Closed-World Semantics for Conjunctive Queries with Negation over $\mathcal{ELH}_\perp$ Ontologies[*]

Stefan Borgwardt and **Walter Forkel**

Chair for Automata Theory, Technische Universität Dresden, Germany

firstname.lastname@tu-dresden.de

Abstract

Ontology-mediated query answering is a popular paradigm for enriching answers to user queries with background knowledge. For querying the *absence* of information, however, there exist only few ontology-based approaches. Moreover, these proposals conflate the closed-domain and closed-world assumption, and therefore are not suited to deal with the anonymous objects that are common in ontological reasoning. We propose a new closed-world semantics for answering conjunctive queries with negation over ontologies formulated in the description logic $\mathcal{ELH}_\perp$, based on the *minimal* canonical model. We propose a rewriting strategy for dealing with negated query atoms, which shows that query answering is possible in polynomial time in data complexity.

1 Introduction

Ontology-mediated query answering (OMQA) allows using background knowledge for answering user queries, supporting data-focused applications offering search, analytics, or data integration functionality. An *ontology* is a logical theory formulated in a decidable fragment of first-order logic, with a trade-off between the expressivity of the ontology and the efficiency of query answering.

Ontology-based systems do not use the *closed-domain* and *closed-world* semantics of databases. Instead, they acknowledge that unknown (*anonymous*) objects may exist (*open domain*) and that facts that are not explicitly stated may still be true (*open world*). Anonymous objects are related to *null* values in databases, but are not used explicitly; for example, if we know that every person has a mother, then first-order models include all mothers, even though they may not be mentioned in the input dataset. The open-world assumption ensures that, if the dataset does not contain an entry on, e.g. whether a person is male or female, then we do not infer that this person is neither male nor female, but rather consider all possibilities.

The biomedical domain is a fruitful area for OMQA, due to the availability of large ontologies[1] and the demand for managing large amounts of patient data, in the form of *electronic health records (EHRs)* [Cresswell and Sheikh, 2017]. 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 [Patel *et al.*, 2007; Besana *et al.*, 2010; Köpcke and Prokosch, 2014; Ni *et al.*, 2015].[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.

One possibility is to introduce (partial) closed-world semantics to ontology languages [Lutz *et al.*, 2013; Ahmetaj *et al.*, 2016]. 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, 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 [Lutz *et al.*, 2013; Ahmetaj *et al.*, 2016] 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.

Epistemic logics are another way to give a closed-world-like semantics to negated formulas; e.g. one can formulate queries like "no cancer diagnosis is *known*" using the epistemic knowledge modality **K**. Such formalisms are also unable to deal with closed-world knowledge over anonymous objects [Wolter, 2000; Calvanese *et al.*, 2006]. Most closely related to our proposal are Datalog-based semantics for negation, based on the (Skolem) chase construction [Hernich *et al.*, 2013]. We compare all these existing semantics in detail in Section 3.

The contribution of this paper is a new closed-world seman-

---

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

[2]https://clinicaltrials.gov

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

[4]An exclusion criterion of the clinical trial described at https://clinicaltrials.gov/ct2/show/NCT01463215

tics to answer *conjunctive queries with (guarded) negation* [Bárány *et al.*, 2012] over ontologies formulated in $\mathcal{ELH}_\perp$, an ontology language that covers many biomedical ontologies. Our semantics is based on the *minimal canonical model*, which encodes all inferences of the ontology in the most concise way possible. As a side effect, this means that standard CQs without negation are interpreted under the standard open-world semantics. In order to properly handle negative knowledge about anonymous objects, however, we have to be careful in the construction of the canonical model, in particular about the number and type of anonymous objects that are introduced. Since in general the minimal canonical model is infinite, we develop a rewriting technique, in the spirit of the combined approach of [Lutz *et al.*, 2009; Kontchakov *et al.*, 2011], and most closely inspired by [Eiter *et al.*, 2012; Bienvenu and Ortiz, 2015], which allows us to evaluate conjunctive queries with negation over a finite part of the canonical model, using traditional database techniques.

An extended version of this paper including proofs can be found at https://tu-dresden.de/inf/lat/papers.

