## The Combined Approach to Query Answering in DL-Lite

## Roman Kontchakov ${ }^{1}$, Carsten Lutz ${ }^{2}$, David Toman ${ }^{3}$, Frank Wolter ${ }^{4}$, and Michael Zakharyaschev ${ }^{1}$

${ }^{1}$ Department of CS and Information Systems, Birkbeck College London, UK
\{roman, michael\}@dcs.bbk. ac.uk
${ }^{2}$ Fachbereich Mathematik und Informatik, Universität Bremen, Germany clu@informatik.uni-bremen.de
${ }^{3}$ D.R. Cheriton School of Computer Science, University of Waterloo, Canada david@cs.uwaterloo.ca
${ }^{4}$ Department of Computer Science, University of Liverpool, UK
frank@csc.liv.ac.uk

## Queries and Ontologies

## Ontology-based Data Access

Enriches (query answers over) explicitly represented data using background knowledge (captured using an ontology.)

## Queries and Ontologies

## Ontology-based Data Access

Enriches (query answers over) explicitly represented data using background knowledge (captured using an ontology.)

## Example

- Bob is a BOSS
- Every BOSS is an EMPloyee

List all EMPloyees $\Rightarrow$ \{Bob $\}$

## Queries and Ontologies

## Ontology-based Data Access

Enriches (query answers over) explicitly represented data using background knowledge (captured using an ontology.)

Problem: answering queries is EXPENSIVE (data complexity)
$\Rightarrow$ large data sets and (relatively) large ontologies.
$\Rightarrow$ need for lightweight ontology and query languages;

## Queries and Ontologies

## Ontology-based Data Access

Enriches (query answers over) explicitly represented data using background knowledge (captured using an ontology.)

Problem: answering queries is EXPENSIVE (data complexity)
$\Rightarrow$ large data sets and (relatively) large ontologies.
$\Rightarrow$ need for lightweight ontology and query languages;

DL-Lite (family) and conjunctive queries.
. . . introduced by [Calvanese et al.]

## Approaches to Ontology-based Data Access

Main Task
INPUT: $\quad \underbrace{\operatorname{Ontology}(\mathcal{T}), \operatorname{Data}(\mathcal{A})}_{\text {Knowledge } \operatorname{Base}(\mathcal{K})}$, and a Query $(Q)$

## OUTPUT: $\quad\{a \mid \mathcal{K} \models Q[a]\}$

## Approaches to Ontology-based Data Access

## Main Task

INPUT: $\quad \underbrace{\operatorname{Ontology}(\mathcal{T}), \operatorname{Data}(\mathcal{A})}_{\text {Knowledge } \operatorname{Base}(\mathcal{K})}$, and a Query $(Q)$

## OUTPUT: $\quad\{a \mid \mathcal{K} \models Q[a]\}$

Approaches:
(1) Reduction to standard reasoning (e.g., satisfiability)
(2) Reduction to querying a relational database
$\Rightarrow$ very good at $\{a \mid \mathcal{A} \models Q[a]\}$ for range restricted $Q$
$\ldots$ what do we do with $\mathcal{T}$ ?

## Definitions\&Background

## Definition (DL-Lite horn ${ }^{\mathcal{N}}$ )

roles: $R::=P \mid P^{-}, \quad$ concepts: $C::=\perp|A| \geq m R$.

$$
\text { where } P \in \mathrm{~N}_{\mathrm{R}}, A \in \mathrm{~N}_{\mathrm{C}} \text { and } m>0 \text {. }
$$

(1) An ontology (TBox) is a finite set $\mathcal{T}$ of concept inclusions $C_{1} \sqcap \cdots \sqcap C_{n} \sqsubseteq C$;
(2) The Data (ABox) is a finite set $\mathcal{A}$ of concept and role assertions $C(a)$ and $R(a, b)$;
(3 A Conjunctive Query (CQ): an existentially quantified finite conjunction of atoms.

## The Master Plan

## IDEA:

(1) Incorporate the background knowledge (i.e., $\mathcal{T}$ ) into the data.
$\Rightarrow$ make implicit knowledge explicit (data completion).
(2) Use the data completion (only) to answer queries
$\Rightarrow$ and use a relational system to do this efficiently.

