Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

First-Order Rewritability of Temporal Ontology-Mediated Queries

A. Artale[1], **R. Kontchakov**[2], **A. Kovtunova**[1], **V. Ryzhikov**[1], **F. Wolter**[3] and **M. Zakharyaschev**[2]

[1]Faculty of Computer Science, Free University of Bozen-Bolzano, Italy

[2]Department of Computer Science and Information Systems, Birkbeck, University of London, UK

[3]Department of Computer Science, University of Liverpool, UK

Abstract

Aiming at ontology-based data access over temporal, in particular streaming data, we design a language of ontology-mediated queries by extending OWL 2 QL and SPARQL with temporal operators, and investigate rewritability of these queries into two-sorted first-order logic with $<$ and PLUS over time.

1 Introduction

Ontology-based data access (OBDA), one of the most promising applications of description logics, has recently been touted as a key ingredient in data stream management systems [Calbimonte *et al.*, 2012; Baader *et al.*, 2013; Özçep and Möller, 2014; Kharlamov *et al.*, 2014]. Its role is to facilitate querying data streams from heterogeneous sensor networks (measuring temperature, vibration, wind speed and direction, blood pressure, etc.) in order to detect, analyse and predict certain events or situations. For example, we may want to detect areas of the country that have been hit by blizzards—severe snowstorms with low temperatures and strong winds for at least three hours—using the weather stations' sensors streaming raw data such as $(\mathit{sensorId}, \mathit{temperature}, \mathit{timestamp})$ [Zhang *et al.*, 2012]. The OBDA component would give us an ontology, $\mathcal{O}$, that defines a high-level vocabulary—possibly including the relation 'area x was hit by a blizzard at time t'—which is linked to the raw data by means of (say, R2RML) mappings. Given a query in the vocabulary of $\mathcal{O}$ such as $\exists t\, \mathit{Blizzard}(x,t)$, the system would first rewrite the pair $(\mathcal{O}, \exists t\, \mathit{Blizzard}(x,t))$, called a *temporal ontology-mediated query* (TOMQ), into a query over the original timestamped data and then evaluate it using standard relational or stream data management systems.

The OBDA scenario outlined above is similar to the classical one [Calvanese *et al.*, 2007a]. An essential difference, however, is that now we query *temporal* data and therefore, require temporal constructs in the ontology and/or query languages, which, on the other hand, should not undermine the rewritability property of TOMQs. As temporal extensions of knowledge representation formalisms are notorious for their bad computational behaviour [Lutz *et al.*, 2008], a first natural step could be to keep standard atemporal OBDA languages (e.g., OWL 2 QL or *DL-Lite* logics), assuming that ontology axioms hold at all moments of time, but add temporal constructs to queries. This approach was taken by [Gutiérrez-Basulto and Klarman, 2012; Baader *et al.*, 2013; Borgwardt *et al.*, 2013; Özcep *et al.*, 2013; Klarman and Meyer, 2014] and shown to preserve query rewritability. Note, however, that the inability to define temporal predicates such as $\mathit{Blizzard}(x,t)$ in ontologies leaves the burden of encoding them within queries to the user, which goes against the OBDA paradigm. Moreover, natural queries such as 'check if a weather station has been serviced every 24 hours' are not expressible in these languages.

The first attempt to introduce linear-time temporal logic (*LTL*) operators to the TOMQs' ontologies was made by Artale *et al.* [2013b], who showed that the operators $\Diamond_F$ and $\Diamond_P$ (sometime in the future and past) on the left-hand side of *DL-Lite* axioms allow rewritings of conjunctive queries into the two-sorted first-order language FO($<$) with variables of sorts 'object' and 'time' and an explicit temporal precedence relation $<$. On the other hand, it was observed that the next- and previous-time operators $\bigcirc_F$ and $\bigcirc_P$—perhaps the most powerful and versatile temporal modelling constructs—do not support FO($<$)-rewritability.

The aim of this paper is to launch a systematic investigation of rewritability of TOMQs with arbitrary temporal operators. Apart from FO($<$), we consider two more target languages for rewritings: FO($<,+$), which complements FO($<$) with a ternary predicate PLUS for 'addition' and still ensures query evaluation in LOGTIME-uniform AC^0 for data complexity; and monadic second-order logic MSO($<$), which guarantees query evaluation over finite linear orders (flows of time) in $\mathrm{NC}^1 \subseteq$ LOGSPACE for data complexity. Axioms of our ontologies are given in clausal normal form

$$\lambda_1 \sqcap \dots \sqcap \lambda_n \ \sqsubseteq\ \lambda_{n+1} \sqcup \dots \sqcup \lambda_{n+m}, \qquad (1)$$

where the λ_i are all either *DL-Lite* concepts or roles, possibly prefixed with the operators $\bigcirc_F$, $\bigcirc_P$, $\Box_F$ and $\Box_P$ (always in the future and the past). Let $\boldsymbol{o} \in \{\Box, \bigcirc, \Box\bigcirc\}$ and $\boldsymbol{c} \in \{\mathit{bool}, \mathit{horn}, \mathit{krom}, \mathit{core}\}$. We denote by $\mathit{DL\text{-}Lite}^{\boldsymbol{o}}_{\boldsymbol{c}}$ the *temporal description logic* with axioms of the form (1), where the λ_i can only use the (future and past) operators indicated in $\boldsymbol{o}$, and $m \le 1$ if $\boldsymbol{c} = \mathit{horn}$; $n+m \le 2$ if $\boldsymbol{c} = \mathit{krom}$; $n+m \le 2$ and $m \le 1$ if $\boldsymbol{c} = \mathit{core}$; and arbitrary n, m if $\boldsymbol{c} = \mathit{bool}$. For example, the $\mathit{DL\text{-}Lite}^{\bigcirc}_{\mathit{horn}}$ axioms

$$\bigsqcap_{0 \le i \le 2} \bigcirc^i_F \mathit{BlizzardCondition} \sqsubseteq \bigcirc^j_F \mathit{Blizzard}, \quad j = 0,1,2,$$

$$\mathit{SevereSnow} \sqcap \mathit{LowTemp} \sqcap \mathit{StrongWind} \sqsubseteq \mathit{BlizzardCondition},$$

where $\bigcirc_F^i$ is a sequence of i-many $\bigcirc_F$, define the temporal concept *Blizzard* (the concepts on the left-hand side of the second axiom are defined via mappings from the sensor data), while the $\textit{DL-Lite}_{core}^{\bigcirc}$ axiom $\textit{Scheduled} \sqsubseteq \bigcirc_F^{24}\textit{Scheduled}$ can be used to detect 'service every 24 hours' (see below).

The queries in our TOMQs range from atoms of the form $A(x,t)$ or $P(x,y,t)$, for A a concept and P a role name, in *atomic* TOMQs (or TOMAQs), to arbitrary positive temporal concepts and roles such as $(\textit{Blizzard}\,\mathcal{U}(\textit{Rain}\sqcup\textit{Flooding}))(x,t)$ in *instance* TOMQs (TOMIQs), where $\mathcal{U}$ is the 'until' operator, and further to two-sorted FO-formulas built from positive temporal concepts and roles. It is well known that, due to the open-world semantics of OBDA, negation in queries results in non-tractable (often undecidable) query evaluation even in the atemporal case. In classical OBDA, a way to retain all FO-constructs in queries was proposed by Calvanese *et al.* [2007b] who interpreted them under an epistemic semantics. We follow in their footsteps: for example, the query $\textit{Scheduled}(x,t) \land \neg\textit{Serviced}(x,t)$ returns weather stations x that are not *known* (or not *certain*) to have been serviced at scheduled times t. With this semantics, TOMQs cover all first-order features of SPARQL 1.1 under the OWL 2 QL entailment regime [Kontchakov *et al.*, 2014]. Although the variables in TOMQs range over the active domain, their positive temporal concepts can express (via qualified existential restrictions and role intersections) tree-shaped CQs with one answer variable.

To only focus on rewritability, we assume that data instances are given as finite ABoxes with timestamped atoms of the form $A(a,n)$ and $P(a,b,n)$, $n \in \mathbb{Z}$ (e.g., generated from streamed data via mappings and the window operator [Calbimonte *et al.*, 2012]). We proceed in two steps. First, we investigate rewritability of TOMQs without roles, which can be regarded as *LTL* TOMQs. Using automata-theoretic techniques, we obtain the following rewritability classification:

	TOMAQs			TOMIQs / TOMQs		
	$\textit{LTL}_\alpha^{\Box}$	$\textit{LTL}_\alpha^{\bigcirc}$	$\textit{LTL}_\alpha^{\Box\bigcirc}$	$\textit{LTL}_\alpha^{\Box}$	$\textit{LTL}_\alpha^{\bigcirc}$	$\textit{LTL}_\alpha^{\Box\bigcirc}$
bool	FO(<)	MSO(<)	MSO(<)	MSO(<)	MSO(<)	MSO(<)
horn	FO(<)	MSO(<)	MSO(<)	FO(<)	MSO(<)	MSO(<)
krom	FO(<)	FO(<,+)	MSO(<)*	MSO(<)	MSO(<)	MSO(<)
core	FO(<)	FO(<,+)	MSO(<)*	FO(<)	FO(<,+)	MSO(<)*

*It is still open whether these can be improved to FO($<,+$); all other results in the table are optimal.