2 Preliminaries

We recall the definitions of $\mathcal{ELH}_\perp$ and conjunctive queries. Let N_C, N_R, N_I be countably infinite sets of *concept*, *role*, and *individual names*, respectively. A *concept* is built according to the syntax rule $C ::= A \mid \top \mid \perp \mid C \sqcap C \mid \exists r.C$, where $A \in N_C$ and $r \in N_R$. An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a,b)$, where $a, b \in N_I$. A *TBox* is a finite set of *concept inclusions* $C \sqsubseteq D$ and *role inclusions* $r \sqsubseteq s$, where C, D are concepts and r, s are roles. In the following we assume that all inclusions are of the form

$$A_1 \sqcap \cdots \sqcap A_n \sqsubseteq B, \quad A \sqsubseteq \exists r.B, \quad \exists r.A \sqsubseteq B, \quad r \sqsubseteq s$$

where $A_{(i)} \in N_C \cup \{\top\}$, $B \in N_C \cup \{\perp\}$, $r, s \in N_R$, and $n \geq 1$. A *knowledge base (KB)* (or *ontology*) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is a pair of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. We write $C \equiv D$ to abbreviate the two inclusions $C \sqsubseteq D, D \sqsubseteq C$.

The semantics of $\mathcal{ELH}_\perp$ is defined in terms of interpretations $I = (\Delta^I, \cdot^I)$ as usual [Baader *et al.*, 2007]. An axiom α is *entailed* by $\mathcal{K}$ (written $\mathcal{K} \models \alpha$) if α is satisfied in all models of $\mathcal{K}$. We abbreviate $\mathcal{K} \models C \sqsubseteq D$ to $C \sqsubseteq_\mathcal{T} D$, and similarly for role inclusions. Entailment in $\mathcal{ELH}_\perp$ can be decided in polynomial time [Baader *et al.*, 2005].

A *conjunctive query* (CQ) $\phi(\mathbf{x})$ is a first-order formula $\exists \mathbf{y}. \varphi(\mathbf{x}, \mathbf{y})$, where φ is a conjunction of concept and role atoms. Let $I = (\Delta, \cdot^I)$ be a first-order interpretation. An *assignment* $\pi \colon \mathrm{Var}(\phi) \to \Delta$ *satisfies* ϕ in I if $I, \pi \models \phi$ under the semantics of first-order logic. Let $\mathcal{K}$ be a KB. A k-tuple $\mathbf{a}$ of individual names from $\mathcal{K}$ is an *answer* to ϕ in I if ϕ has a satisfying assignment π in I with $\pi(\mathbf{x}) = \mathbf{a}$; it is a *certain answer* to q in $\mathcal{K}$ if it is an answer to q in all models of $\mathcal{K}$. We denote the set of all answers to ϕ in I by $\mathrm{ans}(\phi, I)$, and the set of all certain answers to ϕ in $\mathcal{K}$ by $\mathrm{cert}(\phi, \mathcal{K})$.

CQ answering over $\mathcal{ELH}_\perp$ KBs is *combined first-order rewritable* [Lutz *et al.*, 2009]: For any CQ q and consistent KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ one can find a first-order query $q_\mathcal{T}$ and a finite interpretation $I'_\mathcal{K}$ with $\mathrm{cert}(q, \mathcal{K}) = \mathrm{ans}(q_\mathcal{T}, I'_\mathcal{K})$. This is based on the *canonical model* property of $\mathcal{ELH}_\perp$: For any

consistent KB $\mathcal{K}$ one can construct a model $I_\mathcal{K}$ that is homomorphically contained in any other model. This property is very useful since $\mathrm{cert}(q, \mathcal{K}) = \mathrm{ans}(q, I_\mathcal{K})$ holds for all CQs q.

3 Conjunctive Queries With Negation

We are interested in answering queries of the following form.

Definition 1. *Conjunctive queries with (guarded) negation* (NCQs) are constructed by extending CQs with *negated concept atoms* $\neg A(t)$ and *negated role atoms* $\neg r(t, t')$, such that, for any negated atom over terms t (and t') the query contains at least one positive atom over t (and t').

We first discuss different ways of handling negated atoms, and then propose a new semantics that is based on the *minimal canonical model*. We consider an example based on real EHRs (ABoxes) from the MIMIC-III database [Johnson *et al.*, 2016], criteria (NCQs) from clinicaltrials.gov, and the large medical ontology SNOMED CT[5] (the TBox).

We assume that the ABoxes have been extracted from EHRs by a natural language processing tool based, e.g. on existing concept taggers like [Aronson, 2001].

