

Ontology-mediated query answering over temporal data: A Survey

(arXiv.org: <https://arxiv.org/abs/2004.07221>)

Alessandro Artale

*KRDB Research Centre – Faculty of Computer Science
Free University of Bozen-Bolzano, Italy*



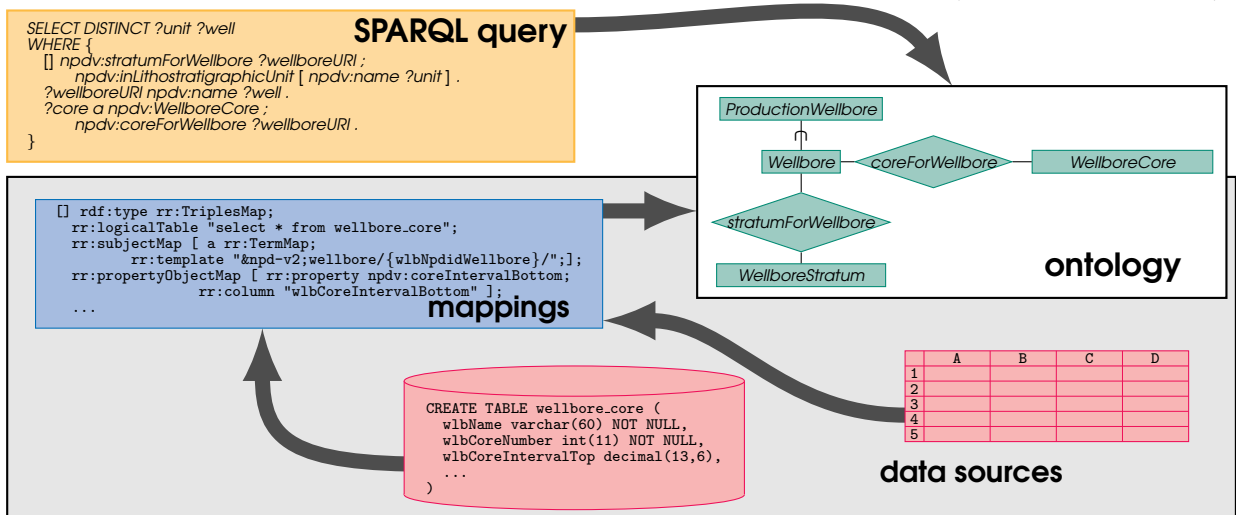
Joint work with

R. Kontchakov, A. Kovtunova, V. Ryzhikov, F. Wolter, M. Zakharyashev

KRDB Summer Online Seminars 2020 - 8 May

Ontology-based data access (OBDA)

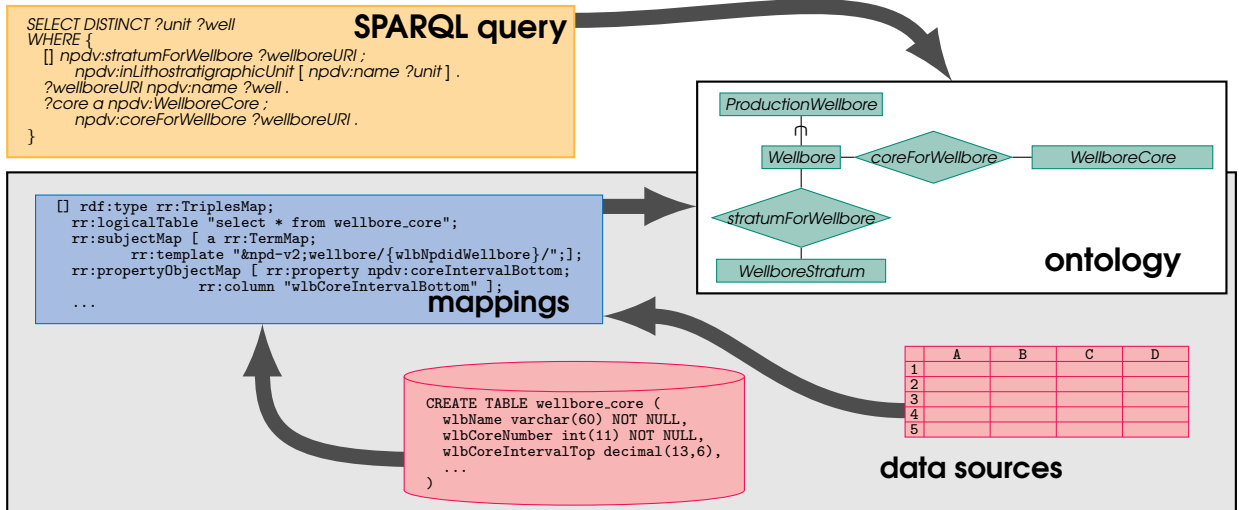
(the Romans \approx 2007)



