Explanations for Query Answers under Existential Rules (Extended Abstract)*

İsmail İlkan Ceylan[1], Thomas Lukasiewicz[1], Enrico Malizia[2], and Andrius Vaicenavičius[1]

[1] University of Oxford, Department of Computer Science, UK
`firstname.lastname@cs.ox.ac.uk`
[2] King's College London, Department of Informatics, UK
`enrico.malizia@kcl.ac.uk`

Ontology-mediated query answering is an extensively studied paradigm, which aims at improving query answers with the use of a logical theory. Ontologies have found applications in *data exchange* [12], *medical diagnosis* [3], and *life sciences* [2], all of which benefit from explanations. As a form of logical entailment, ontology-mediated query answering is fully interpretable, which makes it possible to derive explanations for query answers. Surprisingly, explaining answers for ontology-mediated queries has received little attention for ontology languages based on existential rules. We close this gap, and study the problem of explaining query answers in terms of minimal subsets of database facts.

In this framework, the ontology and the user query are viewed as two components of one composite query, called *ontology-mediated query (OMQ)* [4], and throughout the paper, we assume that the ontology is expressed using *existential rules* [6, 7]. Given a database D and an OMQ (Q, Σ), where Q is a query and Σ is an ontology, we say that a subset E of the database D is a *minimal explanation, MinEX*, if E entails the query (Q, Σ), while no proper subset of E entails it. This definition incorporates the ideas from *axiom pinpointing* [15]. We illustrate minimal explanations on an example for protein design [14, 16].

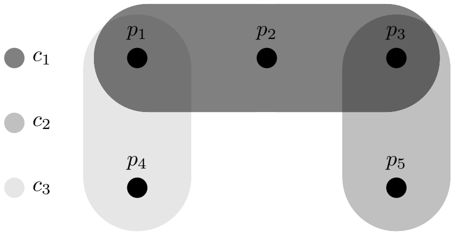

Fig. 1. Protein containment in complexes.

Example 1. Each protein $p_1, \dots, p_5$ is in one of the complexes c_1, c_2 and c_3, as shown in Fig 1. We want to find minimal subsets of proteins that cover all complexes. This can be expressed as a search for a MinEX, where the database is $D = \{prt(p_i) \mid 1 \leq i \leq 5\} \cup \{in(p_i, c_j) \mid p_i \text{ is in } c_j\}$, the ontology is $\Sigma = \{prt(P) \wedge in(P, C) \rightarrow cov(C)\}$, and the query is $Q = (cov(c_1) \wedge cov(c_2) \wedge cov(c_3)) \wedge \phi_{in}$, where ϕ_{in} is a conjuction of all *in* facts in D. By construction, the MinEXs for (Q, Σ) in D are in bijection with the protein covers of complexes. ■

---

* This is an abridged report of the paper that appeared at IJCAI-19 [10].

We study the following problems: (i) given a set, decide whether it is a *minimal explanation* (IS-MINEX), (ii) given a set of sets, decide whether it is the set of *all minimal explanations* (ALL-MINEX), (iii) given a distinguished assertion, decide whether it is *contained in some minimal explanation* (MINEX-REL), (iv) given a set of forbidden sets, decide whether there is a minimal explanation *not containing any forbidden sets* (MINEX-IRREL), (v)/(vi) given a number, decide whether there is a minimal explanation with *less (resp. more) elements than the given number* (SMALL-MINEX) (resp., (LARGE-MINEX)). We illustrate these problems on our running example, and refer to the original paper, for details [10].

Example 2. Consider the following subsets of the database:

$$E_1 = \{prt(p_1), prt(p_3)\} \cup D_{in}, \text{ and } E_2 = \{prt(p_2), prt(p_4), prt(p_5)\} \cup D_{in},$$

where D_{in} is the set of all *in* facts from D. The we have:

- IS-MINEX: Both E_1 and E_2 are MinEXs for (Q, Σ) in D. On the other hand, $E'_1 = \{prt(p_1), prt(p_4)\} \cup D_{in}$ and $E'_2 = \{prt(p_1), prt(p_2), prt(p_3)\} \cup D_{in}$ are not MinEXs as they do not entail the query and are not minimal, respectively.
- ALL-MINEX: Both $E_3 = \{prt(p_1), prt(p_5)\} \cup D_{in}$ and $E_4 = \{prt(p_3), prt(p_4)\} \cup D_{in}$ are MinEXs, and the set $\mathfrak{E} = \{E_1, E_2, E_3, E_4\}$ is the set of all MinEXs for (Q, Σ) in D.
- MINEX-REL: For each $i \in \{1, \ldots, 5\}$, the fact $prt(p_i)$ is contained in some MinEX for (Q, Σ) in D, and thus the fact is relevant.
- MINEX-IRREL: Let $\{\{prt(p_1), prt(p_3)\}, \{prt(p_5)\}\}$ be a set of forbidden sets of facts. Note that there is an explanation that does not contain any of these sets, which is E_4. Notice, however, E_1, E_2 and E_3 contain a forbidden set.
- LARGE-MINEX and SMALL-MINEX: All MinEXs, in this case, are either of size $2 + |D_{in}|$ or $3 + |D_{in}|$. ■

