

First Order Rewritability in Ontology-Mediated Querying in Horn Description Logics (extended abstract of Toman and Weddell 2022)

David Toman and Grant Weddell

Cheriton School of Computer Science, University of Waterloo, Canada

{david,gweddell}@uwaterloo.ca

Abstract

We consider first-order (FO) rewritability for query answering in ontology-mediated querying (OMQ) in which ontologies are formulated in Horn fragments of description logics (DLs). In general, OMQ approaches for such logics rely on non-FO rewriting of the query and/or on non-FO completion of the explicit data in a knowledge base. Specifically, we show how the problem of FO rewritability can be formalized in terms of Beth definability, and comment on how Craig interpolation can then be used to effectively construct the rewritings, when they exist, from Clark’s completion of Datalog-like programs encoding a given DL TBox (and optionally a query).

1 Introduction

Ontology-mediated querying (OMQ) is the problem of answering conjunctive queries (CQs) over a knowledge base (KB) that are typically formulated in Horn fragments of description logics (DLs). Typical OMQ approaches rely on either reformulating a given CQ by incorporating the KB’s terminological knowledge (i.e., its TBox) (Calvanese et al. 2005; Calvanese et al. 2007), and then executing the reformulated query over the explicit data in the KB (i.e., its ABox) as a relational first-order (FO) query, or, for more expressive logics, on a completion of the ABox with respect to the TBox (Lutz, Toman, and Wolter 2009; Kontchakov et al. 2010; Kontchakov et al. 2011; Lutz et al. 2013). In the latter case, data completion can be expressed as a Datalog program. This obtains the *uniform FO rewritability* problem, that is, to determine if, for a particular TBox, any given CQ can be equivalently expressed as an FO query over any given ABox without the need for Datalog completion of the ABox. The existence of such a rewriting enables an OMQ front-end to a relational data source encoding an ABox to operate entirely by query reformulation of a given CQ to an SQL query directly executable over the relational data source, with no requirement to update the contents of tables beforehand.

Our main contribution is a novel approach to deciding uniform FO rewritability of OMQ in two DLs, Horn-*SHIQ* and Horn-*D_LFD*, via Clark’s completion of Datalog programs (Clark 1977) and Beth definability (Beth 1953).

We then extend this result to show how our framework extends to *query specific OMQ*. We also show how a variant of

the perfect rewriting approach to OMQ can be synthesized by appeal again to Beth definability and Craig interpolation (Craig 1957).

FO rewritability for Horn logics in the *ALC* family has been studied by others, e.g., (Bienvenu et al. 2016) and algorithms have been developed for generating such rewritings for logics in the *EL* family (Hansen et al. 2015). Our approach provides an interesting alternative path to detecting rewritability and to generating rewritings utilizing Beth definability and Craig interpolation. Another feature of our approach is its eventual link to interpolation-based query optimization (Hudek, Toman, and Weddell 2015; Toman and Weddell 2011).

2 Background and The Problem

We assume the standard definitions of Horn-*SHIQ* and Horn-*D_LFD* knowledge bases, each consisting of a TBox and an ABox (Hustadt, Motik, and Sattler 2005; McIntyre, Toman, and Weddell 2019) and the associated query answering (OMQ) problems. Our primary concern is then formalized as follows:

Definition 1 ((Uniform) FO Query Rewritability). *Given a TBox \mathcal{T} , the problem of query rewritability is to determine if there is a query reformulation $\varphi_{\mathcal{T}}$ for every CQ φ such that, for every ABox \mathcal{A} and tuple of constant symbols \vec{a} , $(\mathcal{T}, \mathcal{A}) \models \varphi(\vec{a})$ iff $\mathcal{A} \models \varphi_{\mathcal{T}}(\vec{a})$.*

DL Features that Preclude Rewritability. Consider the following:

1. Role Transitivity: Consider a Horn-*SHIQ* knowledge base with a TBox $\{\text{trans}(R)\}$. Then the CQ $\{(x, y) \mid R(x, y)\}$ cannot be FO rewritable since this would allow one to answer the *connectivity* question with respect to any ABox considered as a graph of R -edges.

