Structure of the tutorial

1. Introduction to Ontology-Based Data Access
   1. Introduction to ontologies
   2. Ontology languages

2. Description Logics and the DL-Lite family
   1. A gentle introduction to DLs
   2. DLs as a formal language to specify ontologies
   3. Queries in Description Logics
   4. The DL-Lite family of tractable DLs

3. Reasoning in the DL-Lite family
   1. TBox reasoning
   2. TBox & ABox reasoning
   3. Complexity of reasoning in Description Logics

4. Linking data to ontologies
   1. The Description Logic DL-Lite_A
   2. Connecting ontologies to relational data

5. Hands-on session
Part I

Introduction to Ontology-Based Data Access
Outline

1. Introduction to ontologies
2. Ontology languages
Outline

1. Introduction to ontologies
   - Ontologies in information systems
   - Challenges related to ontologies

2. Ontology languages
Different meanings of “Semantics”

1. Part of linguistics that studies the meaning of words and phrases.
2. Meaning of a set of symbols in some representation scheme. Provides a means to specify and communicate the intended meaning of a set of “syntactic” objects.
3. Formal semantics of a language (e.g., an artificial language). (Meta-mathematical) mechanism to associate to each sentence in a language an element of a symbolic domain that is “outside the language”.

In information systems meanings 2 and 3 are the relevant ones:

- In order to talk about semantics we need a representation scheme, i.e., an ontology.
- ... but 2 makes no sense without 3.
An ontology is a representation scheme that describes a formal conceptualization of a domain of interest.

The specification of an ontology comprises several levels:

- **Meta-level**: specifies a set of modeling categories.
- **Intensional level**: specifies a set of conceptual elements (instances of categories) and of rules to describe the conceptual structures of the domain.
- **Extensional level**: specifies a set of instances of the conceptual elements described at the intensional level.
Ontologies at the core of information systems

The usage of all system resources (data and services) is done through the domain conceptualization.
Desiderata: achieve **logical transparency** in access to data:
- **Hide** to the user where and how data are stored.
- **Present** to the user a **conceptual view** of the data.
- **Use** a **semantically rich formalism** for the conceptual view.

*Similar to Data Integration, but with a rich conceptual description as the global view.*
Ontologies at the core of cooperation

The cooperation between systems is done at the level of the conceptualization.
Three novel challenges

1. Languages
2. Methodologies
3. Tools

... for specifying, building, and managing ontologies to be used in information systems.
The first challenge: ontology languages

- Several proposals for ontology languages have been made.
- **Tradeoff** between expressive power of the language and computational complexity of dealing with (i.e., performing inference over) ontologies specified in that language.
- Usability needs to be addressed.

In this tutorial:

- We propose variants of ontology languages suited for managing ontologies in information systems.
- We discuss in depth the above mentioned tradeoff . . .
- . . . paying particular attention to the aspects related to data management.
The second challenge: methodologies

- Building and dealing with ontologies is a complex and challenging task.
- Building **good ontologies** is even more challenging.
- It requires to master the technologies based on semantics, which in turn requires good knowledge about the languages, their semantics, and the implications it has w.r.t. reasoning over the ontology.

In this tutorial:

- We study in depth the **semantics of ontologies**, with an emphasis on their relationship to data in information sources.
- We thus lay the **foundations for the development of methodologies**, though we do not present specific methodologies here.
The third challenge: tools

- According to the principle that “there is no meaning without a language with a formal semantics”, the formal semantics becomes the solid basis for dealing with ontologies.
- Hence every kind of access to an ontology (to extract information, to modify it, etc.), requires to fully take into account its semantics.
- We need to resort to tools that provide capabilities to perform automated reasoning over the ontology, and the kind of reasoning should be sound and complete w.r.t. the formal semantics.

In this tutorial:
- We discuss the requirements for such ontology management tools.
- We present a tool that has been specifically designed for optimized access to information sources through ontologies.
A challenge across the three challenges: scalability

When we want to use ontologies to access information sources, we have to address the three challenges of languages, methodologies, and tools by taking into account **scalability** w.r.t.:

- the size of (the intensional level of) the ontology
- the number of ontologies
- the size of the information sources that are accessed through the ontology/ontologies.

In this tutorial we pay particular attention to the third aspect, since we work under the realistic assumption that the extensional level (i.e., the data) largely dominates in size the intensional level of an ontology.
Outline

1. Introduction to ontologies

2. Ontology languages
   - Elements of an ontology language
   - Intensional level of an ontology language
   - Extensional level of an ontology language
   - Ontologies and other formalisms
   - Queries
Elements of an ontology language

- **Syntax**
  - Alphabet
  - Languages constructs
  - Sentences to assert knowledge

- **Semantics**
  - Formal meaning

- **Pragmatics**
  - Intended meaning
  - Usage
Static vs. dynamic aspects

The aspects of the domain of interest that can be modeled by an ontology language can be classified into:

- **Static aspects**
  - Are related to the structuring of the domain of interest.
  - Supported by virtually all languages.

- **Dynamic aspects**
  - Are related to how the elements of the domain of interest evolve over time.
  - Supported only by some languages, and only partially (cf. services).

Before delving into the dynamic aspects, we need a good understanding of the static ones.

In this tutorial we concentrate essentially on the static aspects.
An ontology language for expressing the intensional level usually includes:

- Concepts
- Properties of concepts
- Relationships between concepts, and their properties
- Axioms
- Individuals and facts about individuals
- Queries

Ontologies are typically rendered as diagrams (e.g., Semantic Networks, Entity-Relationship schemas, UML Class Diagrams).
Example: ontology rendered as UML Class Diagram

```
Employee
  empCode: Integer
  salary: Integer
Manager
boss
    worksFor
      projectName: String
Project
  manages
    AreaManager
      {disjoint, complete}
    TopManager
```
Concepts

**Concept**

Is an element of the ontology that denotes a collection of instances (e.g., the set of “employees”).

We distinguish between:

- **Intensional definition:**
  specification of name, properties, relations, ...

- **Extensional definition:**
  specification of the instances

Concepts are also called classes, entity types, frames.
Properties

Property

Qualifies an element (e.g., a concept) of an ontology.

Property definition (intensional and extensional):

- **Name**
- **Type**:
  - Atomic (integer, real, string, enumerated, ...)
    - e.g., eye-color → { blu, brown, green, grey }
  - Structured (date, sets, lists, ...)
    - e.g., date → day/month/year
- **Default value**

Properties are also called attributes, features, slots.
Relationships

A relationship expresses an association among concepts. We distinguish between:

- **Intensional definition:** specification of involved concepts.
  - e.g., `worksFor` is defined on `Employee` and `Project`

- **Extensional definition:** specification of the instances of the relationship, called facts.
  - e.g., `worksFor(domenico, TONES)`

Relationships are also called associations, relationship types, roles.
Axioms

Axiom

Is a logical formula that expresses at the intensional level a condition that must be satisfied by the elements at the extensional level.

Different kinds of axioms/conditions:

- subclass relationships, e.g., \( \text{Manager} \sqsubseteq \text{Employee} \)
- equivalences, e.g., \( \text{Manager} \equiv \text{AreaManager} \sqcup \text{TopManager} \)
- disjointness, e.g., \( \text{AreaManager} \sqcap \text{TopManager} \equiv \bot \)
- (cardinality) restrictions,
  e.g., each \( \text{Employee} \) worksFor at least 3 \( \text{Project} \)
- ...

Axioms are also called assertions.
A special kind of axioms are definitions.
At the extensional level we have individuals and facts:

- An instance represents an individual (or object) in the extension of a concept.  
  e.g., `instanceOf(domenico, Employee)`

- A fact represents a relationship holding between instances.  
  e.g., `worksFor(domenico, TONES)`
The three levels of an ontology
Comparison with other formalisms

- **Ontology languages vs. knowledge representation languages:**
  Ontologies are knowledge representation schemas.

- **Ontology vs. logic:**
  Logic is a the tool for assigning semantics to ontology languages.

- **Ontology languages vs. conceptual data models:**
  Conceptual schema are special ontologies, suited for conceptualizing a single logical model (database).

- **Ontology languages vs. programming languages:**
  Class definitions are special ontologies, suited for conceptualizing a single structure for computation.
Classification of ontology languages

- Graph-based
  - Semantic networks
  - Conceptual graphs
  - UML

- Frame based
  - Frame Systems
  - OKBC, XOL

- Logic based
  - Description Logics (e.g., $SHOIQ$, $DLR$, $DL-Lite$, OWL, ...)
  - Rules (e.g., RuleML, LP/Prolog, F-Logic)
  - First Order Logic (e.g., KIF)
  - Non-classical logics (e.g., Nonmonotonic, probabilistic)
An ontology language may also include constructs for expressing queries.

**Query**

In an expression at the intensional level denoting a (possibly structured) collection of individuals satisfying a given condition.

**Meta-Query**

In an expression at the meta level denoting a collection of ontology elements satisfying a given condition.

