Closed-World Semantics for Conjunctive Queries with Negation over $\mathcal{ELH}_\perp$ 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\(^3\) \cite{4,13,16–18}. 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 $\mathcal{ELH}_\perp$, which covers many biomedical ontologies.

Related Work

One possibility to deal with negation is to use (partial) closed-world semantics for ontology languages\(^1\)\(^4\). 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\(^1\)\(^4\) 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](https://tu-dresden.de/inf/lat/papers)
References


