Answering Expressive Conjunctive Queries over RDFS Knowledge Bases (Extended Abstract)

Gianluca Cima, Marco Console, Roberto Maria Delfino, Maurizio Lenzerini and Antonella Poggi

Sapienza University of Rome, via Ariosto 25, Rome, Italy

Abstract

Answering conjunctive queries (CQs) equipped with basic forms of negation is a challenging task, which happens to be undecidable even for lightweight Description Logic (DL) ontologies. Interestingly, the DL counterpart of RDFS seems to be partially unaffected by those negative results, even when equipped with disjointness axioms. This paper summarises our recent work on this subject, where we present a refined complexity analysis of answering CQs with inequality atoms and safe negation posed over such ontologies. We introduce a unified Π_2^p algorithm for the general case, we prove that two inequality atoms already lead to Π_2^p-hardness, and we show similar results for the case of safe negation: answering CQs with one negated atom is in NP, but two negated atoms are enough to reach Π_2^p-hardness. These results close key gaps in the literature and refine the complexity analysis of the query containment problem.

Keywords

Query Answering, DL Ontologies, Negative conditions

Knowledge Base (KB) systems are AI systems that represent domain knowledge symbolically to support both user interaction and the enhancement of other AI components. Knowledge Graphs (KGs) are a notable example, acting as both KBs and the basis for tasks like query answering [1] and boosting generative models [2, 3]. KBs are typically defined using formal semantics, specifying meaning via formal languages such as Description Logics (DLs) [4], i.e., fragments of first-order logic with well-defined syntax and semantics. A DL-based KB, or ontology, comprises axioms divided into a TBox (intensional knowledge) and an ABox (extensional knowledge). A key computational task is query answering [5, 6], which involves verifying whether a query, i.e., an expression defined in a formal language, is satisfied in all models of an ontology. Conjunctive Queries (CQs) and Unions thereof (UCQs) have been extensively studied. However, despite their expressiveness, they cannot express negation, such as filtering individuals lacking certain properties (safe negation) or enforcing distinctions between individuals (inequality atoms), features often crucial when querying heterogeneous KGs. Even basic forms of negation render query answering undecidable, even for lightweight DLs like *DL-Lite*$_\mathcal{R}$ and $\mathcal{EL}$. A notable exception is *DL-Lite*$^{\neg}_{\text{RDFS}}$, the DL counterpart of RDFS [7, 8, 9], extended with disjointness axioms and interpreted without the Unique Name Assumption

The original paper has been accepted and presented at AAAI25.

DL 2025: 38th International Workshop on Description Logics, September 3–6, 2025, Opole, Poland

cima@diag.uniroma1.it (G. Cima); console@diag.uniroma1.it (M. Console); delfino@diag.uniroma1.it (R. M. Delfino); lenzerini@diag.uniroma1.it (M. Lenzerini); poggi@diag.uniroma1.it (A. Poggi)

0000-0003-1783-5605 (G. Cima); 0009-0004-5526-019X (M. Console); 0000-0002-5492-5290 (R. M. Delfino); 0000-0003-2875-6187 (M. Lenzerini); 0000-0002-4030-3458 (A. Poggi)