*Note:* One may also conceive queries that span across levels (object-meta queries), cf. [RDF, Calì & Kifer VLDB’06]
Ontology languages vs. query languages

Ontology languages:

- Tailored for capturing intensional relationships.
- Are quite poor as query languages:
  - Cannot refer to same object via multiple navigation paths in the ontology,
  - i.e., allow only for a limited form of JOIN, namely chaining.

Instead, when querying a data source (either directly, or via the ontology), to retrieve the data of interest, general forms of joins are required.

*It follows that the constructs for queries may be quite different from the constructs used in the ontology to form concepts and relationships.*
Example of query

\[
q(ce, cm, se, sm) \leftarrow \text{worksFor}(e, p) \land \text{manages}(m, p) \land \text{boss}(m, e) \land \text{empCode}(e, ce) \land \text{empCode}(m, cm) \land \text{salary}(e, se) \land \text{salary}(m, sm) \land se \geq sm
\]
Query answering under different assumptions

There are fundamentally different assumptions when addressing query answering in different settings:

- traditional database assumption
- knowledge representation assumption
Query answering under the database assumption

- Data are completely specified (CWA), and typically large.
- Schema/intensional information used in the design phase.
- At runtime, the data is assumed to satisfy the schema, and therefore the schema is not used.
- Queries allow for complex navigation paths in the data (cf. SQL).

Query answering amounts to query evaluation, which is computationally easy.
Query answering under the database assumption (cont’d)
Query answering under the database assumption – Example

For each concept/relationship we have a (complete) table in the DB.

DB:
- Employee = \{ john, mary, nick \}
- Manager = \{ john, nick \}
- Project = \{ prA, prB \}
- worksFor = \{ (john,prA), (mary,prB) \}

Query: \( q(x) \leftarrow \) Manager\( (x) \), Project\( (p) \), worksFor\( (x, p) \)

Answer: \{ john \}
An ontology (or conceptual schema, or knowledge base) imposes constraints on the data.

Actual data may be incomplete or inconsistent w.r.t. such constraints.

The system has to take into account intensional information during query answering, and overcome incompleteness or inconsistency.

Size of the data is not considered critical (comparable to the size of the intensional information).

Queries are typically simple, i.e., atomic (the name of a concept).

Query answering amounts to logical inference, which is computationally more costly.
Query answering under the KR assumption (cont’d)
Query answering under the KR assumption – Example

Partial DB assumption: we have a (complete) table in the database only for some concepts/relationships.

**DB:**
- **Manager** = \{ john, nick \}
- **Project** = \{ prA, prB \}
- **worksFor** = \{ (john,prA), (mary,prB) \}

**Query:**
\[ q(x) \leftarrow \text{Employee}(x) \]

**Answer:** \{ john, nick, mary \}

**Rewritten query:**
\[ q(x) \leftarrow \text{Employee}(x) \lor \text{Manager}(x) \lor \text{worksFor}(x) \]
Query answering under the KR assumption – Example 2

Each person has a father, who is a person

Tables in the DB may be incompletely specified.

DB: Person = \{ john, nick, toni \}
hasFather ⊇ \{ (john,nick), (nick,toni) \}

Queries:

\[ q_1(x,y) \leftarrow \text{hasFather}(x,y) \]
\[ q_2(x) \leftarrow \text{hasFather}(x,y) \]
\[ q_3(x) \leftarrow \text{hasFather}(x,y_1), \text{hasFather}(y_1,y_2), \text{hasFather}(y_2,y_3) \]
\[ q_4(x,y_3) \leftarrow \text{hasFather}(x,y_1), \text{hasFather}(y_1,y_2), \text{hasFather}(y_2,y_3) \]

Answers:

to \( q_1 \): \{ (john,nick), (nick,toni) \}
to \( q_2 \): \{ john, nick, toni \}
to \( q_3 \): \{ john, nick, toni \}
to \( q_4 \): \{ \}

Rewritten queries: see later
QA under the KR assumption – Andrea’s Example

Tables may be *incompletely* specified.

- Employee = \{ andrea, nick, mary, john \}
- Manager = \{ andrea, nick, mary \}
- AreaManager \supseteq \{ nick \}
- TopManager \supseteq \{ mary \}
- supervisedBy = \{ (john, andrea), (john, mary) \}
- officeMate = \{ (mary, andrea), (andrea, nick) \}

Diagram:

- Employee
  - Manager
    - AreaManager
    - TopManager
  - supervisedBy
    - \{ disjoint, complete \}
  - officeMate

Diagram:

- Andrea: Manager
  - supervisedBy
    - John
  - officeMate
    - Mary: TopManager
  - Paul: AreaManager
QA under the KR assumption – Andrea’s Example (cont’d)

To determine this answer, we need to resort to reasoning by cases.

Answer: \{ john \}

Rewritten query? There is none (at least not in SQL).
Query answering in Ontology-Based Data Access

In OBDA, we have to face the difficulties of both assumptions:

- The actual data is stored in external information sources (i.e., databases), and thus its size is typically very large.
- The ontology introduces incompleteness of information, and we have to do logical inference, rather than query evaluation.
- We want to take into account at runtime the constraints expressed in the ontology.
- We want to answer complex database-like queries.
- We may have to deal with multiple information sources, and thus face also the problems that are typical of data integration.

Researchers are starting only now to tackle this difficult and challenging problem. In the rest of this tutorial we provide an insight in state-of-the-art technology in this area.
Part II

Description Logics and the \textit{DL-Lite} family
Outline

3 A gentle introduction to Description Logics

4 DLs as a formal language to specify ontologies

5 Queries in Description Logics

6 The DL-Lite family of tractable Description Logics
Outline

3 A gentle introduction to Description Logics
   - Ingredients of Description Logics
   - Description language
   - Description Logics ontologies
   - Reasoning in Description Logics

4 DLs as a formal language to specify ontologies

5 Queries in Description Logics

6 The DL-Lite family of tractable Description Logics
What are Description Logics?

Description Logics \([\text{BCM}^+03]\) are logics specifically designed to represent and reason on structured knowledge:

The domain is composed of objects and is structured into:

- **concepts**, which correspond to classes, and denote sets of objects
- **roles**, which correspond to (binary) relationships, and denote binary relations on objects

The knowledge is asserted through so-called **assertions**, i.e., logical axioms.
Origins of Description Logics

Description Logics stem from early days Knowledge Representation formalisms (late '70s, early '80s):

- Semantic Networks: graph-based formalism, used to represent the meaning of sentences
- Frame Systems: frames used to represent prototypical situations, antecedents of object-oriented formalisms

Problems: no clear semantics, reasoning not well understood

Description Logics (a.k.a. Concept Languages, Terminological Languages) developed starting in the mid '80s, with the aim of providing semantics and inference techniques to knowledge representation systems.
DLs have evolved from being used “just” in KR.

Novel applications of DLs:
- Databases:
  - schema design, schema evolution
  - query optimization
  - integration of heterogeneous data sources, data warehousing
- Conceptual modeling
- Foundation for the Semantic Web (variants of OWL correspond to specific DLs)
- ...
A **Description Logic** is characterized by:

1. **A description language**: how to form concepts and roles
   
   $$\text{Human} \sqcap \text{Male} \sqcap \exists \text{hasChild} \sqcap \forall \text{hasChild}. (\text{Doctor} \sqcup \text{Lawyer})$$

2. **A mechanism to specify knowledge** about concepts and roles (i.e., a TBox)
   
   $$\mathcal{T} = \{ \text{Father} \equiv \text{Human} \sqcap \text{Male} \sqcap \exists \text{hasChild}, $$
   
   $$\text{HappyFather} \sqsubseteq \text{Father} \sqcap \forall \text{hasChild}. (\text{Doctor} \sqcup \text{Lawyer}) \}$$

3. **A mechanism to specify properties of objects** (i.e., an ABox)
   
   $$\mathcal{A} = \{ \text{HappyFather}(\text{john}), \quad \text{hasChild}(\text{john}, \text{mary}) \}$$

4. **A set of inference services**: how to reason on a given KB
   
   $$\mathcal{T} \models \text{HappyFather} \sqsubseteq \exists \text{hasChild}. (\text{Doctor} \sqcup \text{Lawyer})$$
   
   $$\mathcal{T} \cup \mathcal{A} \models (\text{Doctor} \sqcup \text{Lawyer})(\text{mary})$$
Architecture of a Description Logic system

Ingredients of Description Logics

- A gentle introduction to DLs
- DLs to specify ontologies
- Queries in DLs
- The DL-Lite family

Part 2: Description Logics and the DL-Lite family

D. Calvanese, D. Lembo

Ontology-Based Data Access

ISWC’07 – Nov. 12, 2007
Description language

A description language is characterized by a set of constructs for building complex concepts and roles starting from atomic ones:

- **concepts** correspond to classes: interpreted as sets of objects
- **roles** corr. to relationships: interpreted as binary relations on objects

Formal semantics is given in terms of interpretations.

