First Order-Rewritability and Containment of Conjunctive Queries in Horn Description Logics

Meghyn Bienvenu[1], Peter Hansen[2], Carsten Lutz[2], Frank Wolter[3]

[1]: CNRS, Univ. Montpellier, FR   [2]: Univ. Bremen, DE   [3]: Univ. of Liverpool, UK

1 Introduction

In data access with ontologies [9, 17, 7], query rewriting is a central approach to efficient query anwering using existing database technology. Due to the tight connection between SQL and first-order logic (FO), FO queries are one of the most important target languages for rewriting. In this abstract, we consider FO-rewriting of ontology-mediated queries in the case when the actual query is a conjunctive query and the ontology is formulated in a DL between $\mathcal{EL}$ and Horn-$\mathcal{SHIF}$. In contrast to the case of DL-Lite [10], FO-rewritings are not guaranteed to exist when working with these more expressive DLs. A natural first step to approach the actual construction of FO-rewritings (when they exist) is thus to find an algorithm that decides the existence of a rewriting; in fact, any complete and terminating algorithm for constructing rewritings will implicitly also solve the decision problem and one can expect to learn important lessons already from the latter. It turns out that FO-rewritability is closely related to query containment problems as studied in [5, 8]. For this reason, we also provide results for query containment.

For the case of atomic queries (AQs) and DLs between $\mathcal{EL}$ and Horn-$\mathcal{SHI}$, it has been proved in [6] that FO-rewritability is ExpTime-complete (in the presence of an ABox signature). Our main result is that, when replacing AQs with CQs, this complexity does not change for DLs without inverse roles and jumps to 2ExpTime for DLs with inverse roles. The latter result is relativized by the observation that there is an algorithm that has double exponential runtime only in the size of the actual queries (which tend to be small), but single exponential runtime in the size of the ontology (which tends to be large). We also show NExpTime-completeness for the case where queries are connected and have at least one answer variable, and where the DL includes inverse roles. Our results also yield improved lower bounds for FO-rewritability and containment in monadic Datalog, compared to those in [12, 3, 2].

Related work. As has already been mentioned, the case of AQs was addressed in [6]. Our upper bounds (and also the associated characterizations in terms of tree-like ABoxes) can be viewed as a generalization of those in that paper. It was later shown in [15] that the foundational results in [6] give rise to very efficient and complete algorithms for computing actual rewritings. We hope that the results presented in this abstract will likewise provide the foundation for efficiently computing rewritings of ontology-mediated queries based on CQs. Results about

FO-rewritability in monadic Datalog and in frontier-guarded existential rules can be found in [12, 3], and results about monadic Datalog containment in [2] (also see the references therein).

Pragmatic approaches to OMQ rewriting beyond DL-Lite often consider Datalog as a target language [13, 16, 20–22]. These approaches might produce a non-recursive (thus FO) rewriting if it exists, but there are no guarantees. FO-rewritability of OMQs based on expressive DLs is considered in [4, 14], and based on existential rules in [1]. A problem related to ours is whether *all* queries are FO-rewritable when combined with a given TBox [19, 11].

2 Results

Let an ontology-mediated query (OMQ) be a triple $(\mathcal{T}, \Sigma, q)$ with $\mathcal{T}$ a DL TBox, Σ an ABox signature (set of concept and role names), and q an actual query. We use $(\mathcal{L}, \mathcal{Q})$ to denote the OMQ language that consists of all OMQs where $\mathcal{T}$ is formulated in the description logic $\mathcal{L}$ and q in the query language $\mathcal{Q}$. Our main complexity results concern FO-rewritability and containment for OMQ languages between ($\mathcal{EL}$, AQ) and (Horn-$\mathcal{SHIF}$, CQ).

Theorem 1. *FO-rewritability and containment are*

1. *2*ExpTime*-complete for any OMQ language between ($\mathcal{ELI}$, CQ) and (Horn-$\mathcal{SHIF}$, CQ), and*
2. ExpTime*-complete for any OMQ language between ($\mathcal{EL}$, AQ) and ($\mathcal{ELHF}_\bot$, CQ).*

Moreover, given an OMQ from (Horn-$\mathcal{SHIF}$, CQ) that is FO-rewritable, one can effectively construct a UCQ-rewriting.

