Temporal Description Logic for Ontology-Based Data Access

Alessandro Artale

KRDB Centre, University of Bolzano

joint work with

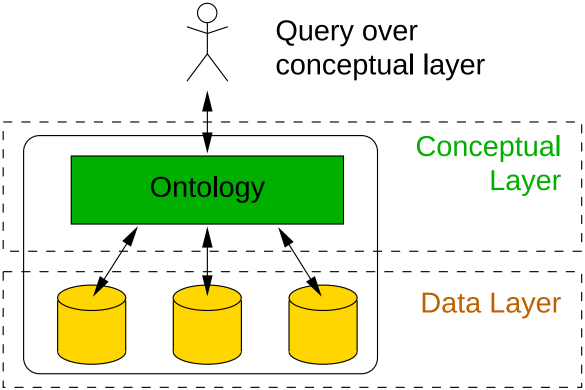
Frank Wolter, Roman Kontchakov, Michael Zakharyaschev

Vladislav Ryzhikov and Alisa Kovtunova

OBDA: Ontology-Based Data Access

Desiderata:

- *Hide* to the user where and how data are stored
- Present to the user a *conceptual view* of the data
- *Query the data sources* through the conceptual model



OBDA: Ontology-Based Data Access

- **ABox** $\mathcal{A}$:

  $$\mathsf{heartpatient}(\mathsf{peter}), \quad \mathsf{diagnose}(\mathsf{sue}, \mathsf{fibrillation}), \quad \mathsf{heartdisease}(\mathsf{fibrillation})$$

OBDA: Ontology-Based Data Access

- **ABox** $\mathcal{A}$:

  $$\textbf{heartpatient}(\textbf{peter}), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}), \quad \textbf{heartdisease}(\textbf{fibrillation})$$

- **Query** q:

  $$q(x) = \exists y.(\textbf{diagnose}(x, y) \land \textbf{heartdisease}(y))$$

  Answer $q(\mathcal{A}) = \{\textbf{sue}\}$.

OBDA: Ontology-Based Data Access

- **ABox** $\mathcal{A}$:

$$\textbf{heartpatient}(\textbf{peter}), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}), \quad \textbf{heartdisease}(\textbf{fibrillation})$$

- **Query** q:

$$q(x) = \exists y.(\textbf{diagnose}(x, y) \wedge \textbf{heartdisease}(y))$$

  Answer $q(\mathcal{A}) = \{\textbf{sue}\}$.

- **Ontology/TBox** $\mathcal{T}$:

$$\textbf{heartpatient} \sqsubseteq \exists \textbf{diagnose}.\textbf{heartdisease}$$

OBDA: Ontology-Based Data Access

- **ABox** $\mathcal{A}$:

$$\textbf{heartpatient}(\textbf{peter}), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}), \quad \textbf{heartdisease}(\textbf{fibrillation})$$

- **Query** q:

$$q(x) = \exists y.(\textbf{diagnose}(x, y) \land \textbf{heartdisease}(y))$$

  Answer $q(\mathcal{A}) = \{\textbf{sue}\}$.

- **Ontology/TBox** $\mathcal{T}$:

$$\textbf{heartpatient} \sqsubseteq \exists \textbf{diagnose}.\textbf{heartdisease}$$

OBDA: Ontology-Based Data Access

- **ABox** $\mathcal{A}$:

$$\textsf{heartpatient}(\textsf{peter}), \quad \textsf{diagnose}(\textsf{sue}, \textsf{fibrillation}), \quad \textsf{heartdisease}(\textsf{fibrillation})$$

- **Query** q:

$$q(x) = \exists y.(\textsf{diagnose}(x, y) \wedge \textsf{heartdisease}(y))$$

Answer $q(\mathcal{A}) = \{\textsf{sue}\}$.