## The Master Plan

## IDEA:

(1) Incorporate the background knowledge (i.e., $\mathcal{T}$ ) into the data.
$\Rightarrow$ make implicit knowledge explicit (data completion).
(2) Use the data completion (only) to answer queries
$\Rightarrow$ and use a relational system to do this efficiently.
Example
$\mathcal{T}=\{B O S S \sqsubseteq E M P\}, \mathcal{A}=\{B O S S(B o b)\}, \quad Q \equiv E M P(x)$
(1) $\mathcal{I}_{\mathcal{K}}=\{B O S S(B o b), E M P(B o b)\}$
(data completion)
(2) $Q\left(\mathcal{I}_{\mathcal{K}}\right)=\{\mathrm{Bob}\}$
(relational query)

## The Master Plan

## IDEA:

(1) Incorporate the background knowledge (i.e., $\mathcal{T}$ ) into the data.
$\Rightarrow$ make implicit knowledge explicit (data completion).
(2) Use the data completion (only) to answer queries
$\Rightarrow$ and use a relational system to do this efficiently.

Issues:
(1) How to complete the data?

Naive unfolding of $\mathcal{T}$ : large/infinite (due to existentials)
$\Rightarrow$ we define a canonical interpretation $\mathcal{I}_{\mathcal{K}}$ (representatives)
(2) Can we then use the original Conjunctive Query?

Not directly: $Q\left(\mathcal{I}_{\mathcal{K}}\right)$ can produce "spurious matches"
$\Rightarrow$ we eliminate the spurious matches by rewriting the query (independently of $\mathcal{T}$ and $\mathcal{A}$ )

## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$

$$
\begin{aligned}
& A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup \\
& P^{\mathcal{I}_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup
\end{aligned}
$$

## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$

$$
\begin{aligned}
& A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup \\
& P^{\mathcal{I}_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup \\
& \left\{\left(d, c_{P}\right) \in \Delta^{\mathcal{I}_{\mathcal{K}}} \times \mathrm{N}_{1}^{\mathcal{T}} \mid d \leadsto c_{P}\right\} \cup\left\{\left(c_{P^{-}}, d\right) \in \mathrm{N}_{1}^{\mathcal{T}} \times \Delta^{\mathcal{I}_{\mathcal{K}}} \mid d \leadsto c_{P^{-}}\right\} \\
& \quad \ldots c_{R^{\prime}} \text { 's only used "when necessary" (for generating roles) }
\end{aligned}
$$

## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$

$$
\begin{aligned}
& A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup\left\{c_{R} \in \Delta^{I_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^{-} \sqsubseteq A\right\}, \\
& P^{I_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup \\
& \quad\left\{\left(d, c_{P}\right) \in \Delta^{\mathcal{I}_{\mathcal{K}}} \times \mathrm{N}_{1}^{\mathcal{T}_{1}} \mid d \leadsto c_{P}\right\} \cup\left\{\left(c_{P^{-}}, d\right) \in N_{1}^{\mathcal{T}} \times \Delta^{\mathcal{I}_{\mathcal{K}}} \mid d \leadsto c_{P-}\right\}
\end{aligned}
$$

$\ldots c_{R}$ 's only used "when necessary" (for generating roles)

## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$
$A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup\left\{c_{R} \in \Delta^{\mathcal{I}_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^{-} \sqsubseteq A\right\}$,
$P^{\mathcal{I}_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup$ $\left\{\left(d, c_{P}\right) \in \Delta^{\mathcal{I}_{\mathcal{K}}} \times N_{1}^{\mathcal{T}} \mid d \leadsto c_{P}\right\} \cup\left\{\left(c_{P^{-}}, d\right) \in N_{1}^{\mathcal{T}} \times \Delta^{\mathcal{I}_{\mathcal{K}}} \mid d \leadsto c_{P-}\right\}$ $\ldots c_{R}$ 's only used "when necessary" (for generating roles)

