Temporal Minimal-World Query Answering over Sparse ABoxes

STEFAN BORGWARDT and WALTER FORKEL and ALISA KOVTUNOVA

Chair for Automata Theory, Technische Universität Dresden, Germany
(*e-mail:* `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. Many real-world applications, like processing electronic health records (EHRs), also contain a temporal dimension, and require efficient reasoning algorithms. Moreover, since medical data is not recorded on a regular basis, reasoners must deal with sparse data with potentially large temporal gaps.

Our contribution consists of two main parts: In the first part we introduce a new closed-world semantics for answering conjunctive queries with negation over ontologies formulated in the description logic $\mathcal{ELH}_\perp$, which is 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. In the second part, we extend this minimal-world semantics for answering metric temporal conjunctive queries with negation over the lightweight temporal logic $\mathcal{TELH}_\perp^{\lozenge,\mathsf{lhs},-}$ and obtain similar rewritability and complexity results.

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, e.g. a description logic (DL) (Baader et al. 2017), with a trade-off between the expressivity of the ontology language and the efficiency of query answering. *Rewritability* is a popular topic of research, the idea being to reformulate ontological queries, such as conjunctive queries (CQs) enhanced by an ontology, into database queries that can be answered by traditional database management systems (Mugnier and Thomazo 2014; Eiter et al. 2012; Bienvenu and Ortiz 2015; Calvanese et al. 2017; Kharlamov et al. 2017).

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; 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 explicitly in the input dataset. In addition, 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.

In the last decade, OMQA has been extended to temporal description logics (DLs) that combine terminological and temporal knowledge representation capabilities (Wolter and Zakharyaschev 2000; Lutz et al. 2008; Artale et al. 2017). To obtain tractable reasoning procedures, lightweight temporal DLs have been developed (Artale et al. 2007; Gutiérrez-Basulto et al. 2016). The idea is to use temporal operators, often from the linear temporal logic LTL, inside DL axioms. For example, $\Diamond\exists$diagnosis.BrokenLeg $\sqsubseteq \exists$treatment.LegCast states that after breaking a leg one has to wear a cast. However, this basic approach cannot represent the distance of events, e.g. that the cast only has to be worn for a fixed amount of time. Recently, metric temporal ontology languages have been investigated (Gutiérrez-Basulto et al. 2016; Baader et al. 2017; Brandt et al. 2018), which allow to replace $\Diamond$ in the above axiom with $\Diamond_{[-8,0]}$, i.e. wearing the cast is required only if the leg was broken ≤ 8 time points (e.g. weeks) ago.

The biomedical domain is a fruitful area for 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)* (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; Tagaris et al. 2014; Ni et al. 2015). In particular, many clinical trials contain temporal eligibility criteria (Crowe and Tao 2015), such as: "type 1 diabetes with duration at least 12 months"[3]; "known history of heart disease or heart rhythm abnormalities"[4]; "CD4+ lymphocytes count > 250/mm3, for at least 6 months"[5]; or "symptomatic recurrent paroxysmal atrial fibrillation (PAF) (> 2 episodes in the last 6 months)"[6]. Moreover, measurements, diagnoses, and treatments in a patients' EHR are clearly valid only for a certain amount of time. To automatically screen patients according to the temporal criteria above, one needs a sufficiently powerful formalism that can reason about biomedical and temporal knowledge.

Additionally, 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,[7] 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. Unfortunately, existing approaches like (partial) closed-world semantics (Lutz et al. 2013; Ahmetaj et al. 2016) or epistemic logics are unable to deal with closed-world knowledge over anonymous objects (Wolter 2000; Calvanese et al. 2006).

In the first part of this paper, we introduce a new closed-world semantics to answer *con-*

---

[1] https://bioportal.bioontology.org
[2] https://clinicaltrials.gov
[3] https://clinicaltrials.gov/ct2/show/NCT02280564
[4] https://clinicaltrials.gov/ct2/show/NCT02873052
[5] https://clinicaltrials.gov/ct2/show/NCT02157311
[6] https://clinicaltrials.gov/ct2/show/NCT00969735
[7] An exclusion criterion in https://clinicaltrials.gov/ct2/show/NCT01463215

junctive queries with (guarded) negation (Bárány et al. 2012) over ontologies formulated in $\mathcal{ELH}_\perp$, an ontology language that covers many biomedical ontologies, e.g. SNOMED CT.[8] Our semantics, called *minimal-world 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 ordinary 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 types 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.

In the second part of the paper, we extend the minimal-world semantics to support also temporal information. When working with EHRs, which contain information for specific points in time only, it is especially important to be able to infer what happened to the patient in the meantime. For example, if a patient is diagnosed with a (currently) incurable disease like Diabetes, they will still have the disease at any future point in time. Similarly, if the EHR contains two entries of CD4Above250 four weeks apart, one may reasonably infer that this was true for the whole four weeks. We use the qualitative temporal DL $\mathcal{TELH}_\perp^{\Diamond,\mathsf{lhs},-}$ (Borgwardt et al. 2019) that can express the former statement by declaring Diabetes as *expanding* via the axiom $\Diamond\text{Diabetes} \sqsubseteq \text{Diabetes}$. Additionally, by a special kind of metric temporal operators it is allowed to write $\Diamond_4\text{CD4Above250} \sqsubseteq \text{CD4Above250}$, making the measurement *convex* for a specified length of time n (e.g. 4 weeks). This means that information is interpolated between time points of distance less than n, thereby computing a convex closure of the available information. The threshold n allows to distinguish the case where two mentions of CD4Above250 are years apart, and are therefore unrelated. Reasoning in $\mathcal{TELH}_\perp^{\Diamond,\mathsf{lhs},-}$ is tractable in data complexity, because $\Diamond$-operators are only allowed on the left-hand side of concept inclusions (Borgwardt et al. 2019; Gutiérrez-Basulto et al. 2016), which is also common for temporal DLs based on *DL-Lite* (Artale et al. 2013; Artale et al. 2015).

Here we consider the problem of answering metric temporal queries over $\mathcal{TELH}_\perp^{\Diamond,\mathsf{lhs},-}$ knowledge bases with the minimal-world semantics introduced in the first part of the paper. Our query language extends the temporal conjunctive queries from (Baader et al. 2015b) by metric temporal operators (Gutiérrez-Basulto et al. 2016; Baader et al. 2017) and negation. For example, we can use queries like $\Box_{[-12,0]}(\exists y.\text{diagnosedWith}(x,y) \land \text{Diabetes}(y))$ to select all patients x for which the first criterion from above is satisfied.

By extending our combined rewriting approach, we show that the data complexity of temporal query answering is not higher than for (positive) atemporal queries in $\mathcal{ELH}_\perp$, and also provide a tight combined complexity result of EXPSPACE. Unlike most research on temporal query answering (Baader et al. 2015b; Artale et al. 2015), we do not assume that input data is given for all time points in a certain interval, but rather at sporadic time points with arbitrarily large gaps (Brandt et al. 2018). The main technical difficulty

---

[8] `https://www.snomed.org/`

is to determine which additional time points are relevant for answering a query, and how to access these time points without having to fill infinitely many gaps.

This paper extends the conference paper (Borgwardt and Forkel 2019) by allowing also non-rooted queries and extends parts of (Borgwardt et al. 2019) by providing full proofs of all results and improving some of the running examples.

2 Preliminaries

We recall the definitions of $\mathcal{ELH}_\perp$ and first-order queries, which are needed for our rewriting of conjunctive queries with negation.

The Description Logic $\mathcal{ELH}_\perp$. Let $\mathbf{C}, \mathbf{R}, \mathbf{I}$ 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 $A \in \mathbf{C}$ and $r \in \mathbf{R}$. An *ABox* is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a,b)$, where $a, b \in \mathbf{I}$. 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 the TBox to be in normal form, i.e. that it contains only inclusions of the form

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

where $A_{(i)} \in \mathbf{C} \cup \{\top\}$, $B \in \mathbf{C} \cup \{\perp\}$, $r, s \in \mathbf{R}$, and $n \geq 1$. A *knowledge base (KB)* (or *ontology*) $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. We refer to the set of individual names occurring in $\mathcal{K}$ by $\mathrm{Ind}(\mathcal{K})$. We write $C \equiv D$ to abbreviate the two inclusions $C \sqsubseteq D$, $D \sqsubseteq C$, and similarly for role inclusions.

The semantics of $\mathcal{ELH}_\perp$ is defined as usual (Baader et al. 2007) in terms of *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where the *domain* $\Delta^{\mathcal{I}}$ is a non-empty set and the functions $\cdot^{\mathcal{I}}$ are extended from concept and role names inductively as follows:

$$\top^{\mathcal{I}} := \Delta^{\mathcal{I}} \qquad\qquad \perp^{\mathcal{I}} := \varnothing$$
$$(C \sqcap D)^{\mathcal{I}} := C^{\mathcal{I}} \cap D^{\mathcal{I}} \qquad\qquad (\exists r.C)^{\mathcal{I}} := \{d \in \Delta^{\mathcal{I}} \mid \exists e \in C^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}}\}$$

$\mathcal{I}$ is a *model* of (or *satisfies*) a concept inclusion $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds, a role inclusion $r \sqsubseteq s$ if $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$ holds, a concept assertion $A(a)$ if $a \in A^{\mathcal{I}}$, a role assertion $r(a,b)$ if $(a,b) \in r^{\mathcal{I}}$, and the KB $\mathcal{K}$ if it satisfies all axioms in $\mathcal{K}$. In the following, we assume all KBs to be consistent and make the standard name assumption, i.e. that for every individual name a in any interpretation $\mathcal{I}$ we have $a^{\mathcal{I}} = a$. 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; note that the ABox does not influence the entailment of inclusions. Entailment in $\mathcal{ELH}_\perp$ can be decided in polynomial time (Baader et al. 2005).

The Temporal Description Logic $\mathcal{TELH}_\perp^{\Diamond,\mathsf{lhs},-}$. We recall the relevant definitions from (Borgwardt et al. 2019), starting with the MTL operators that are used in $\mathcal{TELH}_\perp^{\Diamond,\mathsf{lhs},-}$. LTL formulas are formulated over a finite set P of *propositional variables*. In this section, we consider only formulas built according to the syntax rule $\varphi ::= p \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \Diamond_I \varphi$, where $p \in P$ and I is an interval in $\mathbb{Z}$. The semantics is given by *LTL-structures* $\mathfrak{W} = (w_i)_{i \in \mathbb{Z}}$,
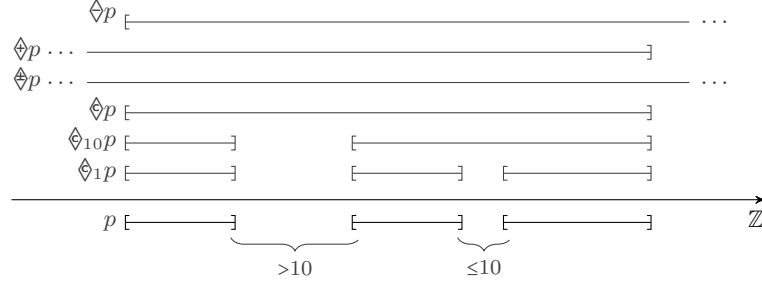
Figure 1. The resulting intervals of the induced functions of the diamond operators when applied to the intervals at which p holds (denoted below the timeline). While $\Diamond$ closes all gaps, $\Diamond_{10}$ does not close the first gap, since it is greater than 10 time units.

where $w_i \subseteq P$. We write

$$\mathfrak{W}, i \vDash p \text{ iff } p \in w_i \text{ if } p \in P, \qquad \mathfrak{W}, i \vDash \varphi \wedge \psi \text{ iff } \mathfrak{W}, i \vDash \varphi \text{ and } \mathfrak{W}, i \vDash \psi,$$
$$\mathfrak{W}, i \vDash \Diamond_I \varphi \text{ iff } \exists k \in I\colon \mathfrak{W}, i + k \vDash \varphi, \qquad \mathfrak{W}, i \vDash \varphi \vee \psi \text{ iff } \mathfrak{W}, i \vDash \varphi \text{ or } \mathfrak{W}, i \vDash \psi.$$

In $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$ the following derived operators are used, where $n \geq 1$.

$$\Diamond\varphi := \Diamond_{(-\infty,\infty)}\varphi \qquad\qquad \Diamond\varphi := \Diamond_{[0,\infty)}\varphi \qquad\qquad \Diamond\varphi := \Diamond_{(-\infty,0]}\varphi$$
$$\Diamond\varphi := \Diamond_{(-\infty,0]}\varphi \wedge \Diamond_{[0,\infty)}\varphi \qquad \Diamond_n\varphi := \bigvee_{\substack{k,m\geq 0 \\ k+m=n-1}} \left(\Diamond_{[-k,0]}\varphi \wedge \Diamond_{[0,m]}\varphi\right) \qquad (1)$$

The operator $\Diamond$ is the "eventually" operator of classical LTL, and $\Diamond, \Diamond$ are two variants that refer to the past, or to both past and future, respectively. The operator $\Diamond$ requires that φ holds *both* in the past and in the future, thereby distinguishing time points that lie within an interval enclosed by time points at which φ holds. This can be used to express the convex closure of time points, as described in the introduction. Finally, the operators $\Diamond_n$ represent a metric variant of $\Diamond$, requiring that different occurrences of φ are at most $n-1$ time points apart, i.e. enclose an interval of length n.

To make the behavior of these operators more clear, we consider their semantics in a more abstract way: given a set of time points where a certain information is available (e.g. a diagnosis), described by a propositional variable p, we consider the resulting set of time points at which $\Diamond p$ holds, where $\Diamond$ is a placeholder for one of the operators defined above (we will similarly use $\Diamond, \Diamond, \Diamond$ as placeholders for different $\Diamond$-operators in the following).

We consider the sets $\mathfrak{D}^{\mathsf{c}} := \{\Diamond\} \cup \{\Diamond_i \mid i \geq 1\}$, $\mathfrak{D}^{\pm} = \{\Diamond, \Diamond, \Diamond\}$, and $\mathfrak{D} := \mathfrak{D}^{\pm} \cup \mathfrak{D}^{\mathsf{c}}$ of *diamond operators*. Each $\Diamond \in \mathfrak{D}$ induces a function $\Diamond\colon 2^{\mathbb{Z}} \to 2^{\mathbb{Z}}$ with $\Diamond(M) := \{i \mid \mathfrak{W}_M, i \vDash \Diamond p\}$ for all $M \subseteq \mathbb{Z}$, with the LTL-structure $\mathfrak{W}_M := (w_i)_{i\in\mathbb{Z}}$ such that $w_i := \{p\}$ if $i \in M$, and $w_i := \varnothing$ otherwise. For an illustration see Figure 1. We will omit the parentheses in $\Diamond(M)$ for a cleaner presentation.

If M is empty, then $\Diamond M$ is also empty, for any $\Diamond \in \mathfrak{D}$. For any non-empty $M \subseteq \mathbb{Z}$, the following expressions are obtained, where $\max M$ may be ∞ and $\min M$ may be $-\infty$.

$$\Diamond M = \mathbb{Z} \qquad \Diamond M = (-\infty, \max M] \quad \Diamond M = [\min M, \infty) \quad \Diamond M = [\min M, \max M]$$
$$\Diamond_1 M = M \qquad \Diamond_n M = \{i \in \mathbb{Z} \mid \exists j, k \in M \text{ with } j \leq i \leq k \text{ and } k - j < n\}$$

Based on the operators in $\mathfrak{D}$ we give the definition of $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$ in the following.

A $\mathcal{TELH}_{\perp}^{\Diamond,\text{lhs},-}$ *concept* is built using the rule $C ::= A \mid \top \mid \perp \mid C \sqcap C \mid \exists r.C \mid \Diamond C$, where $A \in \mathbf{C}$, $\Diamond \in \mathfrak{D}$, and r is a role. Such a C is an $\mathcal{ELH}_{\perp}$ *concept* (or *atemporal concept*) if it does not contain any diamond operators.

A $\mathcal{TELH}_{\perp}^{\Diamond,\text{lhs},-}$ *TBox* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$ and *role inclusions (RIs)* $r \sqsubseteq s$, where C is a $\mathcal{TELH}_{\perp}^{\Diamond,\text{lhs}}$ concept, D is an atemporal concept, and $r, s \in \mathbf{R}$ are roles. Note that temporal roles are not allowed in $\mathcal{TELH}_{\perp}^{\Diamond,\text{lhs},-}$. An *ABox* is a finite set of *concept assertions* $A(a, i)$ and *role assertions* $r(a, b, i)$, where $A \in \mathbf{C}$, $r \in \mathbf{R}$, $a, b \in \mathbf{I}$, and $i \in \mathbb{Z}$. The set of time points $i \in \mathbb{Z}$ occurring in $\mathcal{A}$ we denote as $\text{tem}(\mathcal{A})$. Also we assume each time point is encoded in binary. In the following, we always assume a KB $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ to be given.

A *temporal interpretation* $\mathfrak{I} = (\Delta^{\mathfrak{I}}, (\mathcal{I}_i)_{i \in \mathbb{Z}})$, is a collection of interpretations $\mathcal{I}_i = (\Delta^{\mathfrak{I}}, \cdot^{\mathcal{I}_i})$, $i \in \mathbb{Z}$, over $\Delta^{\mathfrak{I}}$. The functions $\cdot^{\mathcal{I}_i}$ are extended to temporal concepts as follows:

$$(\Diamond C)^{\mathcal{I}_i} := \left\{ d \in \Delta^{\mathfrak{I}} \mid i \in \Diamond\{j \mid d \in C^{\mathcal{I}_j}\} \right\}$$

$\mathfrak{I}$ is a *model* of (or *satisfies*) a concept inclusion $C \sqsubseteq D$ if $C^{\mathcal{I}_i} \subseteq D^{\mathcal{I}_i}$ holds for all $i \in \mathbb{Z}$, a role inclusion $r \sqsubseteq s$ if $r^{\mathcal{I}_i} \subseteq s^{\mathcal{I}_i}$ holds for all $i \in \mathbb{Z}$, a concept assertion $A(a, i)$ if $a \in A^{\mathcal{I}_i}$, a role assertion $r(a, b, i)$ if $(a, b) \in r^{\mathcal{I}_i}$, and the KB $\mathcal{K}$ if it satisfies all axioms in $\mathcal{K}$. This fact is denoted by $\mathfrak{I} \models \alpha$, where α is an *axiom* (i.e. inclusion or assertion) or a KB. Entailment of CIs and assertions in $\mathcal{TELH}_{\perp}^{\Diamond,\text{lhs},-}$ is P-complete (Borgwardt et al. 2019).

