

# Computing CQ Lower-Bounds over OWL 2 through Approximation to RSA <sup>\*</sup>

## Extended Abstract

Federico Igne<sup>[0000-0002-2790-7513]</sup>, Stefano Germano<sup>[0000-0001-6993-0618]</sup>, and  
Ian Horrocks<sup>[0000-0002-2685-7462]</sup>

Department of Computer Science, University of Oxford, Oxford, UK  
`firstname.lastname@cs.ox.ac.uk`

**Abstract** Conjunctive query (CQ) answering over knowledge bases is an important reasoning task. However, with expressive ontology languages such as OWL, query answering is computationally very expensive. The PAGOdA system addresses this issue by using a tractable reasoner to compute lower and upper-bound approximations, falling back to a fully-fledged OWL reasoner only when these bounds don't coincide. The effectiveness of this approach critically depends on the quality of the approximations, and in this paper we explore a technique for computing closer approximations via RSA, an ontology language that subsumes all the OWL 2 profiles while still maintaining tractability of standard reasoning tasks. We present a novel approximation of OWL 2 ontologies into RSA, and an algorithm to compute a closer lower bound approximation using the RSA combined approach. We have implemented these algorithms in our prototype and conducted an extensive evaluation thereof.

**Keywords:** CQ answering · combined approach · ontology approximation · RSA.

This extended abstract<sup>1</sup> is in support of our *poster* for the Posters and Demos Track of ISWC 2021, and related to the paper [7] in the Research Track. In the following sections we will provide an extensive analysis of the evaluation performed on the reference implementation presented alongside [7].

## 1 Introduction

Conjunctive query (CQ) answering is one of the primary reasoning tasks over knowledge bases for many applications. However, when considering expressive

---

<sup>\*</sup> This work was supported by the AIDA project (Alan Turing Institute), the SIR-IUS Centre for Scalable Data Access (Research Council of Norway, project no.: 237889), Samsung Research UK, Siemens AG, and the EPSRC projects AnaLOG (EP/P025943/1), OASIS (EP/S032347/1) and UK FIRES (EP/S019111/1).

<sup>1</sup> Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

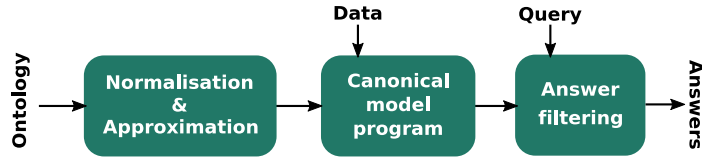


Figure 1. RSAComb architecture and workflow of the system.

description logic languages, query answering is computationally very expensive, even when considering only complexity w.r.t. the size of the data (*data complexity*). Fully-fledged reasoners oriented towards CQ answering over unrestricted OWL 2 ontologies exist but, although heavily optimised, they are only effective on small to medium datasets. In order to achieve tractability and scalability for the problem, two main approaches are often used: either the expressive power of the input ontology or the completeness of the computed answers is sacrificed.

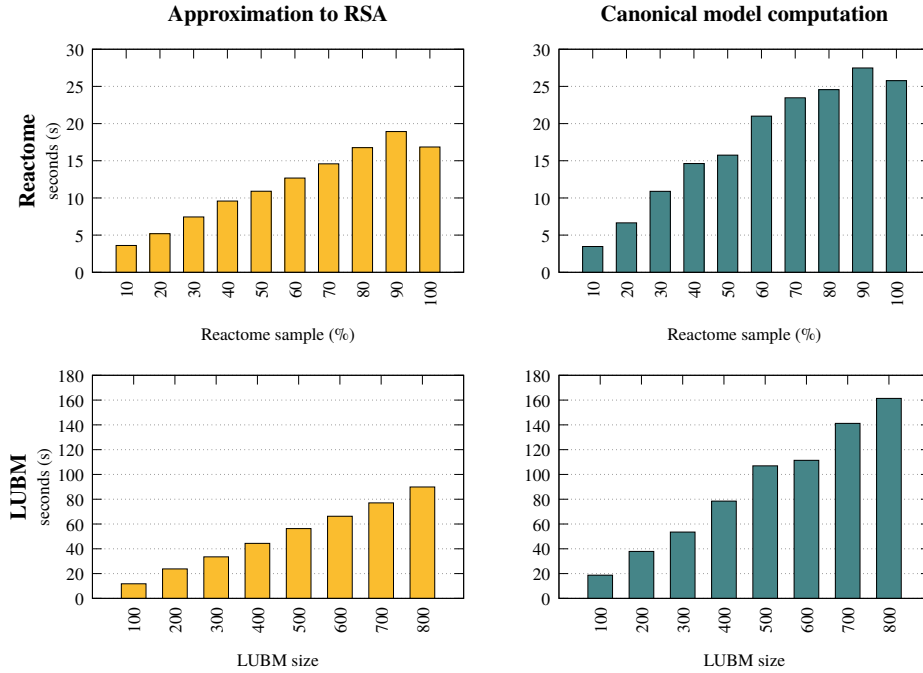
Using the first approach, query answering procedures have been developed for several fragments of OWL 2 for which CQ answering is tractable with respect to data complexity [1]. Three such fragments have been standardised as *OWL 2 profiles*, and CQ answering techniques for these fragments have been shown to be highly scalable at the expense of expressive power [2,9,10,12]. Using the second approach, several algorithms have been proposed to compute an approximation of the set of answers to a given CQ. This usually results in computing *a sound subset* of the answers, sacrificing completeness.

