of a data integration system, answering unions of conjunctive queries can be done through a very efficient technique which reduces this task to standard SQL query evaluation. Finally, we present a management system for ontology-based data integration, called MASTRO-I, which completely implements our approach.