MASTRO: Ontology-Based Data Access at Work (Extended Abstract)

Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Riccardo Rosati, Marco Ruzzi, and Domenico Fabio Savo

Sapienza Università di Roma – Dip. di Ing. Informatica, Automatica e Sistemistica
lastname@dis.uniroma1.it

In this paper we present the current version of MASTRO, a system for ontology-based data access (OBDA) developed at Sapienza Università di Roma. MASTRO allows users for accessing external data sources by querying an ontology expressed in a fragment of the W3C Web Ontology Language (OWL). As in data integration [5], mappings are used in OBDA to specify the correspondence between a unified view of the domain (called global schema in data integration terminology) and the data stored at the sources. The distinguishing feature of the OBDA approach, however, is the fact that the global schema is specified using an ontology language, which typically allows to provide a rich conceptualization of the domain of interest, independently from the source representation.

In the current version of MASTRO, ontologies are specified in *DL-Lite*$_{A,id,den}$, a logic of the *DL-Lite family* of tractable Description Logics, which is specifically tailored for OBDA and is at the basis of OWL 2 QL, one of the profiles of OWL 2, the current W3C standard language for specifying ontologies. *DL-Lite*$_{A,id,den}$ captures the main modeling features of a variety of representation languages, such as basic ontology languages and conceptual data models [6]. Furthermore, it allows for specifying advanced forms of identification and denial assertions [4], which are not part of OWL 2, but are very useful in practice.

Answering unions of conjunctive queries (UCQs) in MASTRO can be done through a technique that reduces this task to standard SQL query evaluation. The technique is purely intensional and is performed in three steps:

1. *TBox rewriting:* The first step rewrites the input UCQ according to the ontology. The rewriting, performed using the Presto algorithm [7], produces as output a non-recursive Datalog program, which encodes the intensional knowledge expressed by the ontology and the user query. The output Datalog program contains the definition of auxiliary predicates, not belonging to the alphabet of the ontology.
2. *Datalog Unfolding:* The output of the first step is then unfolded into a new UCQ by means of the *Datalog Unfolding* algorithm. It consists of a classic rule unfolding technique which eliminates all the auxiliary predicate symbols introduced by the Presto algorithm and produces a final UCQ expressed in terms of ontology concepts, roles, and attributes.
3. *Mapping Unfolding:* The last step takes the unfolded UCQ and the mapping assertions as input and produces an SQL query which can be directly evaluated over the data sources. In particular, the mapping assertions are first *split* into assertions of a simpler form, in which the head of every mapping

P. Herrero et al. (Eds.): OTM 2012 Workshops, LNCS 7567, pp. 667–668, 2012.

assertion contains only a single ontology predicate; then, the final reformulation is produced through a mapping unfolding step, as described in [6].

Notice that the use of Presto is one of the key features of the current version of Mastro wrt previous ones [1]. Presto is an optimization of the well-known PerfectRef [2] algorithm. The latter may indeed produce rewritings consisting of a lot of UCQs, many of which are possibly redundant. Presto instead rewrites the user query into a Datalog program whose rules encode only necessary rewriting steps, thus preventing the generation of useless rules.

The usefulness of OBDA and the efficiency of the Mastro system are showed by several industrial applications in which it has been experimented [3], e.g. in collaboration with SELEX Sistemi Integrati, Banca Monte dei Paschi di Siena and Accenture. As an example, we briefly report on the experiments carried out with the Italian Ministry of Economy and Finance (MEF). The main objectives of the project have been: the design and specification in $DL\text{-}Lite_{A,id,den}$ of an ontology for the domain of the Italian public debt; the realization of the mapping between the ontology and relational data sources that are part of the management accounting system currently in use at the ministry; the definition and execution of queries over the ontology aimed at extracting data of core interest for MEF users. In particular, the information returned by such queries relates to sales of bonds issued by the Italian government and is at the basis of various reports on the overall trend of the national public debt. The Italian public dept ontology is over an alphabet containing 164 atomic concepts, 47 atomic roles, 86 attributes, and comprises around 1440 ontology assertions. The 300 mapping assertions involve around 60 relational tables managed by Microsoft SQLServer. We tested a very high number of queries and produced through Mastro several reports of interest for the ministry.

An extended version of this abstract can be found in [3].

References

1. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodriguez-Muro, M., Rosati, R., Ruzzi, M., Savo, D.F.: The Mastro system for ontology-based data access. Semantic Web J. 2(1), 43–53 (2011)
2. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. J. of Automated Reasoning 39(3), 385–429 (2007)
3. De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rosati, R., Ruzzi, M., Savo, D.F.: Mastro: A reasoner for effective ontology-based data access. In: Proc. of ORE 2012. CEUR, vol. 858 (2012), `ceur-ws.org`
4. Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., Savo, D.F.: Inconsistency-tolerant first-order rewritability of DL-Lite with identification and denial assertions. In: Proc. of DL 2012. CEUR, vol. 846 (2012), `ceur-ws.org`
5. Lenzerini, M.: Data integration: A theoretical perspective. In: Proc. of PODS 2002, pp. 233–246 (2002)
6. Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking Data to Ontologies. In: Spaccapietra, S. (ed.) Journal on Data Semantics X. LNCS, vol. 4900, pp. 133–173. Springer, Heidelberg (2008)
7. Rosati, R., Almatelli, A.: Improving query answering over *DL-Lite* ontologies. In: Proc. of KR 2010, pp. 290–300 (2010)