Example 2. We consider three patients. Patient p_1 (patient 2693 in the MIMIC-III dataset) is diagnosed with breast cancer and an unspecified form of cancer (this often occurs when there are multiple mentions of cancer in a patient's EHR, which cannot be resolved to the same entity). Patient p_2 (MIMIC-III patient 32304) suffers from breast cancer and skin cancer ("Stage IV breast cancer with mets to skin, bone, and liver"). For p_3 (MIMIC-III patient 88432), we know that p_3 has breast cancer that involves the skin ("Skin, left breast, punch biopsy: Poorly differentiated carcinoma").

Since SNOMED CT does not model patients, we add a new role name *diagnosedWith* to connect patients with diagnoses, which can be used in two ways. One can explicitly introduce individual names for diagnoses, e.g. diagnosedWith(p_1, d_1), BreastCancer(d_1), diagnosedWith(p_1, d_2), Cancer(d_2), implying that these diagnoses are treated as distinct entities under the standard name assumption. Alternatively, one can use complex assertions like $\exists$diagnosedWith.Cancer(p_1), which allows the logical semantics to resolve whether two diagnoses refer to the same object. Since ABoxes only contain concept names, in this case one has to introduce auxiliary definitions like CancerPatient $\equiv \exists$diagnosedWith.Cancer into the TBox. We use both variants in our example, to illustrate their different behaviours.

We obtain the KB $\mathcal{K}_C$, containing knowledge about different kinds of cancers and cancer patients, together with information about the three patients. The information about cancers is taken from SNOMED CT (in simplified form):

$$\mathrm{SkinCancer} \equiv \mathrm{Cancer} \sqcap \exists\mathrm{site.SkinStruct}$$
$$\mathrm{BreastCancer} \equiv \mathrm{Cancer} \sqcap \exists\mathrm{site.BreastStruct}$$
$$\mathrm{SkinOfBreastCancer} \equiv \mathrm{Cancer} \sqcap \exists\mathrm{site.SkinOfBreastStruct}$$
$$\mathrm{SkinOfBreastStruct} \sqsubseteq \mathrm{BreastStruct} \sqcap \mathrm{SkinStruct}$$

---

[5]https://www.snomed.org/snomed-ct
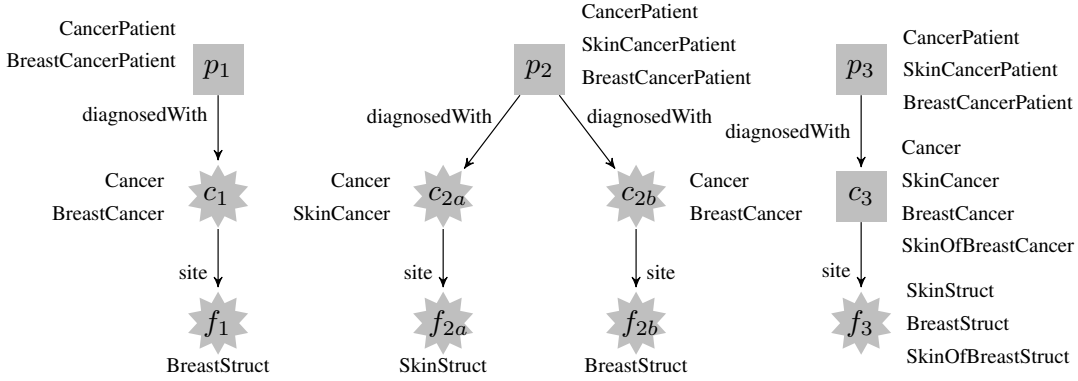
Figure 1: The minimal canonical model $I_{\mathcal{K}_C}$. Named individuals are depicted by squares, anonymous objects by stars.

The EHRs are compiled into several assertions per patient:

Patient p_1: BreastCancerPatient(p_1), CancerPatient(p_1)

Patient p_2: SkinCancerPatient(p_2), BreastCancerPatient(p_2)

Patient p_3: diagnosedWith(p_3, c_3), SkinOfBreastCancer(c_3)

Additionally, we add the following definitions to the TBox:

$$\text{CancerPatient} \equiv \exists \text{diagnosedWith.Cancer}$$
$$\text{SkinCancerPatient} \equiv \exists \text{diagnosedWith.SkinCancer}$$
$$\text{BreastCancerPatient} \equiv \exists \text{diagnosedWith.BreastCancer}$$

For example, skin cancers and breast cancers are cancers that have as (finding) site specific parts of the body ("body structure" in SNOMED CT), and a breast cancer patient is someone who is diagnosed with breast cancer. This means that, in every model of $\mathcal{K}_C$, every object that satisfies BreastCancerPatient (in particular p_2) must have a diagnosedWith-connected object that satisfies BreastCancer, and so on.