As in many lightweight temporal logics, diamonds are not allowed to occur on the right-hand side of CIs, because that would allow to simulate concept disjunction and make the logic intractable (Artale et al. 2007; Gutiérrez-Basulto et al. 2016). As usual, we can simulate CIs involving complex concepts by introducing fresh concept and role names as abbreviations. For example, $\exists r.\Diamond A \sqsubseteq B$ can be split into $\Diamond A \sqsubseteq A'$, and $\exists r.A' \sqsubseteq B$. Hence, we can restrict ourselves w.l.o.g. to inclusions in the following *normal form*:

$$\Diamond A \sqsubseteq B, \ A \sqsubseteq \exists r.B, \ A_1 \sqcap A_2 \sqsubseteq B, \ r \sqsubseteq s, \ \exists r.A \sqsubseteq B, \tag{2}$$

where $\Diamond \in \mathfrak{D}$, $A, A_1, A_2, B \in \mathbf{C} \cup \{\perp, \top\}$, and $r, s \in \mathbf{R}$.

Conjunctive Query Answering. Let $\mathbf{V}$ be a countably infinite set of *variables*. The set of *terms* is $\mathbf{T} := \mathbf{V} \cup \mathbf{I}$. A *first-order query* $\phi(\mathbf{x})$ is a first-order formula built from *concept atoms* $A(t)$ and *role atoms* $r(t, t')$ with $A \in \mathbf{C}$, $r \in \mathbf{R}$, and $t_i \in \mathbf{T}$, using the boolean connectives $(\wedge, \vee, \neg, \rightarrow)$ and universal and existential quantifiers $(\forall x, \exists x)$. The free variables $\mathbf{x}$ of $\phi(\mathbf{x})$ are called *answer variables* and we say that ϕ is k-ary if there are k answer variables. The remaining variables are the *quantified variables*. We use $\text{Var}(\phi)$ to denote the set of all variables in ϕ. A query without any answer variables is called a *Boolean query*.

Let $\mathcal{I} = (\Delta, \cdot^{\mathcal{I}})$ be an interpretation. An *assignment* $\pi: \text{Var}(\phi) \rightarrow \Delta$ *satisfies* ϕ in $\mathcal{I}$, if $\mathcal{I}, \pi \models \phi$ under the standard semantics of first-order logic. We write $\mathcal{I} \models \phi$ if there is a satisfying assignment for ϕ in $\mathcal{I}$. Let $\mathcal{K}$ be a KB. A k-tuple $\mathbf{a}$ of individual names from $\text{Ind}(\mathcal{K})$ is an *answer* to ϕ in $\mathcal{I}$ if ϕ has a satisfying assignment π in $\mathcal{I}$ with $\pi(\mathbf{x}) = \mathbf{a}$; it is a *certain answer* to q over $\mathcal{K}$ if it is an answer to q in all models of $\mathcal{K}$. We denote the set of all answers to ϕ in $\mathcal{I}$ by $\text{ans}(\phi, \mathcal{I})$, and the set of all certain answers to ϕ over $\mathcal{K}$ by $\text{cert}(\phi, \mathcal{K})$.

A *conjunctive query* (CQ) $q(\mathbf{x})$ is a first-order query of the form $\exists \mathbf{y}. \varphi(\mathbf{x}, \mathbf{y})$, where φ is a conjunction of atoms. Abusing notation, we write $\alpha \in q$ if the atom α occurs in q,

and conversely may treat a set of atoms as a conjunction. The *leaf variables x* in q are those that do not occur in any atoms of the form $r(x, y)$. Clearly, q is satisfied in an interpretation if there is a satisfying assignment for $\varphi(\mathbf{x}, \mathbf{y})$, which is often called a *match* for q. A CQ is *rooted* if all variables are connected to an answer variable through (a sequence of) role atoms.

CQ answering over $\mathcal{ELH}_\perp$ KBs is *combined first-order rewritable* (Lutz et al. 2009): For any CQ q and consistent KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ we can find a first-order query $q_\mathcal{T}$ and a finite interpretation $\mathcal{I}_\mathcal{K}^{\text{fin}}$ such that $\text{cert}(q, \mathcal{K}) = \text{ans}(q_\mathcal{T}, \mathcal{I}_\mathcal{K}^{\text{fin}})$. Importantly, $\mathcal{I}_\mathcal{K}^{\text{fin}}$ is independent of q, i.e. can be reused to answer many different queries, while $q_\mathcal{T}$ is independent of $\mathcal{A}$, i.e. each query can be rewritten without using the (possibly large) dataset. The rewritability results are based crucially on the *canonical model* property of $\mathcal{ELH}_\perp$: For any consistent KB $\mathcal{K}$ one can construct a model $\mathcal{I}_\mathcal{K}$ that is homomorphically contained in any other model. This is a very useful property since any match in the canonical model corresponds to matches in all other models of $\mathcal{K}$, and therefore $\text{cert}(q, \mathcal{K}) = \text{ans}(q, \mathcal{I}_\mathcal{K})$ holds for all CQs q. For the complexity analysis, one considers the decision problem of whether a given tuple $\mathbf{a}$ belongs to $\text{cert}(q, \mathcal{K})$. The complexity of query answering is usually viewed in two different ways: In *combined complexity* the knowledge base, the data and the query are considered as input, while in *data complexity*, which is closely related to rewritability, only the data is considered as input. The latter is often viewed as the more appropriate, since the queries and the knowledge base are tend to be fixed and relatively small, compared to potentially huge amounts of data, that are updated frequently. CQ answering for $\mathcal{ELH}_\perp$ is P-complete in data complexity and NP-complete in combined complexity (Rosati 2007b).

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').

An NCQ is *rooted* if its variables are all connected via role atoms to an answer variable (from $\mathbf{x}$) or an individual name. An NCQ is *Boolean* if it does not have answer variables. To determine whether $\mathcal{I} \models \phi$ holds for an NCQ ϕ and an atemporal interpretation $\mathcal{I}$, we use standard first-order semantics.

We first discuss different ways of handling the negated atoms, and then propose a new semantics that is based on a particular kind of *minimal* canonical model. For this, 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[9] (the TBox). We omit here the "role groups" used in SNOMED CT, which do not affect the example. We also simplify the concept names and their definitions for ease of presentation. 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

---

[9] https://www.snomed.org/snomed-ct

2001; Savova et al. 2010); of course, this extraction is an entire research field in itself, which we do not attempt to tackle in this paper.

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 be the same entity). Patient p_2 (patient 32304 in the MIMIC-III dataset) suffers from breast cancer and skin cancer ("[S]tage IV breast cancer with mets to skin, bone, and liver"). For p_3 (patient 88432 in the MIMIC-III dataset), 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 special role name *diagnosedWith* that connects patients with their diagnoses. One can use this to express diagnoses in two ways. First, one can explicitly introduce individual names for diagnoses in assertions like 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 ∃diagnosedWith.Cancer(p_1), which allows the logical semantics to resolve whether two diagnoses actually refer to the same object. Since ABoxes only contain concept names, in this case one has to introduce auxiliary definitions like CancerPatient ≡ ∃diagnosedWith.Cancer into the TBox. We use both variants in our example, to illustrate their different behaviors.

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):

$$\text{SkinCancer} \equiv \text{Cancer} \sqcap \exists \text{findingSite.SkinStructure}$$
$$\text{BreastCancer} \equiv \text{Cancer} \sqcap \exists \text{findingSite.BreastStructure}$$
$$\text{SkinOfBreastCancer} \equiv \text{Cancer} \sqcap \exists \text{findingSite.SkinOfBreastStructure}$$
$$\text{SkinOfBreastStructure} \sqsubseteq \text{BreastStructure} \sqcap \text{SkinStructure}$$

The EHRs are compiled into several assertions per patient:

$$\text{Patient } p_1: \text{BreastCancerPatient}(p_1), \text{ CancerPatient}(p_1)$$
$$\text{Patient } p_2: \text{SkinCancerPatient}(p_2), \text{ BreastCancerPatient}(p_2)$$
$$\text{Patient } p_3: \text{diagnosedWith}(p_3, c_3), \text{ SkinOfBreastCancer}(c_3)$$

Additionally, we add the following auxiliary 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 occurring at 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. Moreover, a cancer may also occur in the
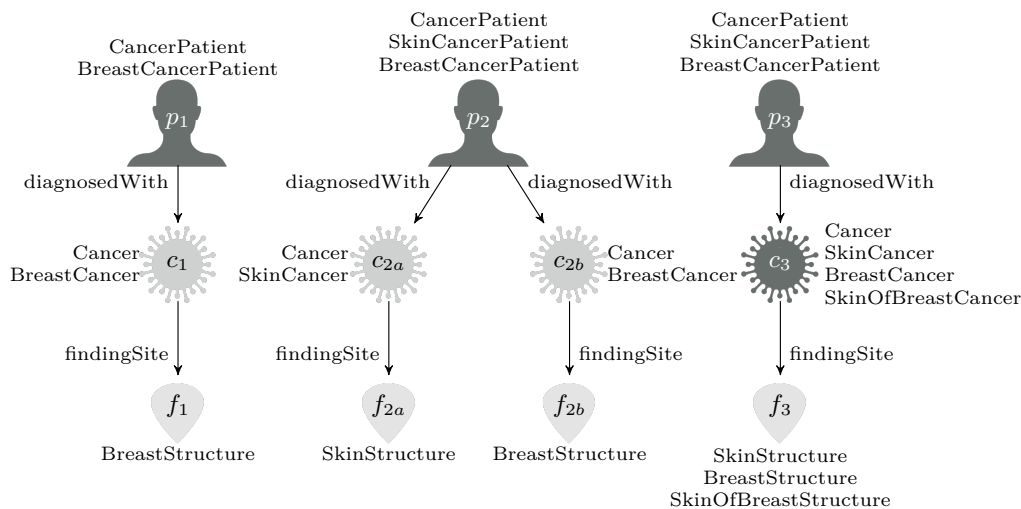
Figure 2. The minimal canonical model $\mathcal{I}_{\mathcal{K}_C}$. The named individuals p_1, p_2, p_3 and c_3 are depicted by dark grey figures, the remaining anonymous objects by light grey.

skin of the breast, i.e. a part of the body that is classified as both a "skin structure" and as a "breast structure".

For a clinical trial,[10] we want to find patients that have "breast cancer", but not "breast cancer that involves the skin." This can be translated into an NCQ:

$$q_B(x) := \exists y, z. \operatorname{diagnosedWith}(x, y) \wedge \operatorname{Cancer}(y) \wedge \operatorname{findingSite}(y, z) \wedge$$
$$\operatorname{BreastStructure}(z) \wedge \neg\operatorname{SkinStructure}(z)$$

We know that p_1 is diagnosed with BreastCancer as well as Cancer. Since the former is more specific, we assume that the latter refers to the same BreastCancer. However, since we have no information about an involvement of the skin, p_1 should be returned as 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.

In the case of p_3, it is explicitly stated that it is the same cancer that is occurring (not necessarily exclusively) 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: The nesting of anonymous objects will be deeper and more branched when using large biomedical ontologies. 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

---

[10] https://clinicaltrials.gov/ct2/show/NCT01960803

anonymous individuals, e.g. in the GALEN ontology[11]. 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.

Standard Certain Answer Semantics as defined in Section 2 is clearly not suited here, because one can easily construct a model of $\mathcal{K}_C$ in which c_1 is also a skin cancer, and hence p_1 is not an element of $\mathrm{cert}(q_B, \mathcal{K}_C)$. Moreover, under certain answer semantics answering CQs with guarded negation is already CONP-complete (Gutiérrez-Basulto et al. 2015), and hence not (combined) rewritable.

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 $\neg\mathrm{SkinStructure}(z)$ as "not known to be a skin structure", i.e. $\neg\mathbf{K}\mathrm{SkinStructure}(z)$. Consider the model $\mathcal{I}_{\mathcal{K}_C}$ in Figure 2 and the assignment $\pi = \{x \mapsto p_3, y \mapsto c_3, z \mapsto f_3\}$, for which we want to check whether it is a match for q_B. Under epistemic semantics, $\neg\mathbf{K}\mathrm{SkinStructure}(z)$ is considered true if $\mathcal{K}$ has a (different) model in which f_3 does not belong to SkinStructure. 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 $\neg\mathbf{K}\mathrm{SkinStructure}(z)$ is true in the original model $\mathcal{I}_{\mathcal{K}_C}$, which is not what we expected. This is a known problem with epistemic first-order logics (Wolter 2000).

Skolemization can enforce a stricter comparison of anonymous objects between models. The inclusion SkinOfBreastCancer $\sqsubseteq$ $\exists$findingSite.SkinOfBreast could be rewritten as the first-order sentence

$$\forall x. \Big(\mathrm{SkinOfBreastCancer}(x) \rightarrow \mathrm{findingSite}\big(x, f(x)\big) \wedge \mathrm{SkinOfBreast}\big(f(x)\big)\Big),$$

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 $\neg\mathbf{K}\mathrm{SkinStructure}(z)$ 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. While this improves the comparison by introducing pseudo-names for all anonymous individuals, it limits us in different ways: 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; Arenas et al. 2014) are closely related to Skolemized ontologies, since their semantics is often based on the so-called *Skolem chase* (Marnette 2009). This is closer to the semantics we propose in Section 3.1, in that a single canonical model is used for all inferences. However, it suffers

---

[11] http://www.opengalen.org/

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. Our minimal canonical model can also be seen as an instance of the approach in (Krötzsch 2020), which identifies conditions under which a chase variant produces a unique core model (for more details on this connection see Section 3.5.3 in (Forkel 2020)). However, (Krötzsch 2020) focuses on cases where this model is finite, which is not the case for general $\mathcal{ELH}_\perp$ ontologies. To the best of our knowledge, there exist no complexity results for Datalog-based languages with negation over these other chase variants.

Closed Predicates are a way to declare, for example, the concept name SkinStructure as "closed", which means that all skin structures must be declared explicitly, and no other SkinStructure object can exist (Lutz et al. 2013; Ahmetaj et al. 2016). This provides a way to give answers to negated atoms as in q_B. However, as explained in the introduction, this mechanism is not suitable for anonymous objects since it means that only named individuals can satisfy SkinStructure. When applied to $\mathcal{K}_C$, the result is even worse: Since there is no (named) SkinStructure object, no skin structures can exist at all and $\mathcal{K}_C$ becomes inconsistent. Closed predicates are appropriate in cases where the KB contains a full list of all instances of a certain concept name, and no other objects should satisfy it; but they are not suitable to infer negative information about anonymous objects. Moreover, CQ answering with closed predicates in $\mathcal{ELH}_\perp$ is already CONP-hard (Lutz et al. 2013).

Summary All of this should not be read as saying that these semantics are bad, just that they have not been developed with our use case involving anonymous objects in mind. Only Skolemization deals with anonymous objects by giving them unique names, but this forces all of them to be distinct. Our new semantics remedies this, but at the same time tries to stay as close as possible to the behavior of the existing semantics when restricted to the known objects in the ABox.

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 "BreastCancer" 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 2, 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 findingSite-successor that satisfies both SkinStructure and BreastStructure.

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

Definition 3 (Structural Subsumption)

Let $\exists r.A$, $\exists t.B$ be concepts with $A, B \in \mathbf{C}$ and $r, t \in \mathbf{R}$. 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 $\exists t.B \sqsubseteq^s_{\mathcal{T}} \exists r.A$.

A CQ $q_1(\mathbf{x})$ is *structurally subsumed* by a CQ $q_2(\mathbf{x})$ with the same answer variables (written $q_1 \sqsubseteq^s_{\mathcal{T}} q_2$) if, for all $x, y \in \mathbf{x}$, it holds that

$$\prod_{\alpha(x) \in q_1} \alpha \sqsubseteq_{\mathcal{T}} \prod_{\alpha(x) \in q_2} \alpha, \text{ and } \prod_{\alpha(x,y) \in q_1} \alpha \sqsubseteq_{\mathcal{T}} \prod_{\alpha(x,y) \in q_2} \alpha,$$

where role conjunction is interpreted in the standard way (Baader et al. 2007).

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. Similarly, structural subsumption for CQs considers all (pairs of) variables separately.

We use this notion to define the minimal canonical model.

Definition 4 (Minimal Canonical Model)

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

1. Set $\Delta^{\mathcal{I}_{\mathcal{K}}} := \mathbf{I}$ and $a^{\mathcal{I}_{\mathcal{K}}} := a$ for all $a \in \mathbf{I}$.
2. Define $A^{\mathcal{I}_{\mathcal{K}}} := \{a \mid \mathcal{K} \models A(a)\}$ for all $A \in \mathbf{C}$ and $r^{\mathcal{I}_{\mathcal{K}}} := \{(a,b) \mid \mathcal{K} \models r(a,b)\}$ for all $r \in \mathbf{R}$.
3. Repeat:
   (a) Select an element $d \in \Delta^{\mathcal{I}_{\mathcal{K}}}$ that has not been selected before and let
       $V := \{\exists r.B \mid d \in A^{\mathcal{I}_{\mathcal{K}}} \text{ and } d \notin (\exists r.B)^{\mathcal{I}_{\mathcal{K}}} \text{ with } A \sqsubseteq_{\mathcal{T}} \exists r.B, A, B \in \mathbf{C}\}$.
   (b) For each $\exists r.B \in V$ that is minimal w.r.t. $\sqsubseteq^s_{\mathcal{T}}$, add a fresh element e to $\Delta^{\mathcal{I}_{\mathcal{K}}}$, for each $B \sqsubseteq_{\mathcal{T}} A$ add e to $A^{\mathcal{I}_{\mathcal{K}}}$, and for each $r \sqsubseteq_{\mathcal{T}} s$ add (d,e) to $s^{\mathcal{I}_{\mathcal{K}}}$.

By $\mathcal{I}_{\mathcal{A}}$ we denote the restriction of $\mathcal{I}_{\mathcal{K}}$ 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 $\mathcal{I}_{\mathcal{K}}$ 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, $\mathcal{I}_{\mathcal{K}}$ 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 $\mathrm{cert}(q, \mathcal{K}) = \mathrm{ans}(q, \mathcal{I}_{\mathcal{K}})$ holds for all CQs q. It turns out that the minimal canonical model is also a *core*: Each endomorphism of $\mathcal{I}_{\mathcal{K}}$, i.e. a homomorphism from $\mathcal{I}_{\mathcal{K}}$ into itself, is injective, surjective and preserves negation (see also (Forkel 2020)).

