Datalog$^\pm$: A Unified Approach to Ontologies and Integrity Constraints*

Andrea Calì[2,1], Georg Gottlob[1,2], and Thomas Lukasiewicz[1,‡]

[1] Computing Laboratory, University of Oxford, UK
firstname.lastname@comlab.ox.ac.uk

[2] Oxford-Man Institute of Quantitative Finance, University of Oxford, UK

In this talk, we survey the results of two recent works [1, 2] on a family of expressive extensions of Datalog, called Datalog$^\pm$, towards query answering over ontologies.

Ontologies are fundamental to the Semantic Web. They also proved to be useful, due to their flexibility and expressive power, in databases, especially in data modeling and data integration. Among ontology formalisms, description logics (DLs) have been playing a prominent role in the last decade, especially in the Semantic Web. Currently, much research on DLs is directed towards scalable and efficient query answering over ontologies. In particular, the DLs of the *DL-Lite* family [5, 7] are the most common DLs in the Semantic Web and databases that allow for tractable query answering.

Rules in Datalog$^\pm$ are rules in Datalog that additionally admit existentially quantified variables in the head, but on which restrictions are enforced on the body to guarantee desirable decidability and tractability properties. More concretely, Datalog$^\pm$ rules are tuple-generating dependencies (TGDs) for which the chase [6] does not terminate, but for which query answering is nonetheless decidable in general and tractable in many cases in the data complexity. Datalog$^\pm$ is divided into the sublanguages of *guarded*, *linear*, and *weakly guarded* Datalog$^\pm$, which have so-called *guarded TGDs (GTGDs)*, *linear TGDs (LTGDs)*, and *weakly guarded TGDs (WGTGDs)* as rules, respectively.

We characterize the complexity of query answering for all three sublanguages of Datalog$^\pm$. The results are summarized in Tables 1 and 2.

We then further enrich Datalog$^\pm$ with additional features, which serve to represent ontology languages. In particular, we add negative constraints, and we show that the introduction of such constraints does not increase the complexity of query answering. As a second extension, we add *non-conflicting keys*, a generalization of the class in [4], which are special equality-generating dependencies (EGDs) that do not interact with TGDs, and thus also do not increase the complexity of query answering in Datalog$^\pm$.

We next show that Datalog$^\pm$ allows for expressing the most common tractable ontology languages. More precisely, linear Datalog$^\pm$ with negative constraints and non-conflicting keys, called Datalog$^\pm_0$, can be used for query answering in *DL-Lite*$_\mathcal{A}$ [7] in a natural and unified way, being also strictly more expressive than *DL-Lite*$_\mathcal{A}$. Furthermore, weakly guarded Datalog$^\pm$ with a single non-conflicting key can be used for query answering in F-Logic Lite ontologies. Other DLs of the *DL-Lite* family [5] (such

* This work is a short version of [3].

‡ Alternative address: Institut für Informationssysteme, Technische Universität Wien, Favoritenstraße 9-11, 1040 Wien, Austria.

BCQ type	LTGDs	GTGDs	WGTGDs
general	PSPACE-complete	2EXPTIME-complete	2EXPTIME-complete
bounded width, fixed, and atomic	PSPACE-complete	2EXPTIME-complete	2EXPTIME-complete

Table 1. Summary of complexity results: variable set of TGDs.

BCQ type	LTGDs	GTGDs	WGTGDs
general	NP-complete	NP-complete	EXPTIME-complete
bounded width, fixed, and atomic	in AC_0	PTIME-complete	EXPTIME-complete

Table 2. Summary of complexity results: fixed set of TGDs.

as *DL-Lite*$_{\mathcal{F}}$ and *DL-Lite*$_{\mathcal{R}}$) can be similarly translated to Datalog$_0^{\pm}$. Since *DL-Lite*$_{\mathcal{R}}$ is able to fully capture (the DL fragment of) RDF Schema, Datalog$_0^{\pm}$ is also able to fully capture (the DL fragment of) RDF Schema.

We finally describe an extension of Datalog$^{\pm}$ with stratified negation, providing a canonical model and a perfect model semantics, and showing that the two coincide. This also provides a natural stratified negation for query answering over ontologies and the *DL-Lite* family, which has been an open problem to date.

Acknowledgments. The work of Andrea Calì and Georg Gottlob was supported by the EPSRC grant EP/E010865/1 "Schema Mappings and Automated Services for Data Integration". G. Gottlob, whose work was partially carried out at the Oxford-Man Institute of Quantitative Finance, gratefully acknowledges support from the Royal Society as the holder of a Royal Society-Wolfson Research Merit Award. Thomas Lukasiewicz's work was supported by the German Research Foundation under the Heisenberg Programme.

References

1. A. Calì, G. Gottlob, and M. Kifer. Taming the infinite chase: Query answering under expressive relational constraints. In *Proc. KR-2008*, pp. 70–80. AAAI Press, 2008. Revised version: http://benner.dbai.tuwien.ac.at/staff/gottlob/CGK.pdf.
2. A. Calì, G. Gottlob, and T. Lukasiewicz. A general Datalog-based framework for tractable query answering over ontologies. In *Proc. PODS-2009*. ACM Press, 2009 (in press).
3. A. Calì, G. Gottlob, and T. Lukasiewicz. Datalog$^{\pm}$: A unified approach to ontologies and integrity constraints. In *Proc. ICDT-2009*, pp. 14–30. ACM Press, 2009.
4. A. Calì, D. Lembo, and R. Rosati. On the decidability and complexity of query answering over inconsistent and incomplete databases. In *Proc. PODS-2003*, pp. 260–271, 2003.
5. 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. Autom. Reasoning*, 39(3):385–429, 2007.
6. D. Maier, A. O. Mendelzon, and Y. Sagiv. Testing implications of data dependencies. *ACM Trans. Database Syst.*, 4(4):455–469, 1979.
7. A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati. Linking data to ontologies. *J. Data Semantics*, 10:133–173, 2008.