- Ontology**
- gives a high-level conceptual view of the data
 - provides a convenient & natural vocabulary for user queries
 - facilitates data integration via a global schema

Ontology-based data access (OBDA)

(the Romans \approx 2007)



- Ontology**
- gives a high-level conceptual view of the data
 - provides a convenient & natural vocabulary for user queries
 - facilitates data integration via a global schema

OWL 2 QL ontology-mediated queries ($\mathcal{O}, q(\vec{x})$) are FO-rewritable

reduction to DB query evaluation

$$\exists q' \forall \mathcal{D}, \vec{a} \quad \mathcal{O}, \mathcal{D} \models q(\vec{a}) \iff \mathcal{D} \models q'(\vec{a})$$

Research problems

- 1 Which temporal logics and query languages are suitable for temporal OBDA as far as their expressive power is concerned?
- 2 Investigate expressivity and computational properties of the temporal Ontology languages (Satisfiability, Logical Implication, etc.)
- 3 Classify temporal languages by the type of rewritability (FO, datalog, etc.) and data complexity of OMQ answering
- 4 Devise algorithms, implementations, experiments, use cases
- 5 Efficiency and scalability of implementations
- 6 ...

LTL Knowledge Bases

timeline

$(\mathbb{Z}, <)$

Basic Temporal Concepts

$C ::= \perp \mid \top \mid A_i \mid \circ_F C \mid \circ_P C \mid \square_F C \mid \square_P C$

where A_i are Atomic Concepts

TBox Axioms

$C_1 \sqcap \dots \sqcap C_k \sqsubseteq C_{n+1} \sqcup \dots \sqcup C_{n+m}$

bool any k and m

krom $k + m \leq 2$

horn $m \leq 1$

core $k + m \leq 2$ and $m \leq 1$

Fragments

LTL_c^o $c \in \{\text{bool}, \text{horn}, \text{krom}, \text{core}\}$ and $o \in \{\square, \circ, \square\circ\}$

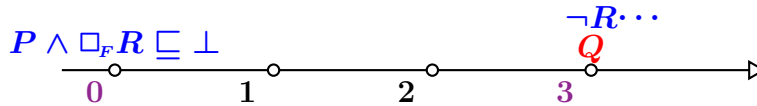
Expressivity of LTL_c^o

$$P \sqsubseteq \diamond_F Q$$

can be expressed in LTL_{krom}^\square as

$$P \sqcap \square_F R \sqsubseteq \perp, \neg R \sqsubseteq Q$$

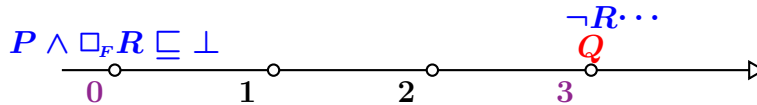
for a fresh r



but cannot be expressed in $LTL_{horn}^{\square\circ}$

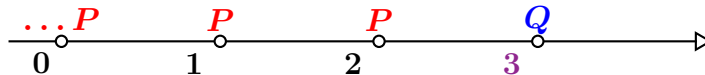
Expressivity of LTL_c^o

$P \sqsubseteq \diamond_F Q$ can be expressed in LTL_{krom}^\square as $P \sqcap \square_F R \sqsubseteq \perp, \neg R \sqsubseteq Q$
 for a fresh r



but **cannot be expressed in $LTL_{horn}^{\square\circ}$**

$\diamond_F Q \sqsubseteq P$ can be expressed in LTL_{core}^\square as $Q \sqsubseteq \square_P P$



Expressivity of LTL_c^o (cont.)