- **Ontology/TBox** $\mathcal{T}$:

$$\textsf{heartpatient} \sqsubseteq \exists \textsf{diagnose}.\textsf{heartdisease}$$

- **Certain Answers**:

$$\textsf{cert}_{\mathcal{T},\mathcal{A}}(q) = \{a \mid \mathcal{T} \cup \mathcal{A} \models q(a)\}$$

In this case

$$\textsf{cert}_{\mathcal{T},\mathcal{A}}(q) = \{\textsf{sue}, \textsf{peter}\}.$$

OBDA for Temporal Data

In applications, data are often time-dependent: employment contracts end, children are born, aircrafts arrive.

OBDA for Temporal Data

In applications, data are often time-dependent: employment contracts end, children are born, aircrafts arrive.

- Temporal data (**temporal ABoxes**) $\mathcal{A}$ are finite sets of pairs consisting of facts and their validity time:

  $$\textbf{atrisk}(\textbf{peter}, 2013), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}, 1982), \quad \textbf{heartdisease}(\textbf{fibrillation})$$

OBDA for Temporal Data

In applications, data are often time-dependent: employment contracts end, children are born, aircrafts arrive.

- Temporal data (**temporal ABoxes**) $\mathcal{A}$ are finite sets of pairs consisting of facts and their validity time:

  $\mathbf{atrisk}(\mathbf{peter}, 2013), \quad \mathbf{diagnose}(\mathbf{sue}, \mathbf{fibrillation}, 1982), \quad \mathbf{heartdisease}(\mathbf{fibrillation})$

- To support querying temporal data, the ontology $\mathcal{T}$ should model **temporal conceptual knowledge** as well:

$$\Diamond_P \exists \mathbf{diagnose}.\mathbf{heartdisease} \sqsubseteq \mathbf{atrisk}$$

$$\forall x, t(((\exists t' < t)\exists y.\mathbf{diagnose}(x, y, t') \land \mathbf{heartdisease}(y, t')) \rightarrow \mathbf{atrisk}(x, t))$$

OBDA for Temporal Data

In applications, data are often time-dependent: employment contracts end, children are born, aircrafts arrive.

- Temporal data (**temporal ABoxes**) $\mathcal{A}$ are finite sets of pairs consisting of facts and their validity time:

$$\textbf{atrisk}(\textbf{peter}, 2013), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}, 1982), \quad \textbf{heartdisease}(\textbf{fibrillation})$$

- To support querying temporal data, the ontology $\mathcal{T}$ should model **temporal conceptual knowledge** as well:

$$\Diamond_P \exists \textbf{diagnose}.\textbf{heartdisease} \sqsubseteq \textbf{atrisk}$$

$$\forall x, t(((\exists t' < t)\exists y.\textbf{diagnose}(x, y, t') \wedge \textbf{heartdisease}(y, t')) \rightarrow \textbf{atrisk}(x, t))$$

- For $q = \textbf{atrisk}(x, 2013)$ we obtain

$$\textbf{cert}_{\mathcal{T},\mathcal{A}}(q) = \{\textbf{peter}, \textbf{sue}\}$$

Our Aims

- Cover **validity time** (no transaction time): ABox assertions of the form

$$A(c,n), \quad P(c,d,n)$$

More succinct intervals $A(c,[n,m])$ not yet considered.

Our Aims

- Cover **validity time** (no transaction time): ABox assertions of the form

$$A(c,n), \quad P(c,d,n)$$

  More succinct intervals $A(c,[n,m])$ not yet considered.

- Ontology language **temporal extension of OWL 2 QL** (OWL standard for OBDA). Axioms time-independent, but model time-dependent classes and properties. E.g.,

$$\Diamond_P \textbf{givesbirth} \sqsubseteq \textbf{mother}$$

Our Aims

- Cover **validity time** (no transaction time): ABox assertions of the form

$$A(c,n), \quad P(c,d,n)$$

  More succinct intervals $A(c,[n,m])$ not yet considered.

- Ontology language **temporal extension of OWL 2 QL** (OWL standard for OBDA). Axioms time-independent, but model time-dependent classes and properties. E.g.,

$$\Diamond_P \mathbf{givesbirth} \sqsubseteq \mathbf{mother}$$

- Queries at least **two sorted conjunctive queries** with variables for individuals and timepoints, and expressions $t < t'$, $A(x,t)$, $P(x,y,t)$.

