Reasoning about *DL-Lite* ontologies

Michael Zakharyaschev

School of Computer Science and Information Systems

Birkbeck College, London

http://www.dcs.bbk.ac.uk/~michael

Joint work with A. Artale, D. Calvanese, R. Kontchakov, F. Wolter

What is **the** ontology language?

- What is **the** natural language? Italian, Latin, English, ... ?
- What is **the** logic? Classical, intuitionistic, ... , 1-order, 2-order, ... ?
- What is **the** modal logic? $\mathcal{S}5$, $\mathcal{S}4$, $\mathcal{K}$, ... , 1-order, ... ?

A naïve 'ontology' of ontology languages

- **'Engineering:'** OWL, RDF/S, OIL, DAML, DAML+OIL, ...
- **Mathematical logic:**
  - Description logic: OWL-DL ($\mathcal{SHOIN}(\mathbf{D})$, $\mathcal{SROIQ}(\mathbf{D})$), OWL-Lite ($\mathcal{SHIF}(\mathbf{D})$), ... , $\mathcal{EL}$ family, *DL-Lite* family
  - First-order logic: Prolog, Prolog + DL, ...
  - Fuzzy logic: ...
  - ...

Open problem: design an ontology of ontology languages and reasoners

DL 'islands of tractability': $\mathcal{EL}$ family

- A number of huge life science ontologies (such as SNOMED CT, NCI) can be represented in (mild extensions of) DL $\mathcal{EL}$ with concepts

$$C \quad ::= \quad \top \quad | \quad A \quad | \quad C_1 \sqcap C_2 \quad | \quad \exists R.C$$

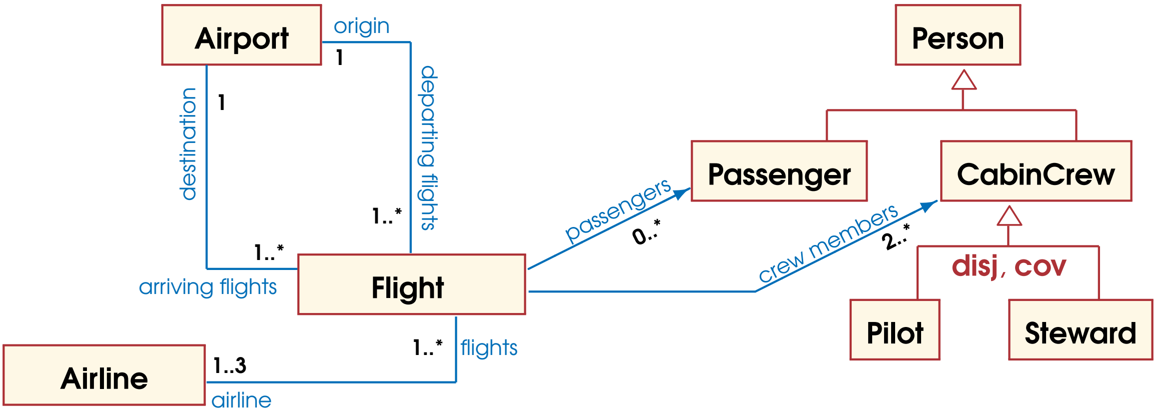
Theorem (Dresden) The problem '$\{C_i \sqsubseteq D_i \mid i \in I\} \models C \sqsubseteq D$?' is in P.