$A \sqsubseteq BUC$ can be expressed in $LTL_{bool}^{\square\circ}$ using the unfolding $BUC \equiv \circ_F C \sqcup (\circ_F B \sqcap \circ_F (BUC))$, which gives rise to four clauses:

$$A \sqsubseteq U, \quad U \sqsubseteq \circ_F C \sqcup \circ_F B, \quad U \sqsubseteq \circ_F C \sqcup \circ_F U, \quad A \sqsubseteq \diamond_F C$$

Expressivity of LTL_c^o (cont.)

$A \sqsubseteq BUC$ can be expressed in $LTL_{bool}^{\square\circ}$ using the unfolding $BUC \equiv \circ_F C \sqcup (\circ_F B \sqcap \circ_F (BUC))$, which gives rise to four clauses:

$$A \sqsubseteq U, \quad U \sqsubseteq \circ_F C \sqcup \circ_F B, \quad U \sqsubseteq \circ_F C \sqcup \circ_F U, \quad A \sqsubseteq \diamond_F C$$

$BUC \sqsubseteq A$ can be expressed in LTL_{horn}° with the following clauses:

$$\circ_F C \sqsubseteq U, \quad \circ_F U \sqcap \circ_F B \sqsubseteq U, \quad U \sqsubseteq A$$

LTL Knowledge Bases

LTL_c^o ontology \mathcal{O} a set of LTL_c^o clauses $c \in \{bool, horn, krom, core\}$, $o \in \{\square, \bigcirc, \square\bigcirc\}$

Data instance \mathcal{D} is a finite set of atoms of the form $A(\ell)$, $\ell \in \mathbb{Z}$

Active Domain of a data instance \mathcal{D} : $\text{tem}(\mathcal{D}) = \{n \in \mathbb{Z} \mid \min \mathcal{D} \leq n \leq \max \mathcal{D}\}$

LTL Knowledge Bases

LTL_c^o ontology \mathcal{O} a set of LTL_c^o clauses $c \in \{bool, horn, krom, core\}$, $o \in \{\square, \bigcirc, \square\bigcirc\}$

Data instance \mathcal{D} is a finite set of atoms of the form $A(\ell)$, $\ell \in \mathbb{Z}$

Active Domain of a data instance \mathcal{D} : $tem(\mathcal{D}) = \{n \in \mathbb{Z} \mid \min \mathcal{D} \leq n \leq \max \mathcal{D}\}$

Example

Ontology

SevereSnow \sqcap LowTemp \sqcap StrongWind \sqsubseteq BlizzardConditions

BlizzardConditions $\sqcap \bigcirc_F$ BlizzardConditions $\sqcap \bigcirc_F \bigcirc_F$ BlizzardConditions \sqsubseteq Blizzard

Data instance

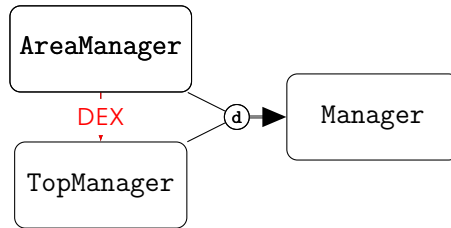
LowTemp(11am19/12/16)

SevereSnow(11am19/12/16)

StrongWind(11am19/12/16)

LTL in Ontologies

Consider the following **Ontology**



It can be captured by the following set of LTL_{krom}^{\square} clauses

$\text{AreaManager} \sqsubseteq \text{Manager}$

$\text{TopManager} \sqsubseteq \text{Manager}$

$\text{AreaManager} \sqsubseteq \neg \text{TopManager}$

$\text{AreaManager} \sqsubseteq \diamond_F \text{TopManager}$

$\text{Manager} \sqsubseteq \text{TopManager} \sqcup \text{AreaManager}$ (Covering, LTL_{bool}^{\square})

