

QUERY ANSWERING WITH DESCRIPTION LOGIC ONTOLOGIES

Meghyn Bienvenu (*CNRS & Université de Montpellier*)

Magdalena Ortiz (*Vienna University of Technology*)

RESEARCH TRENDS IN OMQA

Lots of work on **developing** and **implementing efficient OMQA** algorithms

Focus mostly on **DL-Lite** (and related dialects):

- First algorithm **PerfectRef** proposed in mid-2000's
- Rewrites into **UCQs**, implemented in **QUONTO**
- Improved versions proposed in **REQUIEM, PRESTO, RAPID, ...**
- Some algorithms rewrite into **positive existential queries** or **Datalog programs** instead of **UCQs**
- Resulting queries are **smaller**, can be **easier to evaluate**

Tractable classes, fragments of lower complexity

Rewriting engines for other Horn DLs also developed, e.g.,

- REQUIEM and the related KYRIE cover several \mathcal{EL} dialects
- CLIPPER, and recently RAPID cover Horn-SHIQ

They usually rewrite into Datalog programs

Much attention devoted to understanding the **limits of rewritability** and **size of rewritings**

When are **polynomial rewritings** possible?

Can we **give bounds on the size** of rewritings?

Which non-DL-Lite ontologies can be **rewritten into FO-queries**?

↪ related to **non-uniform** complexity:

- study specific pairs (q, \mathcal{T}) , called **ontology-mediated queries**

Saturate the ABox using the TBox axioms

↔ a **finite version** of the canonical model

and then **evaluate** the query over the **saturated ABox**

Two approaches:

- **modify** the query **before evaluation** to ensure soundness, or
- **evaluate** and then **filter** unsound answers

First proposed for \mathcal{EL} , then also for **DL-Lite**

Extended to other dialects, richer DLs

This course: assume data given as ABox assertions (unary + binary)

Problem: how to query **existing relational data** (arbitrary arity)?

Solution: use **mapping** that specifies **relationship** between the **database relations** and the **concepts / roles** in DL vocabulary

Formally: **mapping assertions** of the form $\varphi \rightarrow \psi$ where:

- φ is a query formulated using DB relations
- ψ is a query in the DL vocabulary

Global-as-view (GAV) mappings: φ CQ, ψ atom (no quantifiers)

Handling mappings:

- apply mappings to **generate ABox, proceed as usual**
- **virtual ABox:** unfolding step to get rewriting over DB relations

OTHER RESEARCH TOPICS (NON-EXHAUSTIVE)

Beyond classical OMQA

- **inconsistency-tolerant** query answering
- **probabilistic** query answering
- **privacy-aware** query answering
- **temporal** query answering

Support for **building and maintaining** OMQA systems

- **module** extraction
- ontology **evolution**
- query **inseparability** and **emptiness**

Improving the **usability** of OMQA systems

- **interfaces** and support for **query formulation**
- **explaining** query (non-)answers
- combining **complete and incomplete information**

ZOOM: INCONSISTENCY-TOLERANT QUERY ANSWERING

In realistic settings, can expect some **errors in the data**

- ABox likely to be **inconsistent** with the TBox

Standard semantics: all tuples are returned - **not informative!**

In realistic settings, can expect some **errors in the data**

- ABox likely to be **inconsistent** with the TBox

Standard semantics: all tuples are returned - **not informative!**

Two approaches to inconsistency handling:

- **resolve the inconsistencies**
 - preferable, but not always applicable!
- live with the inconsistencies - **adopt alternative semantics**
 - **meaningful answers** to queries despite inconsistencies

EXAMPLE: WHICH ANSWERS TO RETURN?

Consider the following TBox \mathcal{T} :

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg\text{Fellow}$
 $\text{Prof} \sqsubseteq \exists\text{Teaches}$ $\exists\text{Teaches} \sqsubseteq \text{Faculty}$ $\exists\text{Teaches}^- \sqsubseteq \text{Course}$

the ABox

$\mathcal{A} = \{\text{Prof}(\text{anna}), \underline{\text{Fellow}(\text{tom})}, \text{Teaches}(\text{tom}, \text{cs101}), \underline{\text{Prof}(\text{tom})}\}$

and the query $q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

EXAMPLE: WHICH ANSWERS TO RETURN?

Consider the following TBox \mathcal{T} :

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

the ABox

$\mathcal{A} = \{\text{Prof}(\text{anna}), \underline{\text{Fellow}(\text{tom})}, \text{Teaches}(\text{tom}, \text{cs101}), \underline{\text{Prof}(\text{tom})}\}$

and the query $q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Which individuals should be returned (or not returned as answers)?