- This theorem holds for the extensions of $\mathcal{EL}$ with:
  - Role inclusions $R_1 \circ R_2 \circ \cdots \circ R_n \sqsubseteq R$
  - Nominals
  - Concrete domains

Atomic negations, inverse roles, (non-qualified) number restrictions, functional roles $\rightsquigarrow$ EXPTIME-complete reasoning

DL 'islands of tractability': *DL-Lite* family

- Reasoning about conceptual database schemas gives rise to an 'orthogonal' family of DLs



Translating into DL:

$\exists \text{passengers}.\top \sqsubseteq \text{Flight}$

$\exists \text{passengers}^{-}.\top \sqsubseteq \text{Passenger}$

$\geq 2\ \text{origin}.\top \sqsubseteq \bot$

$\text{Flight} \sqsubseteq\ \geq 2\ \text{crew-members}.\top$

$\text{CabinCrew} \sqsubseteq \text{Person}$

$\text{CabinCrew} \sqsubseteq \text{Pilot} \sqcup \text{Steward}$

$\text{Pilot} \sqcap \text{Steward} \sqsubseteq \bot$

$\dots$

DL ‘islands of tractability’: *DL-Lite* family (cont.)

1. *DL-Lite*$_{bool}$ (captures full ER) **NP**-complete

$$R ::= P \mid P^-$$
$$B ::= \bot \mid A \mid \geq qR$$
$$C ::= B \mid \neg C \mid C_1 \sqcap C_2$$

TBox axioms $C_1 \sqsubseteq C_2$, ABox assertions: $a : C$, aRb

2. *DL-Lite*$_{horn}$ **P**-complete

TBox axioms $B_1 \sqcap \cdots \sqcap B_n \sqsubseteq B$

3. *DL-Lite*$_{krom}$ (ER without covering constraints, e.g., $B \sqsubseteq B_1 \sqcup B_2$) **NLogSpace**-complete

TBox axioms $B_1 \sqsubseteq B_2$ (subclass) $B_1 \sqsubseteq \neg B_2$ (disjointness) $\neg B_1 \sqsubseteq B_2$

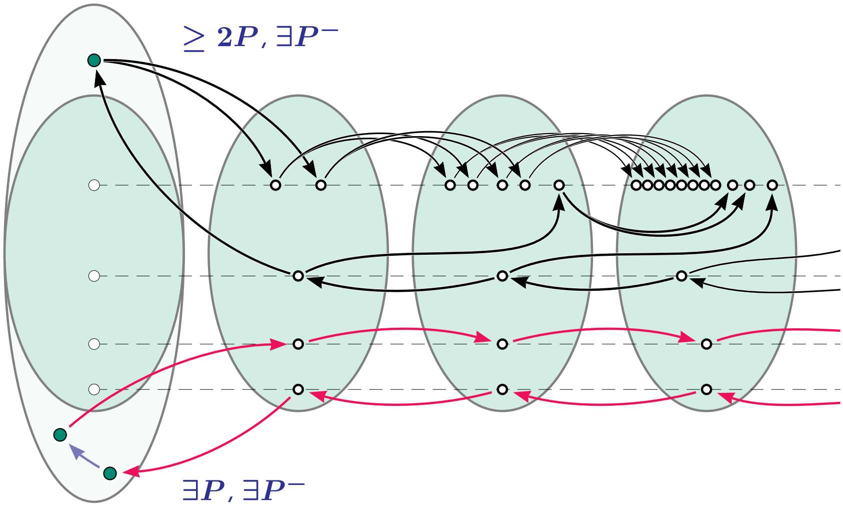
NB: these complexity results are closely related to complexity of reasoning in fragments of propositional logic

DL 'islands of tractability': *DL-Lite* family (cont.)

Theorem The satisfiability problem for *DL-Lite*$_{bool}$ knowledge bases is **NP**-complete

<u>Proof</u> by embedding into the 1-variable fragment of first-order logic (= $\mathcal{S}5$)

DL-Lite$_{bool}$ can only make statements about the whole **domain** and **range** of a role



$(\exists P)^{\mathcal{I}} \neq \emptyset$

iff

$(\exists P^-)^{\mathcal{I}} \neq \emptyset$

No fmp, but only **linear** number of (domain and range) witnesses needed **!**

DL-Lite with role hierarchies

DL-Lite$^{\mathcal{R}}$ = *DL-Lite* + role inclusions $(R_1 \sqsubseteq R_2)$

> **Theorem** Satisfiability of *DL-Lite*$^{\mathcal{R}}_{horn}$ KBs is ExpTime-complete