Thus, replacing AQs with CQs results in an increase of complexity in the presence of inverse roles (indicated by $\mathcal{I}$), but not otherwise. The effect that inverse roles can increase the complexity of querying-related problems was known from expressive DLs of the $\mathcal{ALC}$ family [18], but it has not previously been observed for Horn DLs such as $\mathcal{ELI}$ and Horn-$\mathcal{SHIF}$.

The upper bounds in Theorem 2 rely on characterizations FO-rewritability and containment that are extensions of those in [6], replacing tree-shaped ABox with *pseudo-tree ABoxes* which are tree-shaped except for an initial component whose structure in unrestricted and whose size is bounded by the size of the query. An important observation is that, when deciding FO-rewritability, we can restrict our attention to *connected* queries provided that we have a way of deciding containment (for potentially disconnected queries). We use conCQ to denote the class of connected CQs.

Theorem 2. *Let $\mathcal{L}$ be any of the languages in Theorem 1. Then FO-rewritability in $(\mathcal{L}, CQ)$ can be solved in polynomial time when there is access to oracles for containment in $(\mathcal{L}, \mathcal{Q})$ and for FO-rewritability in $(\mathcal{L}, conCQ)$.*

This allows to avoid rather unpleasant technical complications; note that complications due to disconnectedness have been observed also in monadic Datalog boundedness such as the classical [12] (where they were not overcome).

The jump in complexity from ExpTime to 2ExpTime might appear to be problematic, but this is relativized when analyzing the upper bound in Point 1 of Theorem 1 a bit more carefully:

Theorem 3. *Given OMQs $Q_i = (\mathcal{T}_i, \Sigma_i, q_i)$, $i \in \{1,2\}$, from (Horn-$\mathcal{SHIF}$, CQ), it can be decided*

1. *in time $2^{2^{p(|q_1|+\log(|\mathcal{T}_1|))}}$ whether Q_1 is FO-rewritable and*
2. *in time $2^{2^{p(|q_1|+|q_2|+\log(|\mathcal{T}_1|+|\mathcal{T}_2|))}}$ whether $Q_1 \subseteq Q_2$,*

for some polynomial p.

Note that the runtime is double exponential only in the size of the actual queries q_1 and q_2, while it is only single exponential in the size of the TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$. This is good news since the size of q_1 and q_2 is typically very small compared to the sizes of $\mathcal{T}_1$ and $\mathcal{T}_2$. For this reason, it can even be reasonable to assume that the sizes of q_1 and q_2 are constant, in the same way in which the size of the query is assumed to be constant in data complexity. Note that, under this assumption, Theorem 3 yields ExpTime upper bounds for FO-rewritability and containment in the OMQ languages listed in Point 1 of Theorem 1.

Another interesting way to relativize the high complexity stated in Point 1 of Theorem 1 is to observe that the lower bound proofs require the actual query to be Boolean or disconnected. In practical applications, though, typical queries are connected and have at least one answer variable. We call such CQs *rooted* and use rCQ to denote the class of all rooted CQs. Our last main result states that, when we restrict our attention to rooted CQs, then the complexity drops to coNExpTime.

Theorem 4. *FO-rewritability and containment are* coNExpTime-*complete in any OMQ language between* ($\mathcal{ELI}$, *rCQ) and* (*Horn-*$\mathcal{SHIF}$, *rCQ*).

As mentioned in the introduction, the lower bound proofs underlying Theorem 1 also yield improved lower bounds for FO-rewritability and containment in monadic Datalog.

Corollary 1. *Containment and boundedness of monadic Datalog programs are 2*ExpTime-*hard, even if the arity of EDB relations is bounded by two, rule bodies are connected and tree-shaped, and there are no constants.*

This is already established in [2] for containment, but only in the case where rule bodies are not tree-shaped or there are constants (which, in this case, correspond to nominals in the DL world). We are not aware of existing lower bounds for monadic Datalog boundedness that apply to syntactically restricted programs.