In the second step, we reduce FO-rewritability of $\textit{DL-Lite}_{horn}^{\Box\bigcirc}$ TOMQs to FO-rewritability of certain $\textit{LTL}_{horn}^{\Box\bigcirc}$ TOMQs. In particular, we prove that all $\textit{DL-Lite}_{core}^{\Box}$ TOMQs are FO($<$)-rewritable, while $\textit{DL-Lite}_{core}^{\bigcirc}$ TOMQs are FO($<,+$)-rewritable. On the other hand, we show that some $\textit{DL-Lite}_{horn}^{\Box}$ TOMQs are NC^1-hard for data complexity, and so they cannot be FO-rewritable even using arbitrary numeric predicates. All omitted proofs can be found at tinyurl.com/q6ahvnt.

2 Temporal Ontology-Mediated Queries

First we remind the reader of basic *DL-Lite* logics [Calvanese *et al.*, 2007a; Artale *et al.*, 2009]. Their language contains *object names* $a_0, a_1, \dots$, *concept names* $A_0, A_1, \dots$, and *role names* $P_0, P_1, \dots$. *Roles* R and *basic concepts* B are defined as $R ::= P_k \mid P_k^-$ and $B ::= A_k \mid \exists R$. For $\boldsymbol{c} \in \{\textit{bool}, \textit{horn}, \textit{krom}, \textit{core}\}$, we denote by $\textit{DL-Lite}_{\boldsymbol{c}}$ the description logic with *concept* and *role inclusions* of the form (1) such that the λ_i are all either basic concepts or roles. As usual, we assume that the empty $\sqcap$ is $\top$ and the empty $\sqcup$ is $\bot$. Note that $\textit{DL-Lite}_{horn}$ and $\textit{DL-Lite}_{bool}$ contain role inclusions with monotone Boolean operators.

In temporal *DL-Lite*, we also allow applications of the operators $\bigcirc_F$, $\bigcirc_P$, $\Box_F$, $\Box_P$ to basic concepts and roles. For any $\boldsymbol{o} \in \{\Box, \bigcirc, \Box\bigcirc\}$, $\textit{DL-Lite}_{\boldsymbol{c}}^{\boldsymbol{o}}$ is the *temporal description logic* with concept and role inclusions of the form (1), where each λ_i is a basic concepts or roles are (possibly) prefixed by a string of (future or past) operators indicated in $\boldsymbol{o}$.

A $\textit{DL-Lite}_{\boldsymbol{c}}^{\boldsymbol{o}}$ *TBox* $\mathcal{T}$ (*RBox* $\mathcal{R}$) is a finite set of $\textit{DL-Lite}_{\boldsymbol{c}}^{\boldsymbol{o}}$ concept (role) inclusions; $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$ is a $\textit{DL-Lite}_{\boldsymbol{c}}^{\boldsymbol{o}}$ *ontology*. An *ABox* (or *data instance*), $\mathcal{A}$, is a finite set of atoms of the form $A(a,\ell)$ and $P(a,b,\ell)$, where a, b are object names and $\ell \in \mathbb{Z}$. We denote by $\mathsf{ind}(\mathcal{A})$ the set of object names in $\mathcal{A}$, by $\min \mathcal{A}$ and $\max \mathcal{A}$ the minimal and maximal numbers in $\mathcal{A}$, and set $\mathsf{tem}(\mathcal{A}) = \{n \in \mathbb{Z} \mid \min \mathcal{A} \le n \le \max \mathcal{A}\}$. Without loss of generality, we assume that $\min \mathcal{A} = 0$ and $\max \mathcal{A} \ge 1$.

A *temporal interpretation* is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}(n)})$, where $\Delta^{\mathcal{I}} \neq \emptyset$ and $\mathcal{I}(n) = (\Delta^{\mathcal{I}}, a_0^{\mathcal{I}}, \dots, A_0^{\mathcal{I}(n)}, \dots, P_0^{\mathcal{I}(n)}, \dots)$ is a standard DL interpretation for each time instant $n \in \mathbb{Z}$, that is, $a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, $A_i^{\mathcal{I}(n)} \subseteq \Delta^{\mathcal{I}}$ and $P_i^{\mathcal{I}(n)} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. Thus, we assume that the domain $\Delta^{\mathcal{I}}$ and the interpretations $a_i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ of the object names are the same for all $n \in \mathbb{Z}$. The DL and temporal constructs are interpreted in $\mathcal{I}(n)$ as follows:

$$(P_i^-)^{\mathcal{I}I(n)} = \{(x,y) \mid (y,x) \in P_i^{\mathcal{I}(n)}\},$$
$$(\exists R)^{\mathcal{I}(n)} = \{x \mid (x,y) \in R^{\mathcal{I}(n)}, \text{ for some } y\},$$
$$(\Box_F \lambda)^{\mathcal{I}(n)} = \bigcap\nolimits_{k>n} \lambda^{\mathcal{I}(k)}, \qquad (\bigcirc_F \lambda)^{\mathcal{I}(n)} = \lambda^{\mathcal{I}(n+1)},$$

and symmetrically for $\Box_P$ and $\bigcirc_P$; as usual, $\bot$ is interpreted by $\emptyset$ and $\top$ by $\Delta^{\mathcal{I}}$ for concepts and by $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for roles. Concept and role inclusions are interpreted in $\mathcal{I}$ *globally* in the sense that (1) holds in $\mathcal{I}$ if $\bigcap \lambda_i^{\mathcal{I}(n)} \subseteq \bigcup \lambda_j^{\mathcal{I}(n)}$ for *all* $n \in \mathbb{Z}$. We call $\mathcal{I}$ a *model* of $(\mathcal{O}, \mathcal{A})$ and write $\mathcal{I} \models (\mathcal{O}, \mathcal{A})$ if all concept and role inclusions from $\mathcal{O}$ hold in $\mathcal{I}$, and $a^{\mathcal{I}} \in A^{\mathcal{I}(n)}$ for $A(a,n) \in \mathcal{A}$, and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}(n)}$ for $P(a,b,n) \in \mathcal{A}$.

Remark 1. Note that the *LTL* operators $\Diamond_F$, $\Diamond_P$, $\mathcal{U}$ and $\mathcal{S}$ ('since') can be expressed in $\textit{DL-Lite}_{bool}^{\Box\bigcirc}$ [Fisher *et al.*, 2001; Artale *et al.*, 2013a]. $\textit{DL-Lite}_{horn}^{\Box}$ extends the ontology language *TQL* of Artale *et al.* [2013b] as $\Diamond_P A \sqsubseteq B$ is equivalent to $A \sqsubseteq \Box_F B$. Note also that each of our ontology languages can say that a role or basic concept λ is *expanding* (using $\lambda \sqsubseteq \bigcirc_F \lambda$ in $\textit{DL-Lite}_{core}^{\bigcirc}$ and $\lambda \sqsubseteq \Box_F \lambda$ in $\textit{DL-Lite}_{core}^{\Box}$) or *rigid* (using, in addition, $\lambda \sqsubseteq \bigcirc_P \lambda$ and $\lambda \sqsubseteq \Box_P \lambda$).

To query temporal ontologies with data, we suggest the following language inspired by the recently standardised SPARQL 1.1 entailment regimes (www.w3.org/TR/sparql11-entailment); cf. also [Motik, 2012; Gutierrez *et al.*, 2007].

Positive temporal concepts $\varkappa$ and *positive temporal roles* ϱ are defined by the grammars

$$\varkappa ::= \top \mid A_k \mid \exists R.\varkappa \mid \varkappa_1 \sqcap \varkappa_2 \mid \varkappa_1 \sqcup \varkappa_2 \mid \boldsymbol{op}\,\varkappa \mid \varkappa_1\, \boldsymbol{op}'\, \varkappa_2,$$
$$\varrho ::= R \mid \varrho_1 \sqcap \varrho_2 \mid \varrho_1 \sqcup \varrho_2 \mid \boldsymbol{op}\,\varrho \mid \varrho_1\, \boldsymbol{op}'\, \varrho_2,$$

where $\boldsymbol{op}$ ranges over $\bigcirc_F$, $\Diamond_F$, $\Box_F$, $\bigcirc_P$, $\Diamond_P$, $\Box_P$ and $\boldsymbol{op}'$ over $\mathcal{U}$, $\mathcal{S}$. The extensions of $\varkappa$ and ϱ in a temporal interpretation