(UNA). While less expressive, it retains the core semantic features of RDFS, widely used in KGs, and supports data quality constraints via disjointness. Answering (U)CQs with inequality atoms (UCQ$^{\neq}$) over *DL-Lite*$^{\neg}_{\text{RDFS}}$ ontologies is Π^p_2-complete in combined complexity [10]. The same work shows that answering UCQ$^{\neq}$ with at most one inequality atom in each disjunct over *DL-Lite*$^{\neg}_{\text{RDFS}}$ is an NP-complete problem. These two results led to the question whether there exists a number $k \geq 2$ such that, answering CQ$^{\neq}$ with a fixed number k of inequality atoms over *DL-Lite*$^{\neg}_{\text{RDFS}}$ ontologies is Π^p_2-complete. The results presented in this paper close the gap in the literature about this and other questions related to the computational complexity of answering queries of different languages over *DL-Lite*$^{\neg}_{\text{RDFS}}$ ontologies. We extend the results presented in [10] for UCQs with inequality atoms only, providing a Π^p_2 upper bound in combined complexity and a coNP upper bound in data complexity, where data complexity [11] concerns the case where only the ABox is regarded as the input of the problem. We also provide a matching lower bound for the problem of answering conjunctive queries with at most two inequality atoms, for which a matching lower bound in data complexity was already known. We also show that our result affects the query containment problem, previously proven Π^p_2-complete [12, 13], and studied under various syntactic and semantic restrictions in [14]. To our knowledge, only [10] examined the impact of bounding the number of inequalities in the queries. That work showed NP-completeness when the containing query has at most one inequality atom, matching the complexity of CQs [15]. Here, we complete the analysis by proving that the problem remains Π^p_2-hard even when the containing and contained queries have one and two inequality atoms, respectively. We also provide lower bounds for CQs with safe negations, showing that it presents a behavior similar to that of CQs with inequality atoms, i.e., allowing for two safe negations is enough to obtain Π^p_2-hardness. Finally, we present results for the case where the UNA holds. We denote by CQ$^{\neg_s,\neq}$ the language of conjunctive queries containing both *safe negations* and *inequality atoms*, where a safe negation is a negated atom whose variables occur in at least one positive atom, and an inequality atom is an atom of the form $t \neq t'$, with t and t' being either distinguished variables, existential variables, or constants. We denote the language of conjunctive queries with at most p negated atoms and no inequality atoms by CQ$^{\neg_s,p}$, while the language of conjunctive queries with at most p inequality atoms and no negated atoms is denoted by CQ$^{\neq_p}$. We denote by UCQ^z the language of union of CQ^z, where z is a combination of the symbols denoting the presence of either negations or inequality atoms.

Definition 1. *We denote by ans*$(\mathcal{L}, \mathcal{Q})$ *the problem of deciding,* $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ *in the language* $\mathcal{L}$*, a query* $q \in \mathcal{Q}$ *of arity* n*, and an* n*-tuple* $\bar{c}$ *of constants occurring in* $\mathcal{A}$*, whether* $\bar{c}$ *is an answer to* q *in every model of* $\mathcal{O}$*.*

A local interpretation $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ w.r.t. $\mathcal{O}$ and q has its domain $\Delta^{\mathcal{I}}$ restricted to constants in the ABox and in the query, and interpretations of concepts and roles not appearing in $\mathcal{O}$ and in q are empty. It can be shown that $\bar{c} \notin ans(q, \mathcal{O})$, where $ans(q, \mathcal{O})$ is the set of *certain answers* to q w.r.t. $\mathcal{O}$, i.e., the set of answers satisfying q in every model of $\mathcal{O}$, if and only if there exists a local interpretation $\mathcal{I}$ w.r.t. $\mathcal{O}$ and q such that $\mathcal{I}$ is a model of $\mathcal{O}$ and $(c_1^{\mathcal{I}}, ..., c_n^{\mathcal{I}}) \notin q(\mathcal{I})$. This suggests a nondeterministic algorithm for deciding $\bar{c} \notin ans(q, \mathcal{O})$: guess a local interpretation $\mathcal{I}$ and check if it is a model of $\mathcal{O}$ and if $\bar{c} \notin ans(q, \mathcal{O})$. The size of $\mathcal{I}$ is linear in the size of the input. Checking whether $\mathcal{I}$ is a model of $\mathcal{O}$ is feasible in AC^0 in the size of $\mathcal{I}$, while checking whether $\bar{c} \notin q(\mathcal{I})$ is feasible in NP. Thus, the overall verification can be performed in NP.

This approach leads to the following result.