Complexity Results for Satisfiability Problem of *LTL*

temp. ops. α	$\Box, \Box_F, \Box_P, \bigcirc_F, \bigcirc_P$ $LTL_{\alpha}^{\Box\bigcirc}$	\Box, \Box_F, \Box_P LTL_{α}^{\Box}	$\Box, \bigcirc_F, \bigcirc_P$ LTL_{α}^{\bigcirc}	\Box LTL_{α}^{\Box}
<i>bool</i>	PSPACE (SistlaClarke82)	NP (OnoNakamura80)	PSPACE	NP
<i>horn</i>	PSPACE (ChenLin93)	PTIME	PSPACE	PTIME
<i>krom</i>	NP	NP	NLOGSPACE	NLOGSPACE
<i>core</i>	NP	NLOGSPACE	NLOGSPACE	NLOGSPACE

Artale, Kontchakov, Ryzhikov, Zakharyashev: The Complexity of Clausal Fragments of LTL. LPAR 2013.

Ontology-mediated queries in *LTL*

*LTL*_c^o *ontology-mediated instance query* (OMIQ) is a pair $q = (\mathcal{O}, \mathcal{K})$
where \mathcal{O} is an *LTL*_c^o ontology and

$\mathcal{K} ::= A \mid \neg \mathcal{K} \mid \mathcal{K}_1 \wedge \mathcal{K}_2 \mid \mathcal{K}_1 \vee \mathcal{K}_2 \mid op_1 \mathcal{K} \mid \mathcal{K}_1 op_2 \mathcal{K}_2$

$op_1 \in \{\circ_F, \diamond_F, \square_F, \circ_P, \diamond_P, \square_P\}$ and $op_2 \in \{\mathcal{U}, \mathcal{S}\}$

A is an Atomic Concept appearing in the ontology \mathcal{O}

Ontology-mediated queries in *LTL*

*LTL*_c^o *ontology-mediated instance query* (OMIQ) is a pair $q = (\mathcal{O}, \mathcal{K})$
where \mathcal{O} is an *LTL*_c^o ontology and

$\mathcal{K} ::= A \mid \neg \mathcal{K} \mid \mathcal{K}_1 \wedge \mathcal{K}_2 \mid \mathcal{K}_1 \vee \mathcal{K}_2 \mid op_1 \mathcal{K} \mid \mathcal{K}_1 op_2 \mathcal{K}_2$

$op_1 \in \{\circ_F, \diamond_F, \square_F, \circ_P, \diamond_P, \square_P\}$ and $op_2 \in \{\mathcal{U}, \mathcal{S}\}$

A is an Atomic Concept appearing in the ontology \mathcal{O}

If \mathcal{K} is an atomic concept, $q = (\mathcal{O}, \mathcal{K})$ is an *LTL*_c^o *ontology-mediated atomic query* (OMAQ).

Ontology-mediated queries in *LTL*

*LTL*_c^o *ontology-mediated instance query* (OMIQ) is a pair $q = (\mathcal{O}, \varkappa)$ where \mathcal{O} is an *LTL*_c^o ontology and

$\varkappa ::= A \mid \neg \varkappa \mid \varkappa_1 \wedge \varkappa_2 \mid \varkappa_1 \vee \varkappa_2 \mid op_1 \varkappa \mid \varkappa_1 op_2 \varkappa_2$

$op_1 \in \{\circ_F, \diamond_F, \square_F, \circ_P, \diamond_P, \square_P\}$ and $op_2 \in \{\mathcal{U}, \mathcal{S}\}$

A is an Atomic Concept appearing in the ontology \mathcal{O}

If \varkappa is an atomic concept, $q = (\mathcal{O}, \varkappa)$ is an *LTL*_c^o *ontology-mediated atomic query* (OMAQ).

A *certain answer* to q over a data instance \mathcal{D} , $\text{ans}(q, \mathcal{D})$, is any $\ell \in \text{tem}(\mathcal{D})$ such that, for every model \mathfrak{M}

$$\mathfrak{M} \models (\mathcal{O}, \mathcal{D}) \longrightarrow \mathfrak{M}, \ell \models \varkappa$$

FO-rewriting

An **FO($\langle, +$)-rewriting of q** is an FO($\langle, +$)-formula $q'(t)$ s.t.

