Ontology-Based Query Answering over Datalog Expressible Rule Sets is Undecidable (Extended Abstract)

David Carral[1], Lucas Larroque[2] and Michaël Thomazo[2]

[1]LIRMM, Inria, University of Montpellier, CNRS, Montpellier, France
[2]Inria, DI ENS, ENS, CNRS, PSL University, Paris, France

Abstract

Ontology-based query answering is a problem that takes as input an ontology $\mathcal{R}$ (typically expressed by existential rules), a set $\mathcal{F}$ of facts, and a Boolean conjunctive query (CQ) q, and asks whether $\mathcal{R}, \mathcal{F} \models q$. This problem is undecidable in general, and a widely investigated approach to tackle it in some cases is query rewriting: given some "rule query" $\langle \mathcal{R}, q \rangle$, we compute a Boolean query $q_\mathcal{R}$ such that, for any fact set $\mathcal{F}$, it holds that $\mathcal{R}, \mathcal{F} \models q$ if and only if $\mathcal{F} \models q_\mathcal{R}$. Previous work has mostly focused on output queries $q_\mathcal{R}$ expressed as union of Boolean conjunctive queries (UCQs), and an effective algorithm that computes such a query $q_\mathcal{R}$ whenever it exists has been proposed in the literature. However, UCQ rewritability is not a very general notion and many real-world interesting rule queries do no admit UCQ rewritings. This raises the question whether such a generic algorithm can be designed for a more expressive target language, such as datalog. We solve this question by the negative, by studying the difference between datalog expressibility and datalog rewritability. More precisely, we show that query answering under datalog expressible rule queries is undecidable.

Keywords

OBQA, Existential Rules, Datalog, Query Rewritability, Decidability

1. Introduction

Efficiently accessing data is an important step in many real-world applications. Ontologies have been identified as an important tool to help a user to express their information needs, allowing them to use a vocabulary they are familiar with, while enabling a system to perform automated reasoning, leading to more complete answers. Ontology-based query answering (OBQA) is a core problem therein, where a set of facts is queried while taking into account the domain knowledge expressed in an ontology. These ontologies may be expressed in a variety of formalisms, such as Description Logics or existential rules. The OBQA problem is typically framed as follows: given a fact set $\mathcal{F}$, an ontology $\mathcal{R}$, and a Boolean CQ q, check if $\mathcal{F}, \mathcal{R} \models q$, where $\models$ denotes the classical first-order logic entailment.

This problem is undecidable when the ontology can range over any set of existential rules. Thus, a lot of research has focused on finding decidable and even tractable classes of rule sets; see [1] for an introduction to these. Particularly relevant to us are classes based on the so-called *query rewriting* approach. Given an ontology $\mathcal{R}$ and a Boolean CQ q, one computes a Boolean UCQ $q_\mathcal{R}$ such that for any fact set $\mathcal{F}$, it holds that $\mathcal{F}, \mathcal{R} \models q$ if and only if $\mathcal{F} \models q_\mathcal{R}$. As most data is stored in relational databases, which have been designed to efficiently process CQs, most research has focused on rewriting the output query $q_\mathcal{R}$ as a UCQ. A natural question is then, given an ontology $\mathcal{R}$ and a BCQ q, is there a UCQ rewriting $q_\mathcal{R}$ for $\mathcal{R}$ and q? In other words, is that true that the rule query $\langle \mathcal{R}, q \rangle$ is *UCQ expressible*? This does not always hold, and it is actually undecidable to check whether it is the case [2]. However, there exists an effective algorithm that *computes* a UCQ rewriting when given as input a UCQ expressible rule query [3]. In other words, the UCQ expressibility of every rule query (the existence of a UCQ rewriting) in a class and UCQ rewritability of that class (the computability of UCQ rewritings for all rule queries in that class) are two notions that coincide, which possibly explains why they have not been introduced separately in the literature.

DL 2025: 38th International Workshop on Description Logics, September 3–6, 2025, Opole, Poland
david.carral@inria.fr (D. Carral); lucas.larroque@inria.fr (L. Larroque); michael.thomazo@inria.fr (M. Thomazo)
0000-0001-7287-4709 (D. Carral); 0009-0007-2351-2681 (L. Larroque); 0000-0002-1437-6389 (M. Thomazo)