## Example

$\mathcal{T}=\left\{E M P \sqsubseteq \exists M A N A G E S, \exists M A N A G E S^{-} \sqsubseteq B O S S, B O S S \sqsubseteq E M P\right\}$
$\mathcal{A}=\{E M P($ Bob $), E M P($ Sue $)\}$
Then $E M P^{\mathcal{I}_{\mathcal{K}}}=\left\{\right.$ Bob, Sue, $\left.c_{M}\right\}, B O S S^{\mathcal{I}_{\mathcal{K}}}=\left\{c_{M}\right\}$, and MANAGES $^{\mathcal{I}_{\mathcal{K}}}=\left\{\left(\right.\right.$ Bob, $\left.c_{M}\right),\left(\right.$ Sue,$\left.\left.c_{M}\right),\left(c_{M}, c_{M}\right)\right\}$.

## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$
$A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup\left\{c_{R} \in \Delta^{\mathcal{I}_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^{-} \sqsubseteq A\right\}$,
$P^{\mathcal{I}_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup$

$$
\left\{\left(d, c_{P}\right) \in \Delta^{\mathcal{I}_{\mathcal{K}}} \times \mathbf{N}_{1}^{\mathcal{T}} \mid d \leadsto c_{P}\right\} \cup\left\{\left(c_{P-}, d\right) \in \mathbf{N}_{1}^{\mathcal{T}} \times \Delta^{\mathcal{I}_{\mathcal{K}}} \mid d \leadsto c_{P-}\right\}
$$

$\ldots c_{R}$ 's only used "when necessary" (for generating roles)

## Example

$\mathcal{T}=\left\{E M P \sqsubseteq \exists M A N A G E S, \exists M A N A G E S^{-} \sqsubseteq B O S S, B O S S \sqsubseteq E M P\right\}$
$\mathcal{A}=\{E M P($ Bob $), E M P($ Sue $)\}$
Then $E M P^{\mathcal{I}_{\mathcal{K}}}=\left\{\right.$ Bob, Sue, $\left.c_{M}\right\}, B O S S^{\mathcal{I}_{\mathcal{K}}}=\left\{c_{M}\right\}$, and MANAGES $^{\mathcal{I}_{\mathcal{K}}}=\left\{\left(\right.\right.$ Bob, $\left.c_{M}\right),\left(\right.$ Sue,$\left.\left.c_{M}\right),\left(c_{M}, c_{M}\right)\right\}$.
$\mathcal{I}_{\mathcal{K}}$ is NOT model of $(\mathcal{T}, \mathcal{A})$ in general.

## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$
$A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup\left\{C_{R} \in \Delta^{\mathcal{I}_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^{-} \sqsubseteq A\right\}$,
$P^{\mathcal{I}_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup$
$\left\{\left(d, c_{P}\right) \in \Delta^{\mathcal{I}_{\mathcal{K}}} \times N_{1}^{\mathcal{T}} \mid d \leadsto c_{P}\right\} \cup\left\{\left(c_{P-}, d\right) \in N_{1}^{\mathcal{T}} \times \Delta^{\mathcal{I}_{\mathcal{K}}} \mid d \leadsto c_{P-}\right\}$
$\ldots c_{R}$ 's only used "when necessary" (for generating roles)

## Lemma

There are queries

- $q_{A}^{\mathcal{T}}$ s.t. $\operatorname{ans}\left(q_{A}^{\mathcal{T}}, \mathcal{A}\right)=A^{\mathcal{I}_{\mathcal{K}}}$, and
- $q_{P}^{\mathcal{T}}$ s.t. $\operatorname{ans}\left(q_{P}^{\mathcal{T}}, \mathcal{A}\right)=P^{\mathcal{I}_{\mathcal{K}}}$
for every $K B(\mathcal{T}, \mathcal{A})$ and primitive concept $A$ and role $P$.