$\mathcal{I}$ are computed using the definition above and the following clauses (given only for $\varkappa$ and the future-time operators):

$$(\exists R.\varkappa)^{\mathcal{I}(n)} = \{x \mid (x,y) \in R^{\mathcal{I}(n)}, \text{ for some } y \in \varkappa^{\mathcal{I}(n)}\},$$
$$(\varkappa_1 \sqcap \varkappa_2)^{\mathcal{I}(n)} = \varkappa_1^{\mathcal{I}(n)} \cap \varkappa_2^{\mathcal{I}(n)},$$
$$(\varkappa_1 \sqcup \varkappa_2)^{\mathcal{I}(n)} = \varkappa_1^{\mathcal{I}(n)} \cup \varkappa_2^{\mathcal{I}(n)},$$
$$(\Diamond_{\scriptscriptstyle F} \varkappa)^{\mathcal{I}(n)} = \bigcup_{k>n} \varkappa^{\mathcal{I}(k)},$$
$$(\varkappa_1 \,\mathcal{U}\, \varkappa_2)^{\mathcal{I}(n)} = \bigcup_{k>n} \big(\varkappa_2^{\mathcal{I}(k)} \cap \bigcap_{n<m<k} \varkappa_1^{\mathcal{I}(m)}\big).$$

A *DL-Lite$^{o}_{c}$ ontology-mediated instance query* (TOMIQ) is a pair of the form $(\mathcal{O}, \varkappa)$ or $(\mathcal{O}, \varrho)$, for a *DL-Lite$^{o}_{c}$* ontology $\mathcal{O}$. A *certain answer* to $(\mathcal{O}, \varkappa)$ over an ABox $\mathcal{A}$ is a pair (a, ℓ) such that $a \in \mathsf{ind}(\mathcal{A})$, $\ell \in \mathsf{tem}(\mathcal{A})$ and $a^{\mathcal{I}} \in \varkappa^{\mathcal{I}(\ell)}$ for every $\mathcal{I} \models (\mathcal{O}, \mathcal{A})$. A *certain answer* to $(\mathcal{O}, \varrho)$ over $\mathcal{A}$ is a triple (a, b, ℓ) with $a, b \in \mathsf{ind}(\mathcal{A})$, $\ell \in \mathsf{tem}(\mathcal{A})$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in \varrho^{\mathcal{I}(\ell)}$ for every $\mathcal{I} \models (\mathcal{O}, \mathcal{A})$. Let $\mathsf{ans}(\mathcal{O}, \varkappa, \mathcal{A})$ denote the set of certain answers to $(\mathcal{O}, \varkappa)$ over $\mathcal{A}$, and similarly for $(\mathcal{O}, \varrho)$.

As a technical tool in our proofs, we also require ‘certain answers’ in which ℓ can range over the whole $\mathbb{Z}$ rather than only the *active domain* $\mathsf{tem}(\mathcal{A})$; we denote the set of such certain answers *over $\mathcal{A}$ and $\mathbb{Z}$* by $\mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, \varkappa, \mathcal{A})$ or $\mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, \varrho, \mathcal{A})$.

Example 2. Let $\mathcal{O} = \{\bigcirc_{\scriptscriptstyle P} A \sqsubseteq B, \bigcirc_{\scriptscriptstyle P} B \sqsubseteq A\}$, $\mathcal{A} = \{A(a, 0)\}$ and $\varkappa = \bigcirc_{\scriptscriptstyle F}^2 B$. Then $a^{\mathcal{I}} \in B^{\mathcal{I}(2n+1)}$, for any $n \geq 0$ and $\mathcal{I} \models (\mathcal{O}, \mathcal{A})$. So, $\mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, \varkappa, \mathcal{A}) = \{(a, 2n + 1) \mid n \geq 0\}$ but $\mathsf{ans}(\mathcal{O}, \varkappa, \mathcal{A}) = \{(a, 1)\}$ since $\mathsf{tem}(\mathcal{A}) = \{0, 1\}$.

Given an ABox $\mathcal{A}$ and convex $D \supseteq \mathsf{tem}(\mathcal{A})$, denote by $\mathfrak{S}_{\mathcal{A}}^{D}$ the two-sorted structure with object domain $\mathsf{ind}(\mathcal{A})$ and temporal domain D such that $\mathfrak{S}_{\mathcal{A}}^{D} \models A(a, \ell)$ iff $A(a, \ell) \in \mathcal{A}$ and $\mathfrak{S}_{\mathcal{A}}^{D} \models P(a, b, \ell)$ iff $P(a, b, \ell) \in \mathcal{A}$, for concept and role names A and P. Let $\boldsymbol{q} = (\mathcal{O}, \varkappa)$ be a TOMIQ and $\Phi(x, t)$ a constant-free FO-formula whose signature is the concept and role names in $\boldsymbol{q}$, $<$ and $=$. We call $\Phi(x, t)$ an FO($<$)*-rewriting of $\boldsymbol{q}$* if, for any ABox $\mathcal{A}$, $a \in \mathsf{ind}(\mathcal{A})$ and $\ell \in \mathsf{tem}(\mathcal{A})$, we have $(a, \ell) \in \mathsf{ans}(\boldsymbol{q}, \mathcal{A})$ iff $\mathfrak{S}_{\mathcal{A}}^{\mathsf{tem}(\mathcal{A})} \models \Phi(a, \ell)$. If $\Phi(x, t)$ also uses the predicate $\mathrm{PLUS}(k, n_1, n_2)$ (for $k = n_1 + n_2$) then it is called an FO($<, +$)*-rewriting of $\boldsymbol{q}$*. FO-rewritings of TOMIQs $(\mathcal{O}, \varrho)$ are defined analogously. Note that FO-formulas using $<$ and PLUS as built-in predicates can be evaluated by standard relational or stream database management systems, and so are an appropriate target for query rewriting.

Example 3. Consider $\boldsymbol{q} = (\mathcal{O}, A)$, where $\mathcal{O}$ is the same as in Example 2. It is not hard to see that

$$\exists s, n\,[(A(x, s) \wedge (t - s = 2n \geq 0)) \vee (B(x, s) \wedge (t - s = 2n + 1 \geq 0))]$$

is an FO($<, +$)-rewriting of $\boldsymbol{q}$, where $t - s = 2n \geq 0$ stands for $\exists k\,\big(\mathrm{PLUS}(k, n, n) \wedge \mathrm{PLUS}(t, s, k) \wedge (k \geq 0)\big)$ and $t - s = 2n + 1 \geq 0$ is defined similarly. (The constant 0 is obviously definable.) Note that $\boldsymbol{q}$ is not FO($<$)-rewritable since properties such as ‘t is even’ are not definable by FO($<$)-formulas [Libkin, 2004].

Example 4. For a word $\boldsymbol{e} = (e_0, \dots, e_{n-1}) \in \{0, 1\}^n$, take $\mathcal{A}_{\boldsymbol{e}} = \{B_0(a, n)\} \cup \{A_{e_i}(a, i) \mid i < n\}$ and let $\mathcal{O}'$ contain

$$\bigcirc_{\scriptscriptstyle F} B_k \sqcap A_0 \sqsubseteq B_k \quad \text{and} \quad \bigcirc_{\scriptscriptstyle F} B_k \sqcap A_1 \sqsubseteq B_{1-k}, \quad k = 0, 1.$$

One can check that $(a, 0)$ is a certain answer to $(\mathcal{O}', B_0)$ over $\mathcal{A}_{\boldsymbol{e}}$ iff the number of 1s in $\boldsymbol{e}$ is even (*parity*). It follows that $(\mathcal{O}', B_0)$ is not FO-rewritable even using arbitrary numeric predicates [Arora and Barak, 2009].

If we replace $\mathfrak{S}_{\mathcal{A}}^{\mathsf{tem}(\mathcal{A})}$ in the definition above with $\mathfrak{S}_{\mathcal{A}}^{\mathbb{Z}}$, then we call $\Phi(x, t)$ an FO$^{\mathbb{Z}}$($<$)- or FO$^{\mathbb{Z}}$($<, +$)*-rewriting of $\boldsymbol{q}$*. Finding FO$^{\mathbb{Z}}$-rewritings is often a first step in the construction of FO-rewritings, which are usually more involved.

Example 5. Suppose $\mathcal{O} = \{A \sqsubseteq \bigcirc_{\scriptscriptstyle F}^2 A, B \sqsubseteq \bigcirc_{\scriptscriptstyle F}^3 B\}$ and $\varkappa = \Diamond_{\scriptscriptstyle F}(A \sqcap B)$. Then

$$\exists s, u, v, n, m\,[(t < s) \wedge A(x, u) \wedge (s - u = 2n \geq 0) \wedge B(x, v) \wedge (s - v = 3m \geq 0)]$$

is an FO$^{\mathbb{Z}}$($<, +$)-rewriting of $(\mathcal{O}, \varkappa)$ but not an FO($<, +$)-rewriting because s can be outside the active domain $\mathsf{tem}(\mathcal{A})$.

A *temporal ontology-mediated query* (TOMQ) is a pair $(\mathcal{O}, \psi)$, where ψ is an FO-formula built from atoms $\varkappa(x, t)$, $\varrho(x, y, t)$ and $t < t'$, with $\varkappa$ and ϱ being any positive temporal concept and role, x and y *object variables*, and t and t' *temporal variables*. Given an ABox $\mathcal{A}$, $(\mathcal{O}, \psi)$ is evaluated over the two-sorted structure with the object domain $\mathsf{ind}(\mathcal{A})$ and the temporal domain $\mathsf{tem}(\mathcal{A})$, where $\varkappa(a, \ell)$ holds true iff $(a, \ell) \in \mathsf{ans}(\mathcal{O}, \varkappa, \mathcal{A})$, and likewise for $\varrho(a, b, \ell)$. Thus, similarly to the SPARQL 1.1 entailment regime, we interpret the DL and temporal constructs of TOMIQs in arbitrary temporal models (over $\mathbb{Z}$), while the object and temporal variables of TOMQs range over the active domains only.