$$\ell \in \text{ans}(q, \mathcal{D}) \iff \mathfrak{S}_{\mathcal{D}} \models q'(\ell) \quad \text{for all } \mathcal{D}, \ell \in \text{tem}(\mathcal{D})$$

$$\mathfrak{S}_{\mathcal{D}} = (\mathcal{D}, \langle, \text{tem}(\mathcal{D})), \text{ with } \mathfrak{S}_{\mathcal{D}} \models A(\ell) \quad \text{iff } A(\ell) \in \mathcal{D}$$

FO-rewriting

An **FO($\langle, +$)-rewriting of q** is an FO($\langle, +$)-formula $q'(t)$ s.t.

$$\ell \in \text{ans}(q, \mathcal{D}) \iff \mathfrak{S}_{\mathcal{D}} \models q'(\ell) \quad \text{for all } \mathcal{D}, \ell \in \text{tem}(\mathcal{D})$$

$$\mathfrak{S}_{\mathcal{D}} = (\mathcal{D}, \langle, \text{tem}(\mathcal{D})), \text{ with } \mathfrak{S}_{\mathcal{D}} \models A(\ell) \quad \text{iff } A(\ell) \in \mathcal{D}$$

If $q'(t)$ is an FO(\langle)-formula, then it is an **FO(\langle)-rewriting** of q

FO-rewriting

An **FO($\langle, +$)-rewriting of q** is an FO($\langle, +$)-formula $q'(t)$ s.t.

$$\ell \in \text{ans}(q, \mathcal{D}) \iff \mathfrak{S}_{\mathcal{D}} \models q'(\ell) \quad \text{for all } \mathcal{D}, \ell \in \text{tem}(\mathcal{D})$$

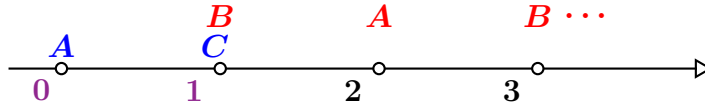
$$\mathfrak{S}_{\mathcal{D}} = (\mathcal{D}, \langle, \text{tem}(\mathcal{D})), \text{ with } \mathfrak{S}_{\mathcal{D}} \models A(\ell) \quad \text{iff } A(\ell) \in \mathcal{D}$$

If $q'(t)$ is an FO(\langle)-formula, then it is an **FO(\langle)-rewriting** of q

Evaluation of FO($\langle, +$)-formulas is in LOGTIME-uniform **AC⁰** for data complexity

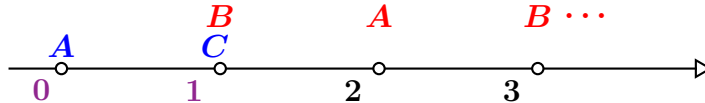
Example - LTL_{core}°

Example 1 $\mathcal{D} = \{A(0), C(1)\}$, $\mathcal{O} = \{\circ_P A \rightarrow B, \circ_P B \rightarrow A\}$, $\mathcal{I} = \circ_F \circ_F B$



Example – LTL_{core}°

Example 1 $\mathcal{D} = \{A(0), C(1)\}$, $\mathcal{O} = \{\circ_P A \rightarrow B, \circ_P B \rightarrow A\}$, $\mathcal{Z} = \circ_F \circ_F B$

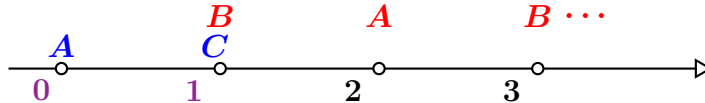


$\text{ans}(q, \mathcal{D}) = \{1\}$ since $\min \mathcal{D} = 0$ and $\max \mathcal{D} = 1$

$\text{ans}^{\mathbb{Z}}(q, \mathcal{D}) = \{2n + 1 \mid n \geq 0\}$

Example – LTL_{core}°

Example 1 $\mathcal{D} = \{A(0), C(1)\}$, $\mathcal{O} = \{\circ_P A \rightarrow B, \circ_P B \rightarrow A\}$, $\mathcal{Z} = \circ_F \circ_F B$



$\text{ans}(q, \mathcal{D}) = \{1\}$ since $\min \mathcal{D} = 0$ and $\max \mathcal{D} = 1$