Lemma 5

Every endomorphism of $\mathcal{I}_{\mathcal{K}}$ is a strong isomorphism.

Proof

To show this, we prove the stronger statement that the only endomorphism on $\mathcal{I}_{\mathcal{K}}$ is the identity. Let $\mathcal{I}_0$, $\mathcal{I}_1$, ... be the interpretations obtained in the construction of $\mathcal{I}_{\mathcal{K}}$ before each application of Step 3. We show by induction on i that there are h_0, h_1, ... such that h_i is the only possible homomorphism from $\mathcal{I}_i$ to $\mathcal{I}_{\mathcal{K}}$ and h_i and h_{i+1} agree on $\Delta^{\mathcal{I}_i}$, that is $h_{i+1}(d) = h_i(d)$ for all $d \in \Delta^{\mathcal{I}_i}$. The endomorphism is then obtained in the limit as $h = \bigcup_{i \geq 0} h_i$. By definition, the homomorphism h_0 has to map all constants onto itself, i.e. $h_0(a) = a$ for all $a \in \mathbf{I}$, and is therefore the identity function. For the induction step, assume that h_i has already been defined. To define h_{i+1}, assume that $d \in \Delta^{\mathcal{I}_i}$ was picked in Step 3(a) and V is the set as defined in Definition 4. For each $\exists r.B \in V$ that is minimal w.r.t. structural subsumption, in Step 3(b) exactly one successor e is introduced. Suppose there would be another successor e' of d, introduced through some minimal $\exists r'.B'$ and e could be mapped to e'. This would imply that $\exists r'.B' \sqsubseteq_{\mathcal{T}}^s \exists r.B$, which is a contradiction since we assumed $\exists r.B$ to be minimal. Hence, such an e' cannot exist and therefore the only possibility is to map e onto itself. Therefore, the only possibility is to define $h_{i+1} := h_i \cup \{e \mapsto e\}$, which is the identity function. $\square$

We can now adopt a result from graph theory and cores, namely Theorem 11 in (Bauslaugh 1995), which shows that, if two structures are homomorphically equivalent, then their cores are isomorphic. Since all canonical models are homomorphically equivalent by definition, this implies that every consistent $\mathcal{ELH}_\perp$-KB has a unique core model (up to isomorphism), which is why a minimal canonical model is also *the* minimal canonical model of $\mathcal{K}$.

We now define the semantics of NCQs as described before, i.e. by evaluating them as first-order formulas over the minimal canonical model $\mathcal{I}_{\mathcal{K}}$, which ensures that our semantics is compatible with the usual certain-answer semantics for CQs.

Definition 6 (*Minimal-World Semantics*)

Let $\mathcal{K}$ be a consistent $\mathcal{ELH}_\perp$ KB. The *(minimal-world) answers* to an NCQ q over $\mathcal{K}$ are $\mathrm{mwa}(q, \mathcal{K}) := \mathrm{ans}(q, \mathcal{I}_{\mathcal{K}})$.

For Example 2, we get $\mathrm{mwa}(q_B, \mathcal{K}_C) = \{p_1, p_2\}$ (see Figure 2), 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 $\mathcal{I}_{\mathcal{A}}$ only.

4 A Combined Rewriting for NCQs

We show that NCQ answering is combined first-order rewritable. As target representation, we obtain first-order queries of a special form.

Definition 7 (*Filtered query*)

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELH}_\perp$ KB. A *filter* on a variable z is a first-order expression $\psi(z)$ of the form

$$\left(\exists z'.\psi^+(z, z') \right) \to \left(\exists z'.\psi^+(z, z') \wedge \psi^-(z, z') \wedge \Psi \right) \tag{3}$$

where $\psi^+(z, z')$ is a conjunction of atoms of the form $A(z')$ or $r(z, z')$, that contains at

least one role atom, $\psi^-(z, z')$ is a conjunction of negated atoms $\neg A(z')$ or $\neg r(z, z')$, and Ψ is a (possibly empty) set of filters on z'.

A *filtered query* ϕ is of the form $\exists \mathbf{y}.\big(\varphi(\mathbf{x}, \mathbf{y}) \wedge \Psi\big)$ where $\exists \mathbf{y}.\varphi(\mathbf{x}, \mathbf{y})$ is an NCQ and Ψ is a set of filters on leaf variables in φ. It is *rooted* if $\exists \mathbf{y}.\varphi(\mathbf{x}, \mathbf{y})$ is rooted.

Note that every NCQ is a filtered query where the set of filters Ψ is empty.

We will use filters to check for the existence of "typical" successors, i.e. role successors that behave like the ones that are introduced by the canonical model construction to satisfy an existential restriction. In particular, a typical successor does not satisfy any superfluous concept or role atoms. For example, in Figure 2 the element c_1 introduced to satisfy $\exists$diagnosedWith.BreastCancer for p_1 is a typical successor, because it satisfies only BreastCancer and Cancer and not, e.g. SkinCancer. In contrast, the diagnosedWith-successor c_3 of p_3 is atypical, since the ontology does not contain an existential restriction $\exists$diagnosedWith.SkinOfBreastCancer that could have introduced such a successor in the canonical model.

The idea of the rewriting procedure is to not only rewrite the positive part of the query, as in (Eiter et al. 2012; Bienvenu and Ortiz 2015), but to also ensure that no critical information is lost. This is accomplished by rewriting the negative parts and by saving the structure of the eliminated part of the query in the filter. A filter on z ensures that the rewritten query can only be satisfied by mapping z to an anonymous individual in the canonical model, or to a named individual that behaves in a similar way.

Definition 8 (Rewriting)
Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a KB and $\phi = \exists \mathbf{y}.\varphi(\mathbf{x}, \mathbf{y}) \wedge \Psi$ be a filtered query. We write $\phi \rightarrow_{\mathcal{T}} \phi'$ if ϕ' can be obtained from ϕ by applying the following steps:

(S1) Select a quantified leaf variable $\hat{x}$ in φ. Let $\hat{y}$ be a fresh variable and select

$$
\begin{aligned}
\mathsf{Pred} &\coloneqq \{y \mid r(y, \hat{x}) \in \varphi\} \cup \{y \mid \neg r(y, \hat{x}) \in \varphi\} &&\text{(predecessors of } \hat{x}\text{)}, \\
\mathsf{Pos} &\coloneqq \{A(\hat{x}) \in \varphi\} \cup \{r(\hat{y}, \hat{x}) \mid r(y, \hat{x}) \in \varphi\} &&\text{(positive atoms for } \hat{x}\text{)}, \\
\mathsf{Neg} &\coloneqq \{\neg A(\hat{x}) \in \varphi\} \cup \{\neg r(\hat{y}, \hat{x}) \mid \neg r(y, \hat{x}) \in \varphi\} &&\text{(negative atoms for } \hat{x}\text{)}.
\end{aligned}
$$

(S2) Select some $M \sqsubseteq_{\mathcal{T}} \exists s.N$ with $M, N \in \mathbf{C}$ that satisfies all of the following:

 (a) $s(\hat{y}, \hat{x}) \wedge N(\hat{x}) \sqsubseteq^s_{\mathcal{T}} \mathsf{Pos}$, and
 (b) $s(\hat{y}, \hat{x}) \wedge N(\hat{x}) \not\sqsubseteq^s_{\mathcal{T}} \alpha$ for all $\neg \alpha \in \mathsf{Neg}$.

(S3) Let $\mathcal{M}'$ be the set of all $M' \in \mathbf{C}$ such that $M' \sqsubseteq_{\mathcal{T}} \exists s'.N'$ with $N' \in \mathbf{C}$,

 (a) $\exists s'.N' \sqsubseteq^s_{\mathcal{T}} \exists s.N$ (where $\exists s.N$ was chosen in (S2)), and
 (b) $s'(\hat{y}, \hat{x}) \wedge N'(\hat{x}) \sqsubseteq^s_{\mathcal{T}} \alpha$ for some $\neg \alpha \in \mathsf{Neg}$.

(S4) Drop from φ every atom that contains $\hat{x}$.
(S5) Replace all variables $y \in \mathsf{Pred}$ in φ with $\hat{y}$.
(S6) Add the atoms $M(\hat{y})$ and $\{\neg M'(\hat{y}) \mid M' \in \mathcal{M}'\}$ to φ.
(S7) Set the new filters to $\Psi' \coloneqq \Psi \cup \{\psi^*(\hat{y})\} \smallsetminus \Psi_{\hat{x}}$, where $\Psi_{\hat{x}} \coloneqq \{\psi(\hat{x}) \in \Psi\}$ and

$$
\psi^*(\hat{y}) \coloneqq \big(\exists \hat{x}.\, s(\hat{y}, \hat{x}) \wedge N(\hat{x})\big) \rightarrow \big(\exists \hat{x}.\, s(\hat{y}, \hat{x}) \wedge N(\hat{x}) \wedge \mathsf{Neg} \wedge \Psi_{\hat{x}}\big).
$$

We write $\phi \rightarrow^*_{\mathcal{T}} \phi'$ if there exists a finite sequence $\phi \rightarrow_{\mathcal{T}} \cdots \rightarrow_{\mathcal{T}} \phi'$. Furthermore, let $\mathrm{rew}_{\mathcal{T}}(\phi) \coloneqq \{\phi' \mid \phi \rightarrow^*_{\mathcal{T}} \phi'\}$ denote the set of all rewritings of ϕ.

Note that $\text{rew}_{\mathcal{T}}(\phi)$ may be infinite. However, for rooted NCQs it is finite and we show later that even for non-rooted NCQs it suffices to consider a finite subset of $\text{rew}_{\mathcal{T}}(\phi)$ (see Lemma 12). To see the former claim, observe that there is only a finite number of possible subsumptions $M \sqsubseteq_{\mathcal{T}} \exists s.N$ that can be used for rewriting steps. Additionally, in every step one variable ($\hat{x}$) is eliminated from the NCQ part of the filtered query. If the query is rooted, there always exists at least one predecessor that is renamed to $\hat{y}$, hence the introduction of $\hat{y}$ never increases the number of variables. Finally, it is easy to see that rewriting a rooted query always yields a rooted query.

The rewriting of Neg to the new negated atoms (via $\mathcal{M}'$ in (S6)) ensures that we do not lose important exclusion criteria, which may result in too many answers. Similarly, the filters exclude atypical successors in the ABox that may result in spurious answers. Both of these constructions are necessary.

Example 9

Consider the query q_B from Example 2. Using Definition 8, we obtain the first-order queries $\phi_B = q_B$, ϕ'_B, and ϕ''_B, where

$$\phi'_B = \exists y.\,\text{diagnosedWith}(x,y) \wedge \text{BreastCancer}(y) \wedge \neg \text{SkinOfBreastCancer}(y) \wedge$$
$$\Big(\big(\exists z.\, \text{findingSite}(y,z) \wedge \text{BreastStructure}(z) \big) \rightarrow$$
$$\big(\exists z.\, \text{findingSite}(y,z) \wedge \text{BreastStructure}(z) \wedge \neg \text{SkinStructure}(z) \big) \Big)$$

results from choosing z in (S1), $\text{BreastCancer} \sqsubseteq_{\mathcal{K}_C} \exists \text{findingSite.BreastStructure}$ in (S2), and computing $\mathcal{M}' = \{\text{SkinOfBreastCancer}\}$ in (S3), and

$$\phi''_B = \text{BreastCancerPatient}(x) \wedge$$
$$\Big(\big(\exists y.\, \text{diagnosedWith}(x,y) \wedge \text{BreastCancer}(y) \big) \rightarrow$$
$$\big(\exists y.\, \text{diagnosedWith}(x,y) \wedge \text{BreastCancer}(y) \wedge \neg \text{SkinOfBreastCancer}(y) \big) \wedge$$
$$\big(\big(\exists z.\, \text{findingSite}(y,z) \wedge \text{BreastStructure}(z) \big) \rightarrow$$
$$\big(\exists z.\, \text{findingSite}(y,z) \wedge \text{BreastStructure}(z) \wedge \neg \text{SkinStructure}(z) \big) \big) \Big)$$

is obtained due to $\text{BreastCancerPatient} \sqsubseteq_{\mathcal{K}_C} \exists \text{diagnosedWith.BreastCancer}$. We omitted the redundant atoms $\text{Cancer}(y)$ for clarity.

The finite interpretation $\mathcal{I}_{\mathcal{A}_C}$ can be seen in Figure 2 by ignoring all light grey figures. When computing the answers over $\mathcal{I}_{\mathcal{A}_C}$, we obtain

$$\text{ans}(\phi_B, \mathcal{I}_{\mathcal{A}_C}) = \varnothing,\ \text{ans}(\phi'_B, \mathcal{I}_{\mathcal{A}_C}) = \varnothing,\ \text{and}\ \text{ans}(\phi''_B, \mathcal{I}_{\mathcal{A}_C}) = \{p_1, p_2\}.$$

For ϕ'_B, the conjunct $\neg \text{SkinOfBreastCancer}(y)$ is necessary to exclude p_3 as an answer. In ϕ''_B, p_3 is excluded due to the filter that detects c_3 as an atypical successor, because it satisfies not only BreastCancer, but also SkinOfBreastCancer. Hence, both (S6) and (S7) are necessary steps in our rewriting. ∎

4.1 Correctness

In Definition 8, the new filter $\psi^*(\hat{y})$ may end up inside another filter expression after applying subsequent rewriting steps, i.e. by rewriting w.r.t. $\hat{y}$. In this case, however, the

original structure of the rewriting is preserved, including all internal filters as well as the atoms $M(\hat{y})$, which are included implicitly by $\exists s.N \sqsubseteq M$, and $\{\neg M'(\hat{y}) \mid M' \in \mathcal{M}'\}$, which are included in Neg. We exploit this behavior to show that, whenever a rewritten query is satisfied in the finite interpretation $\mathcal{I}_{\mathcal{A}}$, then it is also satisfied in $\mathcal{I}_{\mathcal{K}}$. This is the most interesting part of the correctness proof, because it differs from the known constructions for ordinary CQs, for which this step is trivial.

Lemma 10
Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent $\mathcal{ELH}_\perp$ KB and ϕ be an NCQ. Then, for all $\phi' \in \mathrm{rew}_{\mathcal{T}}(\phi)$,

$$\mathrm{ans}(\phi', \mathcal{I}_{\mathcal{A}}) \subseteq \mathrm{mwa}(\phi', \mathcal{K}).$$

Proof
Let $\phi' = \exists \mathbf{y}.(\varphi(\mathbf{x}, \mathbf{y}) \wedge \Psi)$ and π be an assignment of $\mathbf{x}, \mathbf{y}$ to $\mathbf{I}$ such that $\mathcal{I}_{\mathcal{A}}, \pi \models \varphi(\mathbf{x}, \mathbf{y})$. Since $\mathcal{I}_{\mathcal{A}}$ and $\mathcal{I}_{\mathcal{K}}$ coincide on the domain $\mathbf{I}$, we also have $\mathcal{I}_{\mathcal{K}}, \pi \models \varphi(\mathbf{x}, \mathbf{y})$. Consider any filter $\psi(z) = \exists z'.\psi^+(z, z') \rightarrow \exists z'.(\beta(z, z') \wedge \Psi^*)$ in Ψ, where $\beta(z, z') := \psi^+(z, z') \wedge \psi^-(z, z')$. Then $\psi(z)$ was introduced at some point during the rewriting, suppose by selecting $M \sqsubseteq_{\mathcal{T}} \exists s.N$ in (S2). This means that φ contains the atom $M(z)$, and hence $d := \pi(z)$ is a named individual that is contained in $M^{\mathcal{I}_{\mathcal{A}}} \subseteq M^{\mathcal{I}_{\mathcal{K}}}$. By (S2), this means that $\mathcal{I}_{\mathcal{K}}, \pi \models \exists z'.\psi^+(z, z')$, and we have to show that $\mathcal{I}_{\mathcal{K}}, \pi \models \exists z'.(\beta(z, z') \wedge \Psi^*)$:

1. If $\mathcal{I}_{\mathcal{A}}, \pi \models \exists z'.\beta(z, z')$, then $\mathcal{I}_{\mathcal{K}}, \pi \models \exists z'.\beta(z, z')$ by the same argument as for $\varphi(\mathbf{x}, \mathbf{y})$ above, and we can proceed by induction on the structure of the filters to show that the inner filters Ψ^* are satisfied by the assignment π (extended appropriately for z').

2. If $\mathcal{I}_{\mathcal{A}}, \pi \not\models \exists z'.\beta(z, z')$, then we cannot use a named individual to satisfy the filter $\psi(z)$ in $\mathcal{I}_{\mathcal{K}}$. Moreover, since $\mathcal{I}_{\mathcal{A}}$ satisfies $\psi(z)$, we also know that $\mathcal{I}_{\mathcal{A}}, \pi \not\models \exists z'.\psi^+(z, z')$. Since $\psi^+(z, z') = s(z, z') \wedge N(z')$, this implies that $d \notin (\exists s.N)^{\mathcal{I}_{\mathcal{A}}}$. Hence, $\exists s.N$ is included in the set V constructed in Step 3(a) of the canonical model construction for the element $d = \pi(z)$. Thus, there exists $M' \sqsubseteq_{\mathcal{T}} \exists s'.N'$ such that $d \in (M')^{\mathcal{I}_{\mathcal{A}}}$, $d \notin (\exists s'.N')^{\mathcal{I}_{\mathcal{A}}}$, and $\exists s'.N' \sqsubseteq_{\mathcal{T}}^s \exists s.N$. By Step 3(b), $\mathcal{I}_{\mathcal{K}}$ must contain an element d' such that, for any $A \in \mathbf{C}$ and any $r \in \mathbf{R}$, we have $d' \in A^{\mathcal{I}_{\mathcal{K}}}$ iff $N' \sqsubseteq_{\mathcal{T}} A$ and $(d, d') \in r^{\mathcal{I}_{\mathcal{K}}}$ iff $s' \sqsubseteq_{\mathcal{T}} r$. Since $N' \sqsubseteq_{\mathcal{T}} N$ and $s' \sqsubseteq_{\mathcal{T}} s$, we obtain that $\mathcal{I}_{\mathcal{K}}, \pi \cup \{z' \mapsto d'\} \models \psi^+(z, z')$.
   We show that the assignment $\pi \cup \{z' \mapsto d'\}$ also satisfies $\psi^-(z, z') = $ Neg. Assume to the contrary that there is $\neg A(z') \in$ Neg such that $d' \in A^{\mathcal{I}_{\mathcal{K}}}$ (the case of negated role atoms is again analogous). Then we have $N' \sqsubseteq_{\mathcal{T}} A$, which shows that all conditions of (S3) are satisfied, and hence M' must be included in $\mathcal{M}'$. Since the atoms $\{\neg M'(z) \mid M' \in \mathcal{M}'\}$ are contained in φ, we know that they are satisfied by π in $\mathcal{I}_{\mathcal{K}}$, i.e. $d \notin (M')^{\mathcal{I}_{\mathcal{K}}}$ and hence also $d \notin (M')^{\mathcal{I}_{\mathcal{A}}}$, which is a contradiction.
   It remains to show that the inner filters Ψ^* are satisfied by the assignment $\pi \cup \{z' \mapsto d'\}$ in $\mathcal{I}_{\mathcal{K}}$. Since we are now dealing with an anonymous domain element d', we can use similar, but simpler, arguments as above to prove this by induction on the structure of the filters. This is possible because the atoms $s(\hat{y}, \hat{x})$, $N(\hat{x})$ implied by $M(\hat{y})$ and the negated atoms induced by $\mathcal{M}'$ are present in the query even if the filter is integrated into another filter during a subsequent rewriting step. $\square$

We can use this lemma to show correctness of our approach, i.e. the answers returned for the *union* of queries given by $\mathrm{rew}_{\mathcal{T}}(\phi)$ over $\mathcal{I}_{\mathcal{A}}$ are exactly the answers of the original NCQ ϕ over $\mathcal{I}_{\mathcal{K}}$.