Example 6. Suppose $\mathcal{O} = \{\mathit{Scheduled} \sqsubseteq \bigcirc_{\scriptscriptstyle F}^{24} \mathit{Scheduled}\}$, $\psi(x, t) = \mathit{Scheduled}(x, t) \wedge \neg \mathit{Serviced}(x, t)$ and $\mathcal{A}$ contains $\mathit{Scheduled}(a, 0)$ as well as entries $\mathit{Serviced}(a, \ell)$ made by a serviceman. Then $(\mathcal{O}, \psi)$ returns all pairs $(a, 24n)$ such that $24n \leq \max \mathcal{A}$ and there is no entry $\mathit{Serviced}(a, 24n)$ in $\mathcal{A}$.

FO-rewritings of TOMQs are defined similarly to TOMIQs. We generalise [Calvanese *et al.*, 2007b]:

Theorem 7. *If all constituent TOMIQs of a TOMQ $\boldsymbol{q}$ are FO($<$)- or FO($<, +$)-rewritable, then $\boldsymbol{q}$ is also FO($<$)- or, respectively, FO($<, +$)-rewritable.*

From now on we only focus on rewritability of TOMIQs.

3 Rewriting *LTL* TOMQs

We begin our study of FO-rewritability by considering ontologies without role names and assuming that all ABoxes contain a single object name, say a. To simplify notation, we will omit a from the ABox assertions, interpretations, certain answers, etc., and write $A(\ell)$ instead of $A(a, \ell)$ and assume that answers to TOMIQs are subsets of $\mathsf{tem}(A)$ rather than $\{a\} \times \mathsf{tem}(\mathcal{A})$, and that FO-rewritings $\Phi(x, t)$ are one-sorted FO-formulas $\varphi(t)$. Ontologies in this restricted language can be regarded as formulas of the *propositional temporal logic LTL* given in clausal normal form (1), and so we denote the corresponding restrictions of *DL-Lite$^{o}_{c}$* by *LTL$^{o}_{c}$*. In this context, positive temporal concepts are simply negation-free *LTL*-formulas.

As shown by Example 4, some *LTL*$^{\bigcirc}_{horn}$ TOMAQs cannot be FO($<, +$)-rewritable. On the other hand, relying on the well-known fact that the semantics of temporal formulas can be encoded by *monadic second-order* (MSO) formulas (built from atoms of the form $A(t)$, $t = t'$ and $t < t'$ using the Booleans, first-order quantifiers $\forall t$, $\exists t$ and second-order quantifiers $\forall A$ and $\exists A$), one can prove the following:

Theorem 8. *All* $\mathit{LTL}^{\Box\bigcirc}_{\mathit{bool}}$ *TOMQs are MSO*$(<)$*-rewritable.*

It follows that, for any $\mathit{LTL}^{\Box\bigcirc}_{\mathit{bool}}$ TOMQ $\boldsymbol{q}$, one can build an NFA accepting an ABox $\mathcal{A}$ and $\ell \in \mathsf{tem}(\mathcal{A})$ written on its tape iff $\ell \in \mathsf{ans}(\boldsymbol{q}, \mathcal{A})$ [Straubing and Weil, 2010], and that the evaluation problem for $\boldsymbol{q}$ is in NC^1 for data complexity [Ladner and Fischer, 1980]. On the other hand, we also have:

Theorem 9. *There exist an* $\mathit{LTL}^{\bigcirc}_{\mathit{horn}}$ *TOMAQ and* $\mathit{LTL}^{\Box}_{\mathit{krom}}$ *and* $\mathit{LTL}^{\bigcirc}_{\mathit{krom}}$ *TOMIQs that are* NC^1*-hard for data complexity*; *in particular, they are not FO-rewritable even with arbitrary numeric predicates.*

Proof sketch. It is known that there exist NC^1-complete regular languages [Barrington *et al.*, 1992]. Given an NFA $\mathfrak{A}$ and input $\boldsymbol{a} = a_0 \dots a_{n-1}$, we take concept names A_a and B_q for tape symbols a and states q of $\mathfrak{A}$, and then set $\mathcal{A}_{\boldsymbol{a}} = \{B_{q_1}(n)\} \cup \{A_{a_i}(i) \mid i < n\}$, where q_1 is the accepting state, and $\mathcal{O} = \{ \bigcirc_{\!\scriptscriptstyle F} B_{q'} \sqcap A_a \sqsubseteq B_q \mid q \to_a q' \}$. Then $\mathfrak{A}$ accepts $\boldsymbol{a}$ iff $0 \in \mathsf{ans}(\mathcal{O}, B_{q_0}, \mathcal{A}_{\boldsymbol{a}})$, for the initial state q_0. Next, let $\mathcal{O}'$ contain $N_q \sqcap B_q \sqsubseteq \bot$ and $\top \sqsubseteq B_q \sqcup N_q$, for each state q, and let $\varkappa = B_{q_0} \sqcup \bigsqcup_{q \to_a q'} \Diamond_{\scriptscriptstyle P} \Diamond_{\scriptscriptstyle F} (\bigcirc_{\!\scriptscriptstyle F} B_{q'} \sqcap A_a \sqcap N_q)$. Then $\mathfrak{A}$ accepts $\boldsymbol{a}$ iff $0 \in \mathsf{ans}(\mathcal{O}', \varkappa, \mathcal{A}_{\boldsymbol{a}})$. ❑

Next, we show the FO$(<)$- and FO$(<,+)$-rewritability results from the table in the introduction. To begin with, we focus on temporal ontology-mediated *atomic* queries (TOMAQs) of the form $(\mathcal{O}, A)$, where A is a concept name.

Theorem 10. *Any* $\mathit{LTL}^{\bigcirc}_{\mathit{krom}}$ *TOMAQ is FO*$(<,+)$*-rewritable.*

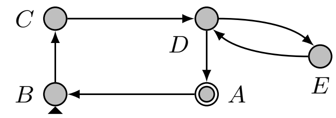
Proof sketch. Suppose $\boldsymbol{q} = (\mathcal{O}, A)$ is an $\mathit{LTL}^{\bigcirc}_{\mathit{krom}}$ TOMAQ. By a *literal*, L, we mean a concept name in $\boldsymbol{q}$ or its negation. We use $\bigcirc^n L$ in place of $\bigcirc_{\!\scriptscriptstyle F}^n L$ if $n > 0$, L if $n = 0$, and $\bigcirc_{\!\scriptscriptstyle P}^{-n} L$ if $n < 0$. We write $\mathcal{O} \models L \sqsubseteq \bigcirc^k L'$ if $\mathcal{I} \models L \sqsubseteq \bigcirc^k L'$ in every model $\mathcal{I}$ of $\mathcal{O}$. For any ABox $\mathcal{A}$ consistent with $\mathcal{O}$, we have:

$$\ell \in \mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, A, \mathcal{A}) \quad \text{iff} \quad \text{either } \mathcal{O} \models \top \sqsubseteq A$$
$$\text{or } \mathcal{O} \models B \sqsubseteq \bigcirc^{\ell - n} A, \text{ for some } B(n) \in \mathcal{A}.$$

Given literals L and L', let $\mathfrak{A}_{L,L'}$ be an NFA whose tape alphabet is $\{0\}$, the states are the literals, with L initial and L' accepting, and whose transitions are of the form $L_1 \to_0 L_2$, for $\mathcal{O} \models L_1 \sqsubseteq \bigcirc L_2$ (without loss of generality we assume that $\mathcal{O}$ does not contain nested $\bigcirc$). It is easy to see that $\mathfrak{A}_{L,L'}$ accepts 0^k ($k > 0$) iff $\mathcal{O} \models L \sqsubseteq \bigcirc^k L'$. By [Chrobak, 1986; To, 2009], there are $N = O(|\mathfrak{A}_{L,L'}|^2)$ arithmetic progressions $a_i + b_i\mathbb{N} = \{a_i + b_i \cdot m \mid m \geq 0\}$, $1 \leq i \leq N$, such that $0 \leq a_i, b_i \leq |\mathfrak{A}_{L,L'}|$ and $\mathfrak{A}_{L,L'}$ accepts 0^k iff $k \in a_i + b_i\mathbb{N}$ for some i, $1 \leq i \leq N$. These progressions give rise to the FO-rewriting we need. To illustrate, suppose $\boldsymbol{q} = (\mathcal{O}, A)$ and

$$\mathcal{O} = \{A \sqsubseteq \bigcirc B,\ B \sqsubseteq \bigcirc C,\ C \sqsubseteq \bigcirc D,\ D \sqsubseteq \bigcirc A,$$
$$D \sqsubseteq \bigcirc E,\ E \sqsubseteq \bigcirc D\}.$$

