A Compilation Technique for Interactive Ontology-mediated Data Exploration

Medina Andresel

Advisor: Ass. Prof. Magdalena Ortiz
Co-advisor: Univ. Ass. Mantas Simkus

EMCL Workshop February 12, 2016

TU WIEN
FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Summary

Summary

Motivation

- **Ontologies on top of data** extensively considered in recent years, they offer many advantages
- **Ontology-based Data Access** [PLC+08] ontologies mediate the access to the data
- **Ontology-mediated Query Answering** evaluating a query in presence of an ontology
  - → central problem in OBDA
  - → representation of data abstracted away
- **Interactive query answering not supported by any OMQA systems** all engines answer queries **one-at-a-time**

Motivation

- **We focus on** how to efficiently answer slightly modified versions of a query
- **Compiling data** we observe the need to compile relevant information in a meaningful way
- **Data exploration neglected so far**
  - → data analysis taken for granted in relational database systems
  - → however in ontology-mediated context not tackled before
- **OMQA Interface systems** focus on helping users build queries *Quelo* [FGTT11], *SemFacet* [ACGK+14]
  - → construct the type of queries we consider

Problem Description

- **Considering**
  - → ontologies formalized in *DL-Lite*[CGL+07]
  - → tree-shaped conjunctive queries with one answer variable
- **Assuming** two upper- and -lower bound queries given
  - → helping the user target the objects of interest
- **Goal** build a compilation for supporting *interactive query answering* and *query modifications*
- we expect a **trade-off** between expensive offline compilation and efficient online evaluation

Summary

Description Logics

- we consider ontologies formalized using *DL-Lite*$_{\mathcal{R}}$ [CGL+07]

Construct	**Syntax**	**Example**	**Semantics**
Atomic role	P	worksFor	$P^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
Inverse role	P^{-}	teacherOf^{-}	$\{(o,o') \mid (o',o) \in P^{\mathcal{I}}\}$
Role negation	$\neg R$	¬memberOf	$(\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}) \setminus R^{\mathcal{I}}$
Atomic concept	A	Employee	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
Existential restriction	$\exists R$	∃advisor	$\{o \mid \exists o' \text{ s.t. } (o,o') \in R^{\mathcal{I}}\}$
Concept negation	$\neg B$	¬∃advisor^{-}	$\Delta^{\mathcal{I}} \setminus B^{\mathcal{I}}$
Top concept	$\top$		$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$
Individual	a	Prof1	$a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$

Table: Syntax and semantics for allowed constructs of concepts and roles

Description Logics

$$B := A \mid \exists R \quad C := B \mid \neg B \quad R := P \mid P^{-} \quad E := R \mid \neg R$$

Axiom	Syntax	Example	Semantics
Concept inclusion	$B \sqsubseteq C$	Employee $\sqsubseteq$ $\exists$worksFor	$B^{\mathcal{I}} \subseteq C^{\mathcal{I}}$
Role inclusion	$R \sqsubseteq E$	worksFor $\sqsubseteq$ memberOf	$R^{\mathcal{I}} \subseteq E^{\mathcal{I}}$
Concept assertion	$A(a)$	Student(Stud1)	$a^{\mathcal{I}} \in A^{\mathcal{I}}$
Role assertion	$P(a,b)$	teacherOf(Prof1, GradCourse1)	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in P^{\mathcal{I}}$
Negated concept assertion	$\neg A(a)$	¬Professor(Stud1)	$a^{\mathcal{I}} \notin A^{\mathcal{I}}$
Negated role assertion	$\neg P(a,b)$	¬worksFor(Stud1, Dep1)	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \notin P^{\mathcal{I}}$

Table: Syntax and semantics for allowed axioms and assertions

Definition

- An **ABox** is a finite set of *assertions*
- A **TBox** is a finite set of *concept or role inclusions*
- A **knowledge base (KB)** $\mathcal{K} = \langle \mathcal{A}, \mathcal{T} \rangle$ consists of an ABox $\mathcal{A}$ and a TBox $\mathcal{T}$.

