Closed-World Semantics for Conjunctive Queries with Negation over $\mathcal{ELH}_\bot$ Ontologies (Abstract)*

Stefan Borgwardt and Walter Forkel

Chair for Automata Theory, Technische Universität Dresden, Germany
firstname.lastname@tu-dresden.de

The biomedical domain is a fruitful area for ontology-mediated query answering (OMQA) methods, due to the availability of large ontologies covering a multitude of topics[1] and the demand for managing large amounts of patient data, in the form of *electronic health records (EHRs)* [7]. For example, for the preparation of clinical trials[2] a large number of patients need to be screened for eligibility, and an important area of current research is how to automate this process [4, 13, 16–18].[3] However, ontologies and EHRs mostly contain *positive* information, while clinical trials also require certain *exclusion criteria* to be absent in the patients. For example, we may want to select only patients that have *not* been diagnosed with cancer,[4] but such information cannot be entailed from the given knowledge. The culprit for this problem is the open-world semantics, which considers a cancer diagnosis possible unless it has been explicitly ruled out.

We consider here the problem of answering *conjunctive queries with (guarded) negation (NCQs)* [3] over ontologies formulated in $\mathcal{ELH}_\bot$, which covers many biomedical ontologies.

Related Work

One possibility to deal with negation is to use (partial) closed-world semantics for ontology languages [1, 14]. For example, one can declare the predicate *human* to be "closed", i.e. if an object is not explicitly listed as *human* in the dataset, then it is considered to be not human. However, such approaches fail to deal with anonymous objects; indeed, they conflate the open-world and open-domain assumptions by requiring that all closed information is restricted to the known objects. For example, even if we don't know the mother of a person A, we still know that she is human, even though this may not be explicitly stated in the ontology (but entailed by it). Using the semantics of [1, 14] would hence enforce a partial *closed-domain* assumption as well, in that A's mother would have to be a known object from the dataset.

* This is an abstract of a paper accepted at JELIA 2019. This work was supported by the DFG grant BA 1122/19-1 (GOASQ) and grant 389792660 (TRR 248) (see https://perspicuous-computing.science).

[1] https://bioportal.bioontology.org

[2] https://clinicaltrials.gov

[3] https://n2c2.dbmi.hms.harvard.edu

[4] An exclusion criterion in https://clinicaltrials.gov/ct2/show/NCT01463215

Epistemic logics are another way to give a closed-world-like semantics to negated formulas. There, one can formulate queries like “no cancer diagnosis is *known*” using the epistemic knowledge modality **K**. The semantics of this operator checks whether the inner formula is true in a specified set of possible worlds (e.g. the set of all models of the given knowledge base [6]). However, such formalisms are also unable to deal with closed-world knowledge over anonymous objects [19], because the behavior of a given anonymous object cannot be compared among multiple models.

One way to fix this is to *Skolemize* the knowledge base, effectively introducing a unique identifier (consisting of nested Skolem functions and an individual name) for each relevant anonymous object, allowing us to keep track of the object when comparing multiple (Herbrand) models. The problem with this approach is that all existential restrictions on the right-hand side of concept inclusions need to be Skolemized, and corresponding objects are forced to exist in all Herbrand models. However, some of these objects may be redundant and may lead to spurious answers to queries containing negation.

Most closely related to our proposal are Datalog-based semantics for negation, based on the *(Skolem) chase* construction [2, 11]. The important feature here is that a single canonical model is used for all inferences. However, using the Skolem chase suffers from the same drawback of Skolemization described above. One can avoid the creation of spurious objects with chase variants such as the *restricted chase* [10] or the *core chase* [8], which, however, do not always produce a unique canonical model. Moreover, to the best of our knowledge, there exist no results on query answering in Datalog-based languages with negation over these other chase variants.

Our Contribution

The contribution of this paper is a new closed-world semantics for answering NCQs over $\mathcal{ELH}_\bot$ ontologies. Our semantics is based on the *minimal canonical model*, which is a unique representation of all inferences of the ontology, in the most concise way possible. In order to properly handle negative knowledge about anonymous objects, we have to be careful in the construction of the minimal canonical model, in particular about the number and types of anonymous objects that are introduced. Instead of open-world semantics, we use a *minimal-world semantics*, in which NCQs are evaluated directly over the minimal canonical model. For the purposes of answering standard CQs without negation, the minimal canonical model behaves as a usual canonical model, which means that our semantics generalizes the standard open-world semantics. Since in general the minimal canonical model is infinite, we develop a rewriting technique, in the spirit of the combined approach of [12, 15], and most closely inspired by [5, 9], which allows us to evaluate rooted NCQs over a finite part of the canonical model using traditional database techniques.

The full version of the paper, including all proofs, can be found at `https://tu-dresden.de/inf/lat/papers`.