The NFA $\mathfrak{A}_{B,A}$ (more precisely, the states reachable from B) is shown below, and for $L \in \{A, C, D, E\}$, $\mathfrak{A}_{L,A}$ is the same



NFA but with the initial state L. It is readily seen that $\mathfrak{A}_{B,A}$ accepts 0^k iff $k \in 3 + 2\mathbb{N}$, which can be described by the formula

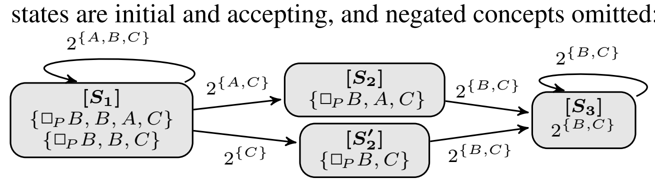
$$\varphi_{B,A}(t) = \exists s, n\, [B(s) \wedge (t - s = a + bn \geq 0)],$$

where with $a = 3$, $b = 2$, $t - s = a + bn \geq 0$ is defined as in Example 3. Similarly, for $\mathfrak{A}_{E,A}$, we have $a = b = 2$. (Note that in general more than one progression is needed to characterise automata $\mathfrak{A}_{L,A}$.) To obtain an FO$(<,+)$- and FO$^{\mathbb{Z}}(<,+)$-rewriting of $\boldsymbol{q}$, we take a disjunction of $\varphi_{L,A}(t)$, for all literals L. ❑

Theorem 11. *Any* $\mathit{LTL}^{\Box}_{\mathit{bool}}$ *TOMAQ is FO*$(<)$*-rewritable.*

Proof sketch. Given an $\mathit{LTL}^{\Box}_{\mathit{bool}}$-ontology $\mathcal{O}$, we construct an NFA $\mathfrak{A}$ that takes as input an ABox $\mathcal{A}$ written as the word $\mathcal{A}_0, \dots, \mathcal{A}_k$, where $k = \max \mathcal{A}$ and $\mathcal{A}_i = \{A \mid A(i) \in \mathcal{A}\}$. Let Σ be the set of temporal concepts in $\mathcal{O}$ and their negations. Each state of $\mathfrak{A}$ is a maximal set $S \subseteq \Sigma$ that is consistent with $\mathcal{O}$; let $\boldsymbol{S}$ be the set of all such states. For $S, S' \in \boldsymbol{S}$ and a tape symbol (set of concept names) X, we set $S \to_X S'$ just in case $X \subseteq S'$, $\Box_{\scriptscriptstyle F}\beta \in S$ iff $\beta, \Box_{\scriptscriptstyle F}\beta \in S'$, and $\Box_{\scriptscriptstyle P}\beta \in S'$ iff $\beta, \Box_{\scriptscriptstyle P}\beta \in S$. A state $S \in \boldsymbol{S}$ is *accepting* if $\mathfrak{A}$ has an infinite 'ascending' chain $S \to_\emptyset S_1 \to_\emptyset \dots$; S is *initial* if $\mathfrak{A}$ has an infinite 'descending' chain $\dots \to_\emptyset S_1 \to_\emptyset S$. The NFA $\mathfrak{A}$ *simulates* $\mathcal{O}$ in the following sense: for any ABox $\mathcal{A}$, concept name A and $\ell \in \mathbb{Z}$, we have $\ell \in \mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, A, \mathcal{A})$ iff $\mathfrak{A}$ does not contain an *accepting path* $S_0 \to_{X_1} \cdots \to_{X_m} S_m$ (S_0 initial and S_m accepting) such that $A \notin S_\ell$, $X_{i+j} = \mathcal{A}_j$ if $0 \leq j \leq k$, and $X_j = \emptyset$ otherwise, for some $0 < i \leq m - k$.

Define an equivalence relation, $\sim$, on $\boldsymbol{S}$ by taking $S \sim S'$ iff $S = S'$ or $\mathfrak{A}$ has a cycle with both S and S'. Let $[S]$ be the $\sim$-equivalence class of S. One can check that $S \to_X S'$ implies $S_1 \to_X S'$, for any $S_1 \in [S]$. Let $\mathfrak{A}'$ be the NFA with states $[S]$, for $S \in \boldsymbol{S}$, and transitions $[S] \to_X [S']$ iff $S_1 \to_X S'_1$, for some $S_1 \in [S]$ and $S'_1 \in [S']$. The initial (accepting) states of $\mathfrak{A}'$ are all $[S]$ with initial (accepting) S. The NFA $\mathfrak{A}'$ also simulates $\mathcal{O}$ and contains no cycles other than trivial loops, which makes it possible to express the simulation condition by an FO$(<)$-formula. For example, $\mathfrak{A}'$ for $\mathcal{O} = \{A \sqsubseteq \Box_{\scriptscriptstyle P} B,\ \Box_{\scriptscriptstyle P} B \sqsubseteq C\}$ is shown below, where all states are initial and accepting, and negated concepts omitted:



Let $\boldsymbol{q} = (\mathcal{O}, C)$. Take all accepting paths π in $\mathfrak{A}'$ with pairwise distinct states at least one of which has a set without C. Thus, for $\pi = [S_1] \to_{\{A\}} [S_2] \to_\emptyset [S_3]$, a set in $[S_3]$ has no C, and the simulation condition for π, which makes sure that $\neg C$ holds at t, can be written as

$$\exists t_1, t_2 \big[\forall t' \big((t' < t_1) \to \mathsf{loop}_{[S_1]}(t')\big) \wedge \mathsf{sym}_{\{A\}}(t_1) \wedge$$
$$\forall t' \, \big((t_1 < t' < t_2) \to \mathsf{loop}_{[S_2]}(t')\big)\big) \wedge \mathsf{sym}_{\emptyset}(t_2) \wedge$$
$$\forall t' \, \big((t' > t_2) \to \mathsf{loop}_{[S_3]}(t')\big) \wedge (t \geq t_2) \wedge \neg C(t)\big],$$

where $\mathsf{sym}_{\{A\}}(t) = A(t) \land \neg B(t) \land \neg C(t)$ and $\mathsf{sym}_\emptyset(t) = \neg A(t) \land \neg B(t) \land \neg C(t)$ define transitions $\to_{\{A\}}$ and $\to_\emptyset$ in π, and $\mathsf{loop}_{[S_1]} = \top$, $\mathsf{loop}_{[S_3]} = \neg A(t)$ and $\mathsf{loop}_{[S_2]} = \bot$ say that $[S_1]$ and $[S_3]$ have loop transitions with any input and any input but A, respectively, but $[S_2]$ has no loop. To obtain an FO($<$)- and FO$^{\mathbb{Z}}$($<$)-rewriting of $\boldsymbol{q}$, we take a disjunction of such formulas for all accepting paths π in $\mathfrak{A}'$ and negate it. ❑

FO-rewritings of $LTL^{\bigcirc}_{core}$ and $LTL^{\Box}_{horn}$ TOMIQs can now be constructed by induction on the structure of concepts using the fact that any consistent $(\mathcal{O}, \mathcal{A})$ in these languages has a single canonical model providing answers to all TOMIQs with the given ontology (it will be formally defined in Section 4). For example, an FO$^{\mathbb{Z}}$($<$)-rewriting of $(\mathcal{O}, A_1 \,\mathcal{U}\, A_2)$ is defined as

$$\exists s \left[\big((s > t) \land \varphi_{A_2}(s)\big) \land \forall u \,\big((t < u < s) \to \varphi_{A_1}(u)\big)\right], \quad (2)$$

where $\varphi_{A_i}(t)$ is an FO$^{\mathbb{Z}}$($<$)-rewriting of $(\mathcal{O}, A_i)$. Note, however, that such FO$^{\mathbb{Z}}$($<$)-rewritings are not in general FO($<$)-rewritings as they may refer to time instants outside the active domain $\mathsf{tem}(\mathcal{A})$; cf. Example 5. To cope with this problem, we require (constant-free) sentences φ^k_A, for any concept A and $k \neq 0$, such that, for every ABox $\mathcal{A}$ consistent with $\mathcal{O}$,

– $\mathfrak{S}^{\mathsf{tem}(\mathcal{A})}_{\mathcal{A}} \models \varphi^k_A$ iff $k + \max \mathcal{A} \in \mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, A, \mathcal{A})$, $k \geq 1$,

– $\mathfrak{S}^{\mathsf{tem}(\mathcal{A})}_{\mathcal{A}} \models \varphi^k_A$ iff $k \in \mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, A, \mathcal{A})$ and $k \leq -1$.