Canonical Model

For a *DL-Lite*$_{\mathcal{R}}$ KB $\langle \mathcal{T}, \mathcal{A} \rangle$ an interpretation $\mathcal{I}^{\mathcal{T},\mathcal{A}}$ such as the domain $\Delta^{\mathcal{I}^{\mathcal{T},\mathcal{A}}}$ consists of all words $aR_1 \ldots R_n$ $(n \geq 0)$, where $a \in \mathbf{Ind}(\mathcal{A})$, R_i - atomic or inverse role. Let $\mathbf{Anon\_Obj} := \Delta^{\mathcal{I}^{\mathcal{T},\mathcal{A}}} \setminus \mathbf{Ind}(\mathcal{A})$ to be the **set of anonymous objects**. The interpretation function, $\cdot^{\mathcal{I}^{\mathcal{T},\mathcal{A}}}$ is defined as follows:

- $a^{\mathcal{I}^{\mathcal{T},\mathcal{A}}} = a$ for all $a \in \mathbf{Ind}(\mathcal{A})$
- $A^{\mathcal{I}^{\mathcal{T},\mathcal{A}}} = \{a \mid \mathcal{T}, \mathcal{A} \vDash A(a)\} \cup \{aR_1 \ldots R_n \mid n \geq 1 \text{ and } \exists R_n^- \sqsubseteq_{\mathcal{T}} A\}$
- $$P^{\mathcal{I}^{\mathcal{T},\mathcal{A}}} = \begin{array}{l} \{(a,b) \mid R(a,b) \in \mathcal{A} \text{ and } R \sqsubseteq_{\mathcal{T}} P\} \cup \\ \{(b,a) \mid R(a,b) \in \mathcal{A} \text{ and } R \sqsubseteq_{\mathcal{T}} P^-\} \cup \\ \{(w_1,w_2) \mid w_2 = w_1 S \text{ and } S \sqsubseteq_{\mathcal{T}} R\} \cup \\ \{(w_2,w_1) \mid w_2 = w_1 S \text{ and } S \sqsubseteq_{\mathcal{T}} R^-\} \end{array}$$

The following result is standard:

Theorem (Canonical model existence [CGL+07])

For any given consistent DL-Lite$_{\mathcal{R}}$ *KB* $\langle \mathcal{T}, \mathcal{A} \rangle$ $\mathcal{I}^{\mathcal{T},\mathcal{A}}$ *can be constructed and is a model of the KB, called* ***canonical model***.

Query answering in *DL-Lite*$_{\mathcal{R}}$

- **Queries**
  - → A **conjunctive query** (CQ)

$$q(\vec{x}) \text{ :- } \varphi(\vec{x}, \vec{y})$$

  where φ is constructed using $\wedge$ from atoms of the form $A(t)$ and $R(t, t')$, where t, t' are *terms* (*individuals* or *variables* from $\vec{x}, \vec{y}$).
  - → $\vec{x}$ - *answer variables*, $\vec{y}$ - *bound variables*
- **Certain answers semantics** the retrieved answers are those that hold in every model
- $\mathbf{cert}(q, \mathcal{K}) = \mathbf{ans}(q, \mathcal{I}^{\mathcal{T}, \mathcal{A}})$, for every CQ q [CGL+07]
- let $\mathcal{K}'$ be the result of removing all negative inclusions and assertions from KB $\mathcal{K}$
  - → $\mathbf{cert}(q, \mathcal{K}) = \mathbf{cert}(q, \mathcal{K}')$, if $\mathcal{K}$ consistant

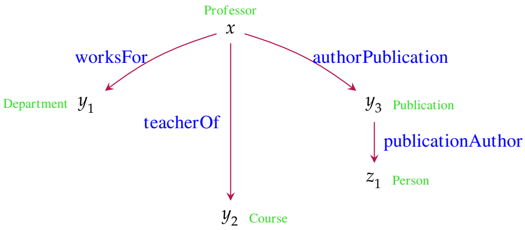
Summary

Tree-shaped CQs

- we call q a **tree-shaped** CQ if its primal graph is a tree
- A CQ q is called **1treeCQ** if it is tree-shaped and has *exactly one answer variable*

Example

$$
\begin{aligned}
q(x) :- \quad & \mathrm{Professor}(x), && \mathrm{worksFor}(y_1), \mathrm{Department}(y_1),\\
& \mathrm{teacherOf}(x, y_2), && \mathrm{Course}(y_2),\\
& \mathrm{authorPublication}(x, y_3), && \mathrm{Publication}(y_3),\\
& \mathrm{publicationAuthor}(y_3, z_1), && \mathrm{Person}(z_1)
\end{aligned}
$$

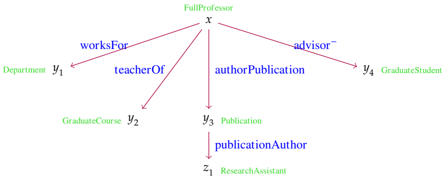


Related Queries

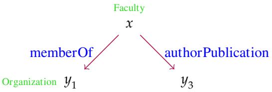
Given a *DL-Lite*$_{\mathcal{R}}$ TBox $\mathcal{T}$ and two 1treeCQs q_1 and q_2, we call q_1 **subquery** of q_2 (w.r.t. $\mathcal{T}$), written $q_1 \trianglelefteq_{\mathcal{T}} q_2$, iff $\mathbf{term}(q_1) \subseteq \mathbf{term}(q_2)$ and

- for each atom $R_1(t_1, t_2) \in q_1$ there exists $R_2(t_1, t_2)$ or $R_2^-(t_2, t_1) \in q_2$ such that $R_2 \sqsubseteq_{\mathcal{T}} R_1$ and
- for each atom $C_1(t) \in q_1$ there exists $C_2(t) \in q_2$ such that $C_2 \sqsubseteq_{\mathcal{T}} C_1$.