Source Language	Target Language	Arbitrary Query Sig.	Implemented	Reference
$\mathcal{SHIQ}$	disj. dat.	$\times$	✓	[9]
$\mathcal{SHIQ}b_{\mathcal{S}}$	disj. dat.	$\times$	$\times$	[10]
Horn-$\mathcal{ALCHOIQ}$	datalog	$\times$	✓	[11]
Horn-$\mathcal{SRIQ}$	datalog	$\times$	✓	[12]
Bounded Detph rules	datalog	$\times$	$\times$	[13]
Frontier guarded rules	datalog	✓	$\times$	[14]
Nearly Guarded Rules	datalog	✓	$\times$	[15]
Guarded disj. rules	disj. datalog	$\times$	$\times$	[16]
Guarded rules	datalog	✓	✓	[17]
Warded rules	datalog	✓	✓	[18]
Linear	non rec. dat.	$\times$	$\times$	[19]
Sticky(-join)	non rec. dat.	$\times$	$\times$	[19]

Table 1
Summary of datalog rewriting approaches applicable for fact entailment

Syntactic conditions such as linearity [4] or stickiness [5] guarantee the existence of UCQ rewritings for any BCQ. Moreover, there is also DL-Lite [6], which is a widely used Description Logic that can be translated into existential rules. However, the expressivity of these languages is too limited for many real-world ontologies. A natural task is to consider a more expressive target query language for the rewritings. It is known that considering first-order queries and conjunctive queries have the same expressive power for rewriting existential rule queries [7]. We then focus on another classical language, namely datalog. Note that all UCQ expressible rule queries are also datalog expressible but the converse is not true.

As discussed in Section 2, there are many known and interesting classes for which specific datalog rewriting algorithms have been designed. However, no generic algorithm, such as in the case of UCQ expressibility, is known so far. The contribution of this paper is to show that, unfortunately, no such algorithm exists. This is done by proving that the problem of checking $\mathcal{R}, \mathcal{F} \models q$ under the assumption that $\langle \mathcal{R}, q \rangle$ is datalog expressible is undecidable, contradicting the existence of a rewriting algorithm for datalog expressible queries. We prove the result by reduction from the halting problem of Turing machines to OBQA, where the difficulty lies in ensuring that rule queries produced by the reduction are datalog expressible.

An extended version of this work with fully detailed proofs can be found in [8].

2. Related Work

Let us first point out that there are several variations around the notion of datalog rewriting. A first dimension is that, rather than rewriting a specific rule query $\langle \mathcal{R}, q \rangle$, one can wish to rewrite $\mathcal{R}$ into a datalog rule set $\mathcal{R}'$, and use $\mathcal{R}'$ to compute answers for a class of queries. Another variation is focused on the fact sets on which the rewriting should output the same answer as the original rule query. In this paper, we consider the strong version where answers should be the same on every fact set over the signature used in the rewritings. Quite often, defined datalog rewritings only preserve answers over fact sets on the original signature – this allows one to introduce fresh predicates which are known not to belong to fact sets on which the datalog program is to be evaluated. There are cases for which there exist datalog rewritings of rule queries for this relaxed definition but not for our restricted one; this is a consequence of Theorem 3 in [20]. Our undecidability result implies undecidability of this more relaxed notion.

The use of datalog as a target language for rewritings has been studied over the last 15 years. The goal was to reduce reasoning task over expressive ontologies towards query answering over datalog, for which optimization techniques have been developed in the database community. This is even more important today, as a variety of efficient datalog reasoners have been implemented [21, 22]. Such an approach has been proposed for providing disjunctive[1] datalog rewritings for $\mathcal{SHIQ}$ for fact entailment over the original signature [9], later generalized to $\mathcal{SHIQ}b_S$ [10]. More recently, such reductions for Horn description logics have been implemented and evaluated [11, 12]. Such datalog rewritings have also been studied for existential rules, for guarded [17], nearly guarded [15], warded [18] and shy [24] rule sets.

Beyond these fragment-specific reductions, the limits of datalog rewritability have been explored. In [13], it is shown that whenever rule queries have bounded depth (meaning that if they are entailed, they are entailed by a portion of the chase that uses only Skolem terms of bounded depth), they are datalog rewritable. This result applies for all syntactic fragment for which the chase is known to terminate [25], but datalog rewritability is not guaranteed (and not always possible) for rule sets having terminating restricted chase [26] – this is proven by a data complexity argument.

Another question of interest is the size of the obtained rewritings. In [16], the authors provide polynomial (disjunctive) datalog rewritings for (disjunctive) guarded rules queries. Non-recursive datalog has also been studied: while it does not increase the expressivity with respect to UCQs, the re-use of predicates allows to significanlty reduce the size of rewritings, reaching polynomiality in some cases [19, 27].

All of these contributions are summarised in Table 1.

3. Future Work

To conclude, we discuss two distinct lines for future research, which naturally follow from our work.