$\text{ans}^{\mathbb{Z}}(q, \mathcal{D}) = \{2n + 1 \mid n \geq 0\}$

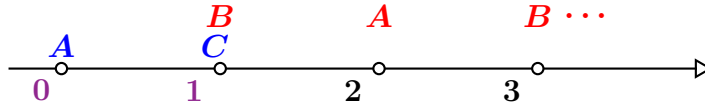
FO($<, +$)-rewriting

$$q'(t) = \exists s < t \exists n [(B(s) \wedge (t - s = 2n)) \vee (A(s) \wedge (t - s = 2n + 1))]$$

$t - s = 2n$ stands for $\exists k ((k = n + n) \wedge (t = s + k))$

Example – LTL_{core}°

Example 1 $\mathcal{D} = \{A(0), C(1)\}$, $\mathcal{O} = \{\circ_P A \rightarrow B, \circ_P B \rightarrow A\}$, $\mathcal{I} = \circ_F \circ_F B$



$\text{ans}(q, \mathcal{D}) = \{1\}$ since $\min \mathcal{D} = 0$ and $\max \mathcal{D} = 1$

$\text{ans}^{\mathbb{Z}}(q, \mathcal{D}) = \{2n + 1 \mid n \geq 0\}$

FO($<$, +)-rewriting

$$q'(t) = \exists s < t \exists n [(B(s) \wedge (t - s = 2n)) \vee (A(s) \wedge (t - s = 2n + 1))]$$

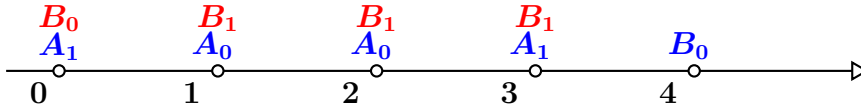
$t - s = 2n$ stands for $\exists k ((k = n + n) \wedge (t = s + k))$

Note. The query is **not** FO($<$)-rewritable since properties such as ***t is even*** are not definable by FO($<$)-formulas (Libkin,04)

Example – LTL_{horn}°

Example 2 $\mathcal{O} = \{\circ_F B_k \wedge A_0 \rightarrow B_k, \circ_F B_k \wedge A_1 \rightarrow B_{1-k} \mid k = 0, 1\}$, $\varkappa = B_0$
for $e = (e_0, \dots, e_{n-1}) \in \{0, 1\}^n$, $\mathcal{D}_e = \{A_{e_i}(i) \mid 0 \leq i < n\} \cup \{B_0(n)\}$

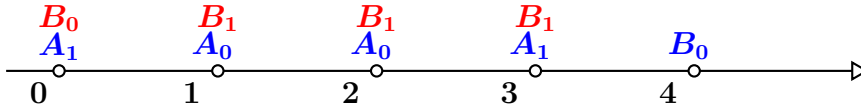
$0 \in \text{ans}(q, \mathcal{D}_e) \iff$ the number of 1s in e is even



Example – LTL_{horn}°

Example 2 $\mathcal{O} = \{\circ_F B_k \wedge A_0 \rightarrow B_k, \circ_F B_k \wedge A_1 \rightarrow B_{1-k} \mid k = 0, 1\}, \varkappa = B_0$
 for $e = (e_0, \dots, e_{n-1}) \in \{0, 1\}^n$, $\mathcal{D}_e = \{A_{e_i}(i) \mid 0 \leq i < n\} \cup \{B_0(n)\}$

$0 \in \text{ans}(q, \mathcal{D}_e) \iff$ the number of 1s in e is even

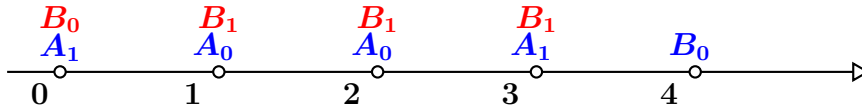


$\text{PARITY} \in \text{NC}^1 \setminus \text{AC}^0$ (Furst,Saxe,Sipser,84) \longrightarrow q is not FO-rewritable with any numeric predicates