The proof is based on existing proofs for ordinary CQs (Eiter et al. 2012; Bienvenu and Ortiz 2015), extended appropriately to deal with the filters.

Lemma 11

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent $\mathcal{ELH}_\perp$ KB and let $\phi(\mathbf{x})$ be an NCQ. Then, for all $\phi' \in \mathrm{rew}_\mathcal{T}(\phi)$,

$$\mathrm{mwa}(\phi, \mathcal{K}) = \bigcup_{\phi' \in \mathrm{rew}_\mathcal{T}(\phi)} \mathrm{ans}(\phi', \mathcal{I}_\mathcal{A}).$$

Proof

($\supseteq$): By Lemma 10, we have $\mathrm{ans}(\phi', \mathcal{I}_\mathcal{A}) \subseteq \mathrm{mwa}(\phi', \mathcal{K}) = \mathrm{ans}(\phi', \mathcal{I}_\mathcal{K})$.

Furthermore, there exists a sequence $\phi_0 \to_\mathcal{T} \cdots \to_\mathcal{T} \phi_n$ $(n > 0)$ with $\phi = \phi_0$ and $\phi' = \phi_n$. Hence it is sufficient to show that $\mathrm{ans}(\phi_i, \mathcal{I}_\mathcal{K}) \subseteq \mathrm{ans}(\phi_{i-1}, \mathcal{I}_\mathcal{K})$ for all $i, 1 \le i \le n$. Suppose the queries are of the following forms:

$$\phi_i = \exists \mathbf{y}_i.(\varphi_i(\mathbf{x}_i, \mathbf{y}_i) \wedge \Psi_i) \tag{4}$$

$$\phi_{i-1} = \exists \mathbf{y}_{i-1}.(\varphi_{i-1}(\mathbf{x}_{i-1}, \mathbf{y}_{i-1}) \wedge \Psi_{i-1}) \tag{5}$$

Let π_i be a satisfying assignment for $\varphi_i(\mathbf{x}_i, \mathbf{y}_i) \wedge \Psi_i$ in $\mathcal{I}_\mathcal{K}$. Suppose $\phi_{i-1} \to_\mathcal{T} \phi_i$ by

1. selecting variable $\hat{x}$ and introducing $\hat{y}$ in (S1) and
2. selecting $M \sqsubseteq_\mathcal{T} \exists s.N$ in (S2).

Let $\pi_i(\hat{y}) = d$. By Step (S6), $M(\hat{y}) \in \varphi_i$ and since π_i satisfies φ_i, it has to hold that $d \in M^{\mathcal{I}_\mathcal{K}}$. This implies that $d \in (\exists s.N)^{\mathcal{I}_\mathcal{K}}$. Since π_i satisfies the new filter $\psi_i^*(\hat{y})$ that is constructed in (S7), and by selecting $M \sqsubseteq_\mathcal{T} \exists s.N$ in (S2) the precondition of $\psi_i^*(\hat{y})$ is satisfied by π_i in $\mathcal{I}_\mathcal{K}$, there has to be an assignment $\pi_i \cup \{\hat{x} \mapsto d'\}$ that satisfies the conclusion of $\psi_i^*(\hat{y})$.

We define the assignment π_{i-1} of the variables of φ_{i-1} as follows

$$\pi_{i-1}(z) := \begin{cases} d' & \text{if } z = \hat{x} \\ d & \text{if } z \in \mathsf{Pred} \\ \pi_i(z) & \text{otherwise.} \end{cases} \tag{6}$$

Then π_{i-1} is a satisfying assignment for ϕ_{i-1} in $\mathcal{I}_\mathcal{K}$. To see this, first consider an atom α in φ_{i-1}. We show that π_{i-1} satisfies α in $\mathcal{I}_\mathcal{K}$.

If α contains $\hat{x}$, it can be of the following forms: $A(\hat{x})$, $\neg A(\hat{x})$, $r(y, \hat{x})$ or $\neg r(y, \hat{x})$ with $y \in \mathsf{Pred}$. For all of these cases, we know by Step (S7) that they are either implied by $s(\hat{y}, \hat{x}) \wedge N(\hat{x})$ or contained in Neg, with y replaced by $\hat{y}$. By the choice of d', we know that π_{i-1} satisfies each such atom.

If α does not contain $\hat{x}$, then φ_i contains the atom α' that is obtained from α by replacing all of the variables from Pred with $\hat{y}$. By construction, we know that $\pi_{i-1}(y) = \pi_i(\hat{y})$ for all $y \in \mathsf{Pred}$ and $\pi_{i-1}(z) = \pi_i(z)$ otherwise. Since α' is satisfied by π_i in $\mathcal{I}_\mathcal{K}$, α is satisfied by π_{i-1} in $\mathcal{I}_\mathcal{K}$.

What remains to show is that π_{i-1} satisfies Ψ_{i-1}. Consider any $\psi(z) \in \Psi_{i-1}$, and distinguish the following cases:

1. If $z = \hat{x}$, then $\psi(\hat{x}) \in \Psi_{\hat{x}}$. Since $\mathcal{I}_\mathcal{K}, \pi_i \cup \{\hat{x} \mapsto d'\} \models \Psi_{\hat{x}}$, we also have $\mathcal{I}_\mathcal{K}, \{\hat{x} \mapsto d'\} \models \psi(\hat{x})$. Therefore, since $\pi_{i-1}(\hat{x}) = d'$, it holds that π_{i-1} satisfies $\psi(\hat{x})$ in $\mathcal{I}_\mathcal{K}$.

2. If $z \in \mathsf{Pred}$ we know that $\pi_{i-1}(z) = \pi_i(\hat{y}) = d$. Since $\mathcal{I}_{\mathcal{K}}, \pi_i \vDash \psi(\hat{y})$, it also holds that $\mathcal{I}_{\mathcal{K}}, \pi_{i-1} \vDash \psi(z)$.

3. Otherwise the filter is present in Ψ_i. In this case we know that $\mathcal{I}_{\mathcal{K}}, \pi_i \vDash \psi(z)$ and $\pi_i(z) = \pi_{i-1}(z)$. Hence, it must also hold that $\mathcal{I}_{\mathcal{K}}, \pi_{i-1} \vDash \psi(z)$.

($\subseteq$): Suppose that $\mathbf{a} \in \mathrm{mwa}(\phi, \mathcal{K}) = \mathrm{ans}(\phi, \mathcal{I}_{\mathcal{K}})$. We have to show that there exists a rewriting $\phi' \in \mathrm{rew}_{\mathcal{T}}(\phi)$ and a satisfying assignment π for ϕ' in $\mathcal{I}_{\mathcal{A}}$ such that $\mathbf{a} = \pi(\mathbf{x})$. To do this, we assign a *degree* (a natural number) to each satisfying assignment (including the existentially quantified variables of the NCQ part) such that a satisfying assignment with degree 0 does not use any anonymous individuals. We then show that for each satisfying assignment with a degree greater than 0, we can find a rewriting for which a satisfying assignment yielding the same answer, but with a lower degree, exists. In addition, for every such assignment π and for all filters $\psi(y)$ in ϕ' it should hold that,

$$\text{if } \pi(y) \in \mathbf{I}, \text{ then } \mathcal{I}_{\mathcal{A}} \vDash \psi(\pi(y)), \tag{†}$$

i.e. all filters (at any stage of the rewriting) are satisfied within the confines of $\mathcal{I}_{\mathcal{A}}$.

For any element $d \in \Delta^{\mathcal{I}_{\mathcal{K}}}$, we denote by $|d|$ the minimal number of role connections required to reach d from an element in $\mathbf{I}$, with $|d| = 0$ iff $d \in \mathbf{I}$. Additionally, for any assignment π' in $\mathcal{I}_{\mathcal{K}}$, let

$$\deg(\pi') := \sum_{y \in \mathrm{dom}(\pi')} |\pi'(y)|. \tag{7}$$

Since $\phi \in \mathrm{rew}_{\mathcal{T}}(\phi)$, to prove the claim it suffices to show that whenever there is $\phi_1 = \exists \mathbf{y}.\varphi_1(\mathbf{x}, \mathbf{y}) \wedge \Psi \in \mathrm{rew}_{\mathcal{T}}(\phi)$ such that φ_1 has a match π_1 in $\mathcal{I}_{\mathcal{K}}$ with $\mathbf{a} = \pi_1(\mathbf{x})$, $\deg(\pi_1) > 0$, and Equation (†) holds for π_1 and the filters in Ψ, then there exist ϕ_2 and π_2 with the same properties, but $\deg(\pi_2) < \deg(\pi_1)$.

Assume $\phi_1 \in \mathrm{rew}_{\mathcal{T}}(\phi)$ as above, and let π_1 be a match of φ_1. Since $\deg(\pi_1) > 0$ by assumption, there must exist a variable $\hat{x}$ of φ_1 such that $\pi_1(\hat{x}) \notin \mathbf{I}$. Select $\hat{x}$ such that it is a leaf node in the subforest of $\mathcal{I}_{\mathcal{K}}$ induced by π_1. Note that $\hat{x}$ cannot be an answer variable.

We know that $\pi_1(\hat{x}) = d_{\hat{x}}$ was induced by some axiom $\alpha = M \sqsubseteq_{\mathcal{T}} \exists s.N$ and element $d_p \in M^{\mathcal{I}_{\mathcal{K}}}$ in Definition 4. By the construction of $\mathcal{I}_{\mathcal{K}}$, we know that

(i) $d_{\hat{x}}$ has just the one predecessor d_p, and
(ii) $d_{\hat{x}} \in A^{\mathcal{I}_{\mathcal{K}}}$ iff $N \sqsubseteq_{\mathcal{T}} A$ and $(d_p, d_{\hat{x}}) \in r^{\mathcal{I}_{\mathcal{K}}}$ iff $s \sqsubseteq_{\mathcal{T}} r$.

We obtain the query ϕ_2 from ϕ_1 through rewriting, by selecting $\hat{x}$ and introducing $\hat{y}$ in (S1), and selecting α in (S2). Let Pred denote the set of predecessor variables of $\hat{x}$ as defined in (S1). To see that this is a valid choice, the conditions in (S2) need to be verified:

(S2a) For any $A(\hat{x}) \in \varphi_1$, we have $d_{\hat{x}} = \pi_1(\hat{x}) \in A^{\mathcal{I}_{\mathcal{K}}}$, and hence $N \sqsubseteq_{\mathcal{T}} A$ by (ii). Consider any role atom $r(y, \hat{x}) \in \varphi_1$. From (i), the construction of $\mathcal{I}_{\mathcal{K}}$ (no inverse edges), and the fact that π_1 is a satisfying assignment for $r(y, \hat{x})$ in $\mathcal{I}_{\mathcal{K}}$, the only possibility is that $\pi_1(y) = d_p$. Therefore $(d_p, d_{\hat{x}}) = (\pi_1(y), \pi_1(\hat{x})) \in r^{\mathcal{I}_{\mathcal{K}}}$. By (ii), this implies that $s \sqsubseteq_{\mathcal{T}} r$.

(S2b) Consider any $\neg A(\hat{x}) \in \varphi_1$, for which we must have $d_{\hat{x}} \notin A^{\mathcal{I}_{\mathcal{K}}}$. From (ii) we know that $N \not\sqsubseteq_{\mathcal{T}} A$. Consider any $\neg r(y, \hat{x}) \in \varphi_1$. Since this is guarded by a positive role atom as above, again the only possibility is that $\pi_1(y) = d_p$. Hence $(d_p, d_{\hat{x}}) \notin r^{\mathcal{I}_{\mathcal{K}}}$. By (ii), this implies that $s \not\sqsubseteq_{\mathcal{T}} r$.

Therefore, we obtain a satisfying assignment π_2 for ϕ_2 in $\mathcal{I}_\mathcal{K}$ such that $\mathbf{a} \in \pi_2(\mathbf{x})$ (and $\deg(\pi_2) < \deg(\pi_1)$) by setting for all $z \in \mathrm{Var}(\varphi_2)$:

$$\pi_2(z) := \begin{cases} \pi_1(z) & \text{if } z \in \mathrm{Var}(\varphi_1) \\ d_p & \text{if } z = \hat{y}. \end{cases}$$

To see that π_2 satisfies ϕ_2, we argue why it satisfies the new atoms and filter from (S6) and (S7); the old atoms (possibly with renamed variables) remain satisfied.

The new atom $M(\hat{y})$ is satisfied since $\pi_2(\hat{y}) = d_p \in M^{\mathcal{I}_\mathcal{K}}$. Consider now an atom $\neg M'(\hat{y})$ with $M' \in \mathcal{M}'$ as specified in (S6); we have to show that $d_p \notin (M')^{\mathcal{I}_\mathcal{K}}$. Assume to the contrary that $d_p \in (M')^{\mathcal{I}_\mathcal{K}}$. By (S3), we know that $M' \sqsubseteq_\mathcal{T} \exists s'.N' \sqsubseteq_\mathcal{T}^s \exists s.N$. Moreover, $\exists s'.N'$ must be included in the set V in Step 3(a) of Definition 4, because otherwise we would already have $d_p \in (\exists s'.N')^{\mathcal{I}_\mathcal{A}}$, i.e. there would be a named individual b such that $(d_p, b) \in (s')^{\mathcal{I}_\mathcal{A}}$ and $b \in (N')^{\mathcal{I}_\mathcal{A}}$. Since $s' \sqsubseteq_\mathcal{T} s$ and $N' \sqsubseteq_\mathcal{T} N$, this would imply $(d_p, b) \in s^{\mathcal{I}_\mathcal{A}}$ and $b \in N^{\mathcal{I}_\mathcal{A}}$, i.e. $d_p \in (\exists s.N)^{\mathcal{I}_\mathcal{A}}$, which shows that the anonymous object $d_{\hat{x}}$ would not have been created. Since $\exists s'.N'$ is included in V and we assumed that $\exists s.N$ is minimal w.r.t. $\sqsubseteq_\mathcal{T}^s$, we must have $s \equiv_\mathcal{T} s'$ and $N \equiv_\mathcal{T} N'$. But then (S3b) directly contradicts (S2b).

We now consider the filters in ϕ_2. Suppose that Equation (†) holds for π_1 and all filters in ϕ_1. For the ones that are only copied from ϕ_1 (modulo renaming some variables to $\hat{y}$), the property is clearly preserved. For the new filter $\psi^*(\hat{y})$, assume that $\pi_2(\hat{y}) \in \mathbf{I}$, and hence we need to show that $\mathcal{I}_\mathcal{A} \vDash \pi_2(\psi^*(\hat{y}))$. Assume that there exists an element $d' \in \mathbf{I}$ such that $(d_p, d') \in s^{\mathcal{I}_\mathcal{A}}$ and $d' \in N^{\mathcal{I}_\mathcal{A}}$. But then in Step 3(a) in Definition 4, $\exists s.N$ could not have been added to V since $d_p \in (\exists s.N)^{\mathcal{I}_\mathcal{K}}$ already holds. Hence, the element $d_{\hat{x}}$ would have never been introduced, which is a contradiction. Therefore, in $\mathcal{I}_\mathcal{A}$ the precondition of $\psi^*(\hat{y})$ is never met, which makes the filter trivially satisfied.

Finally, to show that $\deg(\pi_2) < \deg(\pi_1)$, we make a case distinction on whether the set Pred is empty or not. If $\mathsf{Pred} = \varnothing$, then we essentially replace the variable $\hat{x}$ in φ_1 with a new variable $\hat{y}$ in φ_2 with $|\pi_2(\hat{y})| = |d_p| < |d_{\hat{x}}| = |\pi_1(\hat{x})|$. Since the remaining variables are not affected by the rewriting step, this shows that $\deg(\pi_2) < \deg(\pi_1)$. If $\mathsf{Pred} \neq \varnothing$, then we have $|\pi_2(\hat{y})| = |d_p| = |\pi_1(y)|$ for all $y \in \mathsf{Pred}$. Since the variables in $\mathrm{Var}(\varphi_1) \smallsetminus \{\hat{y}\} = \mathrm{Var}(\varphi_1) \smallsetminus (\mathsf{Pred} \cup \{\hat{x}\})$ are not affected and $|\pi_1(\hat{x})| > 0$, we conclude that

$$\begin{aligned} \deg(\pi_2) &= |\pi_2(\hat{y})| + \sum_{z \in \mathrm{Var}(\varphi_2) \smallsetminus \{\hat{y}\}} |\pi_2(z)| \\ &< |\pi_1(\hat{x})| + \sum_{y \in \mathsf{Pred}} |\pi_1(y)| + \sum_{z \in \mathrm{Var}(\varphi_1) \smallsetminus (\mathsf{Pred} \cup \{\hat{x}\})} |\pi_1(z)| \\ &= \deg(\pi_1). \quad \square \end{aligned}$$

Under data complexity assumptions, ϕ and $\mathcal{T}$, and hence $\mathrm{rew}_\mathcal{T}(\phi)$, are fixed, and $\mathcal{I}_\mathcal{A}$ is of polynomial size in the size of $\mathcal{A}$.