Symmetrically, we call q_2 **superquery** of q_1 (w.r.t. $\mathcal{T}$).



(a) Superquery for q



(b) Subquery for q

$$\text{FullProfessor} \sqsubseteq_{\mathcal{T}} \text{Professor} \quad \text{Professor} \sqsubseteq_{\mathcal{T}} \text{Faculty} \quad \text{worksFor} \sqsubseteq_{\mathcal{T}} \text{memberOf}$$

1treeCQs Family

- For a given *DL-Lite*$_{\mathcal{R}}$ KB $\mathcal{K} := \langle \mathcal{T}, \mathcal{A} \rangle$, let $\mathrm{q_L}$ - **the lower bound query** and $\mathrm{q_U}$ -**the upper bound query** be two 1treeCQs such that $\mathrm{q_L} \trianglelefteq^{\pm}_{\mathcal{T}} \mathrm{q_U}$
- any 1treeCQ q such that $\mathrm{q_L} \trianglelefteq^{\pm}_{\mathcal{T}} q \trianglelefteq^{\pm}_{\mathcal{T}} \mathrm{q_U}$ is called **in-between query**
- **1treeCQs family** is the set of 1treeCQs containing $\mathrm{q_L}$, $\mathrm{q_U}$ together with all the in-between queries.

Corollary

Let $\mathcal{K}$ be a given DL-Lite$_{\mathcal{R}}$ KB. For any two given 1treeCQs q_1, q_2 such that $q_1 \trianglelefteq^{\pm}_{\mathcal{T}} q_2$, the following holds:

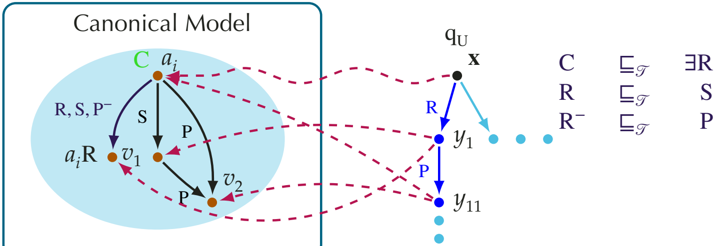
$$\mathbf{cert}(q_2, \mathcal{K}) \subseteq \mathbf{cert}(q_1, \mathcal{K})$$

$\Longrightarrow$ for a given 1treeCQs family, if an individual is an answer to some in-between query, then it must be an answer to $\mathrm{q_L}$

Summary

General Idea

- **Goal** compiling information relevant for 1treeCQs family
- **possible answers** - $a_1, \ldots, a_n$ where each $a_i \in \mathbf{Ind}(\mathscr{A})$, $1 \leq i \leq n$
  - → **storing the relevant information** **matching witness** - w_v^t



matching candidate

$$w_{a_i}^{\mathbf{x}} := \langle\, [\langle \emptyset, \{\mathrm{R}\}\rangle\, a_i\mathrm{R}, \langle \emptyset, \{\mathrm{R}\}\rangle\, v_1]^{y_1}, \ldots \,\rangle$$

$$w_{a_i\mathrm{R}}^{y_1} := \langle\, [\langle \emptyset, \{\mathrm{P}\}\rangle\, v_2, \langle \{\mathrm{C}\}, \{\mathrm{P}\}\rangle\, a_i]^{y_{11}}, \ldots \,\rangle$$

associated labels string

Most Specialized Concepts and Roles

Given a $DL\text{-}Lite_{\mathcal{R}}$ KB $\mathcal{K} := \langle , \mathcal{A} \rangle$ and any $a \in \Delta^{\mathcal{I}^{\mathcal{T},\mathcal{A}}}$, we define:

$$MS_{concept}(a, \mathcal{K}) := \{A \mid \mathcal{I}^{\mathcal{T},\mathcal{A}} \models A(a) \text{ and for each } A' \text{ s.t. } A' \sqsubseteq_{\mathcal{T}} A, \text{ with } A' \neq A, \text{ we have that } \mathcal{I}^{\mathcal{T},\mathcal{A}} \not\models A'(a)\}$$

to be **the set of all most specialized concepts** satisfied by a w.r.t. $\mathcal{K}$
Respectively, for any $a, b \in \Delta^{\mathcal{I}^{\mathcal{T},\mathcal{A}}}$, we define:

$$MS_{role}(a, b, \mathcal{K}) := \{R \mid \mathcal{I}^{\mathcal{T},\mathcal{A}} \models R(a, b) \text{ and for each } R' \text{ s.t. } R' \sqsubseteq_{\mathcal{T}} R, \text{ with } R' \neq R, \text{ we have that } \mathcal{I}^{\mathcal{T},\mathcal{A}} \not\models R'(a, b)\}$$

to be **the set of all most specialized roles** satisfied by (a, b) w.r.t. $\mathcal{K}$.