Theorem 1. *The problem ans($DL\text{-}Lite^{\neg}_{RDFS}$, $UCQ^{\neg s,\neq}$) is in Π^p_2 in combined complexity and in coNP in data complexity.*

This provides an upper bound. Matching lower bounds regarding data complexity were known for the case of two negated atoms [16] and two inequalities [10]. In the following, we show that matching lower bounds also hold for combined complexity for $CQ^{\neg s,2}$ and $CQ^{\neq 2}$.

Theorem 2. *The problem ans($DL\text{-}Lite_{RDFS}$, $UCQ^{\neg s,2}$) is Π^p_2-hard in combined complexity and coNP-hard in data complexity.*

This hardness result is obtained via a LogSpace reduction from ∀∃-CNF [17]. The reduction maps a ∀∃3CNF formula ϕ to a $DL\text{-}Lite^{\neg}_{RDFS}$ ontology $\mathcal{O}_\phi$ and a Boolean $UCQ^{\neg s}$ q_ϕ such that $\mathcal{O}_\phi \models q_\phi$ if and only if ϕ is true. The initial construction uses predicates of arity greater than two. However, the encoding can be transformed into an equivalent one that uses only unary and binary predicates, thus complying with the syntactic restrictions of $DL\text{-}Lite^{\neg}_{RDFS}$. A similar result holds for the query language $CQ^{\neg s,2}$ over $DL\text{-}Lite^{\neg}_{RDFS}$.

Theorem 3. *The problem ans($DL\text{-}Lite^{\neg}_{RDFS}$, $CQ^{\neg s,2}$) is Π^p_2-hard in combined complexity and coNP-hard in data complexity.*

The coNP-hardness comes from [16], while the Π^p_2-hardness is shown by an adaptation of the reduction from ∀∃3CNF used for Theorem 2, involving a fixed TBox with disjointness assertions and a slightly modified query and ABox construction. In contrast to the case of two negated atoms, the complexity drops significantly when only one safe negation atom is allowed per disjunct. In particular, ans($DL\text{-}Lite^{\neg}_{RDFS}$, $UCQ^{\neg s,1}$) can be polynomially reduced to checking entailment of a ground atom for a Datalog program P [18] that can be obtained by means of a transformation from an ontology $\mathcal{O}$, a $UCQ^{\neg s,1}$ q and a tuple $\bar{c}$, and is such that $\bar{c} \in ans(q, \mathcal{O})$ if and only if P entails a specific propositional atom. Combining this property with the known Datalog complexity for predicates with bounded arity [19, 20] results yields the following result.

Theorem 4. *The problem ans($DL\text{-}Lite^{\neg}_{RDFS}$, $UCQ^{\neg s,1}$) is NP-complete in combined complexity and P-complete in data complexity. The hardnesses already hold for ans($DL\text{-}Lite^{\neg}_{RDFS}$, $CQ^{\neg s,1}$).*

Note that NP-hardness follows from the NP-hardness of evaluating CQs over relational databases [21], while P-hardness in data complexity comes from [16].

Regarding the case of queries containing inequalities, previous work thoroughly analysed the case of $UCQ^{\neq}$s over $DL\text{-}Lite^{\neg}_{RDFS}$ ontologies, but only conjectured the Π^p_2-hardness of ans($DL\text{-}Lite^{\neg}_{RDFS}$, $UCQ^{\neq 2}$). The next result confirms the conjecture.

Theorem 5. *The problem ans($DL\text{-}Lite^{\neg}_{RDFS}$, $CQ^{\neq 2}$) is Π^p_2-hard in combined complexity.*

This is proved via a LogSpace reduction from ∀∃3CNF, similar to the safe negation cases. The reduction constructs a $DL\text{-}Lite^{\neg}_{RDFS}$ ontology $\mathcal{O}_\phi$ and a Boolean $CQ^{\neq 2}$ q_ϕ from a ∀∃3CNF formula ϕ such that $\mathcal{O}_\phi \models q_\phi$ if and only if the formula ϕ is true. The ontology simulates propositional assignments, and the query structure checks for satisfiability.

