Querying Data through Ontologies

Andrea Calì[1], Andreas Pieris[2]

[1] Birkbeck, University of London
[2] University of Oxford
andrea@dcs.bbk.ac.uk, andreas.pieris@cs.ox.ac.uk

An *ontology* is an explicit specification of a conceptualization of an area of interest, and consists of a formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. The need for semantically enhancing existing databases with ontological onstraints gave rise to the so-called *ontological database management systems*, that is, a new type of DMBSs equipped with advanced reasoning and query processing mechanisms. In particular, an extensional database D is combined with an ontology Σ which derives new intensional knowledge from D. An input query is not just answered against the database, as in the classical setting, but against the logical theory (a.k.a. ontological database) $D \cup \Sigma$. Database technology providers have recognized the need for combining ontological reasoning and database technology, and have recently started to build ontological reasoning modules on top of their existing software with the aim of delivering effective database management solutions to their customers. For example, Oracle Inc. offers a system, called Oracle Database 11g, enhanced by modules performing ontological reasoning tasks[1]. Enhancing databases with ontologies is also at the heart of several research-based systems such as QuOnto [Acciarri et al. 2005].

Description Logics (DLs) [Baader et al. 2003] are a family of knowledge representation languages widely used in ontological modeling. In fact, DLs model a domain of interest in terms of concepts and roles, which represent classes of individuals and binary relations on classes of individuals, respectively. Interestingly, DLs provide the logical underpinning for the Web Ontology Language (OWL), and its revision OWL 2, as standartized by the W3C[2]. However, in order to achieve favorable computational properties, DLs are able only to describe knowledge for which the underlying relational structure is treelike. Moreover, they usually support only unary and binary relations. Recently, the Datalog$^{\pm}$ family [Calì et al. 2012] of ontology languages has been proposed, with the purpose of overcoming the above limitations of DLs. Datalog$^{\pm}$ languages are based on Datalog rules that allow for the existential quantification of variables in the head, in the same fashion as Datalog with *value invention* [Cabibbo 1998]. The absence of value invention, thoroughly discussed in [Patel-Schneider and Horrocks 2007], is the main shortcoming of Datalog in modeling ontologies. The basic Datalog$^{\pm}$ rules are (function-free) Horn rules extended with existential quantification in the head, known as *tuple-generating dependencies* or *existential rules*. The addition of *negative constraints* of the form $\forall \mathbf{X}\, \varphi(\mathbf{X}) \rightarrow \bot$, where $\bot$ denotes the truth constant *false*, and of restricted classes of *equality-generating dependencies* such as *key dependencies*, makes Datalog$^{\pm}$ expressive enough to capture the most common tractable ontology languages such as the *DL-Lite* family of DLs [Calvanese et al. 2007].

Query answering under Datalog rules extended with existential quantification in the head is undecidable (implicit in [Beeri and Vardi 1981]). Therefore, some syntactic restriction is needed to ensure decidability (hence the symbol "±"). The fundamental restriction paradigms that have been studied so far are as follows:

---

[1] http://www.oracle.com/technetwork/database/enterprise-edition/overview/index.html
[2] http://www.w3.org/TR/owl2-overview/

Weak-Acyclicity. Weakly-acyclic Datalog$^{\pm}$, introduced in the context of data exchange [Fagin et al. 2005], guarantees the finiteness of the universal model of the given ontological database. Notice that a universal model (a.k.a. canonical model) acts as a representative of all the models of the given ontological database, and thus for query answering purposes we can consider only this special model. Therefore, one can just evaluate the given query over the finite universal model.

Guardedness. Decidability of query answering under guarded Datalog$^{\pm}$ follows from the fact that guardedness ensures the existence of a universal model of finite treewidth. Guarded Datalog$^{\pm}$ rules and generalizations thereof were studied in [Calì et al. 2013; Calì et al. 2012; Baget et al. 2011]. Notice that sets of guarded Datalog$^{\pm}$ rules can be rewritten as theories in the guarded fragment of first-order logic [Andréka et al. 1998].