Matching Candidates

Let $\{t_1, t_2\} \subseteq \mathbf{term}(\mathrm{q_U})$ such that there is some $R(t_1, t_2)$ of $\mathrm{q_U}$, and $v_1 \in \mathbf{Ind}(\mathscr{A}) \cup \mathbf{Anon\_Obj}$. The set of **matching candidates** for t_2 relative to $t_1 \mapsto v_1$ is

I if $t_2 \in \mathbf{Ind}(\mathscr{A})$ then

- $\mathbb{MC}^{t_1 \mapsto v_1}(t_2) := \{t_2\}$, if there exists $R \in MS_{role}(v_1, t_2, \mathscr{K})$ s.t.$R' \sqsubseteq_{\mathscr{T}} R$ or $R \sqsubseteq_{\mathscr{T}} R'$ for some $R'(t_1, t_2) \in \mathrm{q_U}$
- $\mathbb{MC}^{t_1 \mapsto v_1}(t_2) := \emptyset$, otherwise.

II if $t_2 \in \mathbf{vars}(\mathrm{q_U})$ then

$$\mathbb{MC}^{t_1 \mapsto v_1}(t_2) := cand_{\mathscr{A}} \cup cand_{vR} \cup cand_{wRS} \cup cand_{w}$$

Matching Candidates

Where:

1 if $v_1 \in \mathbf{Ind}(\mathcal{A})$ then

(i) $cand_{\mathcal{A}} := \{v_2 \mid \mathcal{T}, \mathcal{A} \models R(v_1, v_2)$ where $R' \sqsubseteq_{\mathcal{T}} R$, for some $R'(t_1, t_2) \in \mathrm{q_U}\}$

(ii) $cand_{vR} := \{v_1R \mid \mathcal{T}, \mathcal{A} \models \exists R(v_1)$ where $R \in MS_{role}(v_1, v_1R, \mathcal{K})$ s.t. $R' \sqsubseteq_{\mathcal{T}} R$ or $R \sqsubseteq_{\mathcal{T}} R'$, for some $R'(t_1, t_2) \in \mathrm{q_U}\}$

2 if $v_1 := wR \in \mathbf{Anon\_Obj}$ then

(i) $cand_{wRS} := \{wRS \mid \mathcal{T} \models \exists R^- \sqsubseteq \exists S$, where $S \in MS_{role}(wR, wRS, \mathcal{K})$ s.t. $R' \sqsubseteq_{\mathcal{T}} S$ or $S \sqsubseteq_{\mathcal{T}} R'$, for some $R'(t_1, t_2) \in \mathrm{q_U}\}$

(ii) $cand_{w} := \{w \mid \mathcal{T} \models R^- \sqsubseteq S$, where $R' \sqsubseteq_{\mathcal{T}} S$, for some $R'(t_1, t_2) \in \mathrm{q_U}\}$

Labels String

For a partial match π, we define **the associated labels string** as follows:

I if π is of the form $[t \mapsto v]$, where $t \in \mathbf{term}(q_U)$ and $v \in \mathbf{Ind}(\mathscr{A}) \cup \mathbf{Anon\_Obj}$ then

$$labels(t \mapsto v) := \langle \{C \mid C \in MS_{concept}(v, \mathscr{K}) \text{ s.t. } C' \sqsubseteq_{\mathscr{T}} C \text{ or } C \sqsubseteq_{\mathscr{T}} C', \text{ for some } C'(t) \in q_U\}\rangle$$

II if π is of the form $[t_1 \mapsto a, t_2 \mapsto b]$, where $t_1, t_2 \in \mathbf{term}(q_U)$, $a, b \in Ind(\mathscr{A}) \cup \mathbf{Anon\_Obj}$, such that $b \in \mathbb{MC}^{t_1 \mapsto a}(t_2)$ then

$$labels(t_1 \mapsto a, t_2 \mapsto b) := \langle \mathbb{R}_{Labels}, \mathbb{C}_{Labels}\rangle$$

where $\mathbb{C}_{Labels} := \{C \mid C \in MS_{concept}(b, \mathscr{K}) \text{ s.t. } C' \sqsubseteq_{\mathscr{T}} C \text{ or } C \sqsubseteq_{\mathscr{T}} C',$ for some $C'(t_2) \in q_U\}$