Proof by simulating alternating Turing machines with polynomial tapes:

- $C(q, i, a, j, b)$ — '$\mathcal{M}$ is in state q, scans cell i with symbol a, and cell j contains b'
- **Initialisation:** ABox $\mathcal{A} = \{x : C(q_1, 1, a_1, j, a_j) \mid j \leq p(n)\}$, $\vec{a} = (a_1, \dots, a_n)$
- **Transition** $(q, a) \rightsquigarrow_{\mathcal{M}} (q', a', \rightarrow)$: $C(q, i, a, i, a) \sqsubseteq \exists H_k(q', i+1)$,
  $C(q, i, a, i, a) \sqsubseteq \exists L_k(i, a')$, $C(q, i, a, j, b) \sqsubseteq \exists L_k(j, b)$, for all $j \neq i$, $k = 0, 1$
- **Synchronisation:** $\geq 2\, T_k \sqsubseteq \bot$, $L_k(j, b) \sqsubseteq T_k$, $H_k(q, i) \sqsubseteq T_k$, for $k = 0, 1$
  $\exists L_k^-(i, a) \sqcap \exists L_k^-(j, b) \sqcap \exists H_k^-(q, i) \sqsubseteq C(q, i, a, j, b)$, for all (q, i, a, j, b)
- **Acceptance:** $C(h, i, a, j, b) \sqsubseteq \exists S_\forall^-$, h the halting (existential) state
  $\geq 2\, S_\forall \sqsubseteq \exists S_\exists^-$, $\exists S_\exists \sqsubseteq \exists S_\forall^-$, $S_\exists \sqsubseteq T$, $S_\forall \sqsubseteq T$
  $T_k \sqsubseteq T$, $\geq 2\, T^- \sqsubseteq \bot$

> $\mathcal{M}$ accepts $\vec{a}$ iff $(\mathcal{T}, \mathcal{A} \cup \{x : \neg\exists S_\exists\})$ is not satisfiable

DL-Lite with role hierarchies (cont.)

Theorem Satisfiability of *DL-Lite*$^{\mathcal{R}}_{krom}$ $\cap$ *DL-Lite*$^{\mathcal{R}}_{horn}$ KBs is EXPTIME-complete

Proof role inclusions + number restrictions can simulate $A \sqcap B \sqcap C \sqsubseteq D$:

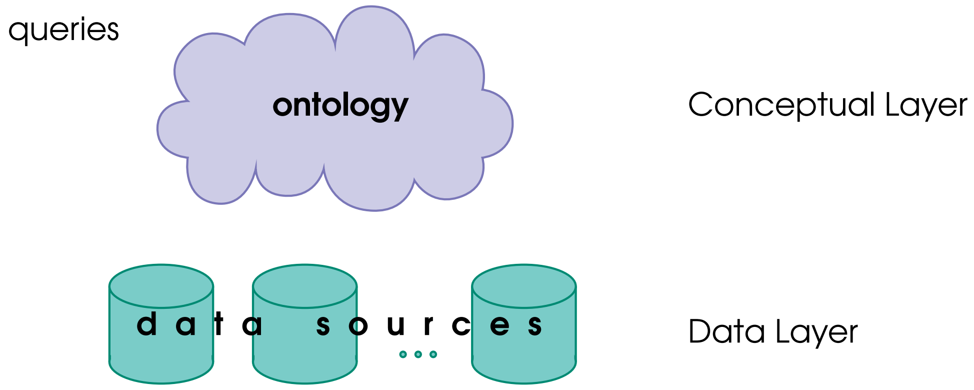
- $A \sqsubseteq \exists R_A, \quad B \sqsubseteq \exists R_B, \quad C \sqsubseteq \exists R_C$
- $\exists R_A^- \sqsubseteq \neg \exists R_B^-, \quad \exists R_A^- \sqsubseteq \neg \exists R_C^-, \quad \exists R_B^- \sqsubseteq \neg \exists R_C^-$
- $R_A \sqsubseteq S, \quad R_B \sqsubseteq S, \quad R_C \sqsubseteq S$
- $\geq 3\, S \sqsubseteq D$