An interpretation $\mathcal{I} = (\Delta^\mathcal{I}, .^\mathcal{I})$ consists of:

- a nonempty set $\Delta^\mathcal{I}$, the domain of $\mathcal{I}$
- an interpretation function $.^\mathcal{I}$, which maps
  - each individual $a$ to an element $a^\mathcal{I}$ of $\Delta^\mathcal{I}$
  - each atomic concept $A$ to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$
  - each atomic role $P$ to a subset $P^\mathcal{I}$ of $\Delta^\mathcal{I} \times \Delta^\mathcal{I}$

The interpretation function is extended to complex concepts and roles according to their syntactic structure.
## Concept constructors

<table>
<thead>
<tr>
<th>Construct</th>
<th>Syntax</th>
<th>Example</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>atomic concept</td>
<td>$A$</td>
<td>Doctor</td>
<td>$A^\mathcal{I} \subseteq \Delta^\mathcal{I}$</td>
</tr>
<tr>
<td>atomic role</td>
<td>$P$</td>
<td>hasChild</td>
<td>$P^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$</td>
</tr>
<tr>
<td>atomic negation</td>
<td>$\neg A$</td>
<td>$\neg$Doctor</td>
<td>$\Delta^\mathcal{I} \setminus A^\mathcal{I}$</td>
</tr>
<tr>
<td>conjunction</td>
<td>$C \sqcap D$</td>
<td>Hum $\sqcap$ Male</td>
<td>$C^\mathcal{I} \cap D^\mathcal{I}$</td>
</tr>
<tr>
<td>(unqual.) exist. res.</td>
<td>$\exists R$</td>
<td>$\exists$hasChild</td>
<td>{ $a$</td>
</tr>
<tr>
<td>value restriction</td>
<td>$\forall R.C$</td>
<td>$\forall$hasChild.Male</td>
<td>{ $a$</td>
</tr>
<tr>
<td>bottom</td>
<td>$\bot$</td>
<td>$\bot$</td>
<td>$\emptyset$</td>
</tr>
</tbody>
</table>

($C$, $D$ denote arbitrary concepts and $R$ an arbitrary role)

The above constructs form the basic language $\mathcal{AL}$ of the family of $\mathcal{AL}$ languages.
### Additional concept and role constructors

<table>
<thead>
<tr>
<th>Construct</th>
<th>AL·</th>
<th>Syntax</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>disjunction</td>
<td>$\mathcal{U}$</td>
<td>$C \sqcup D$</td>
<td>$C^I \cup D^I$</td>
</tr>
<tr>
<td>top</td>
<td>$\top$</td>
<td></td>
<td>$\Delta^I$</td>
</tr>
<tr>
<td>qual. exist. res.</td>
<td>$\mathcal{E}$</td>
<td>$\exists R.C$</td>
<td>${ a \mid \exists b. (a, b) \in R^I \land b \in C^I }$</td>
</tr>
<tr>
<td>(full) negation</td>
<td>$\mathcal{C}$</td>
<td>$\neg C$</td>
<td>$\Delta^I \setminus C^I$</td>
</tr>
<tr>
<td>number</td>
<td>$\mathcal{N}$</td>
<td>$(\geq k R)$</td>
<td>${ a \mid #{b \mid (a, b) \in R^I} \geq k }$</td>
</tr>
<tr>
<td>restrictions</td>
<td></td>
<td>$(\leq k R)$</td>
<td>${ a \mid #{b \mid (a, b) \in R^I} \leq k }$</td>
</tr>
<tr>
<td>qual. number</td>
<td>$\mathcal{Q}$</td>
<td>$(\geq k R.C)$</td>
<td>${ a \mid #{b \mid (a, b) \in R^I \land b \in C^I } \geq k }$</td>
</tr>
<tr>
<td>restrictions</td>
<td></td>
<td>$(\leq k R.C)$</td>
<td>${ a \mid #{b \mid (a, b) \in R^I \land b \in C^I } \leq k }$</td>
</tr>
<tr>
<td>inverse role</td>
<td>$\mathcal{I}$</td>
<td>$R^\leftarrow$</td>
<td>${ (a, b) \mid (b, a) \in R^I }$</td>
</tr>
<tr>
<td>role closure</td>
<td>reg</td>
<td>$\mathcal{R}^*$</td>
<td>$(R^I)^*$</td>
</tr>
</tbody>
</table>

Many different DL constructs and their combinations have been investigated.
Further examples of DL constructs

- Disjunction: \( \forall \text{hasChild}. (\text{Doctor} \sqcup \text{Lawyer}) \)

- Qualified existential restriction: \( \exists \text{hasChild}. \text{Doctor} \)

- Full negation: \( \neg (\text{Doctor} \sqcup \text{Lawyer}) \)

- Number restrictions: \( (\geq 2 \text{hasChild}) \sqcap (\leq 1 \text{sibling}) \)

- Qualified number restrictions: \( (\geq 2 \text{hasChild}. \text{Doctor}) \)

- Inverse role: \( \forall \text{hasChild}^- . \text{Doctor} \)

- Reflexive-transitive role closure: \( \exists \text{hasChild}^* . \text{Doctor} \)
Reasoning on concept expressions

An interpretation $\mathcal{I}$ is a model of a concept $C$ if $C^\mathcal{I} \neq \emptyset$.

Basic reasoning tasks:

1. **Concept satisfiability**: does $C$ admit a model?
2. **Concept subsumption** $C \subseteq D$: does $C^\mathcal{I} \subseteq D^\mathcal{I}$ hold for all interpretations $\mathcal{I}$?

Subsumption used to build the concept hierarchy:

```
Human
  /   \
Man  Woman
  \   /
    Father
  /  \
HappyFather
```

Note: (1) and (2) are mutually reducible if DL is propositionally closed.
Complexity of reasoning on concept expressions

**Complexity of concept satisfiability:** [DLNN97]

<table>
<thead>
<tr>
<th>Description Language</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{AL}$, $\mathcal{ALN}$</td>
<td>PTIME</td>
</tr>
<tr>
<td>$\mathcal{ALU}$, $\mathcal{ALUN}$</td>
<td>NP-complete</td>
</tr>
<tr>
<td>$\mathcal{ALE}$</td>
<td>coNP-complete</td>
</tr>
<tr>
<td>$\mathcal{ALC}$, $\mathcal{ALCN}$, $\mathcal{ALCI}$, $\mathcal{ALCQI}$</td>
<td>PSPACE-complete</td>
</tr>
</tbody>
</table>

**Observations:**

- Two sources of complexity:
  - union ($\cup$) of type NP,
  - existential quantification ($\exists$) of type coNP.

When they are combined, the complexity jumps to PSPACE.

- Number restrictions ($\forall$) do not add to the complexity.
Structural properties vs. asserted properties

We have seen how to build complex concept and roles expressions, which allow one to denote classes with a complex structure.

However, in order to represent real world domains, one needs the ability to assert properties of classes and relationships between them (e.g., as done in UML class diagrams).

The assertion of properties is done in DLs by means of an ontology (or knowledge base).
Description Logics ontology (or knowledge base)

Is a pair $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a TBox and $\mathcal{A}$ is an ABox:

### Description Logics TBox

Consists of a set of **assertions** on concepts and roles:

- Inclusion assertions on concepts: $C_1 \subseteq C_2$
- Inclusion assertions on roles: $R_1 \subseteq R_2$
- Property assertions on (atomic) roles:
  - (transitive $P$)
  - (symmetric $P$)
  - (functional $P$)
  - (reflexive $P$)
  - (domain $P C$)
  - (range $P C$)
  - ...  

### Description Logics ABox

Consists of a set of **membership assertions** on individuals:

- for concepts: $A(c)$
- for roles: $P(c_1, c_2)$ (we use $c_i$ to denote individuals)
Description Logics knowledge base – Example

*Note:* We use $C_1 \equiv C_2$ as an abbreviation for $C_1 \subseteq C_2$, $C_2 \subseteq C_1$.

**TBox assertions:**

- **Inclusion assertions on concepts:**
  - Father $\equiv$ Human $\cap$ Male $\cap \exists$ hasChild
  - HappyFather $\subseteq$ Father $\cap \forall$ hasChild. (Doctor $\sqcup$ Lawyer $\sqcup$ HappyPerson)
  - HappyAnc $\subseteq$ $\forall$ descendant. HappyFather
  - Teacher $\subseteq$ $\neg$ Doctor $\cap$ $\neg$ Lawyer

- **Inclusion assertions on roles:**
  - hasChild $\subseteq$ descendant
  - hasFather $\subseteq$ hasChild$\neg$

- **Property assertions on roles:**
  - (transitive descendant), (reflexive descendant), (functional hasFather)

**ABox membership assertions:**

- Teacher(mary), hasFather(mary, john), HappyAnc(john)
Semantics of a Description Logics knowledge base

The semantics is given by specifying when an interpretation $\mathcal{I}$ satisfies an assertion:

- $C_1 \sqsubseteq C_2$ is satisfied by $\mathcal{I}$ if $C_1^\mathcal{I} \subseteq C_2^\mathcal{I}$.
- $R_1 \sqsubseteq R_2$ is satisfied by $\mathcal{I}$ if $R_1^\mathcal{I} \subseteq R_2^\mathcal{I}$.
- A property assertion ($\text{prop } P$) is satisfied by $\mathcal{I}$ if $P^\mathcal{I}$ is a relation that has the property $\text{prop}$.
  (Note: domain and range assertions can be expressed by means of concept inclusion assertions.)