For a clinical trial,[6] we want to find patients that have "breast cancer", but not "breast cancer that involves the skin." This can be translated into an NCQ:

$$q_B(x) := \exists y, z. \, \text{diagnosedWith}(x, y) \wedge \text{Cancer}(y) \wedge$$
$$\text{site}(y, z) \wedge \text{BreastStruct}(z) \wedge \neg \text{SkinStruct}(z)$$

Since p_1 is diagnosed with BreastCancer as well as Cancer, and the former is more specific, we assume that the latter refers to the same object. However, we have no information about an involvement of the skin, so p_1 should be an answer to q_B.

We know that p_2 suffers from cancer in the skin and the breast, but not if the skin of the breast is also affected. Since neither location is implied by the other, we assume that they refer to distinct areas. p_2 should thus be an answer to q_B.

For p_3, it is explicitly stated that it is the same cancer that is occurring at the skin of the breast. In this case, the ABox assertions override the distinctness assumption we made for p_2. Thus, p_3 should not be an answer to q_B. ∎

In practice, more complicated cases than in our example can occur: For example, in SNOMED CT it is possible to describe many details of a cancer, such as the kind of cancer, whether it is a primary or secondary cancer, and in which part of the body it is found. This means that even a single assertion can lead to the introduction of multiple levels of anonymous objects in the canonical model. In some ontologies there are even cyclic concept inclusions, which lead to infinitely many anonymous individuals, e.g. in the GALEN ontology[7]. We focus on Example 2 in this paper, to illustrate the relevant issues in a clear and easy to follow manner. We now evaluate existing semantics on this example.

Epistemic Logic allows us to selectively apply closed-world reasoning using the modal knowledge operator $\mathbf{K}$. For a formula $\mathbf{K}\varphi$ to be true, it has to hold in all "connected worlds", which is often considered to mean all possible models of the KB, adopting an $S5$-like view [Calvanese *et al.*, 2006]. For q_B, we could read $\neg$SkinStruct(z) as "not known to be a skin structure", i.e. $\neg\mathbf{K}$SkinStruct(z). Consider the model $I_{\mathcal{K}_C}$ in Figure 1 and the assignment $\pi = \{x \mapsto p_3, y \mapsto c_3, z \mapsto f_3\}$, for which we want to check whether it satisfies q_B. Under epistemic semantics, $\neg\mathbf{K}$SkinStruct(z) is considered true if $\mathcal{K}$ has a (different) model in which f_3 does not belong to SkinStruct. However, f_3 is an anonymous object, and hence its name is not fixed. For example, we can easily obtain another model by renaming f_3 to f_1 and vice versa. Then f_3 would not be a skin structure, which means that $\neg\mathbf{K}$SkinStruct(z) is true in the original model $I_{\mathcal{K}_C}$, which is not as expected. This is a known problem with epistemic first-order logics [Wolter, 2000].

Skolemization can enforce a stricter comparison of anonymous objects between models. For example, the inclusion SkinCancer $\sqsubseteq$ $\exists$site.SkinStruct could be rewritten as the first-order sentence

$$\forall x. \Big(\text{SkinCancer}(x) \rightarrow \text{site}\big(x, f(x)\big) \wedge \text{SkinStruct}\big(f(x)\big) \Big),$$

where f is a fresh function symbol. This means that c_3 would be connected to a finding site that has the unique name $f(c_3)$ in every model. Queries would be evaluated over Herbrand models only. Hence, for evaluating $\neg\mathbf{K}$SkinStruct(z) when z is mapped to $f(c_3)$, we would only be allowed to compare the behavior of $f(c_3)$ in other Herbrand models. The general behavior of this anonymous individual is fixed, however, since in all Herbrand models it is *the* finding site of c_3. Then, since p_3

---

is inferred to be a BreastCancerPatient, the Skolemized version of BreastCancerPatient $\sqsubseteq \exists$diagnosedWith.BreastCancer introduces a new successor $g(p_3)$ of p_3 satisfying BreastCancer, which, together with the definition of BreastCancer, means that p_3 is an answer to q_B since there is an additional breast cancer diagnosis that does not involve the skin.