Open problem: What is the complexity of *DL-Lite*$^{\mathcal{R}}_{krom}$ with functionality constraints?
(*DL-Lite*$^{\mathcal{R}}_{horn}$ with functionality constraints is EXPTIME-complete)

Ontology-based data access

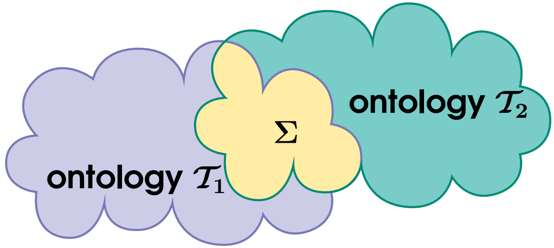
Aim: to achieve **logical transparency** in accessing data

- hide from 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



NB: In fact, this is a form of Data Integration
with a rich conceptual description as the global view

Modularity problem



- If a set of new concepts names, roles, and axioms is added to $\mathcal{T}_1$, does it affect the meaning of a set Σ of concepts and roles names defined in $\mathcal{T}_1$?
- When importing an ontology, do we change the meaning of its vocabulary?
- Do $\mathcal{T}_1$ and $\mathcal{T}_2$ say the same about Σ?

  Can I always import $\mathcal{T}_2$ instead of $\mathcal{T}_1$ into my ontology?

Changing the meaning of Σ could mean:

– changing the classification of Σ-concepts
– changing the set of derivable implications between complex concepts over Σ
– changing the set of answers to (conjunctive) queries over Σ
– ...

Deductive conservative extensions

$\mathcal{T}_1 \cup \mathcal{T}_2$ is a **deductive conservative extension** of $\mathcal{T}_1$ w.r.t. Σ if,
for every general concept inclusion $C_1 \sqsubseteq C_2$ with $\mathbf{sig}(C_1 \sqsubseteq C_2) \subseteq \Sigma$,

$$\mathcal{T}_1 \models C_1 \sqsubseteq C_2 \qquad \text{iff} \qquad \mathcal{T}_1 \cup \mathcal{T}_2 \models C_1 \sqsubseteq C_2$$

Reasoning service: is $\mathcal{T}_1 \cup \mathcal{T}_2$ a deductive conservative extension of $\mathcal{T}_1$ w.r.t. Σ?
If not, return ‘interesting’ counterexamples.

Complexity of deciding deductive conservative extensions

Horn-fragment of propositional logic	—	coNP-complete
Propositional logic	—	Π^p_2-complete ($\forall \vec{p}\, \mathcal{T}_1 \rightarrow \exists \vec{q}\, \mathcal{T}_2$)
DL-Lite$_{horn}$	—	coNP-complete
DL-Lite$_{bool}$	—	Π^p_2-complete (see next slide)
$\mathcal{EL}$	—	ExpTime-complete
$\mathcal{ALC}$, $\mathcal{ALCQI}$	—	2ExpTime-complete
$\mathcal{ALCQIO}$	—	undecidable

Semantic criterion of deductive conservativity for *DL-Lite*$_{bool}$

$\mathcal{T}_1 \cup \mathcal{T}_2$ is a deductive conservative extension of $\mathcal{T}_1$ w.r.t. Σ iff
every $\mathcal{T}_1$-consistent ΣQ-type is also $\mathcal{T}_1 \cup \mathcal{T}_2$-consistent,
where Q is the set of all numerical parameters from $\mathcal{T}_1 \cup \mathcal{T}_2$