Example – LTL_{horn}°

Example 2 $\mathcal{O} = \{\bigcirc_F B_k \wedge A_0 \rightarrow B_k, \bigcirc_F B_k \wedge A_1 \rightarrow B_{1-k} \mid k = 0, 1\}$, $\varkappa = B_0$
for $e = (e_0, \dots, e_{n-1}) \in \{0, 1\}^n$, $\mathcal{D}_e = \{A_{e_i}(i) \mid 0 \leq i < n\} \cup \{B_0(n)\}$

$0 \in \text{ans}(q, \mathcal{D}_e) \iff$ the number of 1s in e is even



$\text{PARITY} \in \text{NC}^1 \setminus \text{AC}^0$ (Furst,Saxe,Sipser,84) \longrightarrow q is not FO-rewritable with any numeric predicates

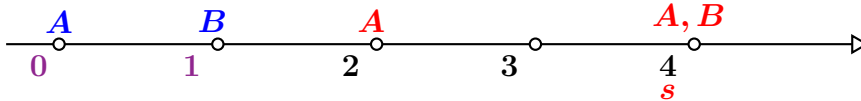
q is FO(RPR)-rewritable

Examples

Example 3 $\mathcal{O} = \{ A \rightarrow \circ_F^2 A, B \rightarrow \circ_F^3 B \}$, $\varkappa = \diamond_F(A \wedge B)$

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

is a rewriting over \mathbb{Z} , but not over $\mathbf{tem}(\mathcal{D})$, indeed, let $\mathcal{D} = \{A(0), B(1)\}$



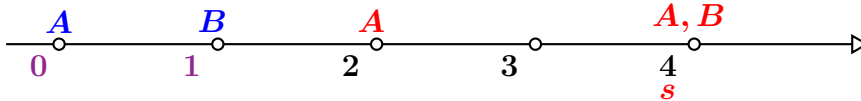
then $s = 4$ is **outside the active domain** $\mathbf{tem}(\mathcal{D}) = \{0, 1\}$ of \mathcal{D} .

Examples

Example 3 $\mathcal{O} = \{ A \rightarrow \circ_F^2 A, B \rightarrow \circ_F^3 B \}, \varkappa = \diamond_F(A \wedge B)$

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

is a rewriting over \mathbb{Z} , but not over $\mathbf{tem}(\mathcal{D})$, indeed, let $\mathcal{D} = \{A(0), B(1)\}$



then $s = 4$ is **outside the active domain** $\mathbf{tem}(\mathcal{D}) = \{0, 1\}$ of \mathcal{D} .

The following is both an $\text{FO}(<)$ - and $\text{FO}^{\mathbb{Z}}(<)$ -rewriting of q :

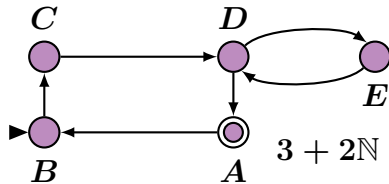
$$\exists u, v [A(u) \wedge B(v)]$$

Automata for proving rewritability of OMAQ – LTL_{krom}°

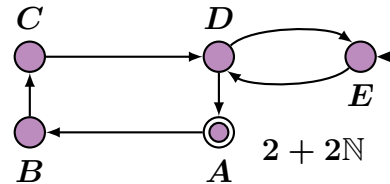
Example – LTL_{core}° $q = (\mathcal{O}, A)$ with

(using unary automata)

$$\mathcal{O} = \{ A \rightarrow \circ B, B \rightarrow \circ C, C \rightarrow \circ D, D \rightarrow \circ A, D \rightarrow \circ E, E \rightarrow \circ D \}$$



$$\exists s (B(s) \wedge (t - s \in 3 + 2\mathbb{N}))$$



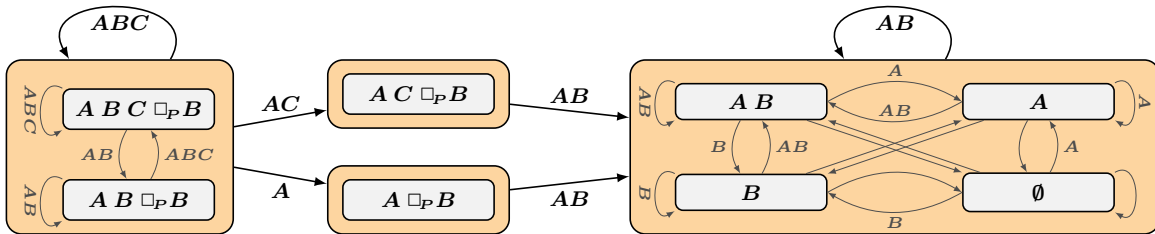
$$\exists s (E(s) \wedge (t - s \in 2 + 2\mathbb{N}))$$