Our Aims

- Cover **validity time** (no transaction time): ABox assertions of the form

$$A(c,n), \quad P(c,d,n)$$

  More succinct intervals $A(c,[n,m])$ not yet considered.

- Ontology language **temporal extension of OWL 2 QL** (OWL standard for OBDA). Axioms time-independent, but model time-dependent classes and properties. E.g.,

$$\Diamond_P \textbf{givesbirth} \sqsubseteq \textbf{mother}$$

- Queries at least **two sorted conjunctive queries** with variables for individuals and timepoints, and expressions $t < t'$, $A(x,t)$, $P(x,y,t)$.

- Every such query should be **SQL/FO-rewritable** (with linear-order $<$ available).

The Ontology Language: TQL

TQL contains **OWL 2 QL**, where OWL 2 QL ontologies consist of inclusions

$$\boldsymbol{B_1 \sqcap B_2 \sqsubseteq \bot, \quad B_1 \sqsubseteq B_2, \quad R_1 \sqsubseteq R_2}$$

with

$$\boldsymbol{R_i} ::= \bot \quad | \quad \boldsymbol{P} \quad | \quad \boldsymbol{P^-},$$

$$\boldsymbol{B_i} ::= \boldsymbol{A} \quad | \quad \exists \boldsymbol{R_i},$$

The Ontology Language: TQL

TQL contains **OWL 2 QL**, where OWL 2 QL ontologies consist of inclusions

$$B_1 \sqcap B_2 \sqsubseteq \bot, \quad B_1 \sqsubseteq B_2, \quad R_1 \sqsubseteq R_2$$

with

$$R_i ::= \bot \quad | \quad P \quad | \quad P^-,$$

$$B_i ::= A \quad | \quad \exists R_i,$$

and should be "maximal" FO-rewritable with:

- rigid concept and roles;
- persistent in the future concepts and roles;
- instantaneous concepts and roles;
- convex concepts and roles.
- etc.

Syntax: OWL 2 QL extended by $\Diamond_F$ and $\Diamond_P$

TQL ontologies/TBox consist of inclusions

$$C \sqsubseteq B, \quad S \sqsubseteq R$$

where

$$\mathrm{R} ::= \bot \quad | \quad P \quad | \quad P^-,$$

$$\mathrm{B} ::= \bot \quad | \quad A \quad | \quad \exists R,$$

Syntax: OWL 2 QL extended by $\Diamond_F$ and $\Diamond_P$

TQL ontologies/TBox consist of inclusions

$$C \sqsubseteq B, \quad S \sqsubseteq R$$

where

$$\mathrm{R} ::= \bot \quad | \quad P \quad | \quad P^-,$$

$$\mathrm{B} ::= \bot \quad | \quad A \quad | \quad \exists R,$$

and C and S are defined by:

$$\mathrm{C} ::= B \quad | \quad C_1 \sqcap C_2 \quad | \quad \Diamond_P C \quad | \quad \Diamond_F C,$$

$$\mathrm{S} ::= R \quad | \quad S_1 \sqcap S_2 \quad | \quad \Diamond_P S \quad | \quad \Diamond_F S,$$

Syntax: OWL 2 QL extended by $\Diamond_F$ and $\Diamond_P$

TQL ontologies/TBox consist of inclusions

$$C \sqsubseteq B, \quad S \sqsubseteq R$$

where

$$\mathrm{R} ::= \bot \quad | \quad P \quad | \quad P^-,$$

$$\mathrm{B} ::= \bot \quad | \quad A \quad | \quad \exists R,$$

and C and S are defined by:

$$\mathrm{C} ::= B \quad | \quad C_1 \sqcap C_2 \quad | \quad \Diamond_P C \quad | \quad \Diamond_F C,$$

$$\mathrm{S} ::= R \quad | \quad S_1 \sqcap S_2 \quad | \quad \Diamond_P S \quad | \quad \Diamond_F S,$$

Thus TQL has a Horn-like TBox with temporal operators only on the left-hand side.