Can be expressed as a **QBF** of the form $\forall \vec{p}\, \exists \vec{q}\, \varphi(\vec{p}, \vec{q})$ (and so is Π_2^p-complete):

$$\forall \vec{t}_\Sigma \left[\mathit{Real}_{\mathcal{T}_1}(\vec{t}_\Sigma) \to \mathit{Real}_{\mathcal{T}_1 \cup \mathcal{T}_2}(\vec{t}_\Sigma) \right]$$

$$\mathit{Real}_{\mathcal{T}}(\vec{t}_\Sigma) = \exists \vec{e}_{\mathcal{T}} \left[\mathit{Witns}_{\mathcal{T},\top}(\vec{t}_\Sigma, \vec{e}_{\mathcal{T}}) \land \bigwedge_{P_i \in \mathit{role}(\mathcal{T})} \Big(\neg e_i \to \mathit{Witns}_{\mathcal{T}, \geq 1\, P_i}(\vec{t}_\Sigma, \vec{e}_{\mathcal{T}}) \land \mathit{Witns}_{\mathcal{T}, \geq 1\, P_i^-}(\vec{t}_\Sigma, \vec{e}_{\mathcal{T}}) \Big) \right]$$

($\vec{e}_{\mathcal{T}}$ — 'empty $\mathcal{T}$-roles)

$$\mathit{Witns}_{\mathcal{T},C}(\vec{t}_\Sigma, \vec{e}_{\mathcal{T}}) = \exists \vec{t}_{\mathcal{T}} \left[\bigwedge_{P_i \in \mathit{role}(\mathcal{T})} (e_i \to \neg(\geq 1 P_i) \land \neg(\geq 1 P_i^-)) \land C \land \bigwedge_{\substack{R_i \in \mathit{role}^{\pm}(\mathcal{T}) \\ q, q' \in Q, q \prec q'}} ((\geq q'\, R_i) \to (\geq q\, R_i)) \land \bigwedge_{C_1 \sqsubseteq C_2 \in \mathcal{T}} (C_1 \to C_2) \right]$$

Query conservative extensions

$\mathcal{T}_1 \cup \mathcal{T}_2$ is a **query conservative extension** of $\mathcal{T}_1$ w.r.t. Σ if,

- for **every** ABox $\mathcal{A}$ with $\mathbf{sig}(\mathcal{A}) \subseteq \Sigma$,
- for **every** positive existential query q with $\mathbf{sig}(q) \subseteq \Sigma$,
- and every tuple $\vec{a}$ of object names from $\mathcal{A}$, we have:

$$(\mathcal{T}_1, \mathcal{A}) \models q(\vec{a}) \qquad \text{iff} \qquad (\mathcal{T}_1 \cup \mathcal{T}_2, \mathcal{A}) \models q(\vec{a}).$$

Every query conservative extension is a deductive conservative extension,
but not the other way round:

Let $\mathcal{T}_1 = \emptyset$, $\mathcal{T}_2 = \{A \sqsubseteq \exists P,\ \exists P^- \sqsubseteq B\}$, $\Sigma = \{A, B\}$.

$\mathcal{T}_2$ is deductive but not query conservative extension of $\mathcal{T}_1$ w.r.t. Σ because

for $\mathcal{A} = \{A(a)\}$, $q = \exists y\, B(y)$, $(\mathcal{T}_1, \mathcal{A}) \not\models q$, but $(\mathcal{T}_2, \mathcal{A}) \models q$

Deciding query conservativity

- Deciding query conservativity is **coNP-complete** for *DL-Lite*$_{horn}$ and Π_2^p**-complete** for *DL-Lite*$_{bool}$

$\mathcal{T}_1 \cup \mathcal{T}_2$ is a query conservative extension of $\mathcal{T}_1$ w.r.t. Σ iff every precisely $\mathcal{T}_1$-realisable set of ΣQ-types is also precisely $\mathcal{T}_1 \cup \mathcal{T}_2$-realisable

