Thesis Proposal

Query Explanation and Revision in Federated Ontology-Mediated Query Answering

Pierre Bourhis <pierre.bourhis@inria.fr>
Antoine Amarilli <a3nm@a3nm.net>
Location: INRIA Lille, Spirals Team; Lille, France

General background. To query large volumes of data, we often need to integrate different heterogeneous systems. The approach called *ontology-mediated query answering* or *OMQA* [1, 3, 5] proposes to address this problem and connect together the data sources with an *ontology*, i.e., a set of logical rules describing the semantics of data. This allows query evaluation to be performed by reasoning with these logical constraints and with the data sources. In particular, we can proceed by *forward chaining*, i.e., deriving the consequences of data using rules; or by *backward chaining*, i.e., rewriting the query so that it can be directly evaluated on the data sources.

Goals of the PhD. The goal of this PhD is to study how we can explain and revise the query answers provided by an OMQA system, by taking into account both the data and the logical rules. To do this, we propose to first develop foundations of these problems, in particular by trying to extend the notion of *provenance* to our context. The formalism of provenance was originally developed for relational databases [2] to describe from where query results originate, i.e., which data was used to produce them. Provenance has already been used to explain and revise the answers to queries in the context of databases. However, explanation and revision of query answers poses new challenges when studied for OMQA systems: we must then also account for the logical rules that have been used when producing the answers, and we must also be able to revise these rules in addition to revising the data. Furthermore, new challenges arise in the notion of *federated OMQA*, where data is integrated from a variety of heterogeneous data sources. In such systems, explanations must be phrased in a way that properly handles the heterogeneity of the data, and revisions to the data must be expressed in a way that is local to each data source.

This PhD can consist of theoretical work on these topics, but could also lead to an implementation of some of the solutions or methods developed as part of the theoretical research. In particular, a prototype could be based on the ProvSQL system [4], which is an implementation of provenance computation as a module of the PostgreSQL relational database engine. Possible application contexts and collaborations can include all members of the ANR CQFD project, which focuses on query evaluation and reasoning over heterogeneous and federated data stores, and which provides the funding of this PhD position.

Environment and supervision. The candidate will be supervised by Pierre Bourhis[1] (CNRS) and will take place in the SPIRALS team of INRIA Lille. The candidate will be co-supervised by Antoine Amarilli[2], associate professor at Télécom Paris, in Palaiseau (south of Paris). The PhD will be funded by the ANR CQFD project for a duration of 3 years, including travel funds for trips

[1] http://cristal.univ-lille.fr/~bourhis/
[2] https://a3nm.net/

between INRIA Lille and Télécom Paris, for conferences and for project meetings. The PhD should start in October 2020 at the latest; it can also start earlier as the funding is already available.

The PhD will preferably be preceded by an internship (for instance a Master's internship) of 4-6 months.

Applications should be sent by email to the advisors and will be reviewed on a rolling basis.

References

[1] J.-F. Baget, M.-L. Mugnier, S. Rudolph, and M. Thomazo. Walking the complexity lines for generalized guarded existential rules. In *IJCAI*, 2011. http://ijcai.org/Proceedings/11/Papers/126.pdf.

[2] T. J. Green, G. Karvounarakis, and V. Tannen. Provenance semirings. In *PODS*, 2007. http://repository.upenn.edu/cgi/viewcontent.cgi?article=1022&context=db_research.

[3] M. König, M. Leclère, M.-L. Mugnier, and M. Thomazo. Sound, complete and minimal UCQ-rewriting for existential rules. *Semantic Web*, 6(5):451–475, 2015. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.764.9248&rep=rep1&type=pdf.

[4] P. Senellart. ProvSQL, 2019. https://github.com/PierreSenellart/provsql.

[5] M. Thomazo. Ontology based query answering with existential rules. In *IJCAI*, 2013. http://ijcai.org/Proceedings/13/Papers/509.pdf.