2. Explicit ABox Equalities: Analogously to transitive roles, allowing equality and inequality between ABox objects, and therefore not adopting the *unique name assumption* (UNA), leads immediately to non-rewritability: Consider a KB in which $\mathcal{T} = \emptyset$ and a CQ $\{(x, y) \mid x = y\}$. Again, this query solves the (undirected) connectivity problem in an ABox with explicit equalities between individuals and thus cannot have an FO rewriting.

Hence, we consider the Horn- \mathcal{ALCHQI} sub-dialect of Horn- \mathcal{SHIQ} without transitive roles, and also adopt UNA for both Horn- \mathcal{SHIQ} and Horn- \mathcal{DLFD} in the rest of the paper.

3 The Solution

Earlier work on OMQ for the FunDL family of DLs (McIntyre, Toman, and Weddell 2019) has presented what was called a *combined combined approach* to OMQ, and has shown that it is essential to preserve tractability of OMQ in the presence of (limited) value restrictions (Toman and Weddell 2013; St. Jacques, Toman, and Weddell 2016; McIntyre et al. 2019). A similar approach has been proposed for Horn- \mathcal{SHIQ} (Eiter et al. 2012). Based on this approach, we study FO rewritability of OMQ for the above-mentioned DLs: Horn- \mathcal{SHIQ} in the \mathcal{ALC} family, and Horn- \mathcal{DLFD} in the FunDL family.

Proposition 2 (The Combined Combined Approach). *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent knowledge base and φ a CQ. Then there is a union of CQs query (UCQ) $\varphi_{\mathcal{T}}$ and a Datalog program $\Pi_{\mathcal{T}}$, both of which can be effectively constructed from \mathcal{T} , such that*

$$\mathcal{K} \models \varphi(\vec{a}) \iff \Pi_{\mathcal{T}}(\mathcal{A}) \models \varphi_{\mathcal{T}}(\vec{a}) \quad (1)$$

for any tuple of constant symbols \vec{a} , and where $\Pi_{\mathcal{T}}(\mathcal{A})$ is the minimal model of $\Pi_{\mathcal{T}}$ when evaluated over \mathcal{A} .

The proposition uses a Datalog program $\Pi_{\mathcal{T}}$ to define an ABox completion over which the query $\varphi_{\mathcal{T}}$, the rewriting of the original user query, is evaluated to compute the certain answers. Note that the existence of $\varphi_{\mathcal{T}}$ indicates that the non-rewritability of CQs is confined to the interaction of the TBox with explicit data given by an ABox which is captured by the Datalog program $\Pi_{\mathcal{T}}$. The program $\Pi_{\mathcal{T}}$ introduces a set of EDB(\mathcal{T}) predicates $P_B(x)$ and $P_R(x, y)$ for every primitive concept B and role R ,¹ to account for *explicit* data in \mathcal{A} , and a set of IDB(\mathcal{T}) predicates $C_B(x)$ and $R_R(x, y)$ with clauses derived from \mathcal{T} corresponding to the *completion* of the ABox w.r.t. \mathcal{T} . For example, a subsumption $A_1 \sqcap A_2 \sqsubseteq B$ is transformed to $C_B(x) \leftarrow C_{A_1}(x), C_{A_2}(x)$. To test for FO definability of the completion (i.e., all the C_A and R_R predicates that stand for the completed ABox instance), we use the following construction based on *Clark's Completion* $\Sigma_{\mathcal{T}}$ of $\Pi_{\mathcal{T}}$, a set of formulae

$$\begin{aligned} C_B(x) &\leftrightarrow P_B(x) \vee (\exists y. \alpha_1) \vee \dots \vee (\exists y. \alpha_n) \\ R_R(x, y) &\leftrightarrow P_R(x, y) \vee \beta_1 \vee \dots \vee \beta_m \end{aligned}$$

that correspond to clauses $C_B(x) \leftarrow \alpha_i$ and $R_R(x, y) \leftarrow \beta_j$ in $\Pi_{\mathcal{T}}$, grouped by the IDB(\mathcal{T}) predicates in the heads of the $\Pi_{\mathcal{T}}$ clauses. Note that the completion is no longer a Datalog program. However, it closes the original Datalog program in the following sense:

1. All IDB(\mathcal{T}) atoms entailed by $\Pi_{\mathcal{T}} \cup \mathcal{A}_{db}$ are also entailed by $\Sigma_{\mathcal{T}} \cup \mathcal{A}_{db}$, and
2. Negations of all IDB(\mathcal{T}) atoms that *finitely fail* over $\Pi_{\mathcal{T}} \cup \mathcal{A}_{db}$ are entailed by $\Sigma_{\mathcal{T}} \cup \mathcal{A}_{db}$,