A particularly interesting approach to CQ answering over unrestricted OWL 2 ontologies is adopted by the reasoner PAGOdA [13]. Its “pay-as-you-go” approach allows to use a Datalog reasoner to handle the bulk of the answer computation, computing lower and upper approximations of the answers to a query, while relying on a fully-fledged OWL 2 reasoner like HerMiT [5] only as necessary to fully answer the query.

This work expands on this “pay-as-you-go” technique; it aims to improve the lower-bound approximation in PAGOdA, tightening the gap between lower and upper bounds and minimising the use of HerMiT. We achieve this by (soundly) approximating the input ontology into RSA [3], an ontology language that subsumes all the OWL 2 profiles and for which a CQ answering algorithm based on the *combined approach* has been proposed in [4]. We present a novel algorithm for approximating the input ontology into RSA, and an implementation [8] of the combined approach CQ answering algorithm adapted to the use of RDFox [11] as a backend Datalog reasoner; this includes the design of an improved version of the filtering step for the combined approach, optimised for RDFox.

## 2 Workflow

Figure 1 summarises the workflow of the system: (i) normalisation and customisable approximation steps approximate an unrestricted OWL 2 ontology to RSA; (ii) the canonical model is then computed for the resulting ontology; (iii) a



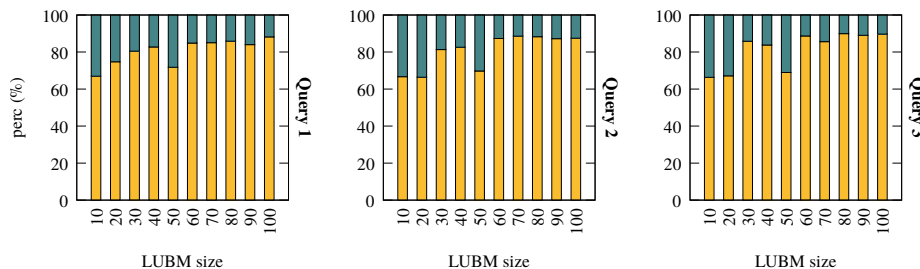
**Figure 2.** Scalability of approximation to RSA and canonical model computation.

Datalog filtering program is derived from the input query and is combined with the canonical model to produce the set of *certain* answers to the input query over the approximated ontology. Depending on how the approximation is performed, the set of answers returned might have different properties (e.g., the approximation currently provided computes a lower bound of the answers to the query over the original ontology).

It is worth noting that, in this scenario, steps (i),(ii) are *query independent*, while step (iii) is *ontology independent*. As such, when multiple queries are submitted, steps (1-2) can be performed “offline” and only the third step needs to be performed for each input query. We took advantage of this and streamlined the execution of the combined approach by factoring out those steps that are *query independent* to make answering multiple queries over the same knowledge base more efficient.

### 3 Evaluation

We have carried out an extensive evaluation to assess the effectiveness of the system and test the scalability and performance of the different steps in the execution of the combined approach. In Figure 2, we compare the scalability of approximation and canonical model computation steps over LUBM [6] and



**Figure 3.** Percent time distribution of canonical model computation (at the bottom, in yellow) and answer filtering (at the top, in blue) in Reactome.