- $A(c)$ is satisfied by $\mathcal{I}$ if $c^\mathcal{I} \in A^\mathcal{I}$.
- $P(c_1, c_2)$ is satisfied by $\mathcal{I}$ if $(c_1^\mathcal{I}, c_2^\mathcal{I}) \in P^\mathcal{I}$.

We adopt the unique name assumption, i.e., $c_1^\mathcal{I} \neq c_2^\mathcal{I}$, for $c_1 \neq c_2$. 
Models of a Description Logics ontology

Model of a DL knowledge base

An interpretation $\mathcal{I}$ is a model of $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ if it satisfies all assertions in $\mathcal{T}$ and all assertions in $\mathcal{A}$.

$\mathcal{O}$ is said to be satisfiable if it admits a model.

The fundamental reasoning service from which all other ones can be easily derived is . . .

Logical implication

$\mathcal{O}$ logically implies an assertion $\alpha$, written $\mathcal{O} \models \alpha$, if $\alpha$ is satisfied by all models of $\mathcal{O}$. 
TBox reasoning

- **Concept Satisfiability:** $C$ is satisfiable wrt $\mathcal{T}$, if there is a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^\mathcal{I}$ is not empty, i.e., $\mathcal{I} \not\models C \equiv \bot$.

- **Subsumption:** $C_1$ is subsumed by $C_2$ wrt $\mathcal{T}$, if for every model $\mathcal{I}$ of $\mathcal{T}$ we have $C_1^\mathcal{I} \subseteq C_2^\mathcal{I}$, i.e., $\mathcal{I} \models C_1 \subseteq C_2$.

- **Equivalence:** $C_1$ and $C_2$ are equivalent wrt $\mathcal{T}$ if for every model $\mathcal{I}$ of $\mathcal{T}$ we have $C_1^\mathcal{I} = C_2^\mathcal{I}$, i.e., $\mathcal{I} \models C_1 \equiv C_2$.

- **Disjointness:** $C_1$ and $C_2$ are disjoint wrt $\mathcal{T}$ if for every model $\mathcal{I}$ of $\mathcal{T}$ we have $C_1^\mathcal{I} \cap C_2^\mathcal{I} = \emptyset$, i.e., $\mathcal{I} \models C_1 \cap C_2 \equiv \bot$.

- **Functionality implication:** A functionality assertion ($\text{funct } R$) is logically implied by $\mathcal{T}$ if for every model $\mathcal{I}$ of $\mathcal{T}$, we have that $(o, o_1) \in R^\mathcal{I}$ and $(o, o_2) \in R^\mathcal{I}$ implies $o_1 = o_2$, i.e., $\mathcal{I} \models (\text{funct } R)$.

Analogous definitions hold for role satisfiability, subsumption, equivalence, and disjointness.
Reasoning over an ontology

- **Ontology Satisfiability**: Verify whether an ontology $\mathcal{O}$ is satisfiable, i.e., whether $\mathcal{O}$ admits at least one model.

- **Concept Instance Checking**: Verify whether an individual $c$ is an instance of a concept $C$ in $\mathcal{O}$, i.e., whether $\mathcal{O} \models C(c)$.

- **Role Instance Checking**: Verify whether a pair $(c_1, c_2)$ of individuals is an instance of a role $R$ in $\mathcal{O}$, i.e., whether $\mathcal{O} \models R(c_1, c_2)$.

- **Query Answering**: see later...
Reasoning in Description Logics – Example

- Inclusion assertions on concepts:
  - Father $\equiv$ Human $\sqcap$ Male $\sqcap$ $\exists$ hasChild
  - HappyFather $\sqsubseteq$ Father $\sqcap$ $\forall$ hasChild.$(\text{Doctor} \sqcup \text{Lawyer} \sqcup \text{HappyPerson})$
  - HappyAnc $\sqsubseteq$ $\forall$ descendant.HappyFather
  - Teacher $\sqsubseteq$ $\neg$ Doctor $\sqcap$ $\neg$ Lawyer

- Inclusion assertions on roles:
  - hasChild $\sqsubseteq$ descendant
  - hasFather $\sqsubseteq$ hasChild$\neg$

- Property assertions on roles:
  - (transitive descendant), (reflexive descendant), (functional hasFather)

The above TBox logically implies: HappyAncestor $\sqsubseteq$ Father.

- Membership assertions:
  - Teacher(mary), hasFather(mary, john), HappyAnc(john)

The above TBox and ABox logically imply: HappyPerson(mary)
Complexity of reasoning over DL ontologies

Reasoning over DL ontologies is much more complex than reasoning over concept expressions:

- **Bad news:**
  - without restrictions on the form of TBox assertions, reasoning over DL ontologies is already **\(\text{ExpTime}\)-hard**, even for very simple DLs (see, e.g., [Don03]).

- **Good news:**
  - We can add a lot of expressivity (i.e., essentially all DL constructs seen so far), while still staying within the **\(\text{ExpTime}\)** upper bound.
  - There are DL reasoners that perform reasonably well in practice for such DLs (e.g, Racer, Pellet, Fact++, . . .) [MH03].
Outline

3 A gentle introduction to Description Logics

4 DLs as a formal language to specify ontologies
   - DLs to specify ontologies
   - DLs vs. OWL
   - DLs vs. UML Class Diagrams

5 Queries in Description Logics

6 The DL-Lite family of tractable Description Logics
Description Logics are nowadays advocated to provide the foundations for ontology languages.

Different versions of the **Ontology Web Language (OWL)** have been defined as syntactic variants of certain Description Logics.

DLs are also ideally suited to capture the fundamental features of conceptual modeling formalisms used in information systems design:

- **Entity-Relationship diagrams**, used in database conceptual modeling
- **UML Class Diagrams**, used in the design phase of software applications

We briefly overview these correspondences, highlighting essential DL constructs, also in light of the tradeoff between expressive power and computational complexity of reasoning.
The Ontology Web Language (OWL) comes in different variants:

- **OWL-Lite** is a variant of the DL $SHIN(D)$, where:
  - $S$ stands for $ALC$ extended with transitive roles
  - $H$ stands for role hierarchies (i.e., role inclusion assertions)
  - $I$ stands for inverse roles
  - $N$ stands for (unqualified) number restrictions
  - $(D)$ stand for data types, which are necessary in any practical knowledge representation language

- **OWL-DL** is a variant of $SHOIQ(D)$, where:
  - $O$ stands for nominals, which means the possibility of using individuals in the TBox (i.e., the intensional part of the ontology)
  - $Q$ stands for qualified number restrictions
### DL constructs vs. OWL constructs

<table>
<thead>
<tr>
<th>OWL constructor</th>
<th>DL constructor</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>intersectionOf</td>
<td>$C_1 \sqcap \cdots \sqcap C_n$</td>
<td>Human $\sqcap$ Male</td>
</tr>
<tr>
<td>unionOf</td>
<td>$C_1 \sqcup \cdots \sqcup C_n$</td>
<td>Doctor $\sqcup$ Lawyer</td>
</tr>
<tr>
<td>complementOf</td>
<td>$\neg C$</td>
<td>$\neg$ Male</td>
</tr>
<tr>
<td>oneOf</td>
<td>${a_1} \sqcup \cdots \sqcup {a_n}$</td>
<td>${\text{john}} \sqcup {\text{mary}}$</td>
</tr>
<tr>
<td>allValuesFrom</td>
<td>$\forall P.C$</td>
<td>$\forall \text{hasChild}.\text{Doctor}$</td>
</tr>
<tr>
<td>someValuesFrom</td>
<td>$\exists P.C$</td>
<td>$\exists \text{hasChild}.\text{Lawyer}$</td>
</tr>
<tr>
<td>maxCardinality</td>
<td>$(\leq n , P)$</td>
<td>$(\leq 1 , \text{hasChild})$</td>
</tr>
<tr>
<td>minCardinality</td>
<td>$(\geq n , P)$</td>
<td>$(\geq 2 , \text{hasChild})$</td>
</tr>
</tbody>
</table>
### DL axioms vs. OWL axioms

<table>
<thead>
<tr>
<th>OWL axiom</th>
<th>DL syntax</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>subClassOf</td>
<td>( C_1 \sqsubseteq C_2 )</td>
<td>Human \sqsubseteq Animal \sqcap Biped</td>
</tr>
<tr>
<td>equivalentClass</td>
<td>( C_1 \equiv C_2 )</td>
<td>Man \equiv Human \sqcap Male</td>
</tr>
<tr>
<td>disjointWith</td>
<td>( C_1 \sqsubseteq \neg C_2 )</td>
<td>Man \sqsubseteq \neg Female</td>
</tr>
<tr>
<td>sameIndividualAs</td>
<td>{ a_1 } \equiv { a_2 }</td>
<td>{ presBush } \equiv { G.W.Bush }</td>
</tr>
<tr>
<td>differentFrom</td>
<td>{ a_1 } \sqsubseteq \neg { a_2 }</td>
<td>{ john } \sqsubseteq \neg { peter }</td>
</tr>
<tr>
<td>subPropertyOf</td>
<td>( P_1 \sqsubseteq P_2 )</td>
<td>hasDaughter \sqsubseteq hasChild</td>
</tr>
<tr>
<td>equivalentProperty</td>
<td>( P_1 \equiv P_2 )</td>
<td>hasCost \equiv hasPrice</td>
</tr>
<tr>
<td>inverseOf</td>
<td>( P_1 \equiv P_2^- )</td>
<td>hasChild \equiv hasParent^-</td>
</tr>
<tr>
<td>transitiveProperty</td>
<td>( P^+ \sqsubseteq P )</td>
<td>ancestor^+ \sqsubseteq ancestor</td>
</tr>
<tr>
<td>functionalProperty</td>
<td>( \top \sqsubseteq (\leq 1 P) )</td>
<td>( \top \sqsubseteq (\leq 1 \text{hasFather}) )</td>
</tr>
<tr>
<td>inverseFunctionalProperty</td>
<td>( \top \sqsubseteq (\leq 1 P^-) )</td>
<td>( \top \sqsubseteq (\leq 1 \text{hasSSN}^-) )</td>
</tr>
</tbody>
</table>
There is a tight correspondence between variants of DLs and UML Class Diagrams [BCDG05].