References

1. Baget, J.F., Leclère, M., Mugnier, M.L., Salvat, E.: On rules with existential variables: Walking the decidability line. Artificial Intelligence 175(9-10), 1620–1654 (2011)
2. Benedikt, M., Bourhis, P., Senellart, P.: Monadic datalog containment. In: Proc. of ICALP. LNCS, vol. 7392, pp. 79–91. Springer (2012)
3. Benedikt, M., ten Cate, B., Colcombet, T., Vanden Boom, M.: The complexity of boundedness for guarded logics. In: Proc. of LICS, pp. 293–304. IEEE (2015)
4. Bienvenu, M., ten Cate, B., Lutz, C., Wolter, F.: Ontology-based data access: a study through Disjunctive Datalog, CSP, and MMSNP. ACM Transactions on Database Systems 39 (2014)
5. Bienvenu, M., Lutz, C., Wolter, F.: Query containment in description logics reconsidered. In: Proc. of KR. pp. 221–231 (2012)
6. Bienvenu, M., Lutz, C., Wolter, F.: First order-rewritability of atomic queries in horn description logics. In: Proc. of IJCAI. pp. 754–760 (2013)
7. Bienvenu, M., Ortiz, M.: Ontology-mediated query answering with data-tractable description logics. In: Proc. of Reasoning Web. LNCS, vol. 9203, pp. 218–307. Springer (2015)
8. Bourhis, P., Lutz, C.: Containment in monadic disjunctive datalog, mmsnp, and expressive description logics. In: Proc. of KR (2016)
9. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rodriguez-Muro, M., Rosati, R.: Ontologies and databases: The DL-Lite approach. In: Proc. of Reasoning Web. LNCS, vol. 5689, pp. 255–356. Springer (2009)
10. Calvanese, D., Giacomo, G.D., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The DL-Lite family. Journal of Automated Reasoning 39(3), 385–429 (2007)
11. Civili, C., Rosati, R.: On the first-order rewritability of conjunctive queries over binary guarded existential rules. In: Proc. of CILC. CEUR Workshop Proceedings, vol. 1459, pp. 25–30. CEUR-WS.org (2015)
12. Cosmadakis, S.S., Gaifman, H., Kanellakis, P.C., Vardi, M.Y.: Decidable optimization problems for database logic programs (preliminary report). In: Proc. of STOC. pp. 477–490. ACM (1988)
13. Eiter, T., Ortiz, M., Simkus, M., Tran, T., Xiao, G.: Query rewriting for Horn-SHIQ plus rules. In: Proc. of AAAI. AAAI Press (2012)
14. Feier, C., Kuusisto, A., Lutz, C.: FO-rewritability of expressive ontology-mediated queries. In: Proc. of DL (2016)
15. Hansen, P., Lutz, C., Seylan, I., Wolter, F.: Efficient query rewriting in the description logic EL and beyond. In: Proc. of IJCAI. pp. 3034–3040. AAAI Press (2015)
16. Kaminski, M., Nenov, Y., Grau, B.C.: Computing datalog rewritings for disjunctive datalog programs and description logic ontologies. In: Proc. of RR. pp. 76–91 (2014)
17. Kontchakov, R., Rodriguez-Muro, M., Zakharyaschev, M.: Ontology-based data access with databases: A short course. In: Reasoning Web. pp. 194–229 (2013)
18. Lutz, C.: The complexity of conjunctive query answering in expressive description logics. In: Proc. of IJCAR. LNCS, vol. 5195, pp. 179–193. Springer (2008)
19. Lutz, C., Wolter, F.: Non-uniform data complexity of query answering in description logics. In: Proc. of KR. (2012)
20. Pérez-Urbina, H., Motik, B., Horrocks, I.: Tractable query answering and rewriting under description logic constraints. Journal of Applied Logic 8(2), 186–209 (2010)

21. Rosati, R.: On conjunctive query answering in $\mathcal{EL}$. In: Proc. of DL. pp. 451–458 (2007)
22. Trivela, D., Stoilos, G., Chortaras, A., Stamou, G.B.: Optimising resolution-based rewriting algorithms for OWL ontologies. Journal of Web Semantics. 33, 30–49 (2015)