The existence of such sentences, called *witnesses* for $(\mathcal{O}, A)$ and k, follows from the proofs of Theorems 10 and 11. Instead of (2), we can now take the *infinite* 'formula'

$$(2) \ \lor \ \Big[\forall u \,((t < u) \to \varphi_{A_1}(u)) \land \bigvee_{k>0} \big(\varphi^k_{A_2} \land \bigwedge_{0<i<k} \varphi^i_{A_1}\big)\Big].$$

It turns out that we can make it finite by observing that the canonical models are ultimately periodical with at most exponential period (see the full paper). Thus, we obtain

Theorem 12. *(i) Any $LTL^{\bigcirc}_{core}$ TOMQ is FO($<, +$)-rewritable. (ii) Any $LTL^{\Box}_{horn}$ TOMQ is FO($<$)-rewritable.*

4 OBDA with Temporal *DL-Lite*

Now we transfer the rewritability results obtained above to certain $DL\text{-}Lite^{\Box\bigcirc}_{horn}$ TOMIQs. To simplify presentation, we assume that our ontologies do not contain nested temporal operators and $\bot$ (and so they are always consistent with any data); a proof without this assumption is given in the full paper.

Observe first that rewritability of TOMIQs of the form $(\mathcal{O}, \varrho)$ can be easily reduced to rewritability of $LTL^{\Box\bigcirc}_{horn}$ TOMIQs because only the RBox $\mathcal{R}$ in $\mathcal{O}$ has to be taken into account. We assume that with every role inclusion (e.g., $\Box_F R \sqcap \bigcirc_F Q^- \sqsubseteq T$) the RBox also contains the corresponding inclusion for the inverse roles (that is, $\Box_F R^- \sqcap \bigcirc_F Q \sqsubseteq T^-$). We treat the roles in $\mathcal{R}$ and ϱ as *concept names* and denote the resulting $LTL^{\Box\bigcirc}_{horn}$ TOMIQ by $(\mathcal{R}^*, \varrho^*)$. If it has an FO-rewriting $\varphi_{\mathcal{R}^*,\varrho^*}(t)$, then we replace every $P(s)$ in it with $P(x, y, s)$, every $P^-(s)$ with $P(y, x, s)$, and denote the resulting FO-formula by $\Phi_{\mathcal{O},\varrho}(x, y, t)$.

Theorem 13. *For every $DL\text{-}Lite^{\Box\bigcirc}_{horn}$ TOMIQ $(\mathcal{O}, \varrho)$ with $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$, if $\varphi_{\mathcal{R}^*,\varrho^*}(t)$ is an F-rewriting of $(\mathcal{R}^*, \varrho^*)$ then $\Phi_{\mathcal{O},\varrho}(x, y, t)$ is an F-rewriting of $(\mathcal{O}, \varrho)$, where F is FO($<$) or FO($<, +$).*

However, this simple reduction to *LTL* does not work for TOMIQs of the form $(\mathcal{O}, \varkappa)$. The main reason is that the RBox in $\mathcal{O}$ can have a strong impact on its TBox.

Example 14. Let $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$, $\mathcal{T} = \{B \sqsubseteq \exists R, \exists Q \sqsubseteq A\}$ and $\mathcal{R} = \{R \sqsubseteq \bigcirc_F Q\}$. It is not hard to see that the following is an FO($<$)-rewriting of $(\mathcal{O}, A)$:

$$A(x, t) \lor \exists y \, Q(x, y, t) \lor \exists y \, R(x, y, t - 1) \lor B(x, t - 1).$$

It can also be constructed by treating $\mathcal{T}$ as an $LTL^{\bigcirc}_{core}$-ontology, but only if the *connection axiom* $\exists R \sqsubseteq \bigcirc_F \exists Q$ is added to $\mathcal{T}$.

To understand what connection axioms are required for any given $DL\text{-}Lite^{\Box\bigcirc}_{horn}$ TOMIQ $(\mathcal{O}, \varkappa)$, we first define a canonical model (or chase) for $(\mathcal{O}, \mathcal{A})$, which will also be used to prove our main Theorems 17 and 21. Let $\mathcal{C}$ be a set of atoms of the form $A(a, n)$, $\exists R(a, n)$ and $R(a, b, n)$, possibly prefixed by temporal operators and such that $P(a, b, n) \in \mathcal{C}$ iff $P^-(b, a, n) \in \mathcal{C}$. Denote by $\mathsf{cl}(\mathcal{C})$ the result of applying non-recursively the following rules to (the same) $\mathcal{C}$:

– if $R(a, b, n) \in \mathcal{C}$ then we add $\exists R(a, n)$ to $\mathcal{C}$;

– if $(\lambda_1 \sqcap \dots \sqcap \lambda_k \ \sqsubseteq \ \lambda) \in \mathcal{O}$ and $\lambda_i(\boldsymbol{a}, n) \in \mathcal{C}$, for all $i = 1, \dots, k$, then we add $\lambda(\boldsymbol{a}, n)$ to $\mathcal{C}$;

– if $\exists R(a, n) \in \mathcal{C}$ then we add $R(a, aR^n, n)$ to $\mathcal{C}$, where aR^n is a new object name (called a *witness* for $\exists R(a, n)$);

– if $\Box_F \lambda(\boldsymbol{a}, n) \in \mathcal{C}$ then add $\lambda(\boldsymbol{a}, m)$ to $\mathcal{C}$ for all $m > n$;

– if $\lambda(\boldsymbol{a}, m) \in \mathcal{C}$ for all $m > n$, then add $\Box_F \lambda(\boldsymbol{a}, n)$ to $\mathcal{C}$;

– if $\bigcirc_F \lambda(\boldsymbol{a}, n) \in \mathcal{C}$ then add $\lambda(\boldsymbol{a}, n + 1)$ to $\mathcal{C}$;

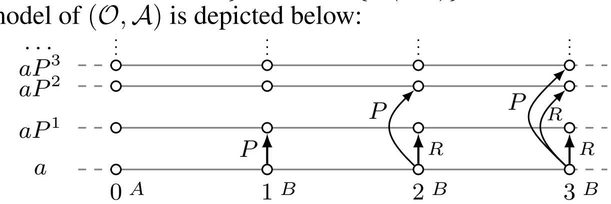
– if $\lambda(\boldsymbol{a}, n) \in \mathcal{C}$ then add $\bigcirc_F \lambda(\boldsymbol{a}, n - 1)$ to $\mathcal{C}$;

and symmetrical rules for $\Box_P$ and $\bigcirc_P$. Then we set $\mathsf{cl}^0(\mathcal{C}) = \mathcal{C}$ and, for any successor ordinal $\xi + 1$ and limit ordinal ζ,

$$\mathsf{cl}^{\xi+1}(\mathcal{C}) = \mathsf{cl}(\mathsf{cl}^{\xi}(\mathcal{C})) \quad \text{and} \quad \mathsf{cl}^{\zeta}(\mathcal{C}) = \bigcup\nolimits_{\xi<\zeta} \mathsf{cl}^{\xi}(\mathcal{C}).$$

Let N be the number of temporal operators in $\mathcal{O}$. We regard $\mathcal{C}_{\mathcal{O},\mathcal{A}} = \mathsf{cl}^{\omega \cdot N}(\mathcal{A})$ as an interpretation whose domain $\Delta^{\mathcal{C}_{\mathcal{O},\mathcal{A}}}$ comprises $\mathsf{ind}(\mathcal{A})$ and the witnesses aR^n and the interpretation function is defined by taking $a^{\mathcal{C}_{\mathcal{O},\mathcal{A}}} = a$, for $a \in \mathsf{ind}(\mathcal{A})$, and $\boldsymbol{a} \in S^{\mathcal{C}_{\mathcal{O},\mathcal{A}}(\ell)}$ iff $S(\boldsymbol{a}, \ell) \in \mathcal{C}_{\mathcal{O},\mathcal{A}}$, for concept or role names S. We call $\mathcal{C}_{\mathcal{O},\mathcal{A}}$ the *canonical model* of $(\mathcal{O}, \mathcal{A})$.

Example 15. Suppose that $\mathcal{O} = \{A \sqsubseteq \Box_F \exists P,\ \Box_F \exists R \sqsubseteq B,\ P \sqsubseteq \bigcirc_F R,\ R \sqsubseteq \bigcirc_F R\}$ and $\mathcal{A} = \{A(a, 0)\}$. The canonical model of $(\mathcal{O}, \mathcal{A})$ is depicted below:



Note that its construction requires $\omega + 1$ applications of cl.

