Internship Proposal

Knowledge Compilation and Enumeration for Efficient Query Evaluation Algorithms

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. Over the past decades, the database and knowledge representation communities have converged to expressive formalisms to describe queries and reasoning. They have studied how to evaluate complex queries on large databases [2] and perform reasoning at scale in the context of *ontology-mediated query answering* [5, 8, 10].

However, even if these approaches are very efficient in the size of the input data, the set of query answers to compute can be prohibitively large. For this reason, the database theory community has focused in recent years on efficient *enumeration* of query results. In this approach, the database is first preprocessed to compute a compact representation of all solutions. Second, the solutions can be enumerated one after the other, with a small *delay* between answers. This fine study of complexity makes it possible to design algorithms that can return a first answer, or the first few answers, as efficiently as possible.

In 2017, we proposed [3] a framework for efficient enumeration algorithms, via the field of *knowledge compilation*. This area studies families of Boolean and arithmetic circuits under different syntactic and semantic restrictions, to ensure the tractability of basic operations on these circuits. We showed that enumeration algorithms could be designed in a modular fashion. First, express the set of all solutions as a circuit in a tractable knowledge compilation formalism. Second, design general-purpose enumeration algorithms on them. The point is that circuits serve as an intermediate representation that neatly separate the algorithmic problems of enumeration from the logical problems of reasoning and query evaluation. Further, circuits can also serve for other tasks, e.g., explanation of query results [7, 4], counting, etc.

Goals of the internship. The goal of this internship is to expand this approach to the latest developments in reasoning and query evaluation. Specifically, one first goal is to design circuit representations and enumeration algorithms for recursive query languages, such as regular path queries (already studied recently in [6]) and Datalog [1]. A second goal is enumeration for query evaluation techniques based on submodular width [2], to generalize the study done for fractional hypertree width in the context of factorized databases [9].

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.

[1] https://pro.univ-lille.fr/pierre-bourhis/
[2] https://a3nm.net/

References

[1] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of databases*. 1995.

[2] M. Abo Khamis, H. Q. Ngo, and D. Suciu. What do Shannon-type inequalities, submodular width, and disjunctive Datalog have to do with one another? In *Proc. PODS*, 2017.

[3] A. Amarilli, P. Bourhis, L. Jachiet, and S. Mengel. A Circuit-based approach to efficient enumeration. In *ICALP*, 2017.

[4] A. Amarilli, P. Bourhis, and P. Senellart. Provenance circuits for trees and treelike instances. In *Proc. ICALP*, 2015. `https://arxiv.org/abs/1511.08723`.

[5] 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`.

[6] K. Casel and M. L. Schmid. Fine-grained complexity of regular path queries. In *Proc. ICDT*, 2021.

[7] D. Deutch, T. Milo, S. Roy, and V. Tannen. Circuits for Datalog provenance. In *ICDT*, 2014.

[8] 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`.

[9] D. Olteanu and J. Závodnỳ. Size bounds for factorised representations of query results. *TODS*, 2015.

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