Datalog-based Ontology Languages with negation [Hernich *et al.*, 2013] are closely related to Skolemized ontologies, since their semantics is often based on the so-called *Skolem chase* [Marnette, 2009]. However, it suffers from the same drawback of Skolemization described above, due to superfluous successors. To avoid this, our semantics uses a special minimal canonical model (see Definition 4), which is similar to the *restricted chase* [Fagin *et al.*, 2005] or the *core chase* [Deutsch *et al.*, 2008], but always produces a unique model without having to merge domain elements.

3.1 Semantics for NCQs

We propose to answer NCQs over a special canonical model of the knowledge base. On the one hand, this eliminates the problem of tracking anonymous objects across different models, and on the other hand enables us to encode our assumptions directly into the construction of the model. In particular, we should only introduce the minimum necessary number of anonymous objects since, unlike in standard CQ answering, the precise shape and number of anonymous objects has an impact on the semantics of negated atoms.

Given $\mathcal{K}_C$, in contrast to the Skolemized semantics, we will not create both a generic "Cancer" and another "Breast-Cancer" successor for p_1, because the BreastCancer is also a Cancer, and hence the first object is redundant. Therefore, in the minimal canonical model of $\mathcal{K}_C$ depicted in Figure 1, for patient p_1 only one successor is introduced to satisfy the definitions of both BreastCancerPatient and CancerPatient at the same time. In contrast, p_2 has two successors, because BreastCancer and SkinCancer do not imply each other. Finally, for p_3 the ABox contains a single successor that is a SkinOfBreastCancer, which implies a single site-successor that satisfies both SkinStruct and BreastStruct.

To detect whether an object required by an existential restriction $\exists r.A$ is redundant, we use the following notion:

Definition 3 (Structural Subsumption). Let $\exists r.A$, $\exists t.B$ be concepts with $A, B \in N_C$ and $r, t \in N_R$. We say that $\exists r.A$ is *structurally subsumed* by $\exists t.B$ (written $\exists r.A \sqsubseteq^s_{\mathcal{T}} \exists t.B$) if $r \sqsubseteq_{\mathcal{T}} t$ and $A \sqsubseteq_{\mathcal{T}} B$. Given a set V of existential restrictions, we say that $\exists r.A \in V$ is *minimal w.r.t.* $\sqsubseteq^s_{\mathcal{T}}$ *(in V)* if there is no $\exists t.B \in V$ such that $\exists t.B \sqsubseteq^s_{\mathcal{T}} \exists r.A$.

In contrast to standard subsumption, $\exists r.A$ is not structurally subsumed by $\exists t.B$ w.r.t. the TBox $\mathcal{T} = \{\exists r.A \sqsubseteq \exists t.B\}$, as neither $r \sqsubseteq_{\mathcal{T}} t$ nor $A \sqsubseteq_{\mathcal{T}} B$ hold.

Definition 4 (Minimal Canonical Model). Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELH}_\perp$ KB. We construct the *minimal canonical model* $I_{\mathcal{K}}$ of $\mathcal{K}$ as follows:

1. Set $\Delta^{I_{\mathcal{K}}} := N_I$ and $a^{I_{\mathcal{K}}} := a$ for all $a \in N_I$.

2. Define $A^{I_{\mathcal{K}}} := \{a \mid \mathcal{K} \models A(a)\}$ for all $A \in N_C$ and $r^{I_{\mathcal{K}}} := \{(a, b) \mid \mathcal{K} \models r(a, b)\}$ for all $r \in N_R$.

3. Repeat:

(a) Select $d \in \Delta^{I_{\mathcal{K}}}$ that has not been selected before and let $V := \{\exists r.B \mid d \in A^{I_{\mathcal{K}}} \text{ and } d \notin (\exists r.B)^{I_{\mathcal{K}}} \text{ with } A \sqsubseteq_{\mathcal{T}} \exists r.B, A, B \in N_C\}$.