Theorem 16. *For any $DL\text{-}Lite^{\Box\bigcirc}_{horn}$ TOMIQs $(\mathcal{O}, \varkappa)$ and $(\mathcal{O}, \varrho)$, any ABox $\mathcal{A}$, $a, b \in \mathsf{ind}(\mathcal{A})$ and $\ell \in \mathbb{Z}$, we have*

$$(a, \ell) \in \mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, \varkappa, \mathcal{A}) \quad \textit{iff} \quad a \in \varkappa^{\mathcal{C}_{\mathcal{O},\mathcal{A}}(\ell)},$$
$$(a, b, \ell) \in \mathsf{ans}^{\mathbb{Z}}(\mathcal{O}, \varrho, \mathcal{A}) \quad \textit{iff} \quad (a, b) \in \varrho^{\mathcal{C}_{\mathcal{O},\mathcal{A}}(\ell)}.$$

Let $(\mathcal{O}, \varkappa)$ be a *DL-Lite*$^{\Box\bigcirc}_{horn}$ TOMIQ with $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$. For any role R in $\mathcal{O}$, we consider the canonical model $\mathcal{C}_{\mathcal{R},\{R(a,b,0)\}}$ and denote by $\boldsymbol{r}_n^R$, $n \in \mathbb{Z}$, the set of roles Q in $\mathcal{O}$ such that $Q(a,b,n) \in \mathcal{C}_{\mathcal{R},\{R(a,b,0)\}}$. It is known from temporal logic [Gabbay *et al.*, 1994] that there are positive numbers $n_r, n_l, k_r, k_l = O(2^{|\mathcal{R}|})$ such that

$$\boldsymbol{r}_i^R = \boldsymbol{r}_{i+k_r}^R \text{ for any } i \geq n_r, \quad \boldsymbol{r}_i^R = \boldsymbol{r}_{i-k_l}^R \text{ for any } i \leq -n_l.$$

We take, for each R and $0 \leq i \leq n_r + k_r$, fresh concept names D_i and *add to* $\mathcal{T}$ the concept inclusions (cf. Example 14):

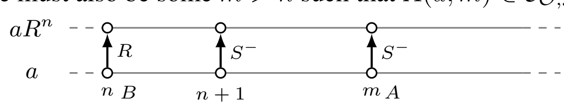
(con) $\exists R \sqsubseteq D_0$, $D_{i-1} \sqsubseteq \bigcirc_F D_i$ (if $i > 0$), $D_{n_r+k_r} \sqsubseteq D_{n_r}$ and $D_i \sqsubseteq \exists Q$, for $0 \leq i \leq n_r + k_r$ and each $Q \in \boldsymbol{r}_i^R$.

and symmetrical inclusions for $-n_l - k_l \leq i \leq 0$. Denote by $\mathcal{T}^*$ the *LTL*$^{\Box\bigcirc}_{horn}$ TBox obtained from $\mathcal{T}$ by replacing the *basic* concepts in it with *concept names*.

Theorem 17. *Suppose $(\mathcal{O}, \varkappa)$ is a DL-Lite$^{\Box\bigcirc}_{horn}$ TOMIQ and $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$ with $\mathcal{T}$ supplemented by **(con)**. If there exist F-rewritings of $(\mathcal{T}^*, A)$ and $(\mathcal{R}^*, P)$, for every concept name A and role name P in $\mathcal{O}$, then $(\mathcal{O}, \varkappa)$ is F-rewritable, where F is FO$(<)$ or FO$(<,+)$*

We illustrate the rather involved proof of this theorem (given in the full paper) by an example.

Example 18. Let $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$, where $\mathcal{T} = \{B \sqsubseteq \exists R\}$ and $\mathcal{R} = \{R \sqsubseteq \Box_F S^-\}$, and $\boldsymbol{q} = (\mathcal{O}, \varkappa)$ with $\varkappa = \exists R. \Diamond_F \exists S.A$. By Theorem 16, certain answers to $\boldsymbol{q}$ over an ABox $\mathcal{A}$ are all $a \in \mathsf{ind}(\mathcal{A})$ and $n \in \mathsf{tem}(\mathcal{A})$ with $a \in \varkappa^{\mathcal{C}_{\mathcal{O},\mathcal{A}}(n)}$. Witnesses for $\exists R$ in $\varkappa$ are either located in $\mathsf{ind}(\mathcal{A})$ or of the form aR^n if $\mathcal{C}_{\mathcal{O},\mathcal{A}}$ contains $B(a,n)$, and so $R(a, aR^n, n)$, in which case there must also be some $m > n$ such that $A(a,m) \in \mathcal{C}_{\mathcal{O},\mathcal{A}}$:



This observation gives the following inductive rewriting:

$$\begin{aligned}
\Phi_{\mathcal{O},\varkappa}(x,t) = {} & \exists y\,[R(x,y,t) \wedge \Phi_{\mathcal{O},\Diamond_F \exists S.A}(y,t)] \quad \vee \\
& [B(x,t) \wedge \exists t'\,((t' > t) \wedge A(x,t'))], \\
\Phi_{\mathcal{O},\Diamond_F \exists S.A}(x,t) = {} & \exists t'\,[(t' > t) \wedge \Phi_{\exists S.A}(x,t')], \\
\Phi_{\mathcal{O},\exists S.A}(x,t) = {} & \exists y\,[A(y,t) \wedge (S(x,y,t) \vee \\
& \exists t'\,((t' < t) \wedge R(y,x,t')))].
\end{aligned}$$

Since the connection axioms **(con)** belong to *DL-Lite*$^{\bigcirc}_{core}$, as a consequence of Theorems 12 (*i*), 13 and 17 we obtain:

Corollary 19. *DL-Lite$^{\bigcirc}_{core}$ TOMIQs are FO$(<,+)$-rewritable.*

On the other hand, the following unexpected result shows, in particular, that **(con)** cannot be expressed in *DL-Lite*$^{\Box}_{horn}$:

Theorem 20. *There is a DL-Lite$^{\Box}_{horn}$ TOMAQ which is* NC^1-*hard for data complexity*; *in particular, it is not FO-rewritable even using arbitrary numeric predicates.*

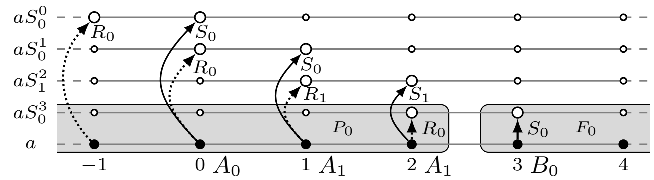
Proof sketch. We first show non-rewritability by modifying the construction of Example 4. Consider the RBox $\mathcal{R}$ with

$$S_k \sqsubseteq F_k,\ S_k \sqsubseteq \Box_F F_k,\ S_k \sqsubseteq \Box_P P_k,\ \Box_F F_k \sqcap P_k \sqsubseteq R_k,$$

and the TBox $\mathcal{T}$ with

$$\exists R_k \sqcap A_0 \sqsubseteq \exists S_k, \quad \exists R_k \sqcap A_1 \sqsubseteq \exists S_{1-k}, \quad B_0 \equiv \exists S_0,$$

where $k = 0, 1$. For $\boldsymbol{e} = (e_0, \dots, e_{n-1}) \in \{0,1\}^n$, we take $\mathcal{A}_{\boldsymbol{e}} = \{B_0(a,n)\} \cup \{\, A_{e_i}(a,i) \mid i < n \,\}$. Then $(a,0)$ is a certain answer to $(\mathcal{T} \cup \mathcal{R}, B_0)$ over $\mathcal{A}_{\boldsymbol{e}}$ iff the number of 1s in $\boldsymbol{e}$ is even—see the picture below and note that $\exists S_k(a,n)$ always generates a *fresh* witness aS_k^n:



The same idea can be used to simulate arbitrary NFAs and show NC^1-hardness of *DL-Lite*$^{\Box}_{horn}$ TOMAQs; see Thm. 9. ❑

Thus, Theorem 12 (*ii*) cannot be lifted to *DL-Lite*$^{\Box}_{horn}$. In the proof above, using *Horn role inclusions* with $\Box_F$ and $\Box_P$, we have encoded $\bigcirc_F$ on concepts: $\mathcal{T} \cup \mathcal{R} \models \bigcirc_F \exists S_k \sqsubseteq \exists R_k$.

We now give a sufficient condition under which **(con)** can be expressed in *DL-Lite*$^{\Box}_{core}$. Call an RBox $\mathcal{R}$ *monotone* if, for any roles R, Q in it, $Q \in \boldsymbol{r}_n^R$ and $n \neq 0$ imply $Q \in \boldsymbol{r}_m^R$ for all $m \geq n$ or all $m \leq n$. Clearly, for a monotone $\mathcal{R}$ and any roles R and Q in it, one of the four options holds:

(*i*) $Q \in \boldsymbol{r}_n^R$ iff $n = 0$,

(*ii*) there is $m_r \in \mathbb{Z}$ such that $Q \in \boldsymbol{r}_n^R$ iff $m_r \leq n$ or $n = 0$,

(*iii*) there is $m_l \in \mathbb{Z}$ such that $Q \in \boldsymbol{r}_n^R$ iff $n \leq m_l$ or $n = 0$,

(*iv*) there are $m_l, m_r \in \mathbb{Z}$, $m_l \leq m_r$, such that $Q \in \boldsymbol{r}_n^R$ iff $n \leq m_l$ or $m_r \leq n$ or $n = 0$.

We encode (*i*) by $\exists R \sqsubseteq \exists Q$; (*ii*) by the following:

$$\begin{array}{ll}
\exists R \sqsubseteq \exists Q, & \text{if } \mathcal{R} \models R \sqsubseteq Q, \\
\exists R \sqsubseteq \Box_F^{m_r} \exists Q, & \text{if } m_r > 0, \\
\exists R \sqsubseteq \Box_F D, \quad \Box_F^{-m_r+1} D \sqsubseteq \exists Q, & \text{if } m_r \leq 0,
\end{array}$$

with fresh D; (*iii*) is symmetrical; (*iv*) combines (*ii*) and (*iii*). The set of all such connection axioms is denoted by **(con$'$)**.