Answer Expressible Rule Sets We intend to study an even more restrictive class of rewritable rule sets: a rule set $\mathcal{R}$ is *answer datalog expressible* if, for every CQ $q[\vec{x}]$, the rule query $\langle \mathcal{R}, q[\vec{x}] \rangle$ admits some *datalog rewriting that preserves all answers.* That is, a datalog query $\langle \mathcal{R}', q'[\vec{x}] \rangle$ such that, for every fact set $\mathcal{F}$ and every list $\vec{a}$ of constants occurring in $\mathcal{F}$, we have that $\langle \mathcal{R}, \mathcal{F} \rangle \models q[\vec{x}/\vec{a}]$ if and only if $\langle \mathcal{R}', \mathcal{F} \rangle \models q[\vec{x}/\vec{a}]$. With this definition in place, we consider the following problem:

Open Problem 1. *Consider a knowledge base $\mathcal{K} = \langle \mathcal{R}, \mathcal{F} \rangle$, a CQ q, and a list $\vec{a}$ of constants occurring in $\mathcal{F}$. Is there a procedure to check if $\vec{a}$ is an answer of q with respect to $\langle \mathcal{R}, \mathcal{F} \rangle$ that is sound, complete, and terminating if $\mathcal{R}$ is answer datalog expressible?*

The above question is quite relevant for our field of research, where we often study the theoretical properties of classes of rule sets and not of rule queries.

At the moment, we believe that the answer to this open problem is negative, which would yield a result strictly stronger than the one of this paper. However, a different proof strategy is required to show this since there are rule sets in the range of the reduction used for this result that are not answer datalog-rewritable.

Alternative Rewriting Languages In this paper, our primary focus is on datalog; in the future, we plan to study alternative query languages for rewritings. For instance, one could consider unions of Boolean conjunctive regular path queries (UBCRPQs) [28] and then consider the following problem:

Open Problem 2. *Is the class of all UBCRPQ expressible queries is UBCRPQ rewritable? Is there a procedure to check if a knowledge base $\langle \mathcal{R}, \mathcal{F} \rangle$ entails a BCQ q that is sound, complete, and terminates if the rule query $\langle \mathcal{R}, q \rangle$ is UCRPQ expressible?*

We can instantiate different versions of this open problem by considering different output rewriting languages. For instance, we could consider as unions of (non-conjunctive) regular path queries, monadic

[1]For disjunctive existential rules and datalog, the reader is invited to consult [23]

datalog, query languages based on context-free grammars [29], or any of the query languages considered by [30].

As a closing remark, note that the answers to the first and second questions in Open Problem 2 might be negative and positive, respectively. That is, it is possible that we can solve entailment for UCRPQ expressible rule queries even if we cannot effectively compute rewritings for these. This is an exciting possibility that may lead us to the discovery of a novel kind of reasoning procedure for this expressive class of rule queries. Or perhaps future research will just result in another undecidability result, which would again be strictly stronger than the main result of this paper. Either way, we look forward to researching (and hopefully settling!) these questions.

Declaration on Generative AI

During the preparation of this work, the authors used Chat GPT in order to do grammar and spelling checks. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Mugnier, M. Thomazo, An introduction to ontology-based query answering with existential rules, in: M. Koubarakis, G. B. Stamou, G. Stoilos, I. Horrocks, P. G. Kolaitis, G. Lausen, G. Weikum (Eds.), Reasoning Web. Reasoning on the Web in the Big Data Era - 10th International Summer School 2014, Athens, Greece, September 8-13, 2014. Proceedings, volume 8714 of *Lecture Notes in Computer Science*, Springer, 2014, pp. 245–278. URL: https://doi.org/10.1007/978-3-319-10587-1_6. doi:10.1007/978-3-319-10587-1_6.

[2] J. Baget, M. Leclère, M. Mugnier, E. Salvat, On rules with existential variables: Walking the decidability line, Artif. Intell. 175 (2011) 1620–1654. URL: https://doi.org/10.1016/j.artint.2011.03.002. doi:10.1016/J.ARTINT.2011.03.002.

[3] M. König, M. Leclère, M. Mugnier, M. Thomazo, Sound, complete and minimal ucq-rewriting for existential rules, Semantic Web 6 (2015) 451–475. URL: https://doi.org/10.3233/SW-140153. doi:10.3233/SW-140153.

[4] A. Calì, G. Gottlob, M. Kifer, Taming the infinite chase: Query answering under expressive relational constraints, J. Artif. Intell. Res. 48 (2013) 115–174. URL: https://doi.org/10.1613/jair.3873. doi:10.1613/JAIR.3873.