(b) For each $\exists r.B \in V$ that is minimal w.r.t. $\sqsubseteq^s_{\mathcal{T}}$, add a fresh element e to $\Delta^{I_{\mathcal{K}}}$, for each $B \sqsubseteq_{\mathcal{T}} A$ add e to $A^{I_{\mathcal{K}}}$, and for each $r \sqsubseteq_{\mathcal{T}} s$ add (d, e) to $s^{I_{\mathcal{K}}}$.

By $I_{\mathcal{A}}$ we denote the restriction of $I_{\mathcal{K}}$ to named individuals, i.e. the result of applying only Steps 1 and 2, but not Step 3.

If Step 3 is applied fairly, i.e. such that each new domain element that is created in (b) is eventually also selected in (a), then $I_{\mathcal{K}}$ is indeed a model of $\mathcal{K}$ (if $\mathcal{K}$ is consistent at all). In particular, all required existential restrictions are satisfied at each domain element, because the existential restrictions that are minimal w.r.t. $\sqsubseteq^s_{\mathcal{T}}$ entail all others.

Moreover, $I_{\mathcal{K}}$ satisfies the properties expected of a canonical model [Lutz *et al.*, 2009; Eiter *et al.*, 2012]: it can be homomorphically embedded into any other model of $\mathcal{K}$, and therefore $\text{cert}(q, \mathcal{K}) = \text{ans}(q, I_{\mathcal{K}})$ holds for all CQs q. We now define the semantics of NCQs as described before, i.e. by evaluating them as first-order formulas over the minimal canonical model $I_{\mathcal{K}}$, which ensures that our semantics is compatible with the usual certain-answer semantics for CQs.

Definition 5 (Minimal-World Semantics). The *(minimal-world) answers* to an NCQ q over a consistent $\mathcal{ELH}_\perp$ KB $\mathcal{K}$ are $\text{mwa}(q, \mathcal{K}) := \text{ans}(q, I_{\mathcal{K}})$.

For Example 2, we get $\text{mwa}(q_B, \mathcal{K}_C) = \{p_1, p_2\}$ (see Figure 1), which is exactly as intended. Unfortunately, in general the minimal canonical model is infinite, and we cannot evaluate the answers directly. Hence, we employ a rewriting approach to reduce NCQ answering over the minimal canonical model to (first-order) query answering over $I_{\mathcal{A}}$ only.

In our full paper [Borgwardt and Forkel, 2019] we show that NCQ answering is combined first-order rewritable, and we obtain the following complexity result.

Theorem 6. *Checking whether a given tuple* **a** *is a closed-world answer to an NCQ* ϕ *over a consistent* $\mathcal{ELH}_\perp$ *KB* $\mathcal{K}$ *can be done in polynomial time in data complexity.*

More important than the complexity is that this approach can be used to evaluate NCQs using standard database methods, e.g. using views to define the finite interpretation $I_{\mathcal{A}}$ based on the input data in $\mathcal{A}$, and SQL queries to evaluate $\text{rew}_{\mathcal{T}}(\phi)$ over these views [Kontchakov *et al.*, 2011].

4 Future Work

We are working on an implementation with the aim to deal with NCQs over large ontologies like SNOMED CT. We will also further develop our approach to represent temporal and numeric information, such as the precise order and duration of a patient's illnesses and treatments, and the dosage of medications. Such information is important for evaluating the eligibility criteria of clinical trials.

Acknowledgments

References

[Ahmetaj *et al.*, 2016] Shqiponja Ahmetaj, Magdalena Ortiz, and Mantas Simkus. Polynomial datalog rewritings for expressive description logics with closed predicates. In *Proc. of the 25th Int. Joint Conf. on Artificial Intelligence (IJCAI'16)*, pages 878–885. AAAI Press, 2016.

[Aronson, 2001] Alan R Aronson. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. In *Proceedings of the AMIA Symposium*, pages 17–21. American Medical Informatics Association, 2001.

[Baader *et al.*, 2005] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI'05)*, pages 364–369. Professional Book Center, 2005.

[Baader *et al.*, 2007] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition, 2007.

[Bárány *et al.*, 2012] Vince Bárány, Balder ten Cate, and Martin Otto. Queries with guarded negation. *Proc. of the VLDB Endowment*, 5(11):1328–1339, 2012.

[Besana *et al.*, 2010] Paolo Besana, Marc Cuggia, Oussama Zekri, Annabel Bourde, and Anita Burgun. Using semantic web technologies for clinical trial recruitment. In *Proc. of the 9th Int. Semantic Web Conf. (ISWC'10)*, pages 34–49. Springer, 2010.