Stickiness. Sticky Datalog$^{\pm}$ has been proposed in [Calì et al. 2012] with the aim of identifying an expressive language that allows for joins in rule-bodies which are expressible only via non-guarded rules. Stickiness enures the termination of resolution-based procedures, which in turn implies that we can construct the (finite) part of the universal model which is needed in order to entail the query.

Shyness. Shy Datalog$^{\pm}$ ensures that during the construction of the universal model only database onstants can participate in a join operation [Leone et al. 2012]. This property allows us to consider only a finite part of the (possibly infinite) universal model which can be effectively constructed.

Towards the identification of even more expressive languages, several attempts have been conducted to consolidate the aforementioned paradigms. Notable formalisms are *glut-guardedness* [Krötzsch and Rudolph 2011] and *weak-stickiness* [Calì et al. 2012], obtained by joining weak-acyclicity with guardedness and stickiness, respectively. A condition, called *tameness*, which allows for a safe combination of guardedness and stickiness has been recently proposed in [Gottlob et al. 2013].

Acknowledgments. Andrea Calì acknowledges support by the EPSRC project "Logic-based Integration and Querying of Unindexed Data" (EP/E010865/1).

REFERENCES

Acciarri, A., Calvanese, D., Giacomo, G. D., Lembo, D., Lenzerini, M., Palmieri, M., and Rosati, R. QuOnto: Querying ontologies. In *Proc. of AAAI.* pp. 1670–1671, 2005.

Andréka, H., van Benthem, J., and Németi, I. Modal languages and bounded fragments of predicate logic. *J. Phil. Log.* vol. 27, pp. 217–274, 1998.

Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., editors. *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge University Press, 2003.

Baget, J.-F., Leclère, M., Mugnier, M.-L., and Salvat, E. On rules with existential variables: Walking the decidability line. *Artif. Intell.* 175 (9-10): 1620–1654, 2011.

Beeri, C. and Vardi, M. Y. The implication problem for data dependencies. In *Proc. of ICALP.* pp. 73–85, 1981.

Cabibbo, L. The expressive power of stratified logic programs with value invention. *Inf. Comput.* 147 (1): 22–56, 1998.

Calì, A., Gottlob, G., and Kifer, M. Taming the infinite chase: Query answering under expressive relational onstraints. *J. Artif. Intell. Res.*, 2013. To appear.

Calì, A., Gottlob, G., and Lukasiewicz, T. A general Datalog-based framework for tractable query answering over ontologies. *J. Web Sem.* vol. 14, pp. 57–83, 2012.

Calì, A., Gottlob, G., and Pieris, A. Ontological query answering under expressive entity-relationship schemata. *Inf. Syst.* 37 (4): 320–335, 2012.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., and Rosati, R. Tractable reasoning and efficient query answering in description logics: The DL-lite family. *J. Autom. Reasoning* 39 (3): 385–429, 2007.

Fagin, R., Kolaitis, P. G., Miller, R. J., and Popa, L. Data exchange: Semantics and query answering. *Theor. Comput. Sci.* 336 (1): 89–124, 2005.

Gottlob, G., Manna, M., and Pieris, A. Combining decidability paradigms for existential rules. *TPLP*, 2013. To appear.

Krötzsch, M. and Rudolph, S. Extending decidable existential rules by joining acyclicity and guardedness. In *Proc. of IJCAI.* pp. 963–968, 2011.

Leone, N., Manna, M., Terracina, G., and Veltri, P. Efficiently computable Datalog$^{\exists}$ programs. In *Proc. of KR*, 2012.

Patel-Schneider, P. F. and Horrocks, I. A comparison of two modelling paradigms in the semantic web. *J. Web Sem.* 5 (4): 240–250, 2007.