- We can devise two transformations:
  - one that associates to each UML Class Diagram $\mathcal{D}$ a DL TBox $\mathcal{T}_\mathcal{D}$.
  - one that associates to each DL TBox $\mathcal{T}$ a UML Class Diagram $\mathcal{D}_\mathcal{T}$.
- The transformations are not model-preserving, but are based on a correspondence between instantiations of the Class Diagram and models of the associated ontology.
- The transformations are satisfiability-preserving, i.e., a class $C$ is consistent in $\mathcal{D}$ iff the corresponding concept is satisfiable in $\mathcal{T}$. 
The ideas behind the encoding of a UML Class Diagram $\mathcal{D}$ in terms of a DL TBox $\mathcal{I}_D$ are quite natural:

- Each class is represented by an atomic concept.
- Each attribute is represented by a role.
- Each binary association is represented by a role.
- Each non-binary association is reified, i.e., represented as a concept connected to its components by roles.
- Each part of the diagram is encoded by suitable assertions.

We illustrate the encoding by means of an example.
Encoding UML Class Diagrams in DLs – Example

Manager ⊑ Employee
AreaManager ⊑ Manager
TopManager ⊑ Manager
Manager ⊑ AreaManager

AreaManager ⊑ ¬TopManager
Employee ⊑ ∃salary
∃salary ⊑ Integer
∃worksFor ⊑ Employee
∃worksFor ⊑ Project
Employee ⊑ ∃worksFor
Project ⊑ (∀ ≥ 3 worksFor)

(funct manages)
(funct manages−)
manages ⊑ worksFor

Note: Domain and range of associations are expressed by means of concept inclusions.
The encoding of an $\mathcal{ALC}$ TBox $\mathcal{T}$ in terms of a UML Class Diagram $\mathcal{T}_D$ is based on the following observations:

- We can restrict the attention to $\mathcal{ALC}$ TBoxes, that are constituted by concept inclusion assertions of a simplified form (single atomic concept on the left, and a single concept constructor on the right).
- For each such inclusion assertion, the encoding introduces a portion of UML Class Diagram, that may refer to some common classes.

Reasoning in the encoded $\mathcal{ALC}$-fragment is already $\text{ExpTime}$-hard. From this, we obtain:

**Theorem**

Reasoning over UML Class Diagrams is $\text{ExpTime}$-hard.
Reasoning on UML Class Diagrams using DLs

- The two encodings show that DL TBoxes and UML Class Diagrams essentially have the **same expressive power**.

- Hence, reasoning over UML Class Diagrams has the same complexity as reasoning over ontologies in expressive DLs, i.e., **ExpTime**-complete.

- The high complexity is caused by:
  1. the possibility to use disjunction (covering constraints)
  2. the interaction between role inclusions and functionality constraints (maximum 1 cardinality)

Without (1) and restricting (2), reasoning becomes simpler [ACK⁺07]:
- **NLogSpace**-complete in combined complexity
- in **LogSpace** in data complexity (see later)
Efficient reasoning on UML Class Diagrams

We are interested in using UML Class Diagrams to specify ontologies in the context of Ontology-Based Data Access.

Questions

- Which is the right combination of constructs to allow in UML Class Diagrams to be used for OBDA?
- Are there techniques for query answering in this case that can be derived from Description Logics?
- Can query answering be done efficiently in the size of the data?
- If yes, can we leverage relational database technology for query answering?
Outline

3 A gentle introduction to Description Logics

4 DLs as a formal language to specify ontologies

5 Queries in Description Logics
   - Queries over Description Logics ontologies
   - Certain answers
   - Complexity of query answering

6 The DL-Lite family of tractable Description Logics
A gentle introduction to DLs

DLs to specify ontologies

Queries in DLs

The DL-Lite family

Part 2: Description Logics and the DL-Lite family

Queries over Description Logics ontologies

We need more complex queries than simple concept (or role) expressions.

A conjunctive query \( q(\vec{x}) \) over an ontology \( \mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle \) has the form:

\[
q(\vec{x}) \leftarrow \exists \vec{y}. \text{conj}(\vec{x}, \vec{y})
\]

where:

- \( \vec{x} \) is a tuple of so-called distinguished variables. The number of variables in \( \vec{x} \) is called the arity of \( q \).
- \( \vec{y} \) is a tuple of so-called non-distinguished variables,
- \( q(\vec{x}) \) is called the head of \( q \).
- \( \text{conj}(\vec{x}, \vec{y}) \), called the body of \( q \), is a conjunction of atoms, where each atom:
  - has as predicate symbol an atomic concept or role of \( \mathcal{T} \),
  - may use variables in \( \vec{x} \) and \( \vec{y} \),
  - may use constants that are individuals of \( \mathcal{A} \).
Queries over Description Logics ontologies (cont’d)

**Note:** we may also use for CQs a simplified notation

\[ q(\vec{x}) \leftarrow \text{body}(\vec{x}, \vec{y}) \]

where \( \text{body}(\vec{x}, \vec{y}) \) is a sequence constituted by the atoms in \( \text{conj}(\vec{x}, \vec{y}) \).

**Example of conjunctive query**

\[ q(x, y) \leftarrow \exists p. \text{Employee}(x) \land \text{Employee}(y) \land \text{Project}(p) \land \text{boss}(x, y) \land \text{worksFor}(x, p) \land \text{worksFor}(y, p) \]

In simplified notation:

\[ q(x, y) \leftarrow \text{Employee}(x), \text{Employee}(y), \text{Project}(p), \text{boss}(x, y), \text{worksFor}(x, p), \text{worksFor}(y, p) \]

**Note:** a CQ corresponds to a select-project-join SQL query.
Certain answers to a query

Let $\mathcal{O} = \langle T, A \rangle$ be an ontology, $\mathcal{I}$ an interpretation for $\mathcal{O}$, and $q(\vec{x}) \leftarrow \exists \vec{y}. \text{conj}(\vec{x}, \vec{y})$ a CQ.

The answer to $q(\vec{x})$ over $\mathcal{I}$, denoted $q^\mathcal{I}$, is the set of tuples $\vec{c}$ of constants of $A$ such that the formula $\exists \vec{y}. \text{conj}(\vec{c}, \vec{y})$ evaluates to true in $\mathcal{I}$.

We are interested in finding those answers that hold in all models of an ontology.

The certain answers to $q(\vec{x})$ over $\mathcal{O} = \langle T, A \rangle$, denoted $\text{cert}(q, \mathcal{O})$, are the tuples $\vec{c}$ of constants of $A$ such that $\vec{c} \in q^\mathcal{I}$, for every model $\mathcal{I}$ of $\mathcal{O}$. 
Query answering over ontologies

**Query answering** over an ontology $\mathcal{O}$

Is the problem of **computing the certain answers** to a query over $\mathcal{O}$.