[Bienvenu and Ortiz, 2015] Meghyn Bienvenu and Magdalena Ortiz. Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web 11th Int. Summer School*, pages 218–307. Springer, 2015.

[Borgwardt and Forkel, 2019] Stefan Borgwardt and Walter Forkel. Closed-world semantics for conjunctive queries with negation over $\mathcal{ELH}_\bot$ ontologies. In *Proc. of the 16th European Conf. on Logics in Artificial Intelligence (JELIA'19)*, Rende, Italy, 2019. Springer.

[Calvanese *et al.*, 2006] Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati. Epistemic first-order queries over description logic knowledge bases. In *Proc. of the 19th Int. Workshop on Description Logics (DL'06)*, pages 51–61, 2006.

[Cresswell and Sheikh, 2017] Kathrin M. Cresswell and Aziz Sheikh. Inpatient clinical information systems. In *Key Advances in Clinical Informatics*, chapter 2, pages 13–29. Academic Press, 2017.

[Deutsch *et al.*, 2008] Alin Deutsch, Alan Nash, and Jeffrey B. Remmel. The chase revisited. In *Proc. of the 27th ACM Symp. on Principles of Database Systems (PODS'08)*, pages 149–158. ACM, 2008.

[Eiter *et al.*, 2012] Thomas Eiter, Magdalena Ortiz, Mantas Šimkus, Trung-Kien Tran, and Guohui Xiao. Query rewriting for horn-$\mathcal{SHIQ}$ plus rules. In *Proc. of the 26th AAAI Conf. on Artificial Intelligence (AAAI'12)*, pages 726–733. AAAI Press, 2012.

[Fagin *et al.*, 2005] Ronald Fagin, Phokion G. Kolaitis, Renée J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. *Theoretical Computer Science*, 336(1):89–124, 2005.

[Hernich *et al.*, 2013] André Hernich, Clemens Kupke, Thomas Lukasiewicz, and Georg Gottlob. Well-founded semantics for extended datalog and ontological reasoning. In *Proc. of the 32nd Symp. on Principles of Database Systems (PODS'13)*, pages 225–236. ACM, 2013.

[Johnson *et al.*, 2016] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(160035):1–9, 2016.

[Kontchakov *et al.*, 2011] Roman Kontchakov, Carsten Lutz, David Toman, Frank Wolter, and Michael Zakharyaschev. The combined approach to ontology-based data access. In *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI'11)*, pages 2656–2661. AAAI Press, 2011.

[Köpcke and Prokosch, 2014] Felix Köpcke and Hans-Ulrich Prokosch. Employing computers for the recruitment into clinical trials: A comprehensive systematic review. *Journal of Medical Internet Research*, 16(7):e161, 2014.

[Lutz *et al.*, 2009] Carsten Lutz, David Toman, and Frank Wolter. Conjunctive query answering in the description logic $\mathcal{EL}$ using a relational database system. In *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI'09)*, pages 2070–2075. AAAI Press, 2009.

[Lutz *et al.*, 2013] Carsten Lutz, Inanc Seylan, and Frank Wolter. Ontology-based data access with closed predicates is inherently intractable (sometimes). In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI'13)*, pages 1024–1030. AAAI Press, 2013.

[Marnette, 2009] Bruno Marnette. Generalized schema mappings: From termination to tractability. In *Proc. of the 28th Symp. on Principles of Database Systems (PODS'09)*, pages 13–22. ACM, 2009.

[Ni *et al.*, 2015] Yizhao Ni, Jordan Wright, John Perentesis, Todd Lingren, Louise Deleger, Megan Kaiser, Isaac Kohane, and Imre Solti. 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.

[Patel *et al.*, 2007] Chintan Patel, James Cimino, Julian Dolby, Achille Fokoue, Aditya Kalyanpur, Aaron Kershenbaum, Li Ma, Edith Schonberg, and Kavitha Srinivas. Matching patient records to clinical trials using ontologies. In *Proc. of the 6th Int. Semantic Web Conf. (ISWC'07)*, pages 816–829. Springer, 2007.

[Wolter, 2000] Frank Wolter. First order common knowledge logics. *Studia Logica*, 65(2):249–271, 2000.