For non-rooted queries the rewriting can get infinite, because we can always find a leaf variable that is not an answer variable. For example, suppose that the TBox $\mathcal{T}$ consists of the two GCIs $A \sqsubseteq \exists r.B$ and $B \sqsubseteq \exists r.A$. Let $\phi() = \exists x. A(x) \wedge \neg B(x)$ be a Boolean query.

The rewriting algorithm would produce an infinite rewriting of the form:

$$\phi_0 = \exists x. A(x) \wedge \neg B(x)$$

$$\phi_1 = \exists x. B(x) \wedge \underbrace{\big(\exists y. r(x,y) \wedge A(y) \rightarrow \exists y. r(x,y) \wedge A(y) \wedge \neg B(y)\big)}_{\psi_1(x)}$$

$$\phi_2 = \exists x. A(x) \wedge \underbrace{\big(\exists y. r(x,y) \wedge B(y) \rightarrow \exists y. r(x,y) \wedge B(y) \wedge \psi_1(y)\big)}_{\psi_2(x)}$$

$$\phi_3 = \exists x. B(x) \wedge \underbrace{\big(\exists y. r(x,y) \wedge A(y) \rightarrow \exists y. r(x,y) \wedge A(y) \wedge \psi_2(y)\big)}_{\psi_3(x)}$$

$$\cdots$$

In the following, we show that such infinite rewritings can be avoided, since after a certain number of rewriting steps queries will not generate any new answers.

For a query ϕ with filters Ψ, where each $\psi(y) \in \Psi$ is of the form

$$\big(\exists y'. \psi^+(y, y')\big) \rightarrow \big(\exists y'. \psi^+(y, y') \wedge \psi^-(y, y') \wedge \Psi'\big),$$

we define the *nested filter depth* as follows:

$$|\phi| := \max_{\psi \in \Psi} |\psi| \qquad\qquad |\psi| := 1 + \max_{\psi' \in \Psi'} |\psi'|,$$

where the second expression is applied recursively to subfilters.

Lemma 12
Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent $\mathcal{ELH}_\perp$ KB and let ϕ be an NCQ. Then

$$\bigcup_{\substack{\phi' \in \mathrm{rew}_{\mathcal{T}}(\phi) \\ |\phi'| \le v + \mathbf{C}_{\mathcal{T}}^2 \cdot \mathbf{R}_{\mathcal{T}}}} \mathrm{ans}(\phi', \mathcal{I}_{\mathcal{A}}) = \bigcup_{\phi' \in \mathrm{rew}_{\mathcal{T}}(\phi)} \mathrm{ans}(\phi', \mathcal{I}_{\mathcal{A}}).$$

where v denotes the number of variables in ϕ, and $\mathbf{C}_{\mathcal{T}}$ and $\mathbf{R}_{\mathcal{T}}$ denote the number of concept and role names in $\mathcal{T}$, respectively.

Proof
In the following, we assume a connected query ϕ with v variables. This is without loss of generality since the non-connected parts in the query can be dealt with separately.

Connectedness is preserved by the rewriting algorithm and hence there are two possible scenarios. In the first one, ϕ has only a finite number of rewritings. That happens if after at most v rewriting steps the rewriting is of the form $\exists \mathbf{z}. \varphi(\mathbf{z}) \wedge \Psi(\mathbf{z})$, where $\varphi(\mathbf{z})$ cannot be rewritten further. Otherwise, ϕ has infinitely many rewritings. In this case, after v rewriting steps, the rewriting and all further rewritings are of the form

$$\exists y. \big(A(y) \wedge \mathsf{Neg} \wedge \psi(y)\big), \tag{$\ddagger$}$$

where $A \in \mathbf{C}$ (recall that we assume a TBox in normal form), Neg is a conjunction of atoms of the form $\neg\hat{A}(y)$ for $\hat{A} \in \mathbf{C}$, and ψ is a filter that is linearly extended in every further rewriting step.

Assume a query ϕ_0 of the form ($\ddagger$) that has been further rewritten to ϕ_n in a sequence of $\phi_0 \rightarrow_{\mathcal{T}} \cdots \rightarrow_{\mathcal{T}} \phi_i \rightarrow_{\mathcal{T}} \cdots \rightarrow_{\mathcal{T}} \phi_n$, where $n \in \mathbb{N}$, every ϕ_j is of the form

$\exists y.A_j(y) \wedge \mathsf{Neg}_j \wedge \psi_j(y)$ for $0 \le j \le n$, and there exists $1 < i < n$ with $A_n = A_i$, $\mathsf{Neg}_n = \mathsf{Neg}_i$ and $\psi_n^+ = \psi_i^+$, i.e. the filters of ϕ_n and ϕ_i have the same body (up to renaming of variables).

We show that if there is a satisfying assignment π_n for ϕ_n over $\mathcal{I}_\mathcal{A}$, then there is also a satisfying assignment for a query ϕ' with $|\phi'| < |\phi_n|$. Suppose the nested filters in ϕ_n are matched up to a nested filter depth of $0 \le d \le n$.

Case 1: If $d = 0$, the body of ψ_n is not satisfied, then π_n is also a match for $A_i \wedge \mathsf{Neg}_i$, since the $A_i = A_n$, $\mathsf{Neg}_i = \mathsf{Neg}_n$ and the body of ψ_i is the same as the body of ψ_n by assumption. Clearly $|\phi_i| < |\phi_n|$.

Case 2: If $d > 0$, then we construct a match for query ϕ_{n-1}. Suppose ϕ_{n-1} has been rewritten to ϕ_n by using a GCI $A \sqsubseteq_\mathcal{T} \exists r.B$. Then we know that

$$\phi_{n-1} \coloneqq \exists y.B(y) \wedge \mathsf{Neg}_{n-1} \wedge \psi(y)$$
$$\phi_n \coloneqq \exists x.A(x) \wedge \mathsf{Neg}_n \wedge \big(\exists y.r(x,y) \wedge B(y) \to \exists y.r(x,y) \wedge B(y) \wedge \mathsf{Neg}_{n-1} \wedge \psi(y)\big)$$

Because $d > 0$, there has to be a satisfying assignment $\{x \mapsto a, y \mapsto b\}$ with $a, b \in \Delta^{\mathcal{I}_\mathcal{A}}$ for $r(x,y) \wedge B(y)$, which also satisfies $\mathsf{Neg}_{n-1} \wedge \psi(y)$. Then the assignment $\pi_{n-1} \coloneqq \{y \mapsto b\}$ satisfies ϕ_{n-1}, and $\psi(y)$ will then be matched up to a depth of $d - 1$.

This result can be used to bound the nested filter depth of queries during rewriting: In each rewriting step the query is rewritten w.r.t. a GCI of the form $A \sqsubseteq_\mathcal{T} \exists r.B$ with $A, B \in \mathbf{C}_\mathcal{T}$ and $r \in \mathbf{R}_\mathcal{T}$. There can be at most $\mathbf{C}_\mathcal{T}^2 \cdot \mathbf{R}_\mathcal{T}$ different GCIs of this form. Suppose a rewriting of ϕ to ϕ' with $|\phi'| > v + \mathbf{C}_\mathcal{T}^2 \cdot \mathbf{R}_\mathcal{T}$, where v denotes the number of variables in ϕ. Then at least two rewritings between ϕ and ϕ' must start with the same expression $\exists x.A(x) \wedge \mathsf{Neg} \wedge (\exists y.r(x,y) \wedge B(y) \to \dots)$. By the arguments above, there are rewritings of ϕ of nested filter depth at most $v + \mathbf{C}_\mathcal{T}^2 \cdot \mathbf{R}_\mathcal{T}$ that yield the same answers as ϕ'. $\square$

Hence, queries need to be rewritten only until this bound on the nested filter depth, which does not depend on $\mathcal{A}$. We obtain the claimed complexity result.

Theorem 13
Answering NCQs under minimal-world semantics over consistent $\mathcal{ELH}_\perp$ KBs is NP-complete, and P-complete in data complexity.

Proof
The lower bounds are inherited from certain answer semantics of CQs over $\mathcal{EL}$ KBs (Rosati 2007b) since for CQs the two semantics coincide. For data complexity, it suffices to observe that the size and number of the rewritten queries does not depend on the ABox, and the size of $\mathcal{I}_\mathcal{A}$ is polynomial in the size of $\mathcal{A}$, hence evaluating the rewriting over $\mathcal{I}_\mathcal{A}$ can be done in polynomial time.

For combined complexity, first we can guess an element ϕ' of $\mathrm{rew}_\mathcal{T}(\phi)$ since by Lemma 12 the number of necessary rewriting steps is bounded polynomially and each rewriting is of polynomial size. Now it remains to evaluate ϕ' over $\mathcal{I}_\mathcal{A}$, which is possible in PSPACE since ϕ' is a first-order query. To obtain a bound of NP, we need to do a more careful evaluation. First, we start by guessing a match of the initial CQ part of ϕ' in $\mathcal{I}_\mathcal{A}$. For each of the polynomially many (nested) filters

$$\big(\exists z'.\psi^+(z,z')\big) \to \big(\exists z'.\psi^+(z,z') \wedge \psi^-(z,z') \wedge \Psi\big)$$

in ϕ', we then enumerate all (polynomially many) possibilities to extend the initial match by mapping z' into $\mathcal{I}_\mathcal{A}$. If none of these satisfies $\psi^+(z,z')$, this (deterministic) check

ends the evaluation successfully. Otherwise, we can continue with recursively guessing an extended match for $\exists z'.\psi^+(z,z') \wedge \psi^-(z,z') \wedge \Psi$, for each nested filter extending the match by one new variable. Overall, this takes only polynomial time, which implies the claim. □

As discussed in Example 2 it can be useful to allow *complex assertions* in the ABox: The assertions of patient p_3 can then be stated without introducing a new constant for the disease by stating

$$\exists \mathrm{diagnosedWith.SkinOfBreastCancer}(p_3).$$

This leads to the introduction of additional acyclic definitions $\mathcal{T}'$, which are not fixed. The complexity nevertheless remains the same: Since $\mathcal{T}$ does not use the new concept names in $\mathcal{T}'$, we can apply the rewriting only w.r.t. $\mathcal{T}$, and extend $\mathcal{I}_{\mathcal{A}}$ by a polynomial number of new elements that result from applying Definition 4 only w.r.t. $\mathcal{T}'$.

What is more important than the complexity result is that this approach can be used to evaluate NCQs using standard database methods, e.g. using views to define the finite interpretation $\mathcal{I}_{\mathcal{A}}$ based on the input data given in $\mathcal{A}$, and SQL queries to evaluate the elements of $\mathrm{rew}_{\mathcal{T}}(\phi)$ over these views (Kontchakov et al. 2011).

5 Temporal Minimal World Semantics in $\mathcal{TELH}_{\perp}^{\lozenge,\mathsf{lhs},-}$

In the previous part, we introduced the minimal-world semantics and showed how NCQs can be answered efficiently over $\mathcal{ELH}_{\perp}$-KBs. In the remaining part of the paper we extend the minimal-world semantics to answering temporal queries over $\mathcal{TELH}_{\perp}^{\lozenge,\mathsf{lhs},-}$.

Example 14
Continuing the previous example of different cancer patients, we can now also model temporal aspects. *Chemotherapy* is a treatment that is usually given in a number of cycles to fight the cancer of a given patient. We formalize this knowledge as follows, where the basic time unit is one day:

$$\mathrm{ChemotherapyPatient} \sqsubseteq \mathrm{CancerPatient} \tag{8}$$

$$\lozenge_{365}\mathrm{CancerPatient} \sqsubseteq \mathrm{CancerPatient} \tag{9}$$

$$\lozenge_{120}\mathrm{ChemotherapyPatient} \sqsubseteq \mathrm{ChemotherapyPatient} \tag{10}$$

We make the assumption that being a cancer patient is 365-days convex, hence if a patient is reported to have cancer at two time points at most 1 year apart, we assume that they were also suffering from cancer in between the two reports. Similarly, we assume that a patient reported to be receiving chemotherapy at most four months apart also received chemotherapy in between.

Suppose the ABox $\mathcal{A}$ contains assertions $\mathrm{ChemotherapyPatient}(p_1, i)$, $i \in \{0, 167, 258\}$, for a patient p_1. The set of entailed assertions, denoted by $\mathcal{A}^*$, is illustrated below, where for simplicity we omit the individual name p_1. Here, CancerPatient and ChemotherapyPatient are abbreviated by C and T, respectively. Representative time points -1, 60, 200, and 259 have been introduced additionally, and the intervals they represent are shown above them.

ϕ	$\mathfrak{I}, i \vDash \phi$ iff
CQ ψ	$\mathcal{I}_i \vDash \psi$
$\top$	true
$\bot$	false
$\neg\phi$	$\mathfrak{I}, i \nvDash \phi$
$\phi \wedge \psi$	$\mathfrak{I}, i \vDash \phi$ and $\mathfrak{I}, i \vDash \psi$
$\phi \vee \psi$	$\mathfrak{I}, i \vDash \phi$ or $\mathfrak{I}, i \vDash \psi$
$\phi \mathcal{U}_I \psi$	$\exists k \in I$ such that $\mathfrak{I}, i + k \vDash \psi$ and $\forall j : 0 \le j < k : \mathfrak{I}, i + j \vDash \phi$
$\phi \mathcal{S}_I \psi$	$\exists k \in I$ such that $\mathfrak{I}, i - k \vDash \psi$ and $\forall j : 0 \le j < k : \mathfrak{I}, i - j \vDash \phi$

Figure 3. Semantics of (Boolean) MTNCQs for $\mathfrak{I} = (\Delta^{\mathfrak{I}}, (\mathcal{I}_i)_{i \in \mathbb{Z}})$ and $i \in \mathbb{Z}$.
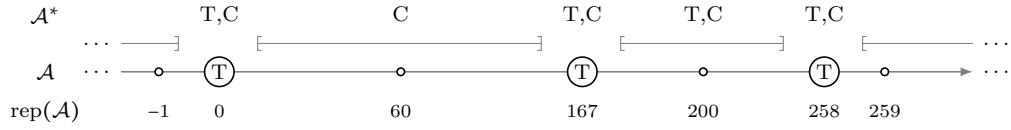
Since there are infinitely many time points, in general $\mathcal{A}^*$ is infinite. However, in (Borgwardt et al. 2019) this infinite set is represented via a finite set of *representative time points* $\mathrm{rep}(\mathcal{A})$, each of which represents a time interval in which the assertions do not change.

5.1 Metric Temporal Conjunctive Queries with Negation

We now consider the reasoning problem of temporal query answering, which generalizes the results from Section 4. We develop a new temporal query language with negation and extend the closed-world semantics for negation from the previous sections to $\mathcal{TELH}_{\bot}^{\lozenge,\mathsf{lhs},-}$.

We combine atemporal NCQs with MTL operators (Alur and Henzinger 1994; Gutiérrez-Basulto et al. 2016; Baader et al. 2017) to obtain Metric Temporal NCQs (MTNCQs) and finally show that such queries can be answered efficiently over $\mathcal{TELH}_{\bot}^{\lozenge,\mathsf{lhs},-}$ KBs when using the minimal-world semantics.

Definition 15
Metric temporal conjunctive queries with negation (MTNCQs) are built by the grammar rule

$$\phi ::= \psi \mid \top \mid \bot \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \mathcal{U}_I \phi \mid \phi \mathcal{S}_I \phi, \tag{11}$$

where ψ is an NCQ, and I is an interval over $\mathbb{N}$. An MTNCQ ϕ is *rooted/Boolean* if all NCQs in it are rooted/Boolean.

We employ the standard semantics shown in Figure 3. One can define the *next* operator as $\bigcirc\phi := \top \mathcal{U}_{[1,1]}\phi$, and similarly $\bigcirc^-\phi := \top \mathcal{S}_{[1,1]}\phi$. We can also express

$$\lozenge_I\phi := (\top \mathcal{S}_{-(I \cap (-\infty,0])}\phi) \vee (\top \mathcal{U}_{I \cap [0,\infty)}\phi)$$

and $\square_I\phi := \neg\lozenge_I\neg\phi$, and hence, by (1), the $\lozenge_n$-operators from Section 2. An *MTCQ* (or *positive MTNCQ*) is an MTNCQ without negation, where we assume that the operator $\square_I$ is nevertheless included as part of the syntax of MTCQs.

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a $\mathcal{TELH}_{\bot}^{\lozenge,\mathsf{lhs},-}$ KB, $\phi(\mathbf{x})$ an MTNCQ, $\mathbf{a}$ a tuple of individual names from $\mathcal{A}$, $i \in \mathrm{tem}(\mathcal{A})$, and $\mathfrak{I}$ a temporal interpretation. The pair $(\mathbf{a}, i)$ is an *answer* to

$\phi(\mathbf{x})$ w.r.t. $\mathfrak{I}$ if $\mathfrak{I}, i \vDash \phi(\mathbf{a})$. The set of all answers for ϕ w.r.t. $\mathfrak{I}$ is denoted $\mathrm{ans}(\phi, \mathfrak{I})$. The tuple $(\mathbf{a}, i)$ is a *certain answer* to ϕ w.r.t. $\mathcal{K}$ if it is an answer in every model of $\mathcal{K}$; all these tuples are collected in the set $\mathrm{cert}(\phi, \mathcal{K})$.

Example 16
Consider the criterion "Patients that received chemotherapy for more than 3 months, but less than 6 months." This can be expressed as an MTNCQ as follows.

$$\phi(x) := \Box_{[-90,0]} \mathrm{ChemotherapyPatient}(x) \wedge \neg \, \Box_{[-180,0]} \mathrm{ChemotherapyPatient}(x)$$

The negated conjunct expresses that the data does not indicate an ongoing chemotherapy for the past 6 months (minimal-world semantics), rather than that such a treatment is categorically ruled out by some TBox axioms (certain-answer semantics).

5.2 Minimal-World Semantics for MTNCQs

To extend the approach from Section 4 we need to find a *minimal canonical model* of a $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$ KB. Note that $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$ does not allow temporal roles, because temporal roles interfere with the *minimality*: they may entail the existence of a generic role successor long after it has been superseded by several more specific successors, and hence the generic successor becomes redundant.

In the definition of the model, we make use of entailment in $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs},-}$, which can be checked in polynomial time (Borgwardt et al. 2019). Thus, we can exclude w.l.o.g. equivalent concept and role names. Also, for simplicity, in the following we assume w.l.o.g. that all TBox inclusions are in normal form (2). In particular, disallowing CIs of the form $\Diamond A \sqsubseteq \exists r.B$ allows us to draw a stronger connection to the original construction in Section 4; see in particular Step 3(a) in Definition 17 below.