(a) if $a, b \in \mathbf{Ind}(\mathscr{A})$ or $a := bR$ then
$\mathbb{R}_{Labels} := \{R \mid R \in MS_{role}(a, b, \mathscr{K}) \text{ s.t. } R' \sqsubseteq_{\mathscr{T}} R \text{ or } R \sqsubseteq_{\mathscr{T}} R'$, for some $R'(t_1, t_2) \in q_U\}$

(b) if $b := aR$ then

$$\mathbb{R}_{Labels} := \{R\}$$

Matching Witness

Definition

For a given $DL\text{-}Lite_{\mathcal{R}}$ KB, lower bound 1treeCQ $\mathrm{q_L}$ and upper bound 1treeCQ $\mathrm{q_U}$ (rooted in $\mathbf{x}$), we define the **root matching witness**

$$w_{root} := \langle[labels(\mathbf{x} \mapsto a_1)a_1, \ldots labels(\mathbf{x} \mapsto a_n)a_n]\rangle,$$

For a given term $t \in \mathbf{term}(\mathrm{q_U})$ and $o \in \mathbf{Ind}(\mathscr{A}) \cup \mathbf{Anon\_Obj}$, a **matching witness** for t and o is defined:

$$w_o^t := \langle values_{t_1}, \ldots, values_{t_k}\rangle,$$

where $\{t_1, \ldots, t_k\} = \{t' \mid \text{there exists } R(t,t') \in \mathrm{q_U}\}$, and $values_{t_i}$, $1 \leq i \leq k$, is constructed as follows:

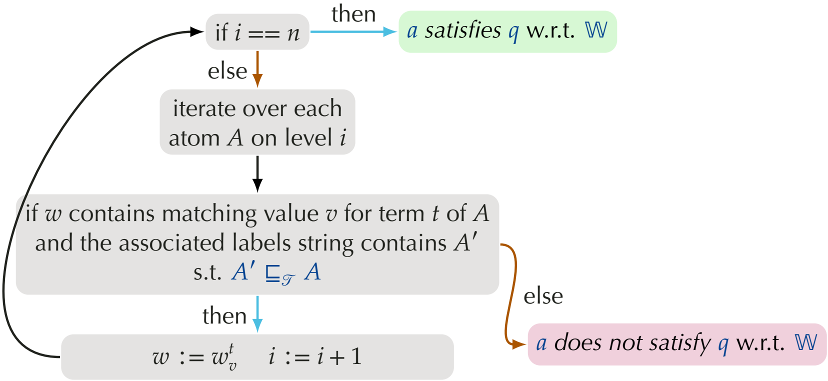
1. $values_{t_i} := [\varepsilon]$, if $\mathbb{MC}^{t \mapsto o}(t_i) = \emptyset$
2. $values_{t_i} := [\phi_1 b_1, \ldots, \phi_m b_m]$, where $b_j \in \mathbb{MC}^{t \mapsto o}(t_i)$, $1 \leq j \leq m$, and $\phi_j := labels(t \mapsto o, t_i \mapsto b_j)$.

Compilation-based Query Answering

- For a given *DL-Lite*$_{\mathcal{R}}$ KB and two 1treeCQs as $\mathrm{q_L}$, $\mathrm{q_U}$ the offline compilation represents the construction of $\mathbb{W}$ - set of all matching witnesses

- **Answering** any query in the 1treeCQs family based on $\mathbb{W}$ **no longer accessing the ABox $\mathscr{A}$**
  - → **Input** q such that $\mathrm{q_L} \trianglelefteq^{+}_{\mathcal{T}} q \trianglelefteq^{+}_{\mathcal{T}} \mathrm{q_U}$, $\mathbb{W}$
  - → **Procedure's starting values** $i := 0$, n is the number of levels in (the tree-structure of) q, $w := w_a^{\mathbf{x}}$ where a is a possible answer and $\mathbf{x}$ is the answer variable of q

Compilation-based Query Answering



Theorem

Given an in-between query q, $a \in \mathbf{Ind}(\mathscr{A})$ *is an answer for q over the KB iff a* satisfies *q w.r.t.* $\mathbb{W}$.