## Canonical Interpretations

ABox completion: the Canonical Interpretation $\mathcal{I}_{\mathcal{K}}$
$A^{\mathcal{I}_{\mathcal{K}}}=\{a \in \operatorname{Ind}(\mathcal{A}) \mid \mathcal{K} \models A(a)\} \cup\left\{C_{R} \in \Delta^{\mathcal{I}_{\mathcal{K}}} \mid \mathcal{T} \models \exists R^{-} \sqsubseteq A\right\}$,
$P^{\mathcal{I}_{\mathcal{K}}}=\{(a, b) \in \operatorname{Ind}(\mathcal{A}) \times \operatorname{Ind}(\mathcal{A}) \mid P(a, b) \in \mathcal{A}\} \cup$

$$
\left\{\left(d, c_{P}\right) \in \Delta^{\mathcal{I}_{\mathcal{K}}} \times \mathbf{N}_{1}^{\mathcal{T}} \mid d \leadsto c_{P}\right\} \cup\left\{\left(c_{P-}, d\right) \in \mathbf{N}_{1}^{\mathcal{T}} \times \Delta^{\mathcal{I}_{\mathcal{K}}} \mid d \leadsto c_{P-}\right\}
$$

$\ldots c_{R}$ 's only used "when necessary" (for generating roles)

## Lemma

There are queries

- $q_{A}^{\mathcal{T}}$ s.t. $\operatorname{ans}\left(q_{A}^{\mathcal{T}}, \mathcal{A}\right)=A^{\mathcal{I}_{\mathcal{K}}}$, and
- $q_{P}^{\mathcal{T}}$ s.t. $\operatorname{ans}\left(q_{P}^{\mathcal{T}}, \mathcal{A}\right)=P^{\mathcal{I}_{\mathcal{K}}}$
for every $K B(\mathcal{T}, \mathcal{A})$ and primitive concept $A$ and role $P$.
free consistency test: $q_{\perp}^{\mathcal{L}}(\mathcal{A})=\emptyset$


## Query Rewriting

## Example

$\mathcal{T}=\left\{E M P \sqsubseteq \exists M A N A G E S, \exists M A N A G E S^{-} \sqsubseteq B O S S, B O S S \sqsubseteq E M P\right\}$ $\mathcal{A}=\{E M P($ Bob $), E M P($ Sue $)\}$

Queries:
(1) $\exists v \cdot \operatorname{MANAGES}(v, v)$
(2 $\exists y \cdot \operatorname{MANAGES}(x, y) \wedge \operatorname{MANAGES}(z, y)$

## Query Rewriting

## Example

$\mathcal{T}=\left\{E M P \sqsubseteq \exists M A N A G E S, \exists M A N A G E S^{-} \sqsubseteq B O S S, B O S S \sqsubseteq E M P\right\}$
$\mathcal{A}=\{E M P($ Bob $), E M P($ Sue $)\}$
Queries:
(1) $\exists v \cdot \operatorname{MANAGES}(v, v)$
(2 $\exists y \cdot \operatorname{MANAGES}(x, y) \wedge \operatorname{MANAGES}(z, y)$


## Query Rewriting

```
Example
T = {EMP\sqsubseteq \existsMANAGES, \existsMANAGES'` GOSS,BOSS\sqsubseteq EMP}
A ={EMP(Bob),EMP(Sue) }
```

Queries:
(1) $\exists v \cdot \operatorname{MANAGES}(v, v)$
(2 $\exists y \cdot \operatorname{MANAGES}(x, y) \wedge \operatorname{MANAGES}(z, y)$

## Query Rewriting

$$
\exists \bar{u} . \varphi \quad \mapsto \quad \exists \bar{u} \cdot \varphi \wedge \varphi_{1} \wedge \varphi_{2} \wedge \varphi_{3}
$$

where $\quad \varphi_{1}$ eliminates answers containing $c_{R}$ 's;
$\varphi_{2}$ eliminates problem (1) above; and
$\varphi_{3}$ eliminates problem (2) above.
selections in SQL

## UNA or not UNA

So far: all results are assuming UNA (the Unique Name Assumption)
$\Rightarrow$ also assumed by the underlying relational technology.

## UNA or not UNA

So far: all results are assuming UNA (the Unique Name Assumption) $\Rightarrow$ also assumed by the underlying relational technology.

BUT OWL does NOT adopt UNA. . .

## UNA or not UNA

So far: all results are assuming UNA (the Unique Name Assumption)
$\Rightarrow$ also assumed by the underlying relational technology.