Definition 17
The *minimal temporal canonical model* $\mathfrak{I}_{\mathcal{K}} = (\Delta^{\mathfrak{I}_{\mathcal{K}}}, (\mathcal{I}_i)_{i \in \mathbb{Z}})$ of a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ is obtained by the following steps.

1. Set $\Delta^{\mathfrak{I}_{\mathcal{K}}} := \mathbf{I}$ and $a^{\mathcal{I}_i} := a$ for all $a \in \mathbf{I}$ and $i \in \mathbb{Z}$.
2. For each time point $i \in \mathbb{Z}$, define $A^{\mathcal{I}_i} := \{a \mid \mathcal{K} \vDash A(a,i)\}$ for all $A \in \mathbf{C}$ and $r^{\mathcal{I}_i} := \{(a,b) \mid \mathcal{K} \vDash r(a,b,i)\}$ for all $r \in \mathbf{R}$.
3. Repeat the following steps:
   (a) Select an element $d \in \Delta^{\mathfrak{I}_{\mathcal{K}}}$ that has not been selected before and, for each $i \in \mathbb{Z}$, let $V_i := \{\exists r.B \mid d \in A^{\mathcal{I}_i}, \ d \notin (\exists r.B)^{\mathcal{I}_i}, \ \mathcal{K} \vDash A \sqsubseteq \exists r.B, \ A, B \in \mathbf{C}\}$.
   (b) For each $\exists r.B$ that is minimal in some V_i, add a fresh element e_{rB} to $\Delta^{\mathfrak{I}_{\mathcal{K}}}$. For all $i \in \mathbb{Z}$ and $\mathcal{K} \vDash B \sqsubseteq A$, add e_{rB} to $A^{\mathcal{I}_i}$.
   (c) For all $i \in \mathbb{Z}$, minimal $\exists r.B$ in V_i, and $\mathcal{K} \vDash r \sqsubseteq s$, add (d, e_{rB}) to $s^{\mathcal{I}_i}$.

We denote by $\mathfrak{I}_{\mathcal{A}}$ the result of executing only Steps 1 and 2 of this definition, i.e. restricting $\mathfrak{I}_{\mathcal{K}}$ to the named individuals. Since there are only finitely many elements of $\mathbf{I}$, $\mathbf{C}$, and $\mathbf{R}$ that are relevant for this definition (i.e. those that occur in $\mathcal{K}$), for simplicity we often treat $\mathfrak{I}_{\mathcal{A}}$ as if it had a finite object (but still infinite time) domain.

In $\mathfrak{I}_{\mathcal{K}}$, there may exist anonymous objects that are not connected to any named individuals in $\mathcal{I}_i$ and are not relevant for the satisfaction of the KB. For this reason, in the following we consider only rooted MTNCQs, which can be evaluated only over the parts

of $\mathfrak{I}_{\mathcal{K}}$ that are connected to the named individuals. We again show that $\mathfrak{I}_{\mathcal{K}}$ is actually a model of $\mathcal{K}$ and is canonical in the usual sense that it can be used to answer *positive* queries over $\mathcal{K}$ under certain answer semantics.

Lemma 18
Let $\mathcal{K}$ be a consistent $\mathcal{TELH}_{\bot}^{\Diamond,\text{lhs},-}$ KB. Then $\mathfrak{I}_{\mathcal{K}}$ is a model of $\mathcal{K}$ and, for every rooted MTCQ ϕ, we have $\text{cert}(\phi,\mathcal{K}) = \text{ans}(\phi,\mathfrak{I}_{\mathcal{K}})$.

Proof
The first claim is easy to prove, and the inclusion $\text{cert}(\phi,\mathcal{K}) \subseteq \text{ans}(\phi,\mathfrak{I}_{\mathcal{K}})$ follows from the fact that $\mathfrak{I}_{\mathcal{K}}$ is a model of $\mathcal{K}$. For the other inclusion, consider any model $\mathfrak{J} = (\Delta^{\mathfrak{J}}, (\mathcal{J}_i)_{i \geq 0})$ of $\mathcal{K}$. We prove that $(\mathbf{a}, i) \in \text{ans}(\phi, \mathfrak{I}_{\mathcal{K}})$ implies $(\mathbf{a}, i) \in \text{ans}(\phi, \mathfrak{J})$ by induction on the structure of ϕ.

- If ϕ is a rooted CQ, then $\mathcal{I}_i \models \phi(\mathbf{a})$. Moreover, since ϕ is rooted, only the rooted part of $\mathcal{I}_i$, consisting of all elements connected to named individuals, is relevant for satisfying $\phi(\mathbf{a})$. It is easy to show that this part can be homomorphically mapped into $\mathcal{J}_i$, hence $\mathfrak{J}, i \models \phi(\mathbf{a})$.
- If $\phi = \phi_1 \vee \phi_2$, then $(\mathbf{a}, i) \in \text{ans}(\phi_1, \mathfrak{I}_{\mathcal{K}})$ or $(\mathbf{a}, i) \in \text{ans}(\phi_2, \mathfrak{I}_{\mathcal{K}})$, hence by induction $(\mathbf{a}, i) \in \text{ans}(\phi_1, \mathfrak{J})$ or $(\mathbf{a}, i) \in \text{ans}(\phi_2, \mathfrak{J})$, either of which implies that $(\mathbf{a}, i) \in \text{ans}(\phi, \mathfrak{J})$.
- The cases of $\mathcal{S}_I$, $\Diamond_I$, and $\Box_I$ are similar, and therefore the claim also extends to $\Diamond_n$, $\bigcirc$, and $\bigcirc^-$. $\quad\square$

Thus, the following *minimal-world* semantics is compatible with certain answer semantics for positive (rooted) queries, while keeping a tractable data complexity.

Definition 19
The set of *minimal-world answers* to an MTNCQ ϕ over a consistent $\mathcal{TELH}_{\bot}^{\Diamond,\text{lhs},-}$ KB $\mathcal{K}$ is $\text{mwa}(\phi,\mathcal{K}) := \text{ans}(\phi,\mathfrak{I}_{\mathcal{K}})$.

6 A Combined Rewriting for MTNCQs

Following the approach used for the atemporal case, we now show that rooted MTNCQ answering under minimal-world semantics is combined first-order rewritable. For rewriting atomic queries we use the results from (Borgwardt et al. 2019). Thus, we proceed as follows. First we rewrite ϕ into a *metric first-order temporal logic (MFOTL)* formula $\phi_{\mathcal{T}}$, which combines first-order formulas via metric temporal operators; for details, see (Basin et al. 2015). The formula $\phi_{\mathcal{T}}$ is obtained by replacing each (rooted) NCQ ψ with the first-order rewriting $\psi_{\mathcal{T}} := \bigvee_{\psi' \in \text{rew}_{\mathcal{T}}(\psi)} \psi'$ from Section 4, and hence $\phi_{\mathcal{T}}$ can be evaluated already over $\mathfrak{I}_{\mathcal{A}}$. Second, we then further rewrite $\phi_{\mathcal{T}}$ into a three-sorted first-order formula (with explicit variables for (a) objects, (b) time points, and (c) bits of the binary representation of time values), which is then evaluated over a restriction $\mathfrak{I}_{\mathcal{A}}^{\text{fin}}$ of $\mathfrak{I}_{\mathcal{A}}$ that contains only finitely many time points.

These two steps produce a valid rewriting for the query ϕ.

Lemma 20
Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent $\mathcal{TELH}_{\bot}^{\Diamond,\text{lhs},-}$ KB and ϕ be a rooted MTNCQ. Then $\text{mwa}(\phi,\mathcal{K}) = \text{ans}(\phi_{\mathcal{T}}, \mathfrak{I}_{\mathcal{A}})$.

Proof

We prove the claim by induction over the structure of ϕ.

Suppose that ϕ is a rooted NCQ. Since ϕ does not contain temporal operators, we can restrict our attention to a single atemporal interpretation $\mathcal{I}_i$ in $\mathfrak{I}_{\mathcal{K}} = (\Delta^{\mathfrak{I}_{\mathcal{K}}}, (\mathcal{I}_i)_{i \in \mathbb{Z}})$. Since ϕ is rooted, only the "rooted" part $\mathcal{I}_i^r$ of $\mathcal{I}_i$, consisting only of those elements connected to named individuals via a sequence of role connections, is relevant for evaluating ϕ (and similarly for $\phi_{\mathcal{T}}$). In the construction of $\mathfrak{I}_{\mathcal{K}}$ (Definition 17), we can observe that $\mathcal{I}_i^r$ is uniquely determined by the definition of $A^{\mathcal{I}_i}$ and $r^{\mathcal{I}_i}$ in Step 2. Moreover, $\mathcal{I}_i^r$ is isomorphic to the (atemporal) minimal canonical model of $(\mathcal{T}', \mathcal{A}_i)$ as in Definition 4, where

- $\mathcal{T}' := \{C \sqsubseteq D \mid \lozenge\!\!\!\lozenge C \sqsubseteq D \in \mathcal{T}, \lozenge\!\!\!\lozenge \in \mathfrak{D}^{\pm}\}$ and
- $\mathcal{A}_i := \{A(a) \mid \mathcal{K} \vDash A(a, i)\} \cup \{r(a, b) \mid \mathcal{K} \vDash r(a, b, i)\}$.

In particular, one can observe that the temporal operators in $\mathcal{T}$ are irrelevant for the behavior of the anonymous elements in $\mathcal{I}_i$ (note that $\lozenge\!\!\!\lozenge C \sqsubseteq D$ entails $C \sqsubseteq D$) and we can restrict the attention to those assertions entailed for time point i. Hence, by Lemma 11 we can conclude that $(\mathbf{a}, i) \in \mathrm{mwa}(\phi, \mathcal{K}) = \mathrm{ans}(\phi, \mathfrak{I}_{\mathcal{K}})$ iff $\mathbf{a} \in \mathrm{ans}(\phi, \mathcal{I}_i) = \mathrm{ans}(\phi, \mathcal{I}_i^r) = \mathrm{ans}(\phi_{\mathcal{T}}, \mathcal{I}_{i,\mathcal{A}})$ iff $(\mathbf{a}, i) \in \mathrm{ans}(\phi_{\mathcal{T}}, \mathfrak{I}_{\mathcal{A}})$, where $\mathcal{I}_{i,\mathcal{A}}$ is the restriction of $\mathcal{I}_i$ to the named individuals.

For the remaining cases, it suffices to observe that ϕ and $\phi_{\mathcal{T}}$ are built on the same structure of temporal operators, which have the same semantics for both MTNCQs and MFOTL formulas. □

For the second rewriting step, we restrict ourselves to finitely many time points. More precisely, we consider the finite structure $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$, which is obtained from $\mathfrak{I}_{\mathcal{A}}$ by restricting the set of time points to $\mathrm{rep}(\mathcal{A})$. By Lemma 4 in (Borgwardt et al. 2019), the information contained in $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$ is already sufficient to answer rooted atomic queries. The idea is that each time point i is assigned a representative time point $|i|$ from $\mathrm{rep}(\mathcal{A})$, at which the same assertions hold. Therefore, the answers to a query at time point i are the same as the answers at $|i|$. We extend this structure a little, by considering the *two* representatives i, j for each maximal interval $[i, j]$ in $\mathbb{Z} \setminus \mathrm{tem}(\mathcal{A})$. In this way, we ensure that the "border" elements are always representatives for their respective intervals. The size of the resulting structure $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$ is polynomial in the size of $\mathcal{K}$.

Example 21

Continuing Example 16, below one can see the finite structure $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$ over the representative time points $\{-1, 0, 1, 166, 167, 168, 257, 258, 259\}$, where for simplicity we omit the individual name.
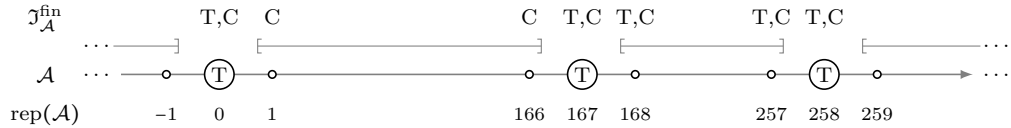
The rewriting from Lemma 20 can refer to time instants outside of $\mathrm{rep}(\mathcal{A})$. However, when we want to evaluate a pure FO formula over the finite structure $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$, this is not possible anymore, because the first-order quantifiers must quantify over the domain of $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$. Moreover, since the query $\phi_{\mathcal{T}}$ can contain metric temporal operators, we need to keep track of the distance between the time points in $\mathrm{tem}(\mathcal{A})$. Hence, in the following we assume that $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$ is given as a first-order structure with the domain $\mathbf{I} \cup \{b_1, \ldots, b_n\} \cup \mathrm{rep}(\mathcal{A})$

and additional predicates $\mathtt{bit}$ and $\mathtt{sign}$ such that $\mathtt{bit}(i,j)$, $1 \le j \le n$, is true iff the jth bit of the binary representation of the time stamp i is 1, and $\mathtt{sign}(i)$ is true iff i is non-negative.

Thus, we now consider three-sorted first-order formulas with the three sorts $\mathbf{I}$ (for objects), $\{b_1, \ldots, b_n\}$ (for bits) and $\mathrm{rep}(\mathcal{A})$ (for time stamps). We denote variables of sort $\mathrm{rep}(\mathcal{A})$ by t, t', t''. To simplify the presentation, we do not explicitly denote the sort of all variables, but this is always clear from the context. Every concept name is now accessed as a binary predicate of sort $\mathbf{I} \times \mathrm{rep}(\mathcal{A})$, e.g. $A(a, i)$ refers to the fact that individual a satisfies A at time point i. Similarly, role names correspond to ternary predicates of sort $\mathbf{I} \times \mathbf{I} \times \mathrm{rep}(\mathcal{A})$.

In the following we show that the expressions $t' \bowtie t$ and even $t' - t \bowtie m$ for some constant m and $\bowtie \in \{\ge, >, =, <, \le\}$ are definable as first-order formulas using the natural order $<$ on $\{1, \ldots, m\}$. More specifically, we define $t' - t \bowtie d$, for some constant $d < \infty$ and $\bowtie \in \{\ge, >, =, <, \le\}$, as first-order formulas with build-in predicates $\mathtt{bit}(t, j)$, $1 \le j \le n$, which is true iff the j-th bit of t is 1, and $\mathtt{sign}(t)$, which is true iff $t \ge 0$. W.l.o.g., let d be a non-negative integer; otherwise, the formula can be reformulated as $t - t' \bowtie k$, where $0 \le k = -d$. First, consider the case of equality. $(t' - t = d)$ is true iff

$$\big(\mathtt{sign}(t') \leftrightarrow \mathtt{sign}^d(t)\big) \wedge \forall j \big(\mathtt{bit}(t', j) \leftrightarrow \mathtt{bit}^d(t, j)\big) \wedge \neg \mathtt{ovf}^d(t),$$

where $\mathtt{sign}^d(t)$ checks whether $t + d$ is non-negative, $\mathtt{bit}^d(t, j)$ says what the jth bit of the binary representation of $t + d$ is, and $\mathtt{ovf}^d(t)$ detects whether the addition of d to t causes an overflow in the n-bit. These three auxiliary predicates are defined inductively as follows:

$$\mathtt{ovf}^0(t) := \bot$$
$$\mathtt{ovf}^{d+1}(t) := \mathtt{ovf}^d(t) \vee \big(\mathtt{sign}^d(t) \wedge \forall j. \, \mathtt{bit}^d(t, j)\big)$$

$$\mathtt{sign}^0(t) := \mathtt{sign}(t)$$
$$\mathtt{sign}^{d+1}(t) := \mathtt{sign}^d(t) \vee \big(\forall j.(\exists j'.(j' < j)) \leftrightarrow \neg \mathtt{bit}^d(t, j)\big)$$

$$\mathtt{bit}^0(t, j) := \mathtt{bit}(t, j)$$
$$\mathtt{bit}^{d+1}(t, j) := \mathtt{sign}^d(t) \wedge \big(\mathtt{bit}^d(t, j) \leftrightarrow \exists j'.(j' < j) \wedge \neg \mathtt{bit}^d(t, j')\big)$$
$$\vee \neg \mathtt{sign}^{d+1}(t) \wedge \big(\mathtt{bit}^d(t, j) \leftrightarrow \exists j'.(j' < j) \wedge \mathtt{bit}^d(t, j')\big)$$

The remaining cases $\bowtie \in \{\ge, >, <, \le\}$ are obtained similarly to the formulas above. For example, $t < t'$ can be expressed by looking at the signs and the most significant bit in which they differ, formally:

$$\big(\neg \mathtt{sign}(t) \wedge \mathtt{sign}(t')\big)$$
$$\vee \Big(\big(\mathtt{sign}(t) \leftrightarrow \mathtt{sign}(t')\big) \wedge \exists j.\big(\forall j'.(j' > j) \rightarrow (\mathtt{bit}(t', j) \leftrightarrow \mathtt{bit}(t, j))\big)$$
$$\wedge \big(\mathtt{bit}(t', j) \leftrightarrow \mathtt{sign}(t)\big) \wedge \big(\mathtt{bit}(t', j) \leftrightarrow \neg \mathtt{bit}(t, j)\big)\Big).$$

Lemma 22
For $\phi_{\mathcal{T}}$ there is a constant $N \in \mathbb{N}$ such that, for every subformula ψ of $\phi_{\mathcal{T}}$, every maximal

interval J in $\mathbb{Z} \smallsetminus \bigcup \{[i - N, i + N] \mid i \in \mathrm{tem}(\mathcal{A})\}$, all $k, \ell \in J$, and all relevant tuples $\mathbf{a}$ over $\mathbf{I}$, we have $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi(\mathbf{a})$ iff $\mathfrak{I}_{\mathcal{A}}, \ell \vDash \psi(\mathbf{a})$.

Proof

We are going to prove a more specific statement. Namely, let N_ψ be the sum of all interval bounds of temporal formulas in a subformula ψ of $\phi_{\mathcal{T}}$ (except ∞). Consequently, for the proof we consider instead every maximal interval J in $\mathbb{Z} \smallsetminus \bigcup \{[i - N_\psi, i + N_\psi] \mid i \in \mathrm{tem}(\mathcal{A})\}$.

We show this by induction on the structure of ψ, but only consider three representative cases; the other cases are similar.

