Query-by-Example for Expressive Horn Description Logics*

Víctor Gutiérrez-Basulto[1], Jean Christoph Jung[2,3], and Leif Sabellek[2]

Cardiff University, UK[1] Universität Bremen, Germany[2] KU Leuven, Belgium[3]

We initiate the study of *query-by-example (QBE)* in the framework of ontology-based data access. Query-by-example is a reverse engineering problem that asks for the existence of a query complying with a set of specified examples. Formally, we study the following problem $\mathsf{QBE}(\mathcal{L},\mathcal{Q})$ for some ontology language $\mathcal{L}$ and a query language $\mathcal{Q}$:

- **Input**: $\mathcal{L}$-knowledge base $(\mathcal{T},\mathcal{A})$, sets S^+, S^-, and a signature Σ
- **Question**: is there a query $q(\mathbf{x}) \in \mathcal{Q}$ using only symbols in Σ such that
  - $(\mathcal{T},\mathcal{A}) \models q(\mathbf{a})$ for all $\mathbf{a} \in S^+$, and
  - $(\mathcal{T},\mathcal{A}) \not\models q(\mathbf{b})$ for all $\mathbf{b} \in S^-$?

As a simple example, consider the knowledge base consisting of

$$\mathcal{T} = \{\mathsf{Human} \sqsubseteq \mathsf{Vertebrate}, \mathsf{Vertebrate} \sqsubseteq \exists\mathsf{hasPart}.\mathsf{Spine}\},$$
$$\mathcal{A} = \{\mathsf{Human}(ax), \mathsf{hasPart}(an, sp), \mathsf{Spine}(sp), \mathsf{Bug}(bug)\}.$$

If the positive examples are $S^+ = \{ax, an\}$ and the negative example is $S^- = \{bug\}$, then $q(x) = \exists y\, \mathsf{hasPart}(x, y) \wedge \mathsf{Spine}(y)$ is a witness query. However, there is no witnessing query in the case $S^+ = \{an, bug\}$ and $S^- = \{ax\}$.

Motivation. Query-by-example is a classical querying paradigm [2] which has, due to 'big data', lately gained new interest since even expert users might find it useful to explore the data in this way. The practical motivation for our paper is to improve the usability of ontology-based systems. In such systems, users access the knowledge base through queries usually formulated in powerful query languages such as (unions of) conjunctive queries (U)CQs. Unfortunately, it has been observed that casual non-expert users are often not able to specify queries in these formalisms, e.g., Statoil geologists [3], clearly hampering the usability of ontology-based systems. Query-by-example suggests an alternative approach to querying: instead of writing the query itself, the user provides examples that she does (not) expect as answers and the system supports the user by providing a conforming query. This is possibly repeated until the user is satisfied.

From the theoretical perspective, our motivation is to continue bridging the gap between DL and machine learning research. Indeed, QBE over description logic knowledge bases can be viewed as an instantiation of the *inductive logic*

* Accepted at IJCAI-ECAI 2018 [1]. Authors were funded by EU's Horizon 2020 programme under the Marie Skłodowska-Curie grant 663830, the Research Foundation - Flanders project G.0428.15, and ERC consolidator grant 647289 CODA, respectively

programming (ILP) framework [4], where typically the knowledge base $\mathcal{P}$ is given in Prolog or a different rule language, and the goal is to learn a set of rules $\mathcal{P}'$ such that $\mathcal{P} \cup \mathcal{P}'$ conforms with the examples. The cases of our query languages correspond to single rule programs $\mathcal{P}'$ for CQs and sets of rules with the same head for UCQs, respectively.

It is worth mentioning that the query-by-example framework has been studied and implemented for $\mathcal{ALC}$-knowledge bases and $\mathcal{ALC}$ *concept queries* as query language [5].