Computing certain answers is a form of **logical implication**:

$$\vec{c} \in \text{cert}(q, \mathcal{O}) \quad \text{iff} \quad \mathcal{O} \models q(\vec{c})$$

*Note*: instance checking is a special case of query answering: it amounts to answering the boolean query $q() \leftarrow A(c)$ (resp., $q() \leftarrow P(c_1, c_2)$) over $\mathcal{O}$ (in this case $\vec{c}$ is the empty tuple).
Query answering over ontologies – Example

\[ \triangleleft \text{hasFather} \]

TBox \( \mathcal{T} \):

\[
\begin{align*}
\exists \text{hasFather} & \sqsubseteq \text{Person} \\
\exists \text{hasFather}^\sim & \sqsubseteq \text{Person} \\
\text{Person} & \sqsubseteq \exists \text{hasFather}
\end{align*}
\]

ABox \( \mathcal{A} \):

\[
\begin{align*}
\text{Person}(\text{john}), \ & \text{Person}(\text{nick}), \ \text{Person}(\text{toni}) \\
\text{hasFather}(\text{john},\text{nick}), \ & \text{hasFather}(\text{nick},\text{toni})
\end{align*}
\]

Queries:

\[
\begin{align*}
q_1(x, y) & \leftarrow \text{hasFather}(x, y) \\
q_2(x) & \leftarrow \exists y. \text{hasFather}(x, y) \\
q_3(x) & \leftarrow \exists y_1, y_2, y_3. \text{hasFather}(x, y_1) \land \text{hasFather}(y_1, y_2) \land \text{hasFather}(y_2, y_3) \\
q_4(x, y_3) & \leftarrow \exists y_1, y_2. \text{hasFather}(x, y_1) \land \text{hasFather}(y_1, y_2) \land \text{hasFather}(y_2, y_3)
\end{align*}
\]

Certain answers:

\[
\begin{align*}
cert(q_1, \langle \mathcal{T}, \mathcal{A} \rangle) & = \{ (\text{john}, \text{nick}), (\text{nick}, \text{toni}) \} \\
cert(q_2, \langle \mathcal{T}, \mathcal{A} \rangle) & = \{ \text{john}, \text{nick}, \text{toni} \} \\
cert(q_3, \langle \mathcal{T}, \mathcal{A} \rangle) & = \{ \text{john}, \text{nick}, \text{toni} \} \\
cert(q_4, \langle \mathcal{T}, \mathcal{A} \rangle) & = \{ \}
\end{align*}
\]
We consider also unions of CQs.

A union of conjunctive queries (UCQ) has the form:

\[ q(\vec{x}) \leftarrow \exists \vec{y}_1. \text{conj}(\vec{x}, \vec{y}_1) \lor \cdots \lor \exists \vec{y}_k. \text{conj}(\vec{x}, \vec{y}_k) \]

where each \( \exists \vec{y}_i. \text{conj}(\vec{x}, \vec{y}_i) \) is the body of a CQ.

Example

\[ q(x) \leftarrow (\text{Manager}(x) \land \text{worksFor}(x, \text{tones})) \lor (\exists y. \text{boss}(x, y) \land \text{worksFor}(y, \text{tones})) \]

The (certain) answers to a UCQ are defined analogously to those for CQs.
Data and combined complexity

When measuring the complexity of answering a query $q(\vec{x})$ over an ontology $\mathcal{O} = \langle T, A \rangle$, various parameters are of importance.

Depending on which parameters we consider, we get different complexity measures:

- **Data complexity**: TBox and query are considered fixed, and only the size of the ABox (i.e., the data) matters.
- **Query complexity**: TBox and ABox are considered fixed, and only the size of the query matters.
- **Schema complexity**: ABox and query are considered fixed, and only the size of the TBox (i.e., the schema) matters.
- **Combined complexity**: no parameter is considered fixed.

In the OBDA setting, the size of the data largely dominates the size of the conceptual layer (and of the query).

$\leadsto$ **Data complexity** is the relevant complexity measure.
Complexity of query answering in DLs

Answering (U)CQs over DL ontologies has been studied extensively:

- **Combined complexity:**
  - NP-complete for plain databases (i.e., with an empty TBox)
  - EXPTIME-complete for ALC [CDGL98, Lut07]
  - 2EXPTIME-complete for very expressive DLs (with inverse roles) [CDGL98, Lut07]

- **Data complexity:**
  - in LOGSPACE for plain databases
  - coNP-hard with disjunction in the TBox [DLNS94, CDGL⁺06b]
  - coNP-complete for very expressive DLs [LR98, OCE06, GHLS07]

**Questions**

- Can we find interesting families of DLs for which the query answering problem can be solved efficiently?
- If yes, can we leverage relational database technology for query answering?
Outline

3. A gentle introduction to Description Logics

4. DLs as a formal language to specify ontologies

5. Queries in Description Logics

6. The \textit{DL-Lite} family of tractable Description Logics
   - The \textit{DL-Lite} family
   - Syntax of \textit{DL-Lite}_\mathcal{F} and \textit{DL-Lite}_\mathcal{R}
   - Semantics of \textit{DL-Lite}
   - Properties of \textit{DL-Lite}
The **DL-Lite** family

- Is a family of DLs optimized according to the tradeoff between expressive power and data complexity of query answering.
- We present now two incomparable languages of this family, $DL-Lite_F$, $DL-Lite_R$ (we use $DL-Lite$ to refer to both).
- We will see that $DL-Lite$ has nice computational properties:
  - $\text{PTIME}$ in the size of the TBox (schema complexity)
  - $\text{LOGSPACE}$ in the size of the ABox (data complexity)
  - enjoys FOL-rewritability
- We will see that $DL-Lite_F$ and $DL-Lite_R$ are in some sense the maximal DLs with these nice computational properties, which are lost with minimal additions of constructs.

Hence, $DL-Lite$ provides a positive answer to our basic questions, and sets the foundations for Ontology-Based Data Access.
**DL-Lite** ontologies

**TBox assertions:**
- Concept inclusion assertions: \[ Cl \sqsubseteq Cr \], with:
  - Cl → A | ∃Q
  - Cr → A | ∃Q | ¬A | ¬∃Q
  - Q → P | P^-
- Functionality assertions: (funct Q)

**ABox assertions:** \[ A(c), P(c_1, c_2) \], with \( c_1, c_2 \) constants

**Observations:**
- Captures all the basic constructs of UML Class Diagrams and ER
- Notable exception: covering constraints in generalizations.
**DL-Lite**\(_R\) ontologies

**TBox assertions:**
- Concept inclusion assertions: \(C_l \sqsubseteq C_r\), with:
  \[
  C_l \rightarrow A | \exists Q \\
  C_r \rightarrow A | \exists Q | \neg A | \neg \exists Q \\
  Q \rightarrow P | P^-
  \]
- Role inclusion assertions: \(Q \sqsubseteq R\), with:
  \[
  R \rightarrow Q | \neg Q
  \]

**ABox assertions:** \(A(c), P(c_1, c_2)\), with \(c_1, c_2\) constants

**Observations:**
- Drops functional restrictions in favor of ISA between roles.
- Extends (the DL fragment of) the ontology language RDFS.
## Semantics of DL-Lite

<table>
<thead>
<tr>
<th>Construct</th>
<th>Syntax</th>
<th>Example</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>atomic conc.</td>
<td>$A$</td>
<td>Doctor</td>
<td>$A^I \subseteq \Delta^I$</td>
</tr>
<tr>
<td>exist. restr.</td>
<td>$\exists Q$</td>
<td>$\exists \text{child}^-$</td>
<td>{d</td>
</tr>
<tr>
<td>at. conc. neg.</td>
<td>$\neg A$</td>
<td>$\neg$Doctor</td>
<td>$\Delta^I \setminus A^I$</td>
</tr>
<tr>
<td>conc. neg.</td>
<td>$\neg \exists Q$</td>
<td>$\neg \exists \text{child}$</td>
<td>$\Delta^I \setminus (\exists Q)^I$</td>
</tr>
<tr>
<td>atomic role</td>
<td>$P$</td>
<td>child</td>
<td>$P^I \subseteq \Delta^I \times \Delta^I$</td>
</tr>
<tr>
<td>inverse role</td>
<td>$P^-$</td>
<td>child$^-$</td>
<td>{(o, o')</td>
</tr>
<tr>
<td>role negation</td>
<td>$\neg Q$</td>
<td>$\neg$manages</td>
<td>$(\Delta^I_o \times \Delta^I_o) \setminus Q^I$</td>
</tr>
<tr>
<td>conc. incl.</td>
<td>$C_l \sqsubseteq C_r$</td>
<td>Father$\sqsubseteq \exists$child</td>
<td>$C_l^I \subseteq C_r^I$</td>
</tr>
<tr>
<td>role incl.</td>
<td>$Q \sqsubseteq R$</td>
<td>hasFather$\sqsubseteq \text{child}^-$</td>
<td>$Q^I \subseteq R^I$</td>
</tr>
<tr>
<td>funct. asser.</td>
<td>(funct $Q$)</td>
<td>(funct succ)</td>
<td>$\forall d, e, e'. (d, e) \in Q^I \land (d, e') \in Q^I \rightarrow e = e'$</td>
</tr>
<tr>
<td>mem. asser.</td>
<td>$A(c)$</td>
<td>Father(bob)</td>
<td>$c^I \in A^I$</td>
</tr>
<tr>
<td>mem. asser.</td>
<td>$P(c_1, c_2)$</td>
<td>child(bob, ann)</td>
<td>$(c_1^I, c_2^I) \in P^I$</td>
</tr>
</tbody>
</table>
Capturing basic ontology constructs in *DL-Lite*

<table>
<thead>
<tr>
<th>Concept</th>
<th>Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISA between classes</td>
<td>( A_1 \sqsubseteq A_2 )</td>
</tr>
<tr>
<td>Disjointness between classes</td>
<td>( A_1 \sqsubseteq \lnot A_2 )</td>
</tr>
<tr>
<td>Domain and range of relations</td>
<td>( \exists P \sqsubseteq A_1 ) ( \exists P^- \sqsubseteq A_2 )</td>
</tr>
<tr>
<td>Mandatory participation</td>
<td>( A_1 \sqsubseteq \exists P ) ( A_2 \sqsubseteq \exists P^- )</td>
</tr>
<tr>
<td>Functionality of relations (in <em>DL-Lite_\mathcal{F}</em> )</td>
<td>( (\text{funct } P) ) ( (\text{funct } P^-) )</td>
</tr>
<tr>
<td>ISA between relations (in <em>DL-Lite_\mathcal{R}</em> )</td>
<td>( Q_1 \sqsubseteq Q_2 )</td>
</tr>
<tr>
<td>Disjointness between relations (in <em>DL-Lite_\mathcal{R}</em> )</td>
<td>( Q \sqsubseteq \lnot Q )</td>
</tr>
</tbody>
</table>
Additionally, in $DL$-Lite$_F$ : (\texttt{funct} manages), (\texttt{funct} manages$^-$), ...  

in $DL$-Lite$_R$ : manages $\sqsubseteq$ worksFor

\textbf{Note:} in $DL$-Lite we cannot capture: – completeness of the hierarchy, 
– number restrictions
Properties of $DL$-Lite

- **The TBox may contain cyclic dependencies** (which typically increase the computational complexity of reasoning).