Automata for proving rewritability of OMAQ – LTL_{bool}^{\square}

Example – LTL_{bool}^{\square} $q = (\mathcal{O}, A)$ with

$$\mathcal{O} = \{C \sqsubseteq \square_P B, \square_P B \sqsubseteq A\}$$

(using partially-ordered automata)



Data Complexities of *LTL* OMQ's

c	OMAQs (Atomic Queries)			OMPIQs / positive MFO(<) queries		
	LTL_c^\square	LTL_c°	$LTL_c^{\square\circ}$	LTL_c^\square	LTL_c°	$LTL_c^{\square\circ}$
<i>bool</i>				MSO(<)		
<i>horn</i>		MSO(<), FO(RPR)		FO(<)	MSO(<), FO(RPR)	
<i>krom</i>	FO(<)			MSO(<)		
<i>core</i>		FO(<, +)		FO(<)		FO(<, +)

$LTL_{bool}^{\square\circ}$ OMAQs are MSO(<)-rewritable \longrightarrow

OMAQ answering is in NC^1 for data complexity

Answering LTL_{horn}° OMAQs is NC^1 -hard for data complexity

Data Complexities of *LTL* OMQ's

c	OMAQs (Atomic Queries)			OMPIQs / positive MFO($<$) queries		
	LTL_c^\square	LTL_c°	$LTL_c^{\square\circ}$	LTL_c^\square	LTL_c°	$LTL_c^{\square\circ}$
<i>bool</i>				MSO($<$)		
<i>horn</i>		MSO($<$), FO(RPR)		FO($<$)	MSO($<$), FO(RPR)	
<i>krom</i>	FO($<$)			MSO($<$)		
<i>core</i>		FO($<$,+)		FO($<$)		FO($<$,+)

$LTL_{bool}^{\square\circ}$ OMAQs are MSO($<$)-rewritable \longrightarrow

OMAQ answering is in NC^1 for data complexity

Answering LTL_{horn}° OMAQs is NC^1 -hard for data complexity

MFO($<$): monadic first-order logic with a built-in linear order

$$\psi(t, t') = \exists x ((t < x < t') \wedge \text{Revise}(x)) \wedge \text{Submission}(t) \wedge \text{Accept}(t').$$

Answering positive MFO queries has the same data complexity as OMQ queries

Conclusions and Future Work

- Investigation of FO-rewritability for temporal ontology-mediated queries based on linear temporal logic *LTL*.
- Classification of the OMQs by the shape of their ontology axioms (core, horn, krom or bool) and by the temporal operators in the ontology axioms.
- **Data Complexity**: Identification of **FO(\langle)**, **FO(\langle , +)**, **FO(RPR)** rewriting results for certain classes of OMQ's.
- Generalise FO rewritings of OMQs to **(2-sorted) CQs**, **positive MFO(\langle)**.

Conclusions and Future Work

- Investigation of FO-rewritability for temporal ontology-mediated queries based on linear temporal logic *LTL*.
- Classification of the OMQs by the shape of their ontology axioms (core, horn, krom or bool) and by the temporal operators in the ontology axioms.
- **Data Complexity**: Identification of **FO(<)**, **FO(<, +)**, **FO(RPR)** rewriting results for certain classes of OMQ's.
- Generalise FO rewritings of OMQs to **(2-sorted) CQs**, **positive MFO(<)**.
- Extend the result to temporal extensions of *DL-Lite*: we are investigating $DL-Lite_{core}$ with LTL_{core}^{\square} , and $DL-Lite_{core}$ with $LTL_{core}^{\square\circ}$.
- Investigate other form of temporal languages (Intervals, Metric, etc.)

Thank you!