¹We use role names R_f to stand for Horn- \mathcal{DLFD} features f in the subsequent constructions in the natural way.

where \mathcal{A}_{db} is the closed world variant of \mathcal{A} , a set of ground facts such that all facts *not in* \mathcal{A}_{db} are false. Clark's completion differs from, e.g., the closed world assumption (CWA) (Reiter 1977), and its variants, in a crucial way. For example, for a clause of the form $p \leftarrow p$ (and for cycles in programs to be completed in general), the completion simply generates a formula $p \leftrightarrow p$ that in turn allows models in which p can be true and models in which p is false. Moreover, had we used $\Pi_{\mathcal{T}}$ instead of $\Sigma_{\mathcal{T}}$, none of the definability results could possibly hold, even in the absence of role/feature subsumptions (such as role hierarchies). This gives us a complete characterization of FO rewritability of the ABox closure of individual primitive concept names with respect to Horn- \mathcal{ALCHQI} and Horn- \mathcal{DLFD} TBoxes:

Theorem 3. *Let \mathcal{T} be a TBox in one of DL dialects considered. Then the completion of an ABox \mathcal{A} w.r.t. \mathcal{T} is FO definable if and only if every predicate in IDB(\mathcal{T}) is Beth definable (Beth 1953) over $\Sigma_{\mathcal{T}}$ w.r.t. the language of EDB(\mathcal{T}).*

Given $\Sigma_{\mathcal{T}}$, one can now reformulate the Beth definability condition as a logical implication problem by making a copy of all formulas of $\Sigma_{\mathcal{T}}$ in which all non-logical symbols *not in* EDB(\mathcal{T}) are starred. Hence, the definability questions for $C_A(x)$ and $R_R(x, y)$ can be expressed as respective logical implication questions of the following forms.

$$\begin{aligned} \Sigma_{\mathcal{T}} \cup \Sigma_{\mathcal{T}}^* &\models \forall x. C_A(x) \rightarrow C_A^*(x) \\ \Sigma_{\mathcal{T}} \cup \Sigma_{\mathcal{T}}^* &\models \forall x, y. R_R(x, y) \rightarrow R_R^*(x, y) \end{aligned} \quad (2)$$

Note that, on closer inspection, all formulas in $\Sigma_{\mathcal{T}}$ can be written as \mathcal{ALCI} subsumptions. Note also that, without role constructors, there is no need to check for the definability of $R_R(x, y)$ atoms since they are always definable. Hence:

Theorem 4. *Let \mathcal{T} be a TBox in one of the dialects considered. Then the existence of the uniform query rewritability over \mathcal{T} is decidable and in EXPTIME.*

Note that the above holds due to the *specific* structure of $\Pi_{\mathcal{T}}$ and is not applicable to general Datalog programs. Indeed, (Chaudhuri and Vardi 1997) show much higher bounds for general programs. A matching lower bound can be obtained for expressive fragments of Horn- \mathcal{ALC} (for which the complexity of reasoning is EXPTIME-complete). However, since the size (and the construction) of rewritings will commonly dominate this cost, even for the simplest ontology languages (Kikot et al. 2012), exact complexity bounds are mostly of academic interest.

4 Further Results in the Full Paper

The following extensions can be easily accommodated in this definability-based framework:

1. One can apply *Craig Interpolation* (Craig 1957) to extract definitions of $C_A(x)$ and $R_R(x, y)$ from proofs of (2);
2. One can make the rewritability test *query specific* by replacing the IDB(\mathcal{T}) predicates in (2) by $\varphi_{\mathcal{T}}$ from (1); and
3. One can integrate the test with *database integrity constraints* that hold over \mathcal{A}_{db} .

Interestingly, the second extension can also be used to automatically synthesize *perfect rewritings* (Calvanese et al. 2007) for TBoxes formulated in DL-Lite.