  Example: $A \sqsubseteq \exists P$, $\exists P^{-} \sqsubseteq A$

- **We have not included in the syntax $\sqcap$ on the right hand-side of inclusion assertions, but it can trivially be added, since**

  
  $Cl \sqsubseteq Cr_1 \sqcap Cr_2$ is equivalent to
  
  $Cl \sqsubseteq Cr_1$
  
  $Cl \sqsubseteq Cr_2$

- **A domain assertion on role $P$ has the form:** $\exists P \sqsubseteq A_1$

  **A range assertion on role $P$ has the form:** $\exists P^{-} \sqsubseteq A_2$
Properties of $DL$-$Lite_F$ does not enjoy the finite model property.

**Example**

**TBox** $\mathcal{T}$:  
$\text{Nat} \sqsubseteq \exists \text{succ} \quad \exists \text{succ}^- \sqsubseteq \text{Nat}$

$\text{Zero} \sqsubseteq \text{Nat} \sqcap \neg \exists \text{succ}^- \quad (\textbf{funct success})$

**ABox** $\mathcal{A}$:  
$\text{Zero}(0)$

$\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ admits only infinite models.
Hence, it is satisfiable, but not finitely satisfiable.

Hence, reasoning w.r.t. arbitrary models is different from reasoning w.r.t. finite models only.
Properties of \( DL-Lite_\mathcal{R} \)

- The TBox may contain cyclic dependencies.

- \( DL-Lite_\mathcal{R} \) does enjoy the finite model property. Hence, reasoning w.r.t. finite models is the same as reasoning w.r.t. arbitrary models.

- With role inclusion assertions, we can simulate qualified existential quantification in the rhs of an inclusion assertion \( A_1 \sqsubseteq \exists Q.A_2 \).

  To do so, we introduce a new role \( Q_{A_2} \) and:
  - the role inclusion assertion \( Q_{A_2} \sqsubseteq Q \)
  - the concept inclusion assertions:
    \[
    A_1 \sqsubseteq \exists Q_{A_2} \quad \exists Q_{A_2} \sqsubseteq A_2
    \]

  In this way, we can consider \( \exists Q.A \) in the right-hand side of an inclusion assertion as an abbreviation.
We have seen that $DL-Lite_F$ and $DL-Lite_R$ can capture the essential features of prominent conceptual modeling formalisms.

In the next part, we will analyze reasoning in $DL-Lite$, and establish the following characterization of its computational properties:

- Ontology satisfiability is polynomial in the size of TBox and ABox.
- Query answering is:
  - $\text{PTime}$ in the size of the TBox.
  - $\text{LogSpace}$ in the size of the ABox, and FOL-rewritable, which means that we can leverage for it relational database technology.

We will also see that $DL-Lite$ is essentially the maximal DL enjoying these nice computational properties.

From (1), (2), and (3) we get the following claim:

$DL-Lite$ is the representation formalism that is best suited to underly Ontology-Based Data Management systems.
Part III

Reasoning in the DL-Lite family
Outline

7. TBox reasoning

8. TBox & ABox Reasoning

9. Complexity of reasoning in Description Logics
Outline

7. TBox reasoning
   - Reducing to subsumption
   - Reducing to ontology satisfiability

8. TBox & ABox Reasoning

9. Complexity of reasoning in Description Logics
Remark

In the following,

- a TBox $T$ that is either a $DL-Lite_R$ or a $DL-Lite_F$ TBox is simply called TBox.
- $C$, possibly with subscript, denotes a general concept, i.e.,
  \[
  C \rightarrow A \mid \neg A \mid \exists Q \mid \neg \exists Q
  \]
  \[
  Q \rightarrow P \mid P^-
  \]
  where $A$ is an atomic concept, $P$ is an atomic role, and $Q$ is a basic role.
- $R$, possibly with subscript, denotes a general role, i.e.,
  \[
  R \rightarrow Q \mid \neg Q
  \]
Reasoning services

- **Concept Satisfiability**: \( C \) is satisfiable wrt \( T \), if there is a model \( \mathcal{I} \) of \( T \) such that \( C^\mathcal{I} \) is not empty, i.e., \( \mathcal{I} \not\models C \equiv \bot \).

- **Subsumption**: \( C_1 \) is subsumed by \( C_2 \) wrt \( T \), if for every model \( \mathcal{I} \) of \( T \) we have \( C_1^\mathcal{I} \subseteq C_2^\mathcal{I} \), i.e., \( \mathcal{I} \models C_1 \subseteq C_2 \).

- **Equivalence**: \( C_1 \) and \( C_2 \) are equivalent wrt \( T \) if for every model \( \mathcal{I} \) of \( T \) we have \( C_1^\mathcal{I} = C_2^\mathcal{I} \), i.e., \( \mathcal{I} \models C_1 \equiv C_2 \).

- **Disjointness**: \( C_1 \) and \( C_2 \) are disjoint wrt \( T \) if for every model \( \mathcal{I} \) of \( T \) we have \( C_1^\mathcal{I} \cap C_2^\mathcal{I} = \emptyset \), i.e., \( \mathcal{I} \models C_1 \cap C_2 \equiv \bot \).

- **Functionality implication**: A functionality assertion \((\text{funct } Q)\) is logically implied by \( T \) if for every model \( \mathcal{I} \) of \( T \), we have that \( (o, o_1) \in Q^\mathcal{I} \) and \( (o, o_2) \in Q^\mathcal{I} \) implies \( o_1 = o_2 \), i.e., \( \mathcal{I} \models (\text{funct } Q) \).

Analogous definitions hold for role satisfiability, subsumption, equivalence, and disjointness.
From TBox reasoning to ontology satisfiability

In the following we will show how to reduce TBox reasoning to ontology satisfiability.

- **Ontology Satisfiability**: Verify whether an ontology $\mathcal{O}$ is satisfiable, i.e., whether $\mathcal{O}$ admits at least one model.

  - We first will show how to reduce TBox reasoning services to concept/role subsumption.
  - Then we will provide reductions from concept/role subsumption to ontology satisfiability.
Concept/role satisfiability, equivalence, and disjointness

**Theorem**

1. $C$ is unsatisfiable wrt $T$ iff $T \models C \sqsubseteq \neg C$.
2. $T \models C_1 \equiv C_2$ iff $T \models C_1 \sqsubseteq C_2$ and $T \models C_2 \sqsubseteq C_1$.
3. $C_1$ and $C_2$ are disjoint iff $T \models C_1 \sqsubseteq \neg C_2$.

**Proof (sketch)**

1. "$\Leftarrow$" if $T \models C \sqsubseteq \neg C$, then $C^I \subseteq \Delta^I \setminus C^I$, for every model $I = \langle \Delta^I, \cdot^I \rangle$ of $T$; but this holds iff $C^I = \emptyset$.
   
   "$\Rightarrow$" if $C$ is unsatisfiable, then $C^I = \emptyset$, for every model $I$ of $T$. Therefore $C^I \subseteq (\neg C)^I$.

2. Trivial.

3. Trivial.

*Analogous reductions for role satisfiability, equivalence and disjointness.*
Reducing to subsumption

From implication of functionalities to subsumption

Theorem

\[\mathcal{T} \models (\text{funt} \ Q) \iff \text{either } (\text{funt} \ Q) \in \mathcal{T} \text{ (only for DL-Lite}_F \text{ ontologies)}, \]

or \( \mathcal{T} \models Q \sqsubseteq \neg Q. \)

Proof (sketch)

“⇐” The case in which \((\text{funt} \ Q) \in \mathcal{T}\) is trivial. Instead, if \(\mathcal{T} \models Q \sqsubseteq \neg Q\), then \(Q^\mathcal{I} = \emptyset\) and hence \(\mathcal{I} \models (\text{funt} \ Q)\), for every model \(\mathcal{I}\) of \(\mathcal{T}\).