BUT OWL does NOT adopt UNA. . .

## What happens without UNA?

DL-Lite ${ }_{\text {core }}^{\mathcal{N}}$ data complexity: coNP (Artale et al. 2009)
DL-Lite ${ }_{\text {horn }}^{\mathcal{F}}$ data complexity: PTIME-complete
$\Rightarrow$ explicit account of equality (via an auxiliary relation eq)
$\Rightarrow$ added to the construction of $\mathcal{I}_{\mathcal{K}}$ (doesn't affect queries)

## Experiments

## Ontologies:

Galen-lite ( 2733 concepts, 207 roles, 4888 axioms)
Core (81 concept names, 58 roles, and 381 axioms) Stockexchange (17 concepts, 12 roles, 62 axioms) University ( 31 concepts, 25 roles, 103 axioms)
System:
DB2-Express version 9.5 running on Intel Core 2 Duo 2.5 GHz CPU, 4GB memory and 500GB storage under Linux 2.6.28.
Queries:
Conjunctive queries with 3-6 atoms in their bodies, e.g.,

```
Q1 (x):-horn(x), hasstate(x,y), cellmorphologystate(y).
Q2 (x):-shortbone(x), hasstate (x,y), cellmorphologystate (y).
Q3(x):-tissue(x), hasstate (x,y),temporalunit (y).
Q4(x):-protozoa(x), contains(x,y),metal(y),
    contains(x,z),steroid(z).
```


## Experiments (results)

|  | $\begin{aligned} & \text { Ind } \\ & \text { (in K) } \end{aligned}$ | ABox size (in M) |  | query |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | original | canonical | Q1 |  |  | Q2 |  |  |
|  |  | CA RA | CA RA | UN | RW | QO | UN | RW | QO |
| Galen-Lite | 20 | 2.02 .0 | 9.93 .7 | 0.02 | 0.04 | 13.69 | 0.02 | 0.08 | 1.65 |
|  | 50 | $5.0 \quad 5.0$ | 24.89 .3 | 0.04 | 0.55 | 14.39 | 0.05 | 0.19 | 2.21 |
|  | 70 | 10.010 .0 | 43.015 .4 | 0.03 | 0.76 | 17.56 | 0.11 | 0.55 | 3.01 |
|  | 100 | 20.020 .0 | 75.025 .8 | 0.05 | 0.87 | 23.86 | 0.14 | 0.76 | 6.55 |


| Q3 |  |  | Q4 |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| UN | RW | QO | UN | RW | QO |
| 0.02 | 0.11 | 1 m | 28 | 0.12 | 0.22 |
| 16 m | 11 |  |  |  |  |
| 0.03 | 0.28 | 51.39 | 0.11 | 0.43 | 13 m 26 |
| 0.06 | 0.73 | 1 m | 11 | 0.15 | 0.63 |
| 13 m 00 |  |  |  |  |  |
| 0.12 | 0.95 | 1 m 31 | 0.18 | 1.52 | 16 m 23 |

Legend:
CA-number of concept assertions; RA-number of role assertions
UN-original query; RW-canonical interpretation; QO-QuOnto system

## Summary of Contributions

## Contributions

(1) Combined approach to query answering in DL-Lite
$\Rightarrow$ efficiency gains in comparison with pure rewriting,
$\Rightarrow$ non-UNA in DL-Lite ${ }^{\mathcal{F}}$ can be supported.
(2) Polynomial rewriting for DL-Lite ${ }_{\text {core }}^{\mathcal{F}}$.

## Summary of Contributions

## Contributions

(1) Combined approach to query answering in DL-Lite
$\Rightarrow$ efficiency gains in comparison with pure rewriting,
$\Rightarrow$ non-UNA in DL-Lite ${ }^{\mathcal{F}}$ can be supported.
(2) Polynomial rewriting for DL-Lite ${ }_{\text {core }}^{\mathcal{F}}$.

## Future Work

(1) Better integration with role hierarchies
$\Rightarrow$ we can do this efficiently (but not by poly-sized query)
(2) Incremental update of the canonical interpretation
$\Rightarrow$ using techniques for incremental view maintenance