Summary

Maximal Queries

q **is maximal** for a (possible answer) if:

(i) a **satisfies** q w.r.t. $\mathbb{W}$

(ii) for each q' s.t. $q \trianglelefteq_{\mathcal{T}} q' \trianglelefteq_{\mathcal{T}} \mathrm{q_U}$ a **does not satisfy** q' w.r.t. $\mathbb{W}$

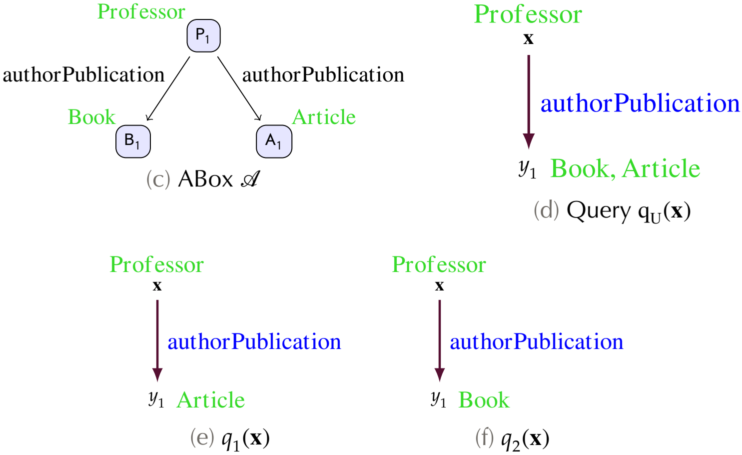


Figure: Queries q_1 and q_2 are maximal for individual $\mathsf{P_1}$

Specializations for 1treeCQs

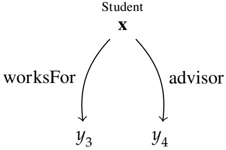
- **query specialization** to refer to any *superquery* of a given q in a 1treeCQ family.
- **Neutral specialization** for q_1 is a 1treeCQ q_2 such that $q_1 \trianglelefteq^{+}_{\mathcal{T}} q_2$ and $\mathbf{ans}(q_2, \mathbb{W}) = \mathbf{ans}(q_1, \mathbb{W})$
  - → **maximal neutral specialization** for q is a neutral specialization q' and for each superquery q'' such that $q' \trianglelefteq^{+}_{\mathcal{T}} q''$ we have $\mathbf{ans}(q'', \mathbb{W}) \neq \mathbf{ans}(q, \mathbb{W})$.
- **Strict specialization** for q_1 is a 1treeCQ q_2 such that $q_1 \trianglelefteq^{+}_{\mathcal{T}} q_2$ and $\mathbf{ans}(q_2, \mathbb{W}) \subsetneq \mathbf{ans}(q_1, \mathbb{W})$, with $\mathbf{ans}(q_2, \mathbb{W}) \neq \emptyset$
  - → **minimal strict specialization** for q is a strict specialization q' such that for each q'', $q \trianglelefteq^{+}_{\mathcal{T}} q'' \trianglelefteq^{+}_{\mathcal{T}} q'$, it holds q'' is a neutral specialization of q.

Generalizations for 1treeCQs

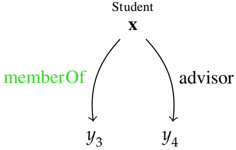
- **query generalization** for q_1 is any q_2 such that $q_2 \trianglelefteq^{+}_{\mathcal{T}} q_1$ and $\mathbf{ans}(q_1, \mathbb{W}) \subsetneq \mathbf{ans}(q_2, \mathbb{W})$
- **Minimal generalization** for q is a generalization q' such that for each q'', that is $q' \trianglelefteq^{+}_{\mathcal{T}} q'' \trianglelefteq^{+}_{\mathcal{T}} q$, it holds that:

$$\mathbf{ans}(q'', \mathbb{W}) = \mathbf{ans}(q, \mathbb{W})$$

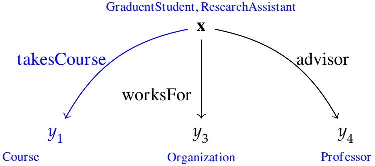
Example of Query Modifications



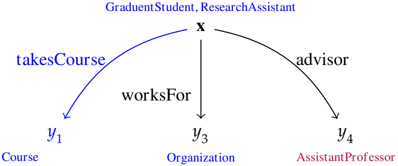
(a) In-between query



(b) Minimal Generalization



(c) Maximal Neutral Specialization



(d) Minimal Strict Specializations

Summary

Implementation and Evaluation

- use **Ontop** [RKZ13] for ABox reasoning and **HermiT** reasoner[1] for TBox reasoning
- As ontology - **LUBM** over the OWL 2 QL profile, that was used in [KRRM+14]

	#D, #U	**Size (MB)**
DS1	6D, 1U	3.3
DS2	11D, 1U	6
DS3	15D, 1U	8.05
DS4	21D, 2U	11.3
DS5	31D, 2U	16.8
DS6	34D, 2U	18.5
DS7	35D, 3U	18.9