Interestingly, our results on hardness for the case of $\mathsf{CQ}^{\neq_2}$s have implications for the *query containment* problem $\mathsf{cont}(\mathcal{Q}, \mathcal{Q}')$, which asks if $q(D) \subseteq q'(D)$ for all databases D, with $q \in \mathcal{Q}$, and $q' \in \mathcal{Q}'$ for some query languages Q and Q'. Through Theorem 5 and a reduction from ontology-mediated query answering to query containment for ontologies without role disjointness assertions, we provide a tight complexity characterization based on the number of inequality atoms present in the queries.

Proposition 1. *Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a DL-Lite$^{\neg}_{\text{RDFS}}$ ontology without role disjointness assertions. It is possible to construct in polynomial time a Boolean CQ$^{\neq}$ $q_{\mathcal{O}}$ such that, for any Boolean CQ$^{\neq}$ q: $\mathcal{O} \models q$ if and only if $q_{\mathcal{O}} \sqsubseteq q$.*

By combining the reduction used for Theorem 5 and Proposition 1, we obtained the following:

Theorem 6. *The problem cont$(CQ^{\neq_1}, CQ^{\neq_2})$ is Π^p_2-complete.*

This is a significant finding, as $\mathsf{cont}(\mathsf{CQ}^{\neq_a}, \mathsf{CQ}^{\neq_b})$ is NP-complete if $a = 0$ or $b \leq 1$ [15, 10].

Finally, we analysed the problem of query answering under the UNA. In this case, the decision problem of interest is denoted by $\mathsf{ansU}(\mathcal{L}, \mathcal{Q})$. The same Π^p_2 combined complexity and coNP data complexity upper bounds as Theorem 1 hold for $\mathsf{ansU}(\textit{DL-Lite}^{\neg}_{\text{RDFS}}, \mathsf{UCQ}^{\neg_s,\neq})$ by using U-local interpretations, i.e., local interpretations under the UNA.

Theorem 7. *ansU$(DL\text{-}Lite^{\neg}_{RDFS}, UCQ^{\neg_s,\neq})$ is in Π^p_2 in combined complexity and in coNP in data complexity.*

For $\mathsf{UCQ}^{\neg_s}$ queries, *DL-Lite*$^{\neg}_{\text{RDFS}}$ is insensitive to the UNA, since it is a sub-logic of *DL-Lite*$_{\mathcal{R}}$, which is insensitive to the UNA for CQ answering [22]. This allows to derive that $\mathsf{ans}(q, \mathcal{O}) = \mathsf{ansU}(q, \mathcal{O})$, where $\mathcal{O}$ is a *DL-Lite*$^{\neg}_{\text{RDFS}}$, and $q \in \mathsf{UCQ}^{\neg_s}$. However, answering $UCQ^{\neq}$s over *DL-Lite*$^{\neg}_{\text{RDFS}}$ is significantly easier under the UNA, becoming tractable in data complexity.

Theorem 8. *ansU$(DL\text{-}Lite^{\neg}_{RDFS}, UCQ^{\neq})$ is NP-complete in combined complexity and in AC^0 in data complexity.*

We provided a thorough analysis of the complexity of answering queries with safe negation and inequalities over *DL-Lite*$^{\neg}_{\text{RDFS}}$ ontologies, confirming its potential as a theoretical foundation for AI systems based on KGs. We completed the picture of the complexity of the answering CQs with fixed numbers of inequality atoms by showing that, for $\mathsf{CQ}^{\neq_2}$, it is Π^p_2-hard, verifying a long-standing conjecture. We showed that CQs with safe negation exhibit a similar complexity jump from one to two negated atoms. These results provide tight complexity bounds and contribute to the understanding of query containment, showing that $\mathsf{cont}(\mathsf{CQ}^{\neq_1}, \mathsf{CQ}^{\neq_2})$ is Π^p_2-complete. The decidability of $\mathsf{ans}(\textit{DL-Lite}_{core}, \mathsf{CQ}^{\neq})$ remains an open problem. Our results suggest that answering $\mathsf{UCQ}^{\neg_s,\neq}$s over *DL-Lite*$^{\neg}_{\text{RDFS}}$ can potentially be implemented using systems designed for Π^p_2-complete problems, such as ASP solvers [23].

