

Interpolation and Domain Independence applied to Databases

Enrico Franconi, Paolo Guagliardo, Volha Kerhet, and Nhung Ngo

KRDB Research Centre
Free University of Bozen-Bolzano, Italy
lastname@inf.unibz.it

1 Query Reformulation

We study a general framework for query rewriting in the presence of a domain independent first-order logic theory (a knowledge base) over a signature including database and non-database predicates, based on Craig’s interpolant and Beth’s definability theorem [4].

In our framework queries are (possibly open) domain independent first-order formulas over the extended signature. The database predicates are meant to be closed, i.e., their extensions are the same in every model since they represent a classical finite relational database, whereas the non-database predicates in the ontology are interpreted classically, i.e., their extensions may vary across different models.

It is important to notice that all the conceptual modelling languages devised for the designing of information and database systems (such as Entity-Relationship schemas, UML Class diagrams, Object-Role Modelling ORM diagrams, etc) are domain independent: they can be formalised as domain independent first order theories.

Given a domain independent knowledge base and a query implicitly definable from the database signature, the framework provides precise semantic conditions to decide the existence of a domain independent first-order logically equivalent reformulation of the query (called exact rewriting) in terms of the database signature, and if so, it provides an effective approach to construct the reformulation based on interpolation using standard theorem proving techniques (e.g., tableau). We are interested in domain independent reformulations of queries because their range-restricted syntax is needed to reduce the original query answering problem to a relational algebra evaluation over the original database, that is, it is effectively executable as an SQL query directly over the database [6].

2 Computational properties

Due to the results on the applicability of Beth’s theorem and Craig’s interpolant, we prove the completeness of our framework in the case of domain independent ontologies and queries expressed in any fragment of first-order logic enjoying finitely controllable determinacy, a stronger property than the finite model property of the logic. If the employed logic does not enjoy finitely controllable determinacy our approach would become sound but incomplete, but still effectively implementable using standard theorem proving techniques.

In order to constructively compute the exact domain independent query reformulation we use the tableau based method to find the Craig’s interpolant from a validity proof of an implication of the form $(\mathcal{KB} \wedge \mathcal{Q}) \rightarrow (\tilde{\mathcal{KB}} \rightarrow \tilde{\mathcal{Q}})$, where $\tilde{\cdot}$ is obtained by uniformly replacing every occurrence of non-database predicates with fresh new predicate symbols.

3 The case of Description Logics

An important fragment of first-order logics for conceptual modelling is represented by description logics, which usually have good computational properties and for which many implemented efficient reasoners exist.

Since description logics knowledge bases are not necessarily domain independent, we have characterised syntactically the domain independent fragment of several very expressive fragments of description logics, and we have proved that these syntactic fragments are equally expressive as the corresponding domain independent fragments – thus extending the Codd’s theorem to description logics fragments of first-order logics. In these syntactic fragments a sort of guarded negation is enforced – recalling the safe-range syntactic property of first-order logic – a very reasonable restriction from the point of view of conceptual modelling [8, 3]. These fragments do also enjoy finitely controllable determinacy.

4 Applications

We have applied this framework not only to query answering under constraints, but also to data exchange – where we want to characterise the case of *lossless* transformations of data [5, 7] – and to view update – where we have characterised the problem of updatability as the determinacy of the inverse of the views [1, 2].

References

- [1] Ingo Feinerer, Enrico Franconi, and Paolo Guagliardo. Lossless horizontal decomposition with domain constraints on interpreted attributes. In *Big Data - 29th British National Conference on Databases (BNCOD 2013)*, volume 7968 of *Lecture Notes in Computer Science*, pages 77–91. Springer, 2013.
- [2] Enrico Franconi and Paolo Guagliardo. Effectively updatable conjunctive views. In *Proceedings of the 7th Alberto Mendelzon International Workshop on Foundations of Data Management*, 2013.
- [3] Enrico Franconi, Yazmin Angelica Ibanez-Garcia, and Inanc Seylan. Query answering with DBoxes is hard. *Electronic Notes in Theoretical Computer Science, Elsevier*, 278(1):71–84, November 2011.
- [4] Enrico Franconi, Volha Kerhet, and Nhung Ngo. Exact query reformulation over databases with first-order and description logics ontologies. *Journal of Artificial Intelligence Research (JAIR)*, 48:885–922, 2013.
- [5] Enrico Franconi, Nhung Ngo, and Evgeny Sherkhonov. The definability abduction problem for data exchange - (abstract). In *Web Reasoning and Rule Systems - 6th International Conference, RR 2012*, pages 217–220, 2012.
- [6] Volha Kerhet and Enrico Franconi. On checking domain independence. In *Proceedings of the 9th Italian Convention on Computational Logic*, number 857 in *CEUR Workshop Proceedings*, pages 246–250, 2012.
- [7] Nhung Ngo and Enrico Franconi. Getting unique solution in data exchange. *Proceedings of the VLDB Endowment*, 6(12):1440–1443, 2013.
- [8] Balder ten Cate, Enrico Franconi, and Inanc Seylan. Beth definability in expressive description logics. *Journal of Artificial Intelligence Research (JAIR)*, 48:347–414, 2013.