Internship Proposal

Provenance for Ontology-Mediated Query Answering

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

General background. To answer queries on large volumes of data, we often need to integrate different heterogeneous systems. The approach called *ontology-mediated query answering* [1, 3, 4] 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. Query evaluation can then 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 internship. The goal of this internship is to study how we can explain 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 extend OMQA using the notion of *provenance*, which was originally developed for relational databases [2]. Provenance is a general formalism to describe from where query results originate, i.e., which data was used to produce them. The challenge is to propose new definitions of provenance for OMQA, and in particular to extend its definition to take into account the logical rules that have been used. One first idea can be to study definitions of the so-called Boolean provenance, and determine the complexity of computing it, e.g., by proposing techniques that adapt forward or backward chaining. A more ambitious goal would be to propose non-Boolean definitions of provenance, e.g., for general semirings. Another possible extension is to study how we can update the query answers and their provenance when the underlying data is modified, and how we can revise data and rules to incorporate user feedback.

Environment and supervision. The internship will last for 4–6 months and be supervised by Pierre Bourhis[1] (CNRS) and Antoine Amarilli[2] (associate professor at Télécom Paris), in the context of the ANR CQFD project. The internship can be located in INRIA Lille or at Télécom Paris in Palaiseau (south of Paris), to be discussed with the applicant. This internship can lead to a PhD thesis funded as part of the CQFD project.

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.

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

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