

The Mastro OBDA plug-in for Protégé

Giacomo Ronconi, Marco Ruzzi,
Valerio Santarelli, Domenico Fabio Savo

Sapienza Università di Roma
(lastname)@dis.uniroma1.it

OBDA Systems
(lastname)@obdasystems.com

Abstract. Ontology-based data access (OBDA) is a recent paradigm for accessing large data repositories through an ontology, that is a formal description of a domain of interest. In this work, we present the Mastro plug-in for the standard ontology editor Protégé, which allows users to create a full OBDA specification and pose SPARQL queries over the ontology language to extract data from the underlying data sources.

1 Introduction

Ontology-based Data Access (OBDA) is a recent paradigm for accessing data sources through the mediation of a conceptual domain view, given in terms of an ontology [12]. OBDA features a three-level architecture composed of the ontology, which provides a formal description of the domain of interest, the data sources used in organizations for storing their information, and the mapping used to specify the semantic relationships between the ontology layer and the data sources. Hence, OBDA can be seen as a form of information integration, where the usual global schema [11] is replaced by the ontology, expressed in a logic-based language, which offers a semantically rich description of the concepts and relationships between them in the domain of interest. Currently the two most popular systems for OBDA are Mastro [3] and Ontop [2].

Mastro is developed by OBDA Systems¹ and Sapienza University of Rome, and has been used in recent years in numerous projects with important business partners from the private and public sectors [1,13]. Mastro comes with its own commercial application, the Mastro Studio system [5], and now, in this demonstration, we officially introduce the Mastro plug-in for the popular Protégé [10] editor for OWL ontologies.

Ontologies in Mastro are specified through languages belonging to the *DL-Lite* [4] family of lightweight Description Logics, and support mappings in the standard R2RML format [8], and also in their native mapping language. Data sources are seen as relational databases and can be accessed through SPARQL queries over the ontology, exploiting the query answering services provided by Mastro.

To illustrate the main features of the Mastro Protégé plug-in, we will invite attendees of the demo to experiment it on an OBDA specification extracted from *Sapientia*,

¹ www.obdasystems.com

the Ontology of Multi-dimensional Research Assessment, developed within a project funded by Sapienza University of Rome [7].

The Mastro Protégé plug-in is available at <http://www.obdasystems.com>.

2 Technical Specifications

In this section, we provide a brief overview of the theoretical background of the OBDA approach, and introduce the main functionalities of the Mastro Protégé plug-in.

An OBDA specification is a triple $\langle \mathcal{O}, \mathcal{M}, \mathcal{D} \rangle$, where \mathcal{O} is a Description Logic ontology; \mathcal{D} is a relational database which models the resources where data are stored; and \mathcal{M} is a set of mapping assertions. In Mastro, \mathcal{O} is expressed in a logic of the *DL-Lite* family [4] of lightweight DLs which is specifically designed for the tractability requirement in OBDA, and is the logic underpinning the OWL 2 QL profile of OWL 2. \mathcal{M} is partitioned into the sets \mathcal{M}_v and \mathcal{M}_o , where \mathcal{M}_v is constituted by a set of *SQL views* over the database, which are assertions of the form $q_{DB}(\mathbf{x}) \rightsquigarrow v(\mathbf{x})$, where $v(\mathbf{x})$ is a view name for the SQL query q_{DB} over the database \mathcal{D} , and \mathcal{M}_o is a set of *ontology predicate mappings*, i.e., assertions of the form $q_v(\mathbf{y}) \rightsquigarrow P(\mathbf{x})$, where $\mathbf{x} \subseteq \mathbf{y}$, P is an ontology predicate, and q_v is a conjunctive query over the set of views in \mathcal{M}_v [9]. Query answering is performed in Mastro via *query reformulation* [4,12]: the user's query over the ontology \mathcal{O} is rewritten into an SQL query that encodes the knowledge expressed in the ontology and is expressed over the database \mathcal{D} .

The partition of \mathcal{M} in ontology predicate mappings and SQL views allows to limit the size of the final rewritten query, and to exploit further optimizations of the rewritten query based on particular assertions that the OBDA designer can specify over the views. For instance, the designer can specify inclusion assertions between (projections of) the view predicates, which are exploited by Mastro in the rewriting process for eliminating queries contained into other queries.

Mastro's plug-in for Protégé provides the user with a fullfledged environment to define an OBDA specification for Mastro, and to access its query answering service. Along with these core features, the plug-in is equipped with other functionalities among which are: (i) the approximation module that is based on the semantic approach presented in [6] for approximating OWL 2 ontologies in *DL-Lite*, the language supported by Mastro. This module is crucial because it allows to load OWL 2 ontologies in Protégé, while using its *DL-Lite* representation for query answering with Mastro; (ii) in order to comply with the current recognized standards of the Semantic Web, Mastro is able to import R2RML mapping into its proprietary format and viceversa, and to export the results of the queries over the ontology into RDF.

3 Overview of the Plug-in

The components of Mastro's plug-in are organized in five sub-tabs under the main Mastro tab, one item in the "Reasoner" menú, and one "Mastro" menú item. Below we briefly describe each component.

Configuration. The Configuration sub-tab is used to create, open, and save a mapping specification, and to define the jdbc parameter connections to the source relational

database, and to a database which Mastro uses handle query executions and to store all their information, e.g., results, execution time, ontology and mapping rewritings, and Mastro configuration parameters.

SQL Views. The SQL views sub-tab is used to create and inspect the views defined in the specification, and to specify assertions over these views, i.e., inclusion assertions, disjointness assertions, and key dependency assertions. Each view is defined by choosing its name and the SQL query code.