Repair: \subseteq -maximal subset of the data consistent with the ontology

- ways to achieve consistency, keeping as much information as possible

Repair: \subseteq -maximal subset of the data consistent with the ontology

- ways to achieve consistency, keeping as much information as possible

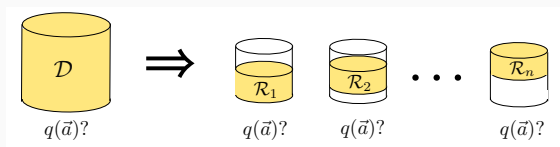
Plausible answers: hold no matter which repair is chosen

Repair: \subseteq -maximal subset of the data consistent with the ontology

- ways to achieve consistency, keeping as much information as possible

Plausible answers: hold no matter which repair is chosen

AR semantics: query each repair separately, intersect results



EXAMPLE: AR SEMANTICS

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \underline{\text{Fellow}(\text{tom})}, \text{Teaches}(\text{tom}, \text{cs101}), \underline{\text{Prof}(\text{tom})}\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

EXAMPLE: AR SEMANTICS

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \underline{\text{Fellow}(\text{tom})}, \text{Teaches}(\text{tom}, \text{cs101}), \underline{\text{Prof}(\text{tom})}\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Two repairs of \mathcal{A} w.r.t. \mathcal{T} :

$\mathcal{R}_1 = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$ drop $\text{Prof}(\text{tom})$

$\mathcal{R}_2 = \{\text{Prof}(\text{anna}), \text{Prof}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$ drop $\text{Fellow}(\text{tom})$

EXAMPLE: AR SEMANTICS

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \underline{\text{Fellow}(\text{tom})}, \text{Teaches}(\text{tom}, \text{cs101}), \underline{\text{Prof}(\text{tom})}\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Two repairs of \mathcal{A} w.r.t. \mathcal{T} :

$\mathcal{R}_1 = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$ drop $\text{Prof}(\text{tom})$

$\mathcal{R}_2 = \{\text{Prof}(\text{anna}), \text{Prof}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$ drop $\text{Fellow}(\text{tom})$

Under AR semantics:

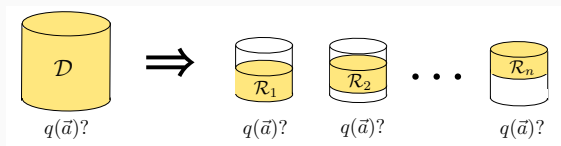
- **anna** and **tom** are both **answers** to q
- **cs101** is **not an answer**

Repair: \subseteq -maximal subset of the data consistent with the ontology

- ways to achieve consistency, keeping as much information as possible

Plausible answers: hold no matter which repair is chosen

AR semantics: query each repair separately, intersect results



Bad news: query answering under AR semantics is **intractable**
(**coNP-hard** in the size of the data)

Worse: intractable **even in very restricted settings** ($\mathcal{T} = \{A \subseteq \neg B\}$)

Idea: **approximate AR semantics** from above and below

Idea: **approximate AR semantics** from above and below

Brave semantics

possible answers

- answer required to **hold w.r.t. at least one repair**

Idea: **approximate AR semantics** from above and below

Brave semantics

possible answers

- answer required to **hold w.r.t. at least one repair**

IAR semantics

surest answers

- query the **intersection of all repairs** (surest facts)

Idea: **approximate AR semantics** from above and below

Brave semantics

possible answers

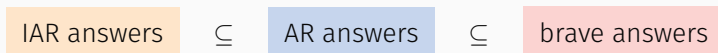
- answer required to **hold w.r.t. at least one repair**

IAR semantics

surest answers

- query the **intersection of all repairs** (surest facts)

Relationship between the semantics:



Idea: **approximate AR semantics** from above and below

Brave semantics

possible answers

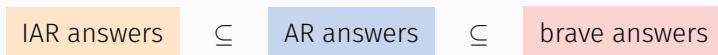
- answer required to **hold w.r.t. at least one repair**

IAR semantics

surest answers

- query the **intersection of all repairs** (surest facts)

Relationship between the semantics:



Good news: these semantics are **tractable** for DL-Lite ontologies

BACK TO THE EXAMPLE

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101}), \text{Prof}(\text{tom})\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Repairs of \mathcal{A} w.r.t. \mathcal{T} :

$\mathcal{R}_1 = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

$\mathcal{R}_2 = \{\text{Prof}(\text{anna}), \text{Prof}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

BACK TO THE EXAMPLE

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101}), \text{Prof}(\text{tom})\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Repairs of \mathcal{A} w.r.t. \mathcal{T} :

$\mathcal{R}_1 = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

$\mathcal{R}_2 = \{\text{Prof}(\text{anna}), \text{Prof}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