References

1. Ahmetaj, S., Ortiz, M., Simkus, M.: Polynomial datalog rewritings for expressive description logics with closed predicates. In: Kambhampati, S. (ed.) Proc. of the 25th Int. Joint Conf. on Artificial Intelligence (IJCAI'16). pp. 878–885. AAAI Press (2016), https://www.ijcai.org/Abstract/16/129
2. Arenas, M., Gottlob, G., Pieris, A.: Expressive languages for querying the semantic web. In: Hull, R., Grohe, M. (eds.) Proc. of the 33rd Symp. on Principles of Database Systems (PODS'14). pp. 14–26. ACM (2014). https://doi.org/10.1145/2594538.2594555
3. Bárány, V., ten Cate, B., Otto, M.: Queries with guarded negation. Proc. of the VLDB Endowment **5**(11), 1328–1339 (2012). https://doi.org/10.14778/2350229.2350250
4. Besana, P., Cuggia, M., Zekri, O., Bourde, A., Burgun, A.: Using semantic web technologies for clinical trial recruitment. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) Proc. of the 9th Int. Semantic Web Conf. (ISWC'10). Lecture Notes in Computer Science, vol. 6497, pp. 34–49. Springer (2010). https://doi.org/10.1007/978-3-642-17749-1_3
5. Bienvenu, M., Ortiz, M.: Ontology-mediated query answering with data-tractable description logics. In: Faber, W., Paschke, A. (eds.) Reasoning Web 11th International Summer School. Lecture Notes in Computer Science, vol. 9203, pp. 218–307. Springer (2015). https://doi.org/10.1007/978-3-319-21768-0_9
6. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Epistemic first-order queries over description logic knowledge bases. In: Parsia, B., Sattler, U., Toman, D. (eds.) Proc. of the 19th Int. Workshop on Description Logics (DL'06). CEUR Workshop Proceedings, vol. 189, pp. 51–61 (2006)
7. Cresswell, K.M., Sheikh, A.: Inpatient clinical information systems. In: Sheikh, A., Cresswell, K.M., Wright, A., Bates, D.W. (eds.) Key Advances in Clinical Informatics, chap. 2, pp. 13–29. Academic Press (2017). https://doi.org/10.1016/B978-0-12-809523-2.00002-9
8. Deutsch, A., Nash, A., Remmel, J.B.: The chase revisited. In: Lenzerini, M., Lembo, D. (eds.) Proc. of the 27th ACM SIGMOD-SIGACT-SIGART Symp. on Principles of Database Systems (PODS'08). pp. 149–158. ACM (2008). https://doi.org/10.1145/1376916.1376938
9. Eiter, T., Ortiz, M., Šimkus, M., Tran, T.K., Xiao, G.: Query rewriting for horn-$\mathcal{SHIQ}$ plus rules. In: Hoffmann, J., Selman, B. (eds.) Proc. of the 26th AAAI Conf. on Artificial Intelligence (AAAI'12). pp. 726–733. AAAI Press (2012), http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/4931
10. Fagin, R., Kolaitis, P.G., Miller, R.J., Popa, L.: Data exchange: Semantics and query answering. Theoretical Computer Science **336**(1), 89–124 (2005). https://doi.org/10.1016/j.tcs.2004.10.033
11. Hernich, A., Kupke, C., Lukasiewicz, T., Gottlob, G.: Well-founded semantics for extended datalog and ontological reasoning. In: Hull, R., Fan, W. (eds.) Proc. of the 32nd Symp. on Principles of Database Systems (PODS'13). pp. 225–236. ACM (2013). https://doi.org/10.1145/2463664.2465229
12. Kontchakov, R., Lutz, C., Toman, D., Wolter, F., Zakharyaschev, M.: The combined approach to ontology-based data access. In: Walsh, T. (ed.) Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI'11). pp. 2656–2661. AAAI Press (2011). https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-442

13. Köpcke, F., Prokosch, H.U.: Employing computers for the recruitment into clinical trials: A comprehensive systematic review. Journal of Medical Internet Research **16**(7), e161 (2014). https://doi.org/10.2196/jmir.3446
14. Lutz, C., Seylan, I., Wolter, F.: Ontology-based data access with closed predicates is inherently intractable (sometimes). In: Rossi, F. (ed.) Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI'13). pp. 1024–1030. AAAI Press (2013), https://www.ijcai.org/Abstract/13/156
15. Lutz, C., Toman, D., Wolter, F.: Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In: Boutilier, C. (ed.) Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09). pp. 2070–2075. AAAI Press (2009)
16. Ni, Y., Wright, J., Perentesis, J., Lingren, T., Deleger, L., Kaiser, M., Kohane, I., Solti, I.: Increasing the efficiency of trial-patient matching: Automated clinical trial eligibility pre-screening for pediatric oncology patients. BMC Medical Informatics and Decision Making **15**, 1–10 (2015). https://doi.org/10.1186/s12911-015-0149-3
17. Patel, C., Cimino, J., Dolby, J., Fokoue, A., Kalyanpur, A., Kershenbaum, A., Ma, L., Schonberg, E., Srinivas, K.: Matching patient records to clinical trials using ontologies. In: Aberer, K., Choi, K.S., Noy, N., Allemang, D., Lee, K.I., Nixon, L., Goldbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) Proc. of the 6th Int. Semantic Web Conf. (ISWC'07). Lecture Notes in Computer Science, vol. 4825, pp. 816–829. Springer (2007). https://doi.org/10.1007/978-3-540-76298-0_59
18. Tagaris, A., Andronikou, V., Chondrogiannis, E., Tsatsaronis, G., Schroeder, M., Varvarigou, T., Koutsouris, D.D.: Exploiting Ontology Based Search and EHR Interoperability to Facilitate Clinical Trial Design, pp. 21–42. Springer (2014). https://doi.org/10.1007/978-3-319-06844-2_3
19. Wolter, F.: First order common knowledge logics. Studia Logica **65**(2), 249–271 (2000). https://doi.org/10.1023/A:1005271815356