References

- Beth, E. W. 1953. On Padoa's method in the theory of definition. *Indagationes Mathematicae* 15:330–339.
- Bienvenu, M.; Hansen, P.; Lutz, C.; and Wolter, F. 2016. First Order-Rewritability and Containment of Conjunctive Queries in Horn Description Logics. In *Proceedings of the 29th International Workshop on Description Logics, Cape Town, South Africa*.
- Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2005. *DL-Lite*: Tractable description logics for ontologies. In *Proc. of the 20th Nat. Conf. on Artificial Intelligence (AAAI 2005)*, 602–607.
- Calvanese, D.; de Giacomo, G.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family. *J. of Automated Reasoning* 39(3):385–429.
- Chaudhuri, S., and Vardi, M. Y. 1997. On the Equivalence of Recursive and Nonrecursive Datalog Programs. *J. Comput. Syst. Sci.* 54(1):61–78.
- Clark, K. L. 1977. Negation as failure. In *Logic and Data Bases, Symposium on Logic and Data Bases, Centre d'études et de recherches de Toulouse, France, 1977*, 293–322.
- Craig, W. 1957. Three uses of the Herbrand-Genzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic* 22:269–285.
- Eiter, T.; Ortiz, M.; Simkus, M.; Tran, T.; and Xiao, G. 2012. Query Rewriting for Horn-SHIQ Plus Rules. In *Proc. 26th AAAI Conference on Artificial Intelligence*.
- Hansen, P.; Lutz, C.; Seylan, I.; and Wolter, F. 2015. Efficient query rewriting in the description logic EL and beyond. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, 3034–3040.
- Hudek, A. K.; Toman, D.; and Weddell, G. E. 2015. On enumerating query plans using analytic tableau. In *Automated Reasoning with Analytic Tableaux and Related Methods - 24th International Conference, TABLEAUX 2015, Wrocław, Poland, September 21-24, 2015. Proceedings*, 339–354.
- Hustadt, U.; Motik, B.; and Sattler, U. 2005. Data complexity of reasoning in very expressive description logics. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 466–471. Professional Book Center.
- Kikot, S.; Kontchakov, R.; Podolskii, V. V.; and Zakharyashev, M. 2012. Long rewritings, short rewritings. In *Proceedings of the 2012 International Workshop on Description Logics, DL-2012, Rome, Italy, June 7-10, 2012*.
- Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyashev, M. 2010. The combined approach to query answering in DL-Lite. In *Proc. KR*, 247–257.
- Kontchakov, R.; Lutz, C.; Toman, D.; Wolter, F.; and Zakharyashev, M. 2011. The combined approach to ontology-based data access. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 2656–2661.
- Lutz, C.; Seylan, I.; Toman, D.; and Wolter, F. 2013. The combined approach to OBDA: Taming role hierarchies using filters. In *ISWC (I)*, 314–330.
- Lutz, C.; Toman, D.; and Wolter, F. 2009. Conjunctive query answering in the description logic \mathcal{EL} using a relational database system. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 2070–2075.
- McIntyre, S.; Borgida, A.; Toman, D.; and Weddell, G. E. 2019. On limited conjunctions and partial features in parameter-tractable feature logics. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019*, 2995–3002.
- McIntyre, S.; Toman, D.; and Weddell, G. E. 2019. FunDL - A family of feature-based description logics, with applications in querying structured data sources. In *Description Logic, Theory Combination, and All That - Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday*, 404–430.
- Reiter, R. 1977. On Closed World Data Bases. In Gallaire, H., and Minker, J., eds., *Logic and Data Bases, Symposium on Logic and Data Bases, Centre d'études et de recherches de Toulouse, France, 1977*, Advances in Data Base Theory, 55–76. New York: Plenum Press.
- St. Jacques, J.; Toman, D.; and Weddell, G. E. 2016. Object-relational queries over \mathcal{CFD}_{nc} knowledge bases: OBDA for the SQL-Literate. In *Proc. IJCAI*, 1258–1264.
- Toman, D., and Weddell, G. E. 2011. *Fundamentals of Physical Design and Query Compilation*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers.
- Toman, D., and Weddell, G. E. 2013. Conjunctive Query Answering in \mathcal{CFD}_{nc} : A PTIME Description Logic with Functional Constraints and Disjointness. In *Australasian Conference on Artificial Intelligence*, 350–361.
- Toman, D., and Weddell, G. E. 2022. First Order Rewritability in Ontology-Mediated Querying in Horn Description Logics. In *AAAI*, 5897–5905. AAAI Press.