Mappings. The Mappings sub-tab is used to create and inspect the ontology predicate mappings in the specification, and to define the IRI templates that Mastro uses to build the answers from the data in the data source. Each ontology predicate mapping is identified by an ID, and contains the ontology predicate that is being mapped, a conjunctive query over the SQL views, and, optionally, a description.

SPARQL Query. The SPARQL Query sub-tab allows to define and execute a SPARQL query through Mastro over the ontology. Once the query has been executed, the user can consult its results and its ontology and mapping rewritings, and export the results in RDF or CSV formats. The user can also save the query into a catalog, together with a description. The catalog is managed through a catalog file which can be loaded, exported, and edited in the sub-tab. Additionally, Mastro saves all query executions into an execution log, allowing the user to visualize those of the selected query.

SPARQL Query Execution Log. The SPARQL Query Execution Log sub-tab provides information regarding all executions of SPARQL queries over an OBDA specification. The data shown in the log can be exported, and the queries re-executed.

The items in the Mastro menú instead allow to see the details of the approximation performed by the semantic approximation module (*Approximation details*), in terms of approximated, translated, and rejected axioms of the original OWL 2 ontology, to access a panel where Mastro's optimizations can be configured (*Mastro properties*), to export Mastro's mappings in R2RML (*Export mappings to R2RML*), and to consult a tutorial of the Mastro plug-in (*Help*).

4 Application Scenario and Demo Session Overview

We demonstrate the Mastro Protégé plug-in through the Ontology of Multi-dimensional Research Assessment, or *Sapientia* [7], an OBDA specification that has been developed and is currently used within a project funded by Sapienza University of Rome.

Sapientia models aspects of assessing research activities and their impact on human knowledge and the economic system. For example, it deals with inter-relationships between research activities, between research activities and peoples personal knowledge, and between research activities and other missions of individuals and institutions. The ontology is composed of fourteen modules, which formalize a wide array of activities, ranging from teaching, publishing, and research, to funding and preservation.

In the project, the adoption of the OBDA approach presents several advantages with respect to traditional data integration, in particular with respect to conceptual access to data, re-usability, documentation and standardization, flexibility, and extensibility.

During the demonstration, attendees will be able to inspect, query and edit the Sapientia OBDA specification through the Mastro plug-in. Initially, we will show how a user

can inspect and edit the ontology through the features offered by Protégé. Then, we will introduce the Mappings and SQL Views tabs, illustrating how the ontology is linked to the data sources through the mapping assertions, and showing how to edit or create new mappings and views. Furthermore, we highlight all the querying functionalities offered by the plugin through the SPARQL query tab and its execution log. Lastly, we will show attendees the details of the approximation procedure of OWL 2 into *DL-Lite*.

Acknowledgments. Work supported by MIUR under the SIR project “MODEUS” grant n. RBSI14TQHQ.

References

1. N. Antonioni, F. Castanò, C. Civili, S. Coletta, S. Grossi, D. Lembo, M. Lenzerini, A. Poggi, D. F. Savo, and E. Virardi. Ontology-based data access: The experience at the italian department of treasury. In *Proc. of CAISE 2013*, volume 1017 of *CEUR Workshop Proceedings*, pages 9–16, 2013.
2. D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodríguez-Muro, and G. Xiao. Ontop: Answering SPARQL queries over relational databases. *Semantic Web*, 8(3):471–487, 2017.
3. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodríguez-Muro, R. Rosati, M. Ruzzi, and D. F. Savo. The Mastro system for ontology-based data access. *Semantic Web J.*, 2(1):43–53, 2011.
4. D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
5. C. Civili, M. Console, G. De Giacomo, D. Lembo, M. Lenzerini, L. Lepore, R. Mancini, A. Poggi, R. Rosati, M. Ruzzi, V. Santarelli, and D. F. Savo. MASTRO STUDIO: Managing ontology-based data access applications. *PVLDB*, 6:1314–1317, 2013.
6. M. Console, J. Mora, R. Rosati, V. Santarelli, and D. F. Savo. Effective computation of maximal sound approximations of description logic ontologies. In *Proc. of ISWC 2014*, volume 8797 of *LNCS*, pages 164–179. Springer, 2014.
7. C. Daraio, M. Lenzerini, C. Leporelli, P. Naggar, A. Bonaccorsi, and A. Bartolucci. The advantages of an ontology-based data management approach: openness, interoperability and data quality. *Scientometrics*, 108(1):441–455, 2016.
8. S. Das, S. Sundara, and R. Cyganiak. R2RML: RDB to RDF mapping language. W3C Recommendation, World Wide Web Consortium, Sept. 2012. Available at <http://www.w3.org/TR/r2rml/>.
9. F. Di Pinto, D. Lembo, M. Lenzerini, R. Mancini, A. Poggi, R. Rosati, M. Ruzzi, and D. F. Savo. Optimizing query rewriting in ontology-based data access. In *Proc. of EDBT 2013*, pages 561–572. ACM Press, 2013.
10. J. H. Gennari, M. A. Musen, R. W. Fergerson, W. E. Grosso, M. Crubézy, H. Eriksson, N. F. Noy, and S. W. Tu. The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-computer studies*, 58(1):89–123, 2003.
11. M. Lenzerini. Data integration: A theoretical perspective. In *Proc. of PODS 2002*, pages 233–246, 2002.
12. A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. *J. on Data Semantics*, X:133–173, 2008.
13. D. F. Savo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodríguez-Muro, V. Romagnoli, M. Ruzzi, and G. Stella. MASTRO at work: Experiences on ontology-based data access. In *Proc. of DL 2010*, volume 573 of *CEUR*, ceur-ws.org, pages 20–31, 2010.