TQL: Expressivity

TQL can express the following temporal constraints:

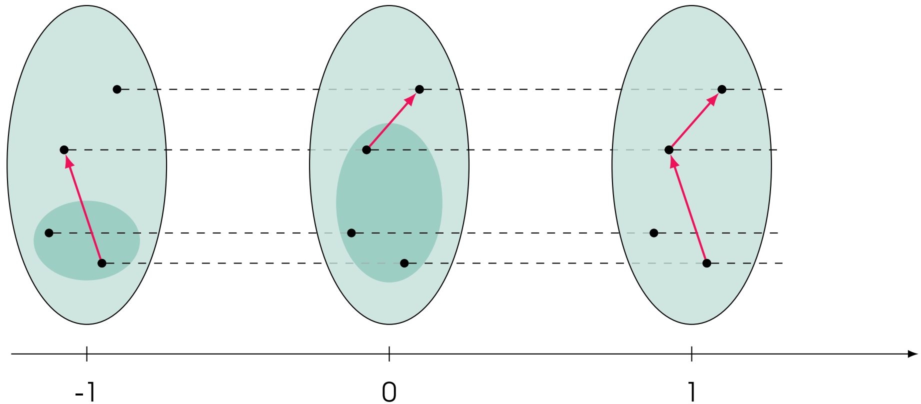
- **person** is **rigid**: $\Diamond_F \Diamond_P \textbf{person} \sqsubseteq \textbf{person}$;
- **mother** is **persistent**: $\Diamond_P \textbf{mother} \sqsubseteq \textbf{mother}$;
- **givesbirth** is **instantaneous**: $\textbf{givesbirth} \sqcap \Diamond_P \textbf{givesbirth} \sqsubseteq \bot$;
- **employed** is **convex**: $\Diamond_P \textbf{employed} \sqcap \Diamond_F \textbf{employed} \sqsubseteq \textbf{employed}$.

Semantics

Temporal interpretations $\mathcal{I}$ are given by $(\mathbb{Z}, <)$ (time points) and standard (atemporal) interpretations

$$\mathcal{I}(n) = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}(n)}),$$

for each $n \in \mathbb{Z}$. We assume constant domain and rigid interpretation of individuals. Thus, interpretations look as follows:

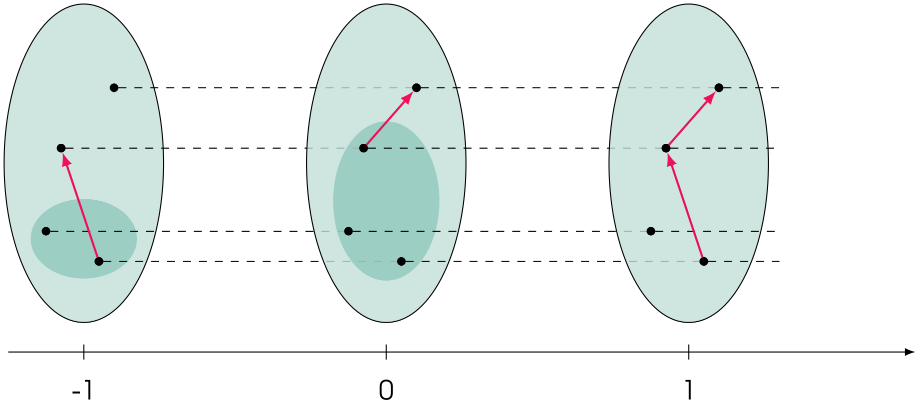


Semantics

Temporal interpretations $\mathcal{I}$ are given by $(\mathbb{Z}, <)$ (time points) and standard (atemporal) interpretations

$$\mathcal{I}(n) = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}(n)}),$$

for each $n \in \mathbb{Z}$. We assume constant domain and rigid interpretation of individuals. Thus, interpretations look as follows:



$$(\Diamond_P C)^{\mathcal{I}(n)} = \{x \mid x \in C^{\mathcal{I}(m)}, \text{ for some } m < n\},$$

$$(\Diamond_F C)^{\mathcal{I}(n)} = \{x \mid x \in C^{\mathcal{I}(m)}, \text{ for some } m > n\}.$$