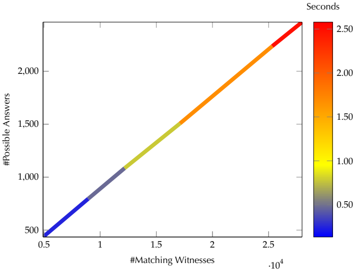
Table: Datasets - Departments(D) distributed over Universities(U)

[1] http://www.hermit-reasoner.com/

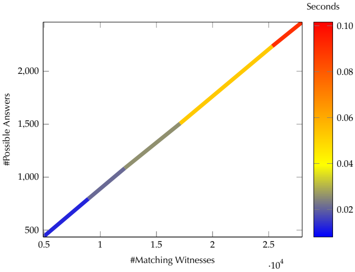
Batch1

$q_L(\mathbf{x}) :- \text{Employee}(\mathbf{x})$

$q_U(\mathbf{x}) :- \text{Dean}(\mathbf{x}), \text{Professor}(\mathbf{x}), \text{FullProfessor}(\mathbf{x}), \text{teacherOf}(\mathbf{x}, y_1), \text{Course}(y_1), \text{headOf}(\mathbf{x}, y_2), \text{Department}(y_2), \text{authorPublication}(\mathbf{x}, y_3), \text{Publication}(y_3)$

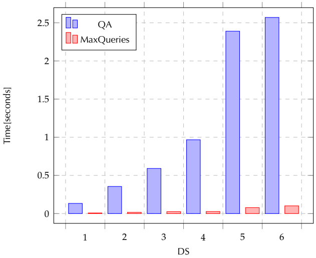


(e) Av. time for QA over Batch1

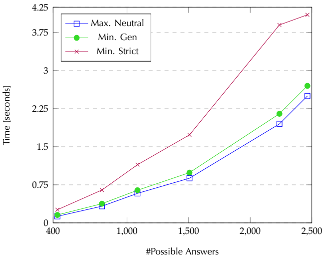


(f) Av. time for computing Maximal Queries over Batch1

Batch1

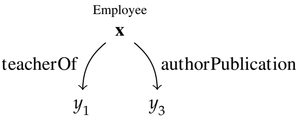


(g) QA and Max. queries performance per Data set for Batch1

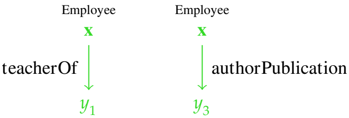


(h) Performance of query modifications for Batch1

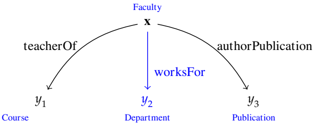
Query Modifications for Batch1



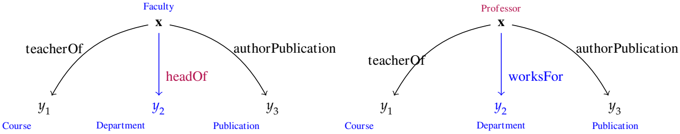
(i) In-between query



(j) Minimal Generalizations

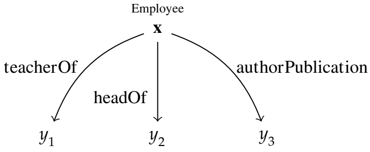


(k) Maximal Neutral Specialization

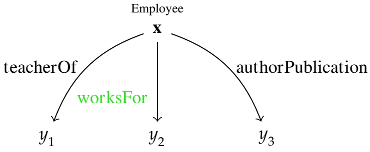


(l) Minimal Strict Specializations

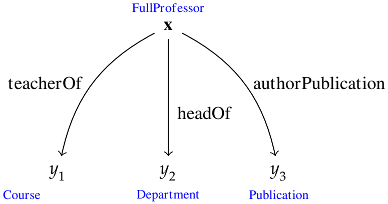
Query Modifications for Batch1



(m) In-between query



(n) Minimal Generalization

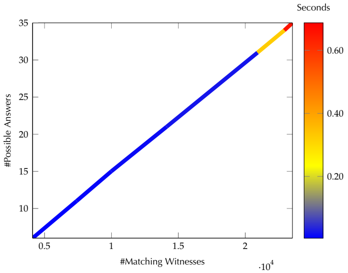


(o) Maximal Neutral Specialization

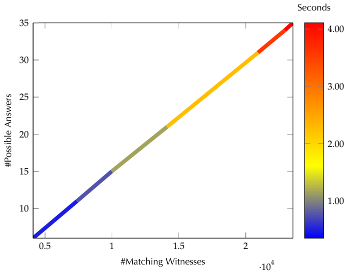
Batch2

$$q_L(\mathbf{x}) :- \text{Organization}(\mathbf{x}), \text{member}(\mathbf{x}, y_1), \text{subOrganizationOf}(\mathbf{x}, y_2)$$