[5] A. Calì, G. Gottlob, A. Pieris, Query answering under non-guarded rules in datalog+/-, in: P. Hitzler, T. Lukasiewicz (Eds.), Web Reasoning and Rule Systems - Fourth International Conference, RR 2010, Bressanone/Brixen, Italy, September 22-24, 2010. Proceedings, volume 6333 of *Lecture Notes in Computer Science*, Springer, 2010, pp. 1–17. URL: https://doi.org/10.1007/978-3-642-15918-3_1. doi:10.1007/978-3-642-15918-3_1.

[6] A. Artale, D. Calvanese, R. Kontchakov, M. Zakharyaschev, The dl-lite family and relations, J. Artif. Intell. Res. 36 (2009) 1–69. URL: https://doi.org/10.1613/jair.2820. doi:10.1613/JAIR.2820.

[7] B. Rossman, Homomorphism preservation theorems, J. ACM 55 (2008) 15:1–15:53. URL: https://doi.org/10.1145/1379759.1379763. doi:10.1145/1379759.1379763.

[8] D. Carral, L. Larroque, M. Thomazo, Ontology-Based Query Answering over Datalog-Expressible Rule Sets is Undecidable, in: Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning, 2024, pp. 232–242. URL: https://doi.org/10.24963/kr.2024/22. doi:10.24963/kr.2024/22.

[9] U. Hustadt, B. Motik, U. Sattler, Reasoning in description logics by a reduction to disjunctive datalog, J. Autom. Reason. 39 (2007) 351–384. URL: https://doi.org/10.1007/s10817-007-9080-3. doi:10.1007/S10817-007-9080-3.

[10] S. Rudolph, M. Krötzsch, P. Hitzler, Type-elimination-based reasoning for the description logic

shiqbs using decision diagrams and disjunctive datalog, Log. Methods Comput. Sci. 8 (2012). URL: https://doi.org/10.2168/LMCS-8(1:12)2012. doi:10.2168/LMCS-8(1:12)2012.

[11] D. Carral, I. Dragoste, M. Krötzsch, The combined approach to query answering in horn-alchoiq, in: M. Thielscher, F. Toni, F. Wolter (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018, AAAI Press, 2018, pp. 339–348. URL: https://aaai.org/ocs/index.php/KR/KR18/paper/view/18076.

[12] D. Carral, L. González, P. Koopmann, From horn-sriq to datalog: A data-independent transformation that preserves assertion entailment, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press, 2019, pp. 2736–2743. URL: https://doi.org/10.1609/aaai.v33i01.33012736. doi:10.1609/AAAI.V33I01.33012736.

[13] B. Marnette, Resolution and Datalog rewriting under value invention and equality constraints, CoRR abs/1212.0254 (2012). URL: http://arxiv.org/abs/1212.0254. arXiv:1212.0254.

[14] V. Bárány, M. Benedikt, B. ten Cate, Rewriting guarded negation queries, in: K. Chatterjee, J. Sgall (Eds.), Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013, Klosterneuburg, Austria, August 26-30, 2013. Proceedings, volume 8087 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 98–110. URL: https://doi.org/10.1007/978-3-642-40313-2_11. doi:10.1007/978-3-642-40313-2_11.

[15] G. Gottlob, S. Rudolph, M. Simkus, Expressiveness of guarded existential rule languages, in: R. Hull, M. Grohe (Eds.), Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS'14, Snowbird, UT, USA, June 22-27, 2014, ACM, 2014, pp. 27–38. URL: https://doi.org/10.1145/2594538.2594556. doi:10.1145/2594538.2594556.

[16] S. Ahmetaj, M. Ortiz, M. Simkus, Rewriting guarded existential rules into small datalog programs, in: B. Kimelfeld, Y. Amsterdamer (Eds.), 21st International Conference on Database Theory, ICDT 2018, March 26-29, 2018, Vienna, Austria, volume 98 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, pp. 4:1–4:24. URL: https://doi.org/10.4230/LIPIcs.ICDT.2018.4. doi:10.4230/LIPICS.ICDT.2018.4.

[17] M. Benedikt, M. Buron, S. Germano, K. Kappelmann, B. Motik, Rewriting the infinite chase, Proc. VLDB Endow. 15 (2022) 3045–3057. URL: https://www.vldb.org/pvldb/vol15/p3045-benedikt.pdf. doi:10.14778/3551793.3551851.

[18] G. Berger, G. Gottlob, A. Pieris, E. Sallinger, The space-efficient core of vadalog, ACM Trans. Database Syst. 47 (2022) 1:1–1:46. URL: https://doi.org/10.1145/3488720. doi:10.1145/3488720.