Contribution. Our main contributions are characterizations, algorithms, and complexity bounds for $\mathsf{QBE}(\mathcal{L}, \mathcal{Q})$ for $\mathcal{L}$ an expressive Horn DL $\mathcal{L} \in \{$Horn-$\mathcal{ALCI}$, Horn-$\mathcal{ALC}\}$ and $\mathcal{Q} \in \{\mathrm{CQ}, \mathrm{UCQ}\}$. We start with providing natural model-theoretic characterizations for $\mathsf{QBE}($Horn-$\mathcal{ALCI}, \mathcal{Q})$ for $\mathcal{Q} \in \{\mathrm{CQ}, \mathrm{UCQ}\}$ by lifting characterizations known from the relational database setting [6] by replacing the database with the universal model of the knowledge base. Unfortunately, our characterizations do not give immediate rise to a decision procedure because the universal model is typically infinite. To overcome this, we exploit the regularity of universal models and provide decision procedures running in 2-ExpTime and coNExpTime for Horn-$\mathcal{ALCI}$ and Horn-$\mathcal{ALC}$, respectively. Having these, we prove matching lower bounds, the most challenging one being a 2-ExpTime-lower bound for $\mathsf{QBE}($Horn-$\mathcal{ALCI}, \mathcal{Q})$, $\mathcal{Q} \in \{\mathrm{CQ}, \mathrm{UCQ}\}$. Interestingly, some results depend on restricting the signature, so we consider also the variant QBE_f of QBE with unrestricted signature. Some of our techniques are inspired from approaches to query conservative extensions [7, 8], but we are not aware of direct reductions. The following table summarizes our results.

$\mathcal{L} \rightarrow$	Horn-$\mathcal{ALCI}$	Horn-$\mathcal{ALC}$
$\mathsf{QBE}(\mathcal{L}, \mathrm{CQ})$	2-ExpTime	coNExpTime
$\mathsf{QBE}(\mathcal{L}, \mathrm{UCQ})$	2-ExpTime	ExpTime
$\mathsf{QBE}_f(\mathcal{L}, \mathrm{CQ})$	2-ExpTime	coNExpTime
$\mathsf{QBE}_f(\mathcal{L}, \mathrm{UCQ})$	ExpTime	ExpTime

Most notably,

- in the cases of $\mathsf{QBE}_*($Horn-$\mathcal{ALC}, \mathrm{CQ})$, the availability of TBoxes does not increase the complexity compared to the relational database setting [9, 6],
- inverse roles lead to a jump in the complexity, and
- 2-ExpTime-hardness already holds for $\mathcal{ELI}$ TBoxes and concept queries.

We obtain the same results for the variant QDEF of QBE, the problem to decide whether some $q \in \mathcal{Q}$ returns *precisely* the positive examples. We also investigate the *size of witness queries* which is vital for practical purposes since eventually the user is interested in obtaining a query to further explore the data. We particularly show that they can be double exponentially large, which is in contrast to the relational database setting. We further show that – if a witness query exists – there is always one of at most double exponential size (Horn-$\mathcal{ALC}$), respectively fourfold exponential size (Horn-$\mathcal{ALCI}$). Such sizes are clearly impractical which motivates the study of approximations in the future.

References

1. Gutiérrez-Basulto, V., Jung, J.C., Sabellek, L.: Reverse engineering queries in ontology-enriched systems: The case of expressive horn description logic ontologies. In: Proc. of IJCAI-ECAI-18, AAAI Press (2018)
2. Zloof, M.M.: Query-by-example: the invocation and definition of tables and forms. In: Proc. of VLDB-75. (1975) 1–24
3. Hovland, D., Kontchakov, R., Skjæveland, M.G., Waaler, A., Zakharyaschev, M.: Ontology-based data access to Slegge. In: Proc. of ISWC-17. (2017) 120–129
4. Nienhuys-Cheng, S., de Wolf, R., eds.: Foundations of Inductive Logic Programming. Volume 1228 of Lecture Notes in Computer Science. Springer (1997)
5. Lehmann, J., Hitzler, P.: Concept learning in description logics using refinement operators. Machine Learning **78**(1) (Sep 2009) 203
6. ten Cate, B., Dalmau, V.: The product homomorphism problem and applications. In: Proc. of ICDT-15. (2015) 161–176
7. Botoeva, E., Konev, B., Lutz, C., Ryzhikov, V., Wolter, F., Zakharyaschev, M.: Inseparability and conservative extensions of description logic ontologies: A survey. In: Proc. of RW-16. (2016)
8. Jung, J., Lutz, C., Martel, M., Schneider, T.: Query conservative extensions in horn description logics with inverse roles. In: Proc. of IJCAI-17. (2017)
9. Barceló, P., Romero, M.: The complexity of reverse engineering problems for conjunctive queries. In: Proc. of ICDT-17. (2017)