Can also be expressed as a **QBF** of the form $\forall \vec{p}\, \exists \vec{q}\, \varphi(\vec{p}, \vec{q})$

(where N is the number of roles in $\mathcal{T}_1 \cup \mathcal{T}_2$)

$$\forall\, \vec{t}^{\,0}_\Sigma, \dots, \vec{t}^{\,N}_\Sigma \;\left[\textit{P-Real}^N_{\mathcal{T}_1}(\vec{t}^{\,0}_\Sigma, \dots, \vec{t}^{\,N}_\Sigma) \rightarrow \textit{P-Real}^N_{\mathcal{T}_1 \cup \mathcal{T}_2}(\vec{t}^{\,0}_\Sigma, \dots, \vec{t}^{\,N}_\Sigma)\right]$$

$$\textit{P-Real}^N_{\mathcal{T}}(\vec{t}^{\,0}_\Sigma, \dots, \vec{t}^{\,N}_\Sigma) \quad = \quad \exists\, \vec{e}_{\mathcal{T}} \;\Big[\bigwedge_{j=0}^{N} \textit{Witns}_{\mathcal{T},\top}(\vec{t}^{\,j}_\Sigma, \vec{e}_{\mathcal{T}}) \quad \wedge$$

$$\bigwedge_{P_i \in \textit{role}(\mathcal{T})} \Big(\neg e_i \;\rightarrow\; \bigvee_{j=0}^{N} \textit{Witns}_{\mathcal{T},\geq 1\, P_i}(\vec{t}^{\,j}_\Sigma, \vec{e}_{\mathcal{T}}) \wedge \bigvee_{j=0}^{N} \textit{Witns}_{\mathcal{T},\geq 1\, P_i^-}(\vec{t}^{\,j}_\Sigma, \vec{e}_{\mathcal{T}})\Big)\Big]$$

Strong deductive/query conservative extensions

Do $\mathcal{T}_1$ and $\mathcal{T}_1 \cup \mathcal{T}_2$ say the same about Σ

whenever they are imported into an ontology?

> $\mathcal{T}_1 \cup \mathcal{T}_2$ is a **strong deductive/query conservative extension** of $\mathcal{T}_1$ w.r.t. Σ if,
>
> **for every TBox** $\mathcal{T}$ with $\mathbf{sig}(\mathcal{T}) \cap \mathbf{sig}(\mathcal{T}_1 \cup \mathcal{T}_2) \subseteq \Sigma$,
>
> $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}$ is a deductive/query conservative extension of $\mathcal{T}_1 \cup \mathcal{T}$ w.r.t. Σ.

Not every deductive conservative extension is a strong one:

Let $\mathcal{T}_1 = \emptyset$, $\mathcal{T}_2 = \{A \sqsubseteq \exists R,\ A \sqcap \exists R^- \sqsubseteq \bot\}$, $\Sigma = \{A\}$

Then $\mathcal{T}_1 \cup \mathcal{T}_2$ is a deductive conservative extension of $\mathcal{T}_1$ w.r.t. Σ.

But for $\mathcal{T} = \{\top \sqsubseteq A\}$,

$$\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T} \models \top \sqsubseteq \bot, \qquad \mathcal{T}_1 \cup \mathcal{T} \not\models \top \sqsubseteq \bot$$

Final results

DL-Lite$_{horn}$: deductive $\lneq$ query $\lneq$ strong deductive $\equiv$ strong query

DL-Lite$_{bool}$: deductive $\lneq$ query $\equiv$ strong deductive $\equiv$ strong query

- All the decision problems for *DL-Lite*$_{horn}$ are **coNP-complete**
- All the decision problems for *DL-Lite*$_{bool}$ are **Π_2^p-complete**

Further work

- Now experimenting with QBF + SAT solvers
- Experiments required for `interesting counterexamples' to conservativity
- Complexity of model conservativity still open
- Applications to data integration
- Add nominals, concrete domains to *DL-Lite*, transitive roles?
- What about finite model reasoning?