Ontology-based data access with MASTRO

Antonella Poggi and Marco Ruzzi

Dipartimento di Informatica e Sistemistica "Antonio Ruberti"
Università di Roma "La Sapienza" - Via Salaria 113, I-00198 Roma, Italy
surname@dis.uniroma1.it

In several areas, such as Enterprise Application Integration, Data Integration [5], and the Semantic Web [4], clients need to access the services exported by the system, and hence require a representation of the intensional level of the application domain in terms of which they can specify the access to the exported services. One of the most interesting usages of such a shared conceptualization is *ontology-based data access*, where a conceptual layer is exported to the client, abstracting away from how actual data is maintained. Therefore, an ontology-based data access system has to adress the "impedance mismatch" issue, existing between data actually present at the sources, and objects represented at the intensional level.

In this demo, we present MASTRO, an ontology-based data access system that extends the QuOnto system [1] with new capabilities and uses the QuOnto engine at its core. Specifically, MASTRO provides tractable reasoning over an ontology, laying over an autonomous relational DBMS managing the data layer. In what follows, we introduce the main services provided by MASTRO, namely (*i*) ontology specification, (*ii*) query answering, (*iii*) ontology satisfiability, and (*iv*) meta-query answering. The present work is one of the outcomes of the European project TONES[1].

Ontology specification MASTRO allows to define an ontology in terms of both an intensional and an extensional level. The former is specified by means of a new Description Logic in the *DL-Lite* family [3], called *DL-Lite*$_A$ [2]. Thus, the domain of interest is represented in terms of (*i*) concepts, denoting set of objects, (*ii*) roles, denoting binary relations between objects, and, notably, (*iii*) domain values, denoting set of values. Interestingly, values allow both concepts and roles to be qualified by attributes. Concerning the ontology extensional level, MASTRO allows for *mappings* specification, that establish how data retrieved from an existing database $\mathcal{DB}$ is related to extensions of terms used in the ontology intensional level. This is achieved by (*i*) allowing *object terms* to denote constants, that are built by applying Skolem functors to data values, (*ii*) defining mappings of the form: $\Psi \rightsquigarrow \varphi$, where Ψ is an arbitrary SQL query over $\mathcal{DB}$, and φ is a *DL-Lite*$_A$ conjunctive query without existential variables, whose atoms may contain *variable object terms*, i.e. terms obtained by applying Skolem functors to variables denoting values.

Query answering The MASTRO query answering service is similar to query answering provided by QuOnto [1]. In particular, MASTRO is able to answer unions of conjunctive queries (UCQs) expressed over the ontology alphabet: the class of UCQs is one of the most important classes of query arising in practical cases. The query answering process is performed by *query rewriting* and strongly separates the intensional level from

[1] http://www.tonesproject.org/

[2] Please, for details about *DL-Lite*$_A$ constructs and assertions, refer to [2, 6].

the extensional one: user queries are first reformulated on the basis of the TBox assertions, and then evaluated directly over the database by means of the mappings. Hence, the TBox assertions are "compiled" into the query, so the TBox can be disregarded during the query evaluation phase. MASTRO also provides two different mapping handling techniques. The first one exploits the SQL engine of the DBMS managing the data layer of the ontology: a view is defined for every general concept, using the SQL query specified into the mapping assertion. The second one unfolds the query by producing an SQL statement that can be directly issued over the source tables.

Ontology satisfiability Instance checking, concepts subsumption and ontology satisfiability are particularly important reasoning tasks. It can be shown that both instance checking and concepts subsumption can be reduced to ontology satisfiability. The MASTRO system provides an efficient ontology satisfiability algorithm which is also used to decide instance checking and concepts subsumption.

Metaquerying MASTRO maintains the ontology metalevel information into a local data structure called *Metadata Repository* (MDR). Such a metalevel is itself represented by means of a *DL-Lite*$_A$ ontology. This enables the MASTRO system to provide some metalevel reasoning services. In particular, a query answering service is provided for answering queries expressed over the ontology metalevel.

Demo Scenario To illustrate the main features of MASTRO and test services provided by the system, we defined an example ontology representing administrative and didactic information about students and teachers, students' master degree thesis, faculties and courses. We then mapped such ontology over an actual large database instance of the University of Rome "La Sapienza": the database stores information within 27 different tables, with an overall size of 200.000 tuples. We tested the satisfiability of the example ontology and evaluated several queries of practical interest: the results clearly show that MASTRO performs well also with very large underlying databases.

References

1. Andrea Acciarri, Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Mattia Palmieri, and Riccardo Rosati. QUONTO: QUerying ONTOlogies. In *Proc. of AAAI 2005*, pages 1670–1671, 2005.
2. Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, and Riccardo Rosati. Linking data to ontologies: The description logic *dl-lite*$_a$. In *Proc. of the 2nd Workshop on OWL: Experiences and Directions (OWLED 2006)*, 2006.
3. Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. DL-Lite: Tractable description logics for ontologies. In *Proc. of AAAI 2005*, pages 602–607, 2005.
4. Jeff Heflin and James Hendler. A portrait of the Semantic Web in action. *IEEE Intelligent Systems*, 16(2):54–59, 2001.
5. Maurizio Lenzerini. Data integration: A theoretical perspective. In *Proc. of PODS 2002*, pages 233–246, 2002.
6. Antonella Poggi. *Structured and Semi-Structured Data Integration*. PhD thesis, Dip. di Inf. e Sist., Univ. di Roma "La Sapienza", 2006.