We provide a detailed complexity analysis in terms of *data*, *fp-combined*, *ba-combined* and *combined* complexity. We allow queries in the form of *unions of conjunctive queries*, which are coupled with existential rules ontologies that range from linear (L) and sticky (S) languages, to the expressive guarded (G) and full (F) fragments. In our data complexity analysis, we show that *all* these problems are *tractable* for FO-rewritable languages L, S, and A. Other tractability results in the data complexity are given for G and F for IS-MINEX. All the other results in the data complexity confirm the *hardness* of deriving explanations, as we always observe an increase in the complexity in comparison to the complexity of OMQA. Similarly, for the *fp*-, *ba*-, and combined complexities, we typically observe an increase of in the complexity compared to the complexity of OMQA.

Our study on explaining OMQA has recently been extended to description logics [11]. Explanations for standard reasoning tasks are extensively studied in DLs [13, 18, 1, 17, 15], but the literature on explaining OMQA remains rather sparse. Explanations for queries in *DL-Lite* has been studied earlier [5], and there is also a complementary work on explaining *negative* answers to OMQs in DLs [8], which was recently also studied in the context of existential rules [9].

References

1. Franz Baader and Boontawee Suntisrivaraporn. Debugging SNOMED CT using axiom pinpointing in the description logic EL+. In *Proc. KR-MED*, 2008.
2. Jonathan B. L. Bard and Seung Y. Rhee. Ontologies in biology: Design, applications and future challenges. *Nat. Rev. Genet.*, 5:213–222, 2004.
3. Valérie Bertaud-Gounot, Régis Duvauferrier, and Anita Burgun. Ontology and medical diagnosis. *Inform. Health Soc. Care*, 37(2):51–61, 2012.
4. Meghyn Bienvenu, Balder Ten Cate, Carsten Lutz, and Frank Wolter. Ontology-based data access: A study through disjunctive Datalog, CSP, and MMSNP. *ACM Trans. Database Syst.*, 39(4):33:1–33:44, 2014.
5. Alexander Borgida, Diego Calvanese, and Mariano Rodriguez-Muro. Explanation in the DL-Lite family of description logics. In *Proc. OTM*, pages 1440–1457, 2008.
6. Andrea Calì, Georg Gottlob, and Michael Kifer. Taming the infinite chase: Query answering under expressive relational constraints. *J. Artif. Intell. Res.*, 48:115–174, 2013.
7. Andrea Calì, Georg Gottlob, and Thomas Lukasiewicz. A general Datalog-based framework for tractable query answering over ontologies. *J. Web Semant.*, 14:57–83, 2012.
8. Diego Calvanese, Magdalena Ortiz, Mantas Šimkus, and Giorgio Stefanoni. Reasoning about explanations for negative query answers in DL-Lite. 48(1), 2013.
9. İsmail İlkan Ceylan, Thomas Lukasiewicz, Enrico Malizia, Cristian Molinaro, and Andrius Vaicenavičius. Explanations for negative query answers under existential rules. In *Proc. KR*, 2020.
10. İsmail İlkan Ceylan, Thomas Lukasiewicz, Enrico Malizia, and Andrius Vaicenavičius. Explanations for query answers under existential rules. In *Proc. IJCAI*, pages 1639–1646, 2019.
11. İsmail İlkan Ceylan, Thomas Lukasiewicz, Enrico Malizia, and Andrius Vaicenavičius. Explanations for ontology-mediated query answering in description logics. In *Proc. ECAI*, 2020.
12. Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: semantics and query answering. *TCS*, 336(1):89–124, 2005.
13. Aditya Kalyanpur, Bijan Parsia, Matthew Horridge, and Evren Sirin. Finding all justifications of OWL DL entailments. In *Proc. ISWC/ASWC*, pages 267–280, 2007.
14. Steffen Klamt, Utz-Uwe Haus, and Fabian Theis. Hypergraphs and cellular networks. *PLOS Comput. Biol.*, 5(5), 2009.
15. Rafael Peñaloza and Barış Sertkaya. Understanding the complexity of axiom pinpointing in lightweight description logics. *Artif. Intell.*, 250:80–104, 2017.
16. Emad Ramadan, Arijit Tarafdar, and Alex Pothen. A hypergraph model for the yeast protein complex network. In *Proc. IPDPS*, 2004.
17. Roberto Sebastiani and Michele Vescovi. Axiom pinpointing in lightweight description logics via Horn-SAT encoding and conflict analysis. In *Proc. CADE*, pages 84–99, 2009.
18. Boontawee Suntisrivaraporn, Guilin Qi, Qiu Ji, and Peter Haase. A modularization-based approach to finding all justifications for OWL DL entailments. In *Proc. ASWC*, pages 1–15, 2008.