Reactome<sup>2</sup>. The two steps are query independent and present a linear growth w.r.t. the dataset size of both knowledge bases.

Another interesting aspect is that most of the time the filtering step takes considerably less time than the canonical model computation. Figure 3 shows the percentage time distribution of three of the tested queries over Reactome. Filtering takes consistently less than 20% of the total execution time, when considering bigger datasets. As mentioned before, we can limit the impact of the canonical model computation by factoring out the task and computing it “offline” whenever we found ourselves in a scenario in which we need to perform query answering over a *fixed* ontology.

Overall, our experimental results show that the new technique yields significant performance improvements in several important application scenarios and solve some critical problems present in the original PAGOdA implementation.

## 4 Ongoing research

We are already working on additional improvements to the approximation algorithm to RSA; the current visit of the dependency graph to detect the axioms to delete might be improved with different heuristics and might in some cases take into account the input query (deleting axioms that are not necessarily involved in the computation of the answers). A similar approach could be introduced to integrate RSA in the upper-bound of the answers to a query, with the ultimate goal of improving this step in PAGOdA as well.

On a different note, we hope to obtain additional improvements in performance in the current implementation of the RSA combined approach by introducing parallel execution of filtering steps for different input queries, using the *named graph* functionality provided by RDFox.

<sup>2</sup> <https://reactome.org/>

## References

1. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Data complexity of query answering in description logics. In: Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning, Lake District of the United Kingdom, June 2-5, 2006. pp. 260–270. AAAI Press (2006)
2. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. Autom. Reasoning* **39**(3), 385–429 (2007). <https://doi.org/10.1007/s10817-007-9078-x>
3. Carral, D., Feier, C., Cuenca Grau, B., Hitzler, P., Horrocks, I.: Pushing the boundaries of tractable ontology reasoning. In: The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II. Lecture Notes in Computer Science, vol. 8797, pp. 148–163. Springer (2014). [https://doi.org/10.1007/978-3-319-11915-1\\_10](https://doi.org/10.1007/978-3-319-11915-1_10)
4. Feier, C., Carral, D., Stefanoni, G., Cuenca Grau, B., Horrocks, I.: The combined approach to query answering beyond the OWL 2 profiles. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015. pp. 2971–2977. AAAI Press (2015)
5. Glimm, B., Horrocks, I., Motik, B., Stoilos, G., Wang, Z.: Hermit: An OWL 2 reasoner. *J. Autom. Reasoning* **53**(3), 245–269 (2014). <https://doi.org/10.1007/s10817-014-9305-1>
6. Guo, Y., Pan, Z., Heflin, J.: LUBM: A benchmark for OWL knowledge base systems. *J. Web Semant.* **3**(2-3), 158–182 (2005). <https://doi.org/10.1016/j.websem.2005.06.005>
7. Igne, F., Germano, S., Horrocks, I.: Computing CQ lower-bounds over OWL 2 through approximation to RSA. In: International Semantic Web Conference (ISWC) (2021), [forthcoming]
8. Igne, F., Germano, S., Horrocks, I.: RSAComb - Combined approach for Conjunctive Query answering in RSA (Jun 2021). <https://doi.org/10.5281/zenodo.5047811>
9. Kontchakov, R., Lutz, C., Toman, D., Wolter, F., Zakharyashev, M.: The combined approach to query answering in dl-lite. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010. AAAI Press (2010)
10. Lutz, C., Toman, D., Wolter, F.: Conjunctive query answering in the description logic EL using a relational database system. In: Boutilier, C. (ed.) IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009. pp. 2070–2075 (2009), <http://ijcai.org/Proceedings/09/Papers/341.pdf>
11. Nenov, Y., Piro, R., Motik, B., Horrocks, I., Wu, Z., Banerjee, J.: Rdflox: A highly-scalable RDF store. In: The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II. Lecture Notes in Computer Science, vol. 9367, pp. 3–20. Springer (2015)
12. Stefanoni, G., Motik, B.: Answering conjunctive queries over EL knowledge bases with transitive and reflexive roles. *CoRR* **abs/1411.2516** (2014)
13. Zhou, Y., Cuenca Grau, B., Nenov, Y., Kaminski, M., Horrocks, I.: Pagoda: Pay-as-you-go ontology query answering using a datalog reasoner. *J. Artif. Intell. Res.* **54**, 309–367 (2015)