Brave answers:

IAR answers:

BACK TO THE EXAMPLE

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101}), \text{Prof}(\text{tom})\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Repairs of \mathcal{A} w.r.t. \mathcal{T} :

$\mathcal{R}_1 = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

$\mathcal{R}_2 = \{\text{Prof}(\text{anna}), \text{Prof}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

Brave answers:

anna

tom

IAR answers:

BACK TO THE EXAMPLE

$\text{Prof} \sqsubseteq \text{Faculty}$ $\text{Fellow} \sqsubseteq \text{Faculty}$ $\text{Prof} \sqsubseteq \neg \text{Fellow}$
 $\text{Prof} \sqsubseteq \exists \text{Teaches}$ $\exists \text{Teaches} \sqsubseteq \text{Faculty}$ $\exists \text{Teaches}^- \sqsubseteq \text{Course}$

$\mathcal{A} = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101}), \text{Prof}(\text{tom})\}$

$q(x) = \exists y. \text{Faculty}(x) \wedge \text{Teaches}(x, y)$

Repairs of \mathcal{A} w.r.t. \mathcal{T} :

$\mathcal{R}_1 = \{\text{Prof}(\text{anna}), \text{Fellow}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

$\mathcal{R}_2 = \{\text{Prof}(\text{anna}), \text{Prof}(\text{tom}), \text{Teaches}(\text{tom}, \text{cs101})\}$

Brave answers:

anna

tom

IAR answers:

anna

CQAPri first system for AR query answering in DL-Lite

Implements **hybrid approach**:

- compute IAR and brave answers **polytime** (data)
 - gives upper and lower **bounds on AR answers**
- use **SAT solvers** to **identify remaining AR answers**
- three categories of answers : **possible, likely**, (almost) **sure**

Interaction with user:

- explaining query results
 - why a possible answer? why not a sure answer?
- query-driven repairing
 - exploit user feedback to improve data quality

ZOOM: COMBINING COMPLETE AND INCOMPLETE INFORMATION

We have seen the classical **certain answer** semantics:

$$\mathcal{T} = \{ \text{BScStud} \sqsubseteq \text{Student}$$

$$\text{Student} \sqsubseteq \exists \text{attends. Course}$$

$$\text{BScStud} \sqsubseteq \forall \text{attends. } \neg \text{GradCourse} \}$$

$$q = \text{attends}(x, y)$$

We have seen the classical **certain answer** semantics:

$$\mathcal{T} = \{\text{BScStud} \sqsubseteq \text{Student}$$

$$\text{Student} \sqsubseteq \exists \text{attends.Course}$$

$$\text{BScStud} \sqsubseteq \forall \text{attends.}\neg \text{GradCourse}\}$$

$$q = \text{attends}(x, y)$$

Consider the following ABox:

$$\mathcal{A} = \{\text{BScStud}(\text{Ann})$$

$$\text{Course}(c_1)$$

$$\text{Course}(c_2)$$

$$\text{GradCourse}(c_2)\}$$

We have seen the classical **certain answer** semantics:

$$\mathcal{T} = \{\text{BScStud} \sqsubseteq \text{Student}$$

$$\text{Student} \sqsubseteq \exists \text{attends. Course}$$

$$\text{BScStud} \sqsubseteq \forall \text{attends. } \neg \text{GradCourse}\}$$

$$q = \text{attends}(x, y)$$

Consider the following ABox:

$$\mathcal{A} = \{\text{BScStud}(\text{Ann})$$

$$\text{Course}(c_1)$$

$$\text{Course}(c_2)$$

$$\text{GradCourse}(c_2)\}$$

Query Answer: \emptyset

Suppose we have **complete knowledge** about existing courses:

$$\mathcal{T} = \{\text{BScStud} \sqsubseteq \text{Student}$$

$$\text{Student} \sqsubseteq \exists \text{attends.Course}$$

$$\text{BScStud} \sqsubseteq \forall \text{attends.}\neg \text{GradCourse}\}$$

$$q = \text{attends}(x, y)$$

$$\Sigma = \{\text{Course}\}$$

Suppose we have **complete knowledge** about existing courses:

$$\begin{aligned} \mathcal{T} = \{ & \text{BScStud} \sqsubseteq \text{Student} \\ & \text{Student} \sqsubseteq \exists \text{attends.Course} \\ & \text{BScStud} \sqsubseteq \forall \text{attends.}\neg \text{GradCourse} \} \end{aligned}$$

$$q = \text{attends}(x, y)$$

$$\Sigma = \{\text{Course}\}$$

Recall the previous ABox:

$$\begin{aligned} \mathcal{A} = \{ & \text{BScStud}(\text{Ann}) \\ & \text{Course}(c_1) \\ & \text{Course}(c_2) \\ & \text{GradCourse}(c_2) \} \end{aligned}$$