Acknowledgments

This work has been supported by MUR under the PNRR project FAIR (PE0000013).
This work has been carried out while Roberto Maria Delfino was enrolled in the Italian National Doctorate on Artificial Intelligence run by Sapienza University of Rome.

References

[1] A. Khan, Knowledge graphs querying 52 (2023) 18–29. URL: https://doi.org/10.1145/3615952.3615956. doi:10.1145/3615952.3615956.

[2] L. Yang, H. Chen, Z. Li, X. Ding, X. Wu, Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 3091–3110. doi:10.1109/TKDE.2024.3360454.

[3] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, X. Wu, Unifying large language models and knowledge graphs: A roadmap, IEEE Transactions on Knowledge and Data Engineering 36 (2024) 3580–3599. doi:10.1109/TKDE.2024.3352100.

[4] F. Baader, I. Horrocks, C. Lutz, U. Sattler, An Introduction to Description Logic, 2017.

[5] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati, Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family 39 (2007) 385–429.

[6] M. Bienvenu, M. Ortiz, Ontology-mediated query answering with data-tractable description logics, volume 9203, 2015, pp. 218–307.

[7] B. Cuenca Grau, A possible simplification of the semantic web architecture, 2004, pp. 704–713.

[8] R. Rosati, The limits of querying ontologies, volume 4353, 2007, pp. 164–178.

[9] G. Cima, A. Poggi, M. Lenzerini, The notion of abstraction in ontology-based data management 323 (2023) 103976.

[10] G. Cima, M. Lenzerini, A. Poggi, Answering conjunctive queries with inequalities in $DL\text{-}Lite_R$, 2020, pp. 2782–2789.

[11] M. Y. Vardi, The complexity of relational query languages (extended abstract), 1982, pp. 137–146.

[12] A. C. Klug, On conjunctive queries containing inequalities, J. ACM 35 (1988) 146–160.

[13] R. van der Meyden, The complexity of querying indefinite data about linearly ordered domains, J. Comput. Syst. Sci. 54 (1997) 113–135.

[14] P. G. Kolaitis, D. L. Martin, M. N. Thakur, On the complexity of the containment problem for conjunctive queries with built-in predicates, 1998, pp. 197–204.

[15] A. K. Chandra, P. M. Merlin, Optimal implementation of conjunctive queries in relational data bases, 1977, pp. 77–90.

[16] V. Gutiérrez-Basulto, Y. A. Ibáñez-García, R. Kontchakov, E. V. Kostylev, Queries with negation and inequalities over lightweight ontologies 35 (2015) 184–202.

[17] L. J. Stockmeyer, The polynomial-time hierarchy, Theor. Comput. Sci. 3 (1976) 1–22.

[18] S. Ceri, G. Gottlob, L. Tanca, What you always wanted to know about datalog (and never dared to ask) 1 (1989) 146–166.

[19] E. Dantsin, T. Eiter, G. Gottlob, A. Voronkov, Complexity and expressive power of logic programming 33 (2001) 374–425.

[20] G. Gottlob, T. Schwentick, Rewriting ontological queries into small nonrecursive Datalog programs, 2011.

[21] S. Abiteboul, R. Hull, V. Vianu, Foundations of Databases, 1995.

[22] A. Artale, D. Calvanese, R. Kontchakov, M. Zakharyaschev, The *DL-Lite* Family and Relations, Technical Report BBKCS-09-03, School of Computer Science and Information

Systems, Birbeck College, London, 2009. Available at http://www.dcs.bbk.ac.uk/research/techreps/2009/bbkcs-09-03.pdf.
[23] V. Lifschitz, Answer Set Programming, Springer, 2019.