- If ψ is the rewriting of an NCQ, then $N_\psi = 0$ and the semantics of ψ depends only on the interpretation at a single time point. Since k and ℓ belong to the same maximal interval in $\mathbb{Z} \smallsetminus \mathrm{tem}(\mathcal{A})$, by Lemma 4 in (Borgwardt et al. 2019) and the construction of $\mathfrak{I}_{\mathcal{A}}$, this interpretation behaves in the same way at k and at ℓ.

- If ψ is of the form $\psi_1 \mathcal{U}_{[c_1, c_2]} \psi_2$, then $N_{\psi_1} \le N_\psi - c_2$ and $N_{\psi_2} \le N_\psi - c_2$. Assume that $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi(\mathbf{a})$. Then there exists $j \in [c_1, c_2]$ such that

$$\mathfrak{I}_{\mathcal{A}}, k + j \vDash \psi_2(\mathbf{a}) \text{ and } \mathfrak{I}_{\mathcal{A}}, m \vDash \psi_1(\mathbf{a}), \text{ for all } m \text{ with } k \le m < k + j. \qquad (12)$$

  In case that $j = c_1 = 0$, we have $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi_2(\mathbf{a})$. Since k and ℓ are farther than $N_\psi \ge N_{\psi_2}$ from the nearest element of $\mathrm{rep}(\mathcal{A})$, by induction we also have $\mathfrak{I}_{\mathcal{A}}, \ell \vDash \psi_2(\mathbf{a})$ and thus $\mathfrak{I}_{\mathcal{A}}, \ell \vDash \psi(\mathbf{a})$ in this case. Hence, we can assume in the following that $j \ge c_1 > 0$, and thus in particular $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi_1(\mathbf{a})$.
  Since both $k + j$ and $\ell + c_2$ are farther than $N_\psi - c_2 \ge N_{\psi_2}$ from the nearest element of $\mathrm{rep}(\mathcal{A})$, by induction we have $\mathfrak{I}_{\mathcal{A}}, \ell + c_2 \vDash \psi_2(\mathbf{a})$. Moreover, since $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi_1(\mathbf{a})$ and k as well as all elements in $[\ell, \ell + c_2]$ are farther than $N_\psi - c_2 \ge N_{\psi_1}$ from the nearest element of $\mathrm{rep}(\mathcal{A})$, by induction we have $\mathfrak{I}_{\mathcal{A}}, m \vDash \psi_1(\mathbf{a})$ for all m with $\ell \le m \le \ell + c_2$. Hence, $\mathfrak{I}_{\mathcal{A}}, \ell \vDash \psi(\mathbf{a})$.

- If ψ is of the form $\psi_1 \mathcal{U}_{[c_1, \infty)} \psi_2$, then we have a similar situation as above, except that j is not bounded by c_2. We can again assume that $j > 0$ and $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi_1(\mathbf{a})$.
  Let p be the maximal element of J. If $k + j > p + c_1$, then $k + j > \ell$ and the distance between ℓ and $k + j$ must be at least c_1. Moreover, by assumption 12 we have $\mathfrak{I}_{\mathcal{A}}, m \vDash \psi_1(\mathbf{a})$ for all m with $p < m < k + j$. Since $\mathfrak{I}_{\mathcal{A}}, k \vDash \psi_1(\mathbf{a})$ and all elements in J are farther than $N_\psi \ge N_{\psi_1}$ from the nearest element of $\mathrm{rep}(\mathcal{A})$, by induction we also have $\mathfrak{I}_{\mathcal{A}}, m \vDash \psi_1(\mathbf{a})$ for all m with $\ell \le m \le p$. Thus, $\mathfrak{I}_{\mathcal{A}}, \ell \vDash \psi(\mathbf{a})$.
  We now consider the remaining case that $k + j \le p + c_1$. Then both $k + j$ and $\ell + c_1$ are farther than $N_\psi - c_1 \ge N_{\psi_2}$ from the nearest element of $\mathrm{rep}(\mathcal{A})$, and thus by induction we have $\mathfrak{I}_{\mathcal{A}}, \ell + c_1 \vDash \psi_2(\mathbf{a})$.
  By similar arguments as above, we obtain $\mathfrak{I}_{\mathcal{A}}, \ell \vDash \psi(\mathbf{a})$. $\qquad \square$

Hence, for evaluating subformulas of $\phi_{\mathcal{T}}$, it suffices to keep track of time points up to N steps away from the elements of $\mathrm{rep}(\mathcal{A})$; this includes at least one element from each of the intervals J mentioned in Lemma 22, since every element of $\mathrm{tem}(\mathcal{A})$ is immediately surrounded by two elements of $\mathrm{rep}(\mathcal{A})$.

We exploit Lemma 22 in the following definition of the three-sorted first-order formula $[\psi]^n(\mathbf{x}, t)$ that simulates the behavior of $\psi(\mathbf{x})$ at the "virtual" time point $t + n$, where $n \in [-N, N]$. Whenever we use a formula $[\psi]^n(\mathbf{x}, t)$, we require that t denotes a representative for $t + n$. Due to our assumption that each maximal interval from $\mathbb{Z} \smallsetminus \mathrm{tem}(\mathcal{A})$

is represented by its endpoints (see Example 21), we know that t is a representative for $t + n$ iff there is no element of $\mathrm{rep}(\mathcal{A})$ between t and $t + n$. We can encode this check in an auxiliary formula:

$$\mathtt{rep}^n(t) \coloneqq \neg \exists t'. (t + n \leq t' < t) \vee (t < t' \leq t + n).$$

Example 23

In Example 21, time points 1 and 166 are representatives for the missing time points 2–165, and we have $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}} \vDash \mathtt{rep}^1(1)$ (with $N = 1$). However, for $\phi_{\mathcal{T}} = \bigcirc \neg \mathrm{T}(x)$, we have $\mathfrak{I}_{\mathcal{A}}, 1 \vDash \phi_{\mathcal{T}}(p_1)$, but $\mathfrak{I}_{\mathcal{A}}, 166 \nvDash \phi_{\mathcal{T}}(p_1)$, i.e. the behavior at 1 and 166 differs. To distinguish this, we need to refer to the "virtual" time point 2 (see the gray circled "v"s below) that is not included in $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}$, via the formula $[\neg \mathrm{T}(x)]^1(p_1, 1)$. By Lemma 22, it is sufficient to consider time point 2, because this determines the behavior at 3–165 .
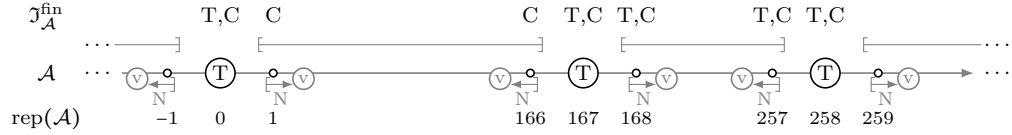
We now define $[\psi]^n(\mathbf{x}, t)$ recursively, for each subformula ψ of $\phi_{\mathcal{T}}$. If ψ is a single rewritten NCQ, then $[\psi]^n(\mathbf{x}, t)$ is obtained by replacing each atemporal atom $A(x)$ by $A(x, t)$, and similarly for role atoms. The parameter n can be ignored here, because we assumed that t is a representative for $t + n$, and hence the time points t and $t + n$ are interpreted in $\mathfrak{I}_{\mathcal{A}}$ equally. For conjunctions, we set $[\psi_1 \wedge \psi_2]^n(\mathbf{x}, t) \coloneqq [\psi_1]^n(\mathbf{x}, t) \wedge [\psi_2]^n(\mathbf{x}, t)$ and similarly for the other Boolean constructors. Finally, we demonstrate the translation for $\mathcal{U}$-formulas (the case of $\mathcal{S}$-formulas is analogous). We define $[\psi_1 \mathcal{U}_{[c_1, c_2]} \psi_2]^n(\mathbf{x}, t)$ as

$$\exists t'. \bigvee_{n' \in [-N, N]} \Bigg((t + n + c_1 \leq t' + n' \leq t + n + c_2) \wedge \mathtt{rep}^{n'}(t') \wedge [\psi_2]^{n'}(\mathbf{x}, t') \wedge$$

$$\forall t''. \bigwedge_{n'' \in [-N, N]} \Big(\big((t + n \leq t'' + n'' < t' + n') \wedge \mathtt{rep}^{n''}(t'') \big) \to [\psi_1]^{n''}(\mathbf{x}, t'') \Big) \Bigg),$$

where c_2 may be ∞, in which case the upper bound of $t + n + c_2$ can be removed.

Lemma 24

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a consistent $\mathcal{TELH}_{\perp}^{\Diamond, \mathsf{lhs}, -}$ KB and ϕ be an MTNCQ. Then $\mathrm{ans}([\phi_{\mathcal{T}}]^0(\mathbf{x}, t), \mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}}) = \mathrm{ans}(\phi_{\mathcal{T}}, \mathfrak{I}_{\mathcal{A}})$.

Proof

We show the following claim by induction on the structure of ϕ: For all $i \in \mathrm{rep}(\mathcal{A})$, all $n \in [-N, N]$, all relevant tuples $\mathbf{a}$, and all MTNCQs ϕ such that if $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}} \vDash \mathtt{rep}^n(i)$ then

$$\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}} \vDash [\phi_{\mathcal{T}}]^n(\mathbf{a}, i) \text{ iff } \mathfrak{I}_{\mathcal{A}}, i + n \vDash \phi_{\mathcal{T}}(\mathbf{a}).$$

Since this includes the case where $i \in \mathrm{tem}(\mathcal{A})$, $n = 0$, for which $\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}} \vDash \mathtt{rep}^0(i)$ holds, the statement of the lemma follows.

If ϕ is an NCQ, then

$$\mathfrak{I}_{\mathcal{A}}^{\mathrm{fin}} \vDash [\phi_{\mathcal{T}}]^n(\mathbf{a}, i) \text{ iff } \mathfrak{I}_{\mathcal{A}}, i \vDash \phi_{\mathcal{T}}(\mathbf{a}) \text{ iff } \mathfrak{I}_{\mathcal{A}}, i + n \vDash \phi_{\mathcal{T}}(\mathbf{a})$$

since i is a representative for $i + n$ and a single temporal variable t is used in $[\phi_{\mathcal{T}}]^n(\mathbf{x}, t)$ to denote "current" time point in $\phi_{\mathcal{T}}$.

For the Boolean constructors, the claim follows immediately from the semantics of first-order logic.

We now consider a formula of the form $\phi \mathcal{U}_I \psi$. By induction, we know that the following hold: (i) $\mathfrak{I}_{\mathcal{A}}^{\text{fin}} \vDash [\psi_{\mathcal{T}}]^{n'}(\mathbf{a}, i')$ iff $\mathfrak{I}_{\mathcal{A}}, i' + n' \vDash \psi_{\mathcal{T}}(\mathbf{a})$, for any time point i' with $i' + n' \geq i + n$, and (ii) $\mathfrak{I}_{\mathcal{A}}^{\text{fin}} \vDash [\phi_{\mathcal{T}}]^{n''}(\mathbf{a}, i'')$ iff $\mathfrak{I}_{\mathcal{A}}, i'' + n'' \vDash \phi_{\mathcal{T}}(\mathbf{a})$ for all time points i'' and offsets n'' such that $i + n \leq i'' + n'' < i' + n'$ (assuming w.l.o.g. that ϕ and ψ have the same answer variables).

Hence, the formula $[\phi_{\mathcal{T}} \mathcal{U}_I \psi_{\mathcal{T}}]^n(\mathbf{a}, i)$ checks the conditions required for the satisfaction of the $\mathcal{U}_I$-expression for all time points in $\bigcup\{[i - N, i + N] \mid i \in \text{rep}(\mathcal{A})\}$. However, Lemma 22 tells us that, if $\psi_{\mathcal{T}}$ is satisfied in $\mathfrak{I}_{\mathcal{A}}$ at some time point $i' + n'$ with $n' > N$, then this is also the case for $n' = N$. Similarly, to check whether $\phi_{\mathcal{T}}$ is satisfied at all time points between $i + n$ and $i' + n'$, it suffices to consider the time points up to N away from some element of $\text{rep}(\mathcal{A})$. Hence, $\mathfrak{I}_{\mathcal{A}}^{\text{fin}} \vDash [\phi_{\mathcal{T}} \mathcal{U}_I \psi_{\mathcal{T}}]^n(\mathbf{a}, i)$ iff $\mathfrak{I}_{\mathcal{A}}, i + n \vDash (\phi_{\mathcal{T}} \mathcal{U}_I \psi_{\mathcal{T}})(\mathbf{a})$. □

This lemma allows us to compute in polynomial time that patient p_1 from Example 14 is an answer to $\phi(x)$ from Example 16 exactly at time points 257 and 258. Below we summarize our tight complexity results, which by Lemma 18 also hold for rooted MTCQs under certain answer semantics.

Theorem 25

Answering rooted MTNCQs under minimal-world semantics over $\mathcal{TELH}_{\perp}^{\Diamond, \text{lhs}, -}$ KBs is ExpSpace-complete, and P-complete in data complexity.

Proof

ExpSpace-hardness is inherited from propositional MTL (Alur and Henzinger 1994; Furia and Spoletini 2008). Moreover, first-order formulas over finite structures can be evaluated in PSpace (Vardi 1982). Finally, the size of $[\phi_{\mathcal{T}}]^0(\mathbf{x}, t)$ is bounded exponentially in the size of ϕ and $\mathcal{T}$: Each rewritten NCQ $\psi_{\mathcal{T}}$ may be exponentially larger than ψ, and each $[\psi_1 \mathcal{U}_I \psi_2]^n(\mathbf{x}, t)$ introduces exponentially many disjuncts and conjuncts (but the nesting depth of constructors in this formula is linear in the nesting depth of $\psi_1 \mathcal{U}_I \psi_2$).

For data complexity, hardness is inherited from atemporal $\mathcal{EL}$ (Calvanese et al. 2013). The evaluation of FO($<$, `bit`)-formulas is in DLogTime-uniform AC^0 in data complexity (Lindell 1992), and the size of our rewriting only depends on the query and the TBox. By Lemmas 20 and 24 and since $\mathfrak{I}_{\mathcal{A}}^{\text{fin}}$ is of size polynomial in the size of $\mathcal{A}$, deciding whether a tuple $\mathbf{a}$ is a minimal-world answer of an MTNCQ w.r.t. a $\mathcal{TELH}_{\perp}^{\Diamond, \text{lhs}, -}$ KB is possible in P. □

7 Related Work

Until a decade ago, the work on combining ontology and temporal languages was mostly focused on the main reasoning tasks such as concept subsumption (whether an inclusion between concepts is entailed by a temporal ontology) and knowledge base satisfiability (whether a given knowledge base consisting of a temporal data instance and a temporal ontology has a model); we refer the reader to (Artale and Franconi 2005; Baader et al. 2003; Lutz et al. 2008; Artale et al. 2014) for details. Particularly, in the presence of a

single rigid role, allowing the operator ⬦ on both sides of $\mathcal{EL}$ CIs makes subsumption undecidable (Artale et al. 2007). In (Gutiérrez-Basulto et al. 2016), a variety of restrictions are investigated to regain decidability, such as acyclic TBoxes and restrictions on the occurrences of temporal operators. In particular, allowing the qualitative operators ⬦, ⬦, ⬦, ⬦ only on the left-hand side of CIs makes the logic tractable. Less closely related work considers temporalized DLs where LTL operators are used to combine axioms, but are not allowed within concepts (Baader et al. 2012).

Since atomic query answering (whether a concept or role name is entailed by a knowledge base) can be reduced to the satisfiability checking problem, further developments of query languages and the complexities of ontology-mediated query answering over temporal data have recently appeared in the literature. In the following we briefly summarize recent research on combining ontology and query languages with negation and temporal formalisms. For a general overview of temporal ontology and query languages, see (Lutz et al. 2008; Artale et al. 2017).

In this research we focus on a discrete timeline (over $\mathbb{Z}$) and data facts stamped with a single time point. However, in the literature there are other approaches how to incorporate temporal formalisms in an ontology and the data, e.g. dense timelines, like $\mathbb{Q}$ or $\mathbb{R}$ (Brandt et al. 2018; Ryzhikov et al. 2019); or interval-based data models, where facts are stamped with a pair of time points denoting the interval in which they are true (Kontchakov et al. 2016; Brandt et al. 2018). Discussing this work is beyond the scope of this article. We only want to mention here that choosing $\mathbb{Z}$ rather than $\mathbb{Q}$ or $\mathbb{R}$ is not a restriction in our setting since (i) our formalism allows arbitrarily large gaps between time points in the input and (ii) there are no inferences that would require a dense model of time, e.g. computations approaching some time point only in the limit.

Within the discrete time point-based approach, one can distinguish between formalisms with LTL temporal constructs and formalisms employing more refined Metric Temporal Logic (MTL) operators (Alur et al. 1996). MTL extends LTL with temporal intervals for the modal operators, restricting them to a specific time range. Combining ontology-mediated query answering with LTL operators has been investigated in depth. In particular, similarly to the query language adapted in this paper, there is a multitude of works (Baader et al. 2013; Borgwardt and Thost 2015b; Baader et al. 2015a; Baader et al. 2015b) investigating the complexity of answering LTL-CQs that are obtained from LTL formulas by replacing occurrences of propositional variables by arbitrary conjunctive queries (CQs). Moreover, as in this article, the research (Borgwardt et al. 2013; Borgwardt et al. 2015) focuses on the rewritability properties of LTL-CQs. An orthogonal approach for query rewriting over a temporalized *DL-Lite* ontology was proposed in (Artale et al. 2013; Artale et al. 2015). Here the focus mainly lies on increasing expressivity of an ontology language by allowing a concept or a role to be prefixed with LTL constructs. Only recently, also metric variants of LTL-CQs have been considered (Gutiérrez-Basulto et al. 2016; Baader et al. 2017; Thost 2018).

Negation in queries with the classical open-world semantics results in non-tractable (mostly CONP or even undecidable) query evaluation (Rosati 2007a; Gutiérrez-Basulto et al. 2015). Moreover, prior work (Borgwardt and Thost 2015a; Borgwardt and Thost 2020) on temporalized ontology-mediated query answering with negation shows that the high complexity of temporal query answering with negation is mostly due to the open-world assumption for negation in a query language. There are several approaches how to

introduce negation in ontology-mediated query answering without losing tractability, e.g. closed predicates, epistemic semantics, Skolemization. For details, we refer the reader to Section 3, where we motivate the semantics adopted in the paper. Indeed, as we show by this work, by changing semantics for negation, we can apply efficient (in data complexity) algorithms for temporal query answering with negation.