Suppose we have **complete knowledge** about existing courses:

$$\begin{aligned} \mathcal{T} = \{ & \text{BScStud} \sqsubseteq \text{Student} \\ & \text{Student} \sqsubseteq \exists \text{attends.Course} \\ & \text{BScStud} \sqsubseteq \forall \text{attends.}\neg \text{GradCourse} \} \end{aligned}$$

$$q = \text{attends}(x, y)$$

$$\Sigma = \{\text{Course}\}$$

Recall the previous ABox:

$$\begin{aligned} \mathcal{A} = \{ & \text{BScStud}(\text{Ann}) \\ & \text{Course}(c_1) \\ & \text{Course}(c_2) \\ & \text{GradCourse}(c_2) \} \end{aligned}$$

$$\text{Query Answer: } \{(\text{Ann}, c_1)\}$$

DLs adopt an **open-world view**

- Standard FOL-semantics
- expresses **incomplete knowledge**, many models

A **partial closed-world view** is desirable

- use **partial completeness** to infer more answers
- meaningful when data comes from complete DB tables

Closed Predicates in DLs

- We can enrich a KB with a **set Σ of concept/role names**
- We assume those predicates are **complete**
- **Models of $(\mathcal{T}, \mathcal{A}, \Sigma)$** are models of $(\mathcal{T}, \mathcal{A})$
- Additionally, the extensions of closed predicates must be **exactly as given in \mathcal{A}**

Observation: **we need reasoning by cases**

- Even for instance checking in lightweight DLs

Observation: **we need reasoning by cases**

- Even for instance checking in lightweight DLs

Closed predicates **make reasoning harder**

Observation: **we need reasoning by cases**

- Even for instance checking in lightweight DLs

Closed predicates **make reasoning harder**

- but how much harder?

Observation: **we need reasoning by cases**

- Even for instance checking in lightweight DLs

Closed predicates **make reasoning harder**

- but how much harder?

What do we know?

- **Expressible if the DL is expressive enough**
- We can use a construct called **nominals**

$$\text{Course} \sqsubseteq \{c_1\} \sqcup \{c_2\}$$

We need a DL called *ALCO*

We need a DL called *ALCO*

$$\mathcal{L} + \text{closed predicates} \subseteq \mathcal{L} \cup \text{ALCO}$$

We need a DL called *ALCO*

$$\mathcal{L} + \text{closed predicates} \subseteq \mathcal{L} \cup \text{ALCO}$$

This allows us to infer upper bounds in **combined complexity**

We need a DL called *ALCO*

$$\mathcal{L} + \text{closed predicates} \subseteq \mathcal{L} \cup \text{ALCO}$$

This allows us to infer upper bounds in **combined complexity** but they are not so good:

- **Standard reasoning:** consistency, fact entailment **in EXP**
- **Conjunctive query answering:** **in 2EXP**

We need a DL called *ALCO*

$$\mathcal{L} + \text{closed predicates} \subseteq \mathcal{L} \cup \text{ALCO}$$

This allows us to infer upper bounds in **combined complexity** but they are not so good:

- **Standard reasoning:** consistency, fact entailment **in EXP**
- **Conjunctive query answering:** **in 2EXP**

Optimal?

COMPLEXITY OF STANDARD REASONING

	Without closed predicates	With closed predicates
DL-Lite	NLOGSPACE	NP
DL-Lite _R	NLOGSPACE	NP
\mathcal{EL}	P	EXP
\mathcal{ALCO}	EXP	EXP
\mathcal{SHOQ} , \mathcal{SHOI}	EXP	EXP

all are completeness results

COMPLEXITY OF CONJUNCTIVE QUERY ANSWERING

	Without closed predicates	With closed predicates
DL-Lite	NP	coNEXP-hard
DL-Lite _R	NP	2EXP
\mathcal{EL}	NP	2EXP
\mathcal{ALCO}	2EXP	2EXP
\mathcal{SHOQ} , \mathcal{SHOI}	2EXP	2EXP

all but red are completeness results

(q, \mathcal{T}, Σ) a query q ,
 a TBox \mathcal{T} ,
 set of closed predicates Σ

Viewed as a query language, it is **non-monotonic**

(q, \mathcal{T}, Σ) a query q ,
a TBox \mathcal{T} ,
set of closed predicates Σ

Viewed as a query language, it is **non-monotonic**

For OMQs with closed predicates, there are
polynomial rewritings into
disjunctive Datalog with stratified negation

QUESTIONS ?