[19] G. Gottlob, T. Schwentick, Rewriting ontological queries into small nonrecursive datalog programs, in: G. Brewka, T. Eiter, S. A. McIlraith (Eds.), Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012, AAAI Press, 2012. URL: http://www.aaai.org/ocs/index.php/KR/KR12/paper/view/4510.

[20] M. Krötzsch, Efficient rule-based inferencing for OWL EL, in: T. Walsh (Ed.), IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011, IJCAI/AAAI, 2011, pp. 2668–2673. URL: https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-444. doi:10.5591/978-1-57735-516-8/IJCAI11-444.

[21] Y. Nenov, R. Piro, B. Motik, I. Horrocks, Z. Wu, J. Banerjee, Rdfox: A highly-scalable RDF store, in: M. Arenas, Ó. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, S. Staab (Eds.), The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II, volume 9367 of *Lecture Notes in Computer Science*, Springer, 2015, pp. 3–20. URL: https://doi.org/10.1007/978-3-319-25010-6_1. doi:10.1007/978-3-319-25010-6_1.

[22] J. Urbani, M. Krötzsch, C. J. H. Jacobs, I. Dragoste, D. Carral, Efficient model construction for horn logic with vlog - system description, in: D. Galmiche, S. Schulz, R. Sebastiani (Eds.), Automated Reasoning - 9th International Joint Conference, IJCAR 2018, Held as Part of the Federated Logic Conference, FloC 2018, Oxford, UK, July 14-17, 2018, Proceedings, volume 10900 of *Lecture Notes in*

Computer Science, Springer, 2018, pp. 680–688. URL: https://doi.org/10.1007/978-3-319-94205-6_44. doi:10.1007/978-3-319-94205-6_44.

[23] A. Deutsch, V. Tannen, Reformulation of XML queries and constraints, in: D. Calvanese, M. Lenzerini, R. Motwani (Eds.), Database Theory - ICDT 2003, 9th International Conference, Siena, Italy, January 8-10, 2003, Proceedings, volume 2572 of *Lecture Notes in Computer Science*, Springer, 2003, pp. 225–241. URL: https://doi.org/10.1007/3-540-36285-1_15. doi:10.1007/3-540-36285-1_15.

[24] N. Leone, M. Manna, G. Terracina, P. Veltri, Fast query answering over existential rules, ACM Trans. Comput. Log. 20 (2019) 12:1–12:48. URL: https://doi.org/10.1145/3308448. doi:10.1145/3308448.

[25] B. C. Grau, I. Horrocks, M. Krötzsch, C. Kupke, D. Magka, B. Motik, Z. Wang, Acyclicity notions for existential rules and their application to query answering in ontologies, J. Artif. Intell. Res. 47 (2013) 741–808. URL: https://doi.org/10.1613/jair.3949. doi:10.1613/JAIR.3949.

[26] M. Krötzsch, M. Marx, S. Rudolph, The power of the terminating chase (invited talk), in: P. Barceló, M. Calautti (Eds.), 22nd International Conference on Database Theory, ICDT 2019, March 26-28, 2019, Lisbon, Portugal, volume 127 of *LIPIcs*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 3:1–3:17. URL: https://doi.org/10.4230/LIPIcs.ICDT.2019.3. doi:10.4230/LIPICS.ICDT.2019.3.

[27] G. Gottlob, S. Kikot, R. Kontchakov, V. V. Podolskii, T. Schwentick, M. Zakharyaschev, The price of query rewriting in ontology-based data access, Artif. Intell. 213 (2014) 42–59. URL: https://doi.org/10.1016/j.artint.2014.04.004. doi:10.1016/J.ARTINT.2014.04.004.

[28] D. Florescu, A. Y. Levy, D. Suciu, Query containment for conjunctive queries with regular expressions, in: A. O. Mendelzon, J. Paredaens (Eds.), Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 1-3, 1998, Seattle, Washington, USA, ACM Press, 1998, pp. 139–148. URL: https://doi.org/10.1145/275487.275503. doi:10.1145/275487.275503.

[29] C. M. Medeiros, M. A. Musicante, U. S. da Costa, Querying graph databases using context-free grammars, J. Comput. Lang. 68 (2022) 101089. URL: https://doi.org/10.1016/j.cola.2021.101089. doi:10.1016/J.COLA.2021.101089.

[30] P. Bourhis, M. Krötzsch, S. Rudolph, Reasonable highly expressive query languages - IJCAI-15 distinguished paper (honorary mention), in: Q. Yang, M. J. Wooldridge (Eds.), Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, AAAI Press, 2015, pp. 2826–2832. URL: http://ijcai.org/Abstract/15/400.