8 Conclusion

Dealing with the absence of information is an important and at the same time challenging task. In many real-world scenarios, it is not clear whether a piece of information is missing because it is unknown or because it is false. EHRs mostly talk about positive diagnoses and it would be impossible to list all the negative diagnoses, i.e. the diseases a patient does not suffer from. Moreover, EHRs and clinical trial criteria contain an inherent temporal component. We showed that such a setting cannot be handled adequately by existing logic-based approaches, mostly because they do not deal with closed-world negation over anonymous objects. We introduced a novel semantics for answering metric temporal conjunctive queries with negation and showed that it is well-behaved also for anonymous objects. Moreover, we demonstrated combined first-order rewritability, which allows us to answer MTNCQs by using conventional relational database technologies. The data complexity result of P shows that MTNCQ answering over $\mathcal{TELH}_{\bot}^{\Diamond,\mathsf{lhs}}$ KBs is not inherently more difficult than CQ answering in $\mathcal{EL}$. Similarly, the combined complexity does not increase over the lower bound of ExpSpace inherited from metric temporal logic.

We are working on an optimized implementation of this method with the aim to deal with queries over large ontologies such as SNOMED CT. First results show that this is feasible, in spite of the theoretical complexity of ExpSpace.[12] There is already a prototype system that can translate simple clinical trial criteria from clinicaltrials.gov into MTNCQs (Xu et al. 2019). On the theoretical side, we will further develop our approach to also represent numeric information, such as measurements and dosages of medications, which are important for evaluating eligibility criteria of clinical trials (Crowe and Tao 2015; Bonomi and Jiang 2018). Similar to negation, inequalities in CQs also lead to an increase in complexity under certain answer semantics (Gutiérrez-Basulto et al. 2015). It would be interesting to investigate their behavior under closed-world semantics, also because inequalities are related to counting queries, e.g. counting the number of diseases or treatments of a patient. Another open problem is to extend our results to temporal logics that support *temporal roles*, i.e. where temporal operators can also be applied to role names.

Competing interests The author(s) declare none.

---

[12] `https://github.com/wko/quelk`

References

AHMETAJ, S., ORTIZ, M., AND SIMKUS, M. 2016. Polynomial datalog rewritings for expressive description logics with closed predicates. In *Proc. of the 25th Int. Joint Conf. on Artificial Intelligence (IJCAI'16)*, S. Kambhampati, Ed. AAAI Press, 878–885.

ALUR, R., FEDER, T., AND HENZINGER, T. A. 1996. The benefits of relaxing punctuality. *J. ACM 43,* 1, 116–146.

ALUR, R. AND HENZINGER, T. A. 1994. A really temporal logic. *J. ACM 41,* 1, 181–204.

ARENAS, M., GOTTLOB, G., AND PIERIS, A. 2014. Expressive languages for querying the semantic web. In *Proc. of the 33rd Symp. on Principles of Database Systems (PODS'14)*, R. Hull and M. Grohe, Eds. ACM, 14–26.

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

ARTALE, A. AND FRANCONI, E. 2005. Temporal description logics. In *Handbook of Temporal Reasoning in Artificial Intelligence*. Foundations of Artificial Intelligence, vol. 1. Elsevier, 375–388.

ARTALE, A., KONTCHAKOV, R., KOVTUNOVA, A., RYZHIKOV, V., WOLTER, F., AND ZA-KHARYASCHEV, M. 2015. First-order rewritability of ontology-mediated temporal queries. In *Proc. IJCAI*. AAAI Press, 2706–2712.

ARTALE, A., KONTCHAKOV, R., KOVTUNOVA, A., RYZHIKOV, V., WOLTER, F., AND ZA-KHARYASCHEV, M. 2017. Ontology-mediated query answering over temporal data: A survey (invited talk). In *Proc. of the 24th Int. Symp. on Temporal Representation and Reasoning (TIME'17)*. Leibniz International Proceedings in Informatics, vol. 90. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 1:1–1:37.

ARTALE, A., KONTCHAKOV, R., LUTZ, C., WOLTER, F., AND ZAKHARYASCHEV, M. 2007. Temporalising tractable description logics. In *Proc. TIME*. IEEE Press, 11–22.

ARTALE, A., KONTCHAKOV, R., RYZHIKOV, V., AND ZAKHARYASCHEV, M. 2014. A cookbook for temporal conceptual data modelling with description logics. *ACM Trans. Comput. Log. 15,* 3, 25:1–25:50.

ARTALE, A., KONTCHAKOV, R., WOLTER, F., AND ZAKHARYASCHEV, M. 2013. Temporal description logic for ontology-based data access. In *Proc. IJCAI*. AAAI Press, 711–717.

BAADER, F., BORGWARDT, S., KOOPMANN, P., OZAKI, A., AND THOST, V. 2017. Metric temporal description logics with interval-rigid names. In *Proc. of the 11th Int. Symp. on Frontiers of Combining Systems (FroCoS'17)*. Springer, 60–76.

BAADER, F., BORGWARDT, S., AND LIPPMANN, M. 2013. Temporalizing ontology-based data access. In *Proc. of the 24th Int. Conf. on Automated Deduction (CADE'13)*. Springer, 330–344.

BAADER, F., BORGWARDT, S., AND LIPPMANN, M. 2015a. Temporal conjunctive queries in expressive description logics with transitive roles. In *Proc. of the 28th Australasian Joint Conf. on Advances in Artificial Intelligence, AI'15*. LNCS, vol. 9457. Springer, 21–33.

BAADER, F., BORGWARDT, S., AND LIPPMANN, M. 2015b. Temporal query entailment in the description logic $\mathcal{SHQ}$. *J. Web Sem. 33*, 71–93.

BAADER, F., BRANDT, S., AND LUTZ, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI'05)*. Professional Book Center, 364–369.

BAADER, F., CALVANESE, D., MCGUINNESS, D. L., NARDI, D., AND PATEL-SCHNEIDER, P. F., Eds. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*, 2nd ed. Cambridge University Press.

BAADER, F., GHILARDI, S., AND LUTZ, C. 2012. LTL over description logic axioms. *ACM Trans. Comput. Logic 13,* 3, 21:1–21:32.

BAADER, F., HORROCKS, I., LUTZ, C., AND SATTLER, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

BAADER, F., KÜSTERS, R., AND WOLTER, F. 2003. Extensions to description logics. In *The Description Logic Handbook*. Cambridge University Press, 219–261.

BÁRÁNY, V., TEN CATE, B., AND OTTO, M. 2012. Queries with guarded negation. *Proc. of the VLDB Endowment 5,* 11, 1328–1339.

BASIN, D., KLAEDTKE, F., MÜLLER, S., AND ZĂLINESCU, E. 2015. Monitoring metric first-order temporal properties. *J. ACM 62,* 2, 15:1–15:45.

BAUSLAUGH, B. 1995. Core-like properties of infinite graphs and structures. *Discrete mathematics 138,* 1-3, 101–111.

BESANA, P., CUGGIA, M., ZEKRI, O., BOURDE, A., AND BURGUN, A. 2010. Using semantic web technologies for clinical trial recruitment. In *Proc. of the 9th Int. Semantic Web Conf. (ISWC'10)*, P. F. Patel-Schneider, Y. Pan, P. Hitzler, P. Mika, L. Zhang, J. Z. Pan, I. Horrocks, and B. Glimm, Eds. Lecture Notes in Computer Science, vol. 6497. Springer, 34–49.

BIENVENU, M. AND ORTIZ, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web 11th Int. Summer School*, W. Faber and A. Paschke, Eds. Lecture Notes in Computer Science, vol. 9203. Springer, 218–307.

BONOMI, L. AND JIANG, X. 2018. Patient ranking with temporally annotated data. *Journal of Biomedical Informatics 78*, 43–53.

BORGWARDT, S. AND FORKEL, W. 2019. Closed-world semantics for conjunctive queries with negation over $\mathcal{ELH}_\perp$ ontologies. In *Proc. of the 16th European Conf. on Logics in Artificial Intelligence (JELIA'19)*. Springer, Rende, Italy, 371–386. To appear.

BORGWARDT, S., FORKEL, W., AND KOVTUNOVA, A. 2019. Finding new diamonds: Temporal minimal- world query answering over sparse aboxes. In *Proc. of the 3rd International Joint Conference on Rules and Reasoning (RuleML+RR'19)*, P. Foder, M. Montali, D. Calvanese, and D. Roman, Eds. Springer, Bolzano, Italy. To appear.

BORGWARDT, S., LIPPMANN, M., AND THOST, V. 2013. Temporal query answering in the description logic *DL-Lite*. In *Proc. of the 9th Int. Symposium on Frontiers of Combining Systems, FroCoS'13*. LNCS, vol. 8152. Springer, 165–180.

BORGWARDT, S., LIPPMANN, M., AND THOST, V. 2015. Temporalizing rewritable query languages over knowledge bases. *Web Semantics: Science, Services and Agents on the World Wide Web 33*, 50–70.

BORGWARDT, S. AND THOST, V. 2015a. Temporal query answering in *DL-Lite* with negation. In *Proc. of the 1st Global Conf. on Artificial Intelligence (GCAI'15)*. EasyChair, 51–65.

BORGWARDT, S. AND THOST, V. 2015b. Temporal query answering in the description logic $\mathcal{EL}$. In *Proc. of the 24h Int. Joint Conf. on Artificial Intelligence, IJCAI'15*. AAAI Press, 2819–2825.

BORGWARDT, S. AND THOST, V. 2020. Temporal conjunctive query answering in the extended *DL-Lite* family. In *arXiv:2003.09508*.

BRANDT, S., KALAYCI, E. G., RYZHIKOV, V., XIAO, G., AND ZAKHARYASCHEV, M. 2018. Querying log data with metric temporal logic. *J. Artif. Intell. Res. 62*, 829–877.

CALVANESE, D., COGREL, B., KOMLA-EBRI, S., KONTCHAKOV, R., LANTI, D., REZK, M., RODRIGUEZ-MURO, M., AND XIAO, G. 2017. Ontop: Answering SPARQL queries over relational databases. *Semantic Web 8*, 471–487.

CALVANESE, D., DE GIACOMO, G., LEMBO, D., LENZERINI, M., AND ROSATI, R. 2006. Epistemic first-order queries over description logic knowledge bases. In *Proc. of the 19th Int. Workshop on Description Logics (DL'06)*, B. Parsia, U. Sattler, and D. Toman, Eds. CEUR Workshop Proceedings, vol. 189. 51–61.

CALVANESE, D., DE GIACOMO, G., LEMBO, D., LENZERINI, M., AND ROSATI, R. 2013. Data complexity of query answering in description logics. *Art. Intell. 195*, 335–360.

CRESSWELL, K. M. AND SHEIKH, A. 2017. Inpatient clinical information systems. In *Key

Advances in Clinical Informatics, A. Sheikh, K. M. Cresswell, A. Wright, and D. W. Bates, Eds. Academic Press, Chapter 2, 13–29.

CROWE, C. L. AND TAO, C. 2015. Designing ontology-based patterns for the representation of the time-relevant eligibility criteria of clinical protocols. *AMIA Joint Summits on Translational Science Proc. 2015*, 173–177.

DEUTSCH, A., NASH, A., AND REMMEL, J. B. 2008. The chase revisited. In *Proc. of the 27th ACM Symp. on Principles of Database Systems (PODS'08)*. ACM, 149–158.

EITER, T., ORTIZ, M., ŠIMKUS, M., TRAN, T.-K., AND XIAO, G. 2012. Query rewriting for Horn-$\mathcal{SHIQ}$ plus rules. In *Proc. of the 26th AAAI Conf. on Artificial Intelligence (AAAI'12)*, J. Hoffmann and B. Selman, Eds. AAAI Press, 726–733.

FAGIN, R., KOLAITIS, P. G., MILLER, R. J., AND POPA, L. 2005. Data exchange: Semantics and query answering. *Theoretical Computer Science 336,* 1, 89–124.

FORKEL, W. 2020. Closed-world semantics for query answering in temporal description logics. Ph.D. thesis, Technische Universität Dresden, Germany.

FURIA, C. A. AND SPOLETINI, P. 2008. Tomorrow and all our yesterdays: MTL satisfiability over the integers. In *Proc. ICTAC*. Springer, 126–140.

GUTIÉRREZ-BASULTO, V., IBÁÑEZ-GARCÍA, Y., KONTCHAKOV, R., AND KOSTYLEV, E. V. 2015. Queries with negation and inequalities over lightweight ontologies. *Journal of Web Semantics 35*, 184–202.

GUTIÉRREZ-BASULTO, V., JUNG, J. C., AND KONTCHAKOV, R. 2016. Temporalized $\mathcal{EL}$ ontologies for accessing temporal data: Complexity of atomic queries. In *Proc. IJCAI*. AAAI Press, 1102–1108.

GUTIÉRREZ-BASULTO, V., JUNG, J. C., AND OZAKI, A. 2016. On metric temporal description logics. In *Proc. ECAI*. IOS Press, 837–845.

HERNICH, A., KUPKE, C., LUKASIEWICZ, T., AND GOTTLOB, G. 2013. Well-founded semantics for extended datalog and ontological reasoning. In *Proc. of the 32nd Symp. on Principles of Database Systems (PODS'13)*, R. Hull and W. Fan, Eds. ACM, 225–236.

JOHNSON, A. E. W., POLLARD, T. J., SHEN, L., LEHMAN, L.-w. H., FENG, M., GHASSEMI, M., MOODY, B., SZOLOVITS, P., CELI, L. A., AND MARK, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data 3,* 160035, 1–9.

KHARLAMOV, E., HOVLAND, D., SKJÆVELAND, M. G., BILIDAS, D., JIMÉNEZ-RUIZ, E., XIAO, G., SOYLU, A., LANTI, D., REZK, M., ZHELEZNYAKOV, D., GIESE, M., LIE, H., IOANNIDIS, Y., KOTIDIS, Y., KOUBARAKIS, M., AND WAALER, A. 2017. Ontology based data access in Statoil. *J. Web Semantics 44*, 3–36.

KONTCHAKOV, R., LUTZ, C., TOMAN, D., WOLTER, F., AND ZAKHARYASCHEV, M. 2011. The combined approach to ontology-based data access. In *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI'11)*, T. Walsh, Ed. AAAI Press, 2656–2661.

KONTCHAKOV, R., PANDOLFO, L., PULINA, L., RYZHIKOV, V., AND ZAKHARYASCHEV, M. 2016. Temporal and spatial OBDA with many-dimensional Halpern-Shoham logic. In *Proc. of the 25th Int. Joint Conf. on Artificial Intelligence (IJCAI'16)*. AAAI Press, 1160–1166.

KÖPCKE, F. AND PROKOSCH, H.-U. 2014. Employing computers for the recruitment into clinical trials: A comprehensive systematic review. *Journal of Medical Internet Research 16,* 7, e161.

KRÖTZSCH, M. 2020. Computing cores for existential rules with the standard chase and ASP. In *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR 2020)*. IJCAI.

LINDELL, S. 1992. A purely logical characterization of circuit uniformity. In *Proc. of the 7th Annual Structure in Complexity Theory Conf.* 185–192.

LUTZ, C., SEYLAN, I., AND WOLTER, F. 2013. Ontology-based data access with closed predicates is inherently intractable (sometimes). In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI'13)*, F. Rossi, Ed. AAAI Press, 1024–1030.

Lutz, C., Toman, D., and Wolter, F. 2009. 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)*. AAAI Press, 2070–2075.

Lutz, C., Wolter, F., and Zakharyaschev, M. 2008. Temporal description logics: A survey. In *Proc. of the 15th Int. Symp. on Temporal Representation and Reasoning (TIME'08)*. IEEE Press, 3–14.

Marnette, B. 2009. Generalized schema mappings: From termination to tractability. In *Proc. of the 28th Symp. on Principles of Database Systems (PODS'09)*, J. Paredaens and J. Su, Eds. ACM, 13–22.

Mugnier, M.-L. and Thomazo, M. 2014. An introduction to ontology-based query answering with existential rules. In *Reasoning Web International Summer School*. 245–278.

Ni, Y., Wright, J., Perentesis, J., Lingren, T., Deleger, L., Kaiser, M., Kohane, I., and Solti, I. 2015. 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.

Patel, C., Cimino, J., Dolby, J., Fokoue, A., Kalyanpur, A., Kershenbaum, A., Ma, L., Schonberg, E., and Srinivas, K. 2007. Matching patient records to clinical trials using ontologies. In *Proc. of the 6th Int. Semantic Web Conf. (ISWC'07)*, K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Goldbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, Eds. Lecture Notes in Computer Science, vol. 4825. Springer, 816–829.

Rosati, R. 2007a. The limits of querying ontologies. In *Proc. of the 11th Int. Conf. on Database Theory (ICDT'07)*, T. Schwentick and D. Suciu, Eds. 164–178.

Rosati, R. 2007b. On conjunctive query answering in $\mathcal{EL}$. In *Proc. DL*. 451–458.

Ryzhikov, V., Walega, P. A., and Zakharyaschev, M. 2019. Data complexity and rewritability of ontology-mediated queries in metric temporal logic under the event-based semantics. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 1851–1857.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., and Chute, C. G. 2010. Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association 17,* 5, 507–513.

Tagaris, A., Andronikou, V., Chondrogiannis, E., Tsatsaronis, G., Schroeder, M., Varvarigou, T., and Koutsouris, D.-D. 2014. *Exploiting Ontology Based Search and EHR Interoperability to Facilitate Clinical Trial Design.* Springer, 21–42.

Thost, V. 2018. Metric temporal extensions of *DL-Lite* and interval-rigid names. In *Proc. KR*. AAAI Press, 665–666.

Vardi, M. Y. 1982. The complexity of relational query languages (extended abstract). In *Proc. STOC*. ACM, 137–146.

Wolter, F. 2000. First order common knowledge logics. *Studia Logica 65,* 2, 249–271.

Wolter, F. and Zakharyaschev, M. 2000. Temporalizing description logics. In *Frontiers of Combining Systems 2*. Research Studies Press/Wiley, 379–402.

Xu, C., Forkel, W., Borgwardt, S., Baader, F., and Zhou, B. 2019. Automatic translation of clinical trial eligibility criteria into formal queries. In *Proc. of the 9th Workshop on Ontologies and Data in Life Sciencs (ODLS'19), part of The Joint Ontology Workshops (JOWO'19)*, M. Boeker, L. Jansen, F. Loebe, and S. Schulz, Eds. CEUR Workshop Proceedings.