$$q_U(\mathbf{x}) :- \text{Department}(\mathbf{x}), \text{Institute}(\mathbf{x}), \text{ResearchGroup}(\mathbf{x}), \text{College}(\mathbf{x}), \text{subOrganizationOf}(\mathbf{x}, y_2), \text{University}(y_2), \text{member}(\mathbf{x}, y_1), \text{FullProfessor}(y_1), \text{teacherOf}(y_1, z_1), \text{orgPublication}(\mathbf{x}, y_3), \text{publicationAuthor}(y_3, z_2), \text{ResearchAssistant}(z_2), \text{worksFor}(z_3, u_1), \text{ResearchGroup}(u_1), \text{takesCourse}(z_3, u_2), \text{GraduateCourse}(u_2), \text{Course}(u_2)$$

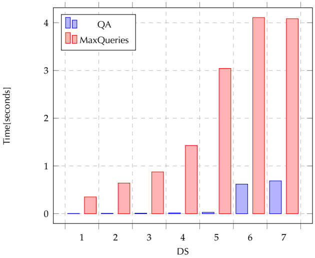


(p) Av. time for QA over Batch2

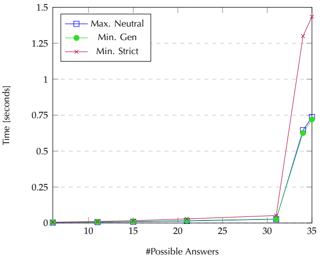


(q) Av. time for computing Maximal Queries over Batch2

Batch2

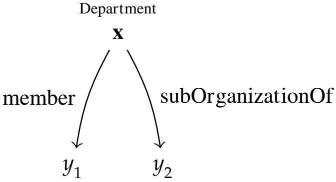


(r) QA and Max. queries performance per Data set for Batch2

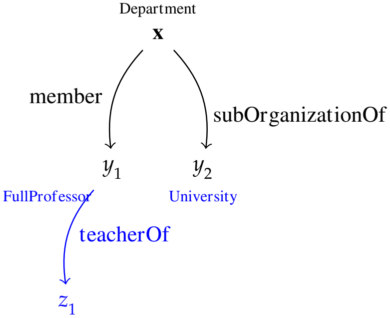


(s) Performance of query modifications for Batch2

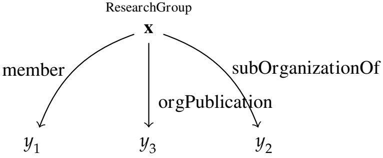
Query Modifications for Batch2



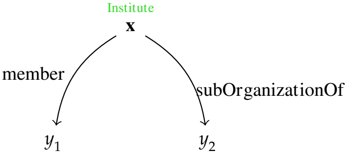
(t) In-between query



(u) Maximal Neutral Specialization



(v) In-between query



(w) Minimal Generalization

Summary

Conclusions

- Implementation can be optimized
- Many questions that remain open
    - → extend solution for other DLs
    - → extend solution for queries not exactly tree-shaped
    - → refocusing - provide the user the possibility to choose a different answer variable
- Key contributions
    - → novel perspective on query answering
    - → interesting obtained query modifications that offer more insights on the data
    - → experiments overall good performance of query answering procedure
    - → first steps towards ontology-mediated data exploration

References I

Marcelo Arenas, Bernardo Cuenca Grau, Evgeny Kharlamov, Sarunas Marciuska, and Dmitriy Zheleznyakov.
Faceted search over ontology-enhanced rdf data.
In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 939--948, 2014.

Diego Calvanese, G. De Giacomo, Domenico Lembo, Maurizio Lenzerini, and Riccardo Rosati.
Tractable reasoning and efficient query answering in description logics: The dl-lite family.
J. OF AUTOMATED REASONING, page 2007, 2007.

Enrico Franconi, Paolo Guagliardo, Sergio Tessaris, and Marco Trevisan.
Quelo: an ontology-driven query interface, 2011.

References II

Roman Kontchakov, Martin Rezk, Mariano Rodriguez-Muro, Guohui Xiao, and Michael Zakharyaschev.
Answering SPARQL queries over databases under OWL 2 QL entailment regime.
In *Proc. of International Semantic Web Conference (ISWC 2014)*, Lecture Notes in Computer Science. Springer, 2014.

Antonella Poggi, Domenico Lembo, Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati.
Linking data to ontologies.
In Stefano Spaccapietra, editor, *Journal on Data Semantics X*, volume 4900 of *Lecture Notes in Computer Science*, pages 133--173. Springer Berlin Heidelberg, 2008.

Mariano Rodriguez-Muro, Roman Kontchakov, and Michael Zakharyaschev.
Ontology-based data access: Ontop of databases.
In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*, pages 558--573, 2013.

Questions?