Temporal SQL/FO-Rewritability

Consider again

- $\mathcal{A}$:

$$\textbf{atrisk}(\textbf{peter}, 2013), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}, 1982), \quad \textbf{heartdisease}(\textbf{fibrillation})$$

- $\mathcal{T}$:

$$\Diamond_P \exists \textbf{diagnose}.\textbf{heartdisease} \sqsubseteq \textbf{atrisk}$$

Temporal SQL/FO-Rewritability

Consider again

- $\mathcal{A}$:

  $\textbf{atrisk}(\textbf{peter}, 2013), \quad \textbf{diagnose}(\textbf{sue}, \textbf{fibrillation}, 1982), \quad \textbf{heartdisease}(\textbf{fibrillation})$

- $\mathcal{T}$:

  $$\Diamond_P \exists \textbf{diagnose}.\textbf{heartdisease} \sqsubseteq \textbf{atrisk}$$

- Then $q = \textbf{atrisk}(x, 2013)$ can be rewritten into

  $q_{\mathcal{T}} = \textbf{atrisk}(x, 2013) \vee \exists t' < 2013. \exists y.\textbf{diagnose}(x, y, t') \wedge \textbf{heartdisease}(y, t')$

  and

  $$(\mathcal{T}, \mathcal{A}) \models q(a, 2013) \text{ iff } \mathcal{A} \models q_{\mathcal{T}}(a, 2013)$$

Temporal Datalog$_\exists$ Formulation

Let

$$B \quad = \quad A \quad | \quad \exists R$$

TBoxes consist of "datalog" rules of the form

$$B(x, t) \leftarrow \textbf{Body}(x, \vec{t})$$

where $\textbf{Body}(x, \vec{t})$ is a conjunction of atoms of the form $B'(x, t')$ and $t' < t''$ and

$$P(x, y, t) \leftarrow \textbf{Body}(x, y, \vec{t})$$

where $\textbf{Body}(x, y, \vec{t})$ is a conjunction of atoms of the form $B'(x, y, t')$ and $t' < t''$.

Note: Link between rules for unary and binary predicates only via $\exists R$.

Main Result

Queries are two-sorted conjunctive queries (CQs):

$$\exists \vec{y}\, \vec{t}\, \underbrace{\varphi(\vec{x}, \vec{y}, \vec{s}, \vec{t})}_{\text{conjunction of atoms}}$$

where the atoms are of the form

$$A(x, t), \quad P(x, y, t), \quad (t_1 = t_2), \quad (t_1 < t_2)$$

Main Result

Queries are two-sorted conjunctive queries (CQs):

$$\exists \vec{y}\, \vec{t}\, \underbrace{\varphi(\vec{x}, \vec{y}, \vec{s}, \vec{t})}_{\text{conjunction of atoms}}$$

where the atoms are of the form

$$A(x, t), \quad P(x, y, t), \quad (t_1 = t_2), \quad (t_1 < t_2)$$

Theorem. Let $q(\vec{x}, \vec{t})$, be a CQ and $\mathcal{T}$ a TQL ontology.
Then one can construct a disjunction of CQs $q_{\mathcal{T}}(\vec{x}, \vec{t})$ such that, for any $\mathcal{A}$, any $\vec{a} \subseteq \mathsf{ind}(\mathcal{A})$, and any $\vec{n} \subseteq \mathsf{tem}(\mathcal{A})$, we have

$$(\mathcal{T}, \mathcal{A}) \models q(\vec{a}, \vec{n}) \quad \text{iff} \quad \mathcal{A} \models q_{\mathcal{T}}(\vec{a}, \vec{n})$$

Extensions not tractable and not FO-rewritable

- Mixing concepts and roles: $\exists R.A \sqsubseteq A$ not FO-rewritable.

Extensions not tractable and not FO-rewritable

- Mixing concepts and roles: $\exists R.A \sqsubseteq A$ not FO-rewritable.

  $\Diamond_P A \sqsubseteq A$ is rewritable only because $<$ is transitive.

Extensions not tractable and not FO-rewritable

- Mixing concepts and roles: $\exists R.A \sqsubseteq A$ not FO-rewritable.

