MSc Internship

Query Answering for Expressive Frontier-One Constraints

Antoine Amarilli, Pierre Senellart

Télécom ParisTech

Topic description

Traditional database management techniques usually assume that all the data of interest is reflected in the database. In many contexts, however, this assumption is not true, and one must work under the *open-world assumption*: the data that we have is correct but may not be complete. In these cases, it may still be useful to reason about *necessary consequences* of the data that we have. In other words, we assume some logical constraints about the world (e.g., "each website is maintained by one or several sysadmins", "each domain hosts at most one website", etc.), and we complete our database by what is implied by these constraints.

The *open-world query answering* (QA) problem is accordingly defined as follows: determine, from your data and from the logical constraints, whether a query is implied by the constraints on the data, i.e., whether the query holds on all completions of the data that satisfy the constraints. Depending on which logical constraints are allowed, this problem may be undecidable, and the problem of designing expressive constraint languages with *decidable* QA has been studied in several communities.

One first family of languages is that of *expressive description logics* [6]. The decidability of QA for such languages is (mostly) captured by the result of QA decidability for the logic GC^2, the guarded two-variable fragment of first-order logic with counting quantifiers [5]. These logics allow most of the expressive power of first-order logic, but their main restriction is that the vocabulary has *arity two*. In other words, we can represent interactions between two objects, but we cannot express directly more complex relationships, e.g., "John Doe is responsible for the website `www.example.com` from 2040 to 2042".

A second family of languages are based on *tuple-generating dependencies*, which intuitively express that some data patterns imply the existence of more data. One example of such languages are *existential rules* [3], which do not suffer from the arity-two limitation. However, these languages are limited in other ways, for instance they cannot express *disjunction* (e.g., "each website is either maintained by a sysadmin or delegated to a third party"), or *negative constraints* such as disjointness (e.g., "no website is both deployed on a local domain and managed by a third-party").

With Michael Benedikt (University of Oxford), we have shown [1] that combinations of constraints from these two families of languages could enjoy decidable QA, under some conditions. However, the resulting decidable language is not very homogeneous, because it consists of constraints expressed in two totally different formalisms.

The goal of the internship is to study whether one can really "bridge the gap" between the two representations. This involves the study of whether the frontier-one existential rules that we allow could be generalized to allow some of the GC^2 constructs such as negation and disjunction, and whether they can still be extended with *non-conflicting functional dependencies* in some sense, or some generalization thereof. Another question is in which cases the arity-two constraints can apply

on higher-arity predicates, e.g., still following a two-variable restriction. The ideal goal would be to propose a logical fragment with a natural definition, that would capture some of the specific features of existential rules and of GC^2, and for which one could show that QA is decidable.

A related question (but a challenging one!) is the study of *finite* QA, namely, QA over *finite completions* of the data. Our recent result [2] shows that this problem is decidable for a severely restricted fragment of frontier-one existential rules and functional dependencies. The intuition is that finite QA amounts to standard QA once an implication closure has been performed. The implication closure process, however, may generalize to more expressive operators, though it has not been studied for now outside of the arity-two setting [4]. Thus, if one designs a logical fragment that extends frontier-one existential rules with more expressive operators, it may also be possible to show the decidability of *finite* QA for this language.

Supervision and Environment

This Master's internship will have a duration of between 4 and 6 months and will be supervised by Pierre Senellart[1], professor at Télécom ParisTech and senior research fellow at the National University of Singapore and Antoine Amarilli[2], third-year PhD candidate at Télécom ParisTech. The internship will be based at Télécom ParisTech, in Paris.

References

[1] A. Amarilli and M. Benedikt. Combining existential rules and description logics. In *Proc. IJCAI*, 2015. `http://arxiv.org/abs/1505.00326`.

[2] A. Amarilli and M. Benedikt. Finite open-world query answering with number restrictions. In *Proc. LICS*, 2015. `http://arxiv.org/abs/1505.04216`.

[3] J. Baget, M. Leclère, M. Mugnier, and E. Salvat. On rules with existential variables: Walking the decidability line. *Artif. Intell.*, 2011. `http://www.lirmm.fr/~mugnier/ArticlesPostscript/BLMSDecidabilityLine-PreAIJ2011.pdf`.

[4] Y. Ibáñez-García, C. Lutz, and T. Schneider. Finite model reasoning in horn description logics. In *Proc. KR*, 2014. `http://www.informatik.uni-bremen.de/tdki/research/papers/2014/ILS-KR14.pdf`.

[5] I. Pratt-Hartmann. Data-complexity of the two-variable fragment with counting quantifiers. *Inf. Comput.*, 2009. `http://arxiv.org/abs/0806.1636`.

[6] S. Tobies. *Complexity results and practical algorithms for logics in knowledge representation.* PhD thesis, 2001. `http://arxiv.org/abs/cs/0106031`.

[1] `http://pierre.senellart.com/`
[2] `http://a3nm.net/`