Theorem 21. *Theorem 17 holds true for $\mathcal{O} = \mathcal{T} \cup \mathcal{R}$ with monotone $\mathcal{R}$ and **(con$'$)** in place of **(con)**.*

One can show that all *DL-Lite*$^{\Box}_{core}$ RBoxes as well as *DL-Lite*$^{\Box}_{horn}$ RBoxes without $\Box$-operators on the left-hand side of role inclusions are monotone. We denote by *DL-Lite*$^{mon\Box}_{horn}$ the fragment of *DL-Lite*$^{\Box}_{horn}$ whose role inclusions do not contain negative occurrences of $\Box_F$ and $\Box_P$; this language can be regarded as an extension of *TQL* [Artale *et al.*, 2013b].

Corollary 22. *All DL-Lite$^{\Box}_{core}$ as well as DL-Lite$^{mon\Box}_{horn}$ TOMIQs are FO$(<)$-rewritable.*

5 Conclusions

We have developed a two-step approach to analysing FO-rewritability of temporal ontology-mediated queries. First, we classified the FO-rewritability properties of TOMQs in fragments of *LTL*. Second, we proved two general transfer theorems identifying conditions under which FO-rewritability is preserved for combinations of *LTL* with *DL-Lite*. The transfer results show that, although temporal DLs are notorious for their high complexity, one can nevertheless find large 'islands' of tractable and expressive TOMQs. Many interesting and challenging research questions can be tackled based on the results of this paper:

– complexity and succinctness of rewritings (the size of our rewritings of TOMAQs varies from polynomial for $LTL_{krom}^{\bigcirc}$ to single, double and triple exponential in the size $|\mathcal{T}|$ of the TBox $\mathcal{T}$ for various temporalised versions of *DL-Lite*; rewritings of TOMQs are of size $S^{|\boldsymbol{q}|}$ and $S^{|\boldsymbol{q}||\mathcal{T}|}$ for, respectively, *LTL* and *DL-Lite* ontologies, where $|\boldsymbol{q}|$ is the size of the query and S the size of the rewritings for the underlying TOMAQs);

– generalising our FO rewritings of TOMIQs to (two-sorted) CQs using the methods of [Artale *et al.*, 2013b];

– data complexity of evaluating $\textit{DL-Lite}_{horn}^{\Box\bigcirc}$ TOMQs; based on our proofs, we conjecture that it is NC^1-complete;

– extension of the transfer results for Horn logics to $\sqcup$ and $\neg$, possibly via a non-uniform analysis;

– our rewriting algorithms need to be optimised and evaluated in real-world applications.

References

[Arora and Barak, 2009] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, USA, 1st edition, 2009.

[Artale *et al.*, 2009] A. Artale, D. Calvanese, R. Kontchakov, and M. Zakharyaschev. The DL-Lite family and relations. *J. Artif. Intell. Res. (JAIR)*, 36:1–69, 2009.

[Artale *et al.*, 2013a] A. Artale, R. Kontchakov, V. Ryzhikov, and M. Zakharyaschev. The complexity of clausal fragments of LTL. In *Proc. of the 19th Int. Conf. on Logic for Programming, Artificial Intelligence and Reasoning (LPAR)*, vol. 8312 of *LNCS*, pages 35–52. Springer, 2013.

[Artale *et al.*, 2013b] A. Artale, R. Kontchakov, F. Wolter, and M. Zakharyaschev. Temporal description logic for ontology-based data access. In *Proc. of the 23rd Int. Joint Conf. on Artificial Intelligence (IJCAI)*. IJCAI/AAAI, 2013.

[Baader *et al.*, 2013] F. Baader, S. Borgwardt, and M. Lippmann. Temporalizing ontology-based data access. In *Proc. of the 24th Int. Conf. on Automated Deduction (CADE)*, volume 7898 of *LNCS*, pages 330–344. Springer, 2013.

[Barrington *et al.*, 1992] D. Barrington, K. Compton, H. Straubing, and D. Thérien. Regular languages in NC^1. *J. Comput. Syst. Sci.*, 44(3):478–499, 1992.

[Borgwardt *et al.*, 2013] S. Borgwardt, M. Lippmann, and V. Thost. Temporal query answering in the description logic DL-Lite. In *Proc. of the 9th Int. Symp. on Frontiers of Combining Systems (FroCoS)*, volume 8152 of *LNCS*, pages 165–180. Springer, 2013.

[Calbimonte *et al.*, 2012] J.-P. Calbimonte, H. Jeung, Ó. Corcho, and K. Aberer. Enabling query technologies for the semantic sensor web. *Int. J. Semantic Web Inf. Syst.*, 8(1):43–63, 2012.

[Calvanese *et al.*, 2007a] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. Automat. Reasoning*, 39(3):385–429, 2007.

[Calvanese *et al.*, 2007b] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. EQL-Lite: Effective first-order query processing in description logics. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pages 274–279, 2007.

[Chrobak, 1986] M. Chrobak. Finite automata and unary languages. *Theor. Comp. Sci.*, 47(2):149–158, 1986.

[Fisher *et al.*, 2001] M. Fisher, C. Dixon, and M. Peim. Clausal temporal resolution. *ACM Trans. Comput. Logic*, 2(1):12–56, 2001.

[Gabbay *et al.*, 1994] D. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal Logic: Mathematical Foundations and Computational Aspects*, 1. Oxford University Press, 1994.

[Gutiérrez-Basulto and Klarman, 2012] V. Gutiérrez-Basulto and S. Klarman. Towards a unifying approach to representing and querying temporal data in description logics. In *Proc. of the 6th Int. Conf. on Web Reasoning and Rule Systems (RR)*, vol. 7497 of *LNCS*, pages 90–105. Springer, 2012.

[Gutierrez *et al.*, 2007] C. Gutierrez, C. A. Hurtado, and A. A. Vaisman. Introducing time into RDF. *IEEE Trans. Knowl. Data Eng.*, 19(2):207–218, 2007.

[Kharlamov *et al.*, 2014] E. Kharlamov, N. Solomakhina, Ö. Özçep, D. Zheleznyakov, T. Hubauer, S. Lamparter, M. Roshchin, A. Soylu, and S. Watson. How semantic technologies can enhance data access at Siemens Energy. In *Proc. of the 13th Int. Semantic Web Conf. (ISWC), Part I*, volume 8796 of *LNCS*, pages 601–619. Springer, 2014.

[Klarman and Meyer, 2014] S. Klarman and T. Meyer. Querying temporal databases via OWL 2 QL. In *Proc. of the 8th Int. Conf. on Web Reasoning and Rule Systems (RR)*, vol. 8741 of *LNCS*, pages 92–107. Springer, 2014.

[Kontchakov *et al.*, 2014] R. Kontchakov, M. Rezk, M. Rodriguez-Muro, G. Xiao, and M. Zakharyaschev. Answering SPARQL queries over databases under OWL 2 QL entailment regime. In *Proc. of the 13th Int. Semantic Web Conf. (ISWC), Part I*, volume 8796 of *LNCS*, pages 552–567. Springer, 2014.

[Ladner and Fischer, 1980] R. E. Ladner and M. J. Fischer. Parallel prefix computation. *J. ACM*, 27(4):831–838, 1980.

[Libkin, 2004] L. Libkin. *Elements Of Finite Model Theory*. Springer, 2004.

[Lutz *et al.*, 2008] C. Lutz, F. Wolter, and M. Zakharyaschev. Temporal description logics: A survey. In *Proc. of the 15th Int. Symp. on Temporal Representation and Reasoning (TIME)*, pages 3–14. IEEE Computer Society, 2008.

[Motik, 2012] B. Motik. Representing and querying validity time in RDF and OWL: A logic-based approach. *J. Web Sem.*, 12:3–21, 2012.

[Özçep and Möller, 2014] Ö. Özçep and R. Möller. Ontology based data access on temporal and streaming data. In *Proc. of the 10th Int. Summer School on Reasoning Web (RW)*, vol. 8714 of *LNCS*, pages 279–312. Springer, 2014.

[Özcep *et al.*, 2013] Ö. Özcep, R. Möller, C. Neuenstadt, D. Zheleznyakov, and E. Kharlamov. A semantics for temporal and stream-based query answering in an OBDA context. Tech. report, Del. D5.1, FP7-318338, EU, 2013.

[Straubing and Weil, 2010] H. Straubing and P. Weil. An introduction to finite automata and their connection to logic. *CoRR*, abs/1011.6491, 2010.

[To, 2009] A. W. To. Unary finite automata vs. arithmetic progressions. *Inf. Proc. Lett.*, 109(17):1010–1014, 2009.

[Zhang *et al.*, 2012] Y. Zhang, M.-D. Pham, Ó. Corcho, and J.-P. Calbimonte. SRBench: A streaming RDF/SPARQL benchmark. In *Proc. of the 11th Int. Semantic Web Conf. (ISWC), Part I*, volume 7649 of *LNCS*, pages 641–657. Springer, 2012.