  $\Diamond_P A \sqsubseteq A$ is rewritable only because $<$ is transitive.

- NEXT-operators: $\bigcirc_P A \sqsubseteq B$ and $\bigcirc_P B \sqsubseteq A$ can be used to express even distance between time points.

Extensions not tractable and not FO-rewritable

- Mixing concepts and roles: $\exists R.A \sqsubseteq A$ not FO-rewritable.

  $\Diamond_P A \sqsubseteq A$ is rewritable only because $<$ is transitive.

- NEXT-operators: $\bigcirc_P A \sqsubseteq B$ and $\bigcirc_P B \sqsubseteq A$ can be used to express even distance between time points.

- CQ answering for $\{A \sqsubseteq \Diamond_P B\}$ NP-hard—by reduction of $2 + 2$-SAT.

Extensions with NEXT - $\bigcirc_F$

The TQL language with nextime, $\bigcirc_F$, Atomic Concepts and Horn axioms is not in AC^0.

Extensions with NEXT - $\bigcirc_F$

The TQL language with nextime, $\bigcirc_F$, Atomic Concepts and Horn axioms is not in AC^0.

Parity problem: Given a binary string output 1 iff the number of 1s is even.

We reduce the Parity problem which is not computable in AC^0 (Furst,Saxe and Sipser,1984) to query answering in TQL TBox with $\bigcirc_F$.

Extensions with NEXT - $\bigcirc_F$

The TQL language with nextime, $\bigcirc_F$, Atomic Concepts and Horn axioms is not in AC^0.

Parity problem: Given a binary string output 1 iff the number of 1s is even.

We reduce the Parity problem which is not computable in AC^0 (Furst,Saxe and Sipser,1984) to query answering in TQL TBox with $\bigcirc_F$.

TBox

$$\mathcal{T} = \{C_1 \sqcap \bigcirc_F C_{\mathit{even}} \sqsubseteq C_{\mathit{odd}},\ C_1 \sqcap \bigcirc_F C_{\mathit{odd}} \sqsubseteq C_{\mathit{even}}$$

$$C_0 \sqcap \bigcirc_F C_{\mathit{even}} \sqsubseteq C_{\mathit{even}},\ C_0 \sqcap \bigcirc_F C_{\mathit{odd}} \sqsubseteq C_{\mathit{odd}}\}$$

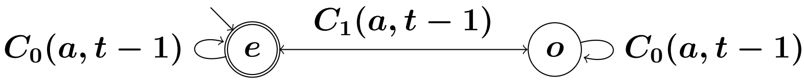
ABox. Encodes the binary strings and terminates with $C_{\mathit{even}}(a, n+1)$. E.g., the binary string $w = 01001$ is encoded as:

$$\mathcal{A}_w = \{C_0(a,0), C_1(a,1), C_0(a,2), C_0(a,3), C_1(a,4), C_{\mathit{even}}(a,5)\}$$

$(\mathcal{T}, \mathcal{A}_w) \models C_{\mathit{even}}(a,0)$ iff w has an even number of 1's

Extensions with NEXT and Automata

We can construct a Non-Deterministic Finite Automata (NFA) to compute query answers. E.g., the automaton $\mathfrak{A}_\mathcal{T}$ for the parity TBox starting at $t = n$ is:



$\mathfrak{A}_\mathcal{T}$ accepts $\mathcal{A}$ iff $(\mathcal{T}, \mathcal{A}) \models C_{even}(a, 0)$.

- **Upper Bound.** The problem whether an automata accepts a word is tractable: it belongs to complexity class NC^1 (contained in LogSpace).
- **Future Work.** The automata encoding without roles is obvious: We intend to extend it to languages with roles.

Future Work

- Investigate **efficient** rewritings, implementation.
- Consider **datalog-rewritability**: then NEXT-operator should be ok.
- The TQL languages with $\bigcirc_F$ seems to be still FO-rewritable with arithmetic predicates, e.i., TQL$_{core,\bigcirc_F}$ is conjectured to be in $\boldsymbol{FO}(+, \times)$.