“⇒” Starting from the assumption that neither \((\text{funt} \ Q) \in \mathcal{T}\) nor \(\mathcal{T} \models Q \sqsubseteq \neg Q\), we can construct a model of \(\mathcal{T}\) that is not a model of \((\text{funt} \ Q)\).
From concept subsumption to ontology satisfiability

**Theorem**

Let $\hat{A}$ be an atomic concept not in $\mathcal{T}$, and $c$ a constant. $\mathcal{T} \models C_1 \sqsubseteq C_2$ iff the ontology $\mathcal{O}_{C_1 \sqsubseteq C_2} = \langle \mathcal{T} \cup \{\hat{A} \sqsubseteq C_1, \hat{A} \sqsubseteq \neg C_2\}, \{\hat{A}(c)\}\rangle$ is unsatisfiable.

Intuitively, $C_1$ is subsumed by $C_2$ iff the smallest ontology containing $\mathcal{T}$ and implying both $C_1(c)$ and $\neg C_2(c)$ is unsatisfiable.

**Proof (sketch)**

“$\Leftarrow$” Suppose that $\mathcal{O}_{C_1 \sqsubseteq C_2}$ is unsatisfiable, but $\mathcal{T} \not\models C_1 \sqsubseteq C_2$, i.e., there exists a model $\mathcal{I}$ of $\mathcal{T}$ such that $C_1^\mathcal{I} \not\subseteq C_2^\mathcal{I}$. From $\mathcal{I}$ we construct a model for $\mathcal{O}_{C_1 \sqsubseteq C_2}$, thus getting a contradiction.

“$\Rightarrow$” Suppose that $\mathcal{O}_{C_1 \sqsubseteq C_2}$ is satisfiable, and let $\mathcal{I}$ be a model of $\mathcal{O}_{C_1 \sqsubseteq C_2}$. Then $\mathcal{I} \models \mathcal{T}$, and $\mathcal{I} \models C_1(c)$ and $\mathcal{I} \models \neg C_2(c)$, i.e., $\mathcal{I} \not\models C_1 \sqsubseteq C_2$, i.e., $\mathcal{T} \not\models C_1 \sqsubseteq C_2$. 

D. Calvanese, D. Lembo

Ontology-Based Data Access

ISWC’07 – Nov. 12, 2007

(105/217)
From role subsumption to ontology satisfiability

**Theorem**

Let $\mathcal{T}$ be a $\text{DL-Lite}_R$ TBox, $Q_1$ and $Q_2$ two general roles, $\hat{P}$ an atomic role not in $\mathcal{T}$, and $c_1, c_2$ two constants. $\mathcal{T} \models R_1 \sqsubseteq R_2$ iff the ontology $\mathcal{O}_{R_1 \sqsubseteq R_2} = \langle \mathcal{T} \cup \{\hat{P} \sqsubseteq R_1, \hat{P} \sqsubseteq \neg R_2\}, \{\hat{P}(c_1, c_2)\}\rangle$ is unsatisfiable.

Intuitively, $R_1$ is subsumed by $R_2$ iff the smallest ontology containing $\mathcal{T}$ and implying both $R_1(c_1, c_2)$ and $\neg R_2(c_1, c_2)$ is unsatisfiable.

**Proof (sketch)**

Analogous to above.

Notice that $\mathcal{O}_{Q_1 \sqsubseteq Q_2}$ is inherently a $\text{DL-Lite}_R$ ontology.
Theorem

Let $\mathcal{T}$ be a $\text{DL-Lite}_\mathcal{F}$ TBox, $Q_1$ and $Q_2$ two basic roles such that $Q_1 \neq Q_2$. Then,

1. $\mathcal{T} \models Q_1 \sqsubseteq Q_2$ iff $Q_1$ is unsatisfiable.
2. $\mathcal{T} \models \neg Q_1 \sqsubseteq Q_2$ iff $\mathcal{T}$ is unsatisfiable.
3. $\mathcal{T} \models Q_1 \sqsubseteq \neg Q_2$ iff either
   
   (a) $\exists Q_1$ and $\exists Q_2$ are disjoint, or
   
   (b) $\exists Q_1^\neg$ and $\exists Q_2^\neg$ are disjoint.

Notice that an inclusion of the form $\neg Q_1 \sqsubseteq \neg Q_2$ is equivalent to $Q_2 \sqsubseteq Q_1$, and therefore is considered in the first item.
From role subsumption to ontology satisfiability (cont’d)

Theorem

Let $\mathcal{T}$ be a $DL$-$Lite_\mathcal{F}$ TBox, $Q_1$ and $Q_2$ two basic roles such that $Q_1 \neq Q_2$, $\hat{A}$ an atomic concept not in $\mathcal{T}$, and $c$ a constant. Then,

1. $\mathcal{T} \models Q_1 \sqsubseteq Q_2$ iff either
   - (a) the ontology $\mathcal{O}_{\exists Q_1 \sqsubseteq \neg \exists Q_2} = \langle \mathcal{T} \cup \{ \hat{A} \sqsubseteq \exists Q_1 \}, \{ \hat{A}(c) \} \rangle$ is unsatisfiable, or
   - (b) the ontology $\mathcal{O}_{\exists Q_1^{-} \sqsubseteq \neg \exists Q_1^{-}} = \langle \mathcal{T} \cup \{ \hat{A} \sqsubseteq \exists Q_1^{-} \}, \{ \hat{A}(c) \} \rangle$ is unsatisfiable.

2. $\mathcal{T} \models \neg Q_1 \sqsubseteq Q_2$ iff $\mathcal{T}$ is unsatisfiable.

3. $\mathcal{T} \models Q_1 \sqsubseteq \neg Q_2$ iff either
   - (a) the ontology $\mathcal{O}_{\exists Q_1 \sqsubseteq \exists Q_2} = \langle \mathcal{T} \cup \{ \hat{A} \sqsubseteq \exists Q_1, \hat{A} \sqsubseteq \exists Q_2 \}, \{ \hat{A}(c) \} \rangle$ is unsatisfiable, or
   - (b) the ontology $\mathcal{O}_{\exists Q_1^{-} \sqsubseteq \exists Q_2^{-}} = \langle \mathcal{T} \cup \{ \hat{A} \sqsubseteq \exists Q_1^{-}, \hat{A} \sqsubseteq \exists Q_2^{-} \}, \{ \hat{A}(c) \} \rangle$ is unsatisfiable.
The results above say us that we can support TBox reasoning services relying on ontology satisfiability services.

Ontology satisfiability is a form of reasoning over both the TBox and the ABox of the ontology.

In the following, we first consider other TBox & ABox reasoning services, in particular query answering, and then turn back to ontology satisfiability.
Outline

7 TBox reasoning

8 TBox & ABox Reasoning
   - Query answering
   - Ontology satisfiability

9 Complexity of reasoning in Description Logics
Query answering and instance checking

- **Concept Instance Checking**: Verify whether an individual \( c \) is an instance of a concept \( C \) in an ontology \( O \), i.e., whether \( O \models C(c) \).
- **Role Instance Checking**: Verify whether a pair \((c_1, c_2)\) of individuals is an instance of a role \( Q \) in an ontology \( O \), i.e., whether \( O \models Q(c_1, c_2) \).
- **Query Answering**: Given a query \( q \) over an ontology \( O \), find all tuples \( \vec{c} \) of constants such that \( O \models q(\vec{c}) \).

Notice that instance checking is a special case of query answering: it amounts to answering the boolean query \( C(c) \) (resp., \( Q(c_1, c_2) \)) over \( O \) (in this case \( \vec{c} \) is the empty tuple).
Certain answers

We recall that

Query answering over a ontology $O = \langle T, A \rangle$ is a form of logical implication:

find all tuples $\vec{c}$ of constants s.t. $O \models q(\vec{c})$

A.k.a. **certain answers** in databases, i.e., the tuples that are answers to $q$ in all models of $O = \langle T, A \rangle$:

$$\text{cert}(q, O) = \{ \vec{c} \mid \vec{c} \in q^I, \text{ for every model } I \text{ of } \langle T, A \rangle \}$$
Data complexity of query answering

**Recognition problem:** Given an ontology $\mathcal{O}$, a query $q$ over $\mathcal{O}$, a tuple of constants $\vec{c}$, check whether $\vec{c} \in \text{cert}(q, \mathcal{O})$.

We consider a setting where the size of the data largely dominates the size of the conceptual layer. We look at data complexity of query answering, i.e., complexity of the recognition problem computed w.r.t. the size of the ABox only.

**Basic questions:**

1. For which ontology languages can we answer queries over an ontology efficiently?
2. How complex becomes query answering over an ontology when we consider more expressive ontology languages?
To study data complexity, we need to separate the contribution of $\mathcal{A}$ from the contribution of $q$ and $\mathcal{T}$.

$\leadsto$ Study $Q$-rewritability for query language $Q$. 
Query answering can always be thought as done in two phases:

1. **Perfect reformulation**: producing the query \( r_{q,T} \), namely the function \( \text{cert}[q,T](\cdot) \)

2. **Query evaluation**: evaluating \( r_{q,T} \) over the ABox \( \mathcal{A} \) seen as a complete database, and forgetting about the TBox \( T \)

Produces \( \text{cert}(q,\langle T,A \rangle) \)

Let \( Q \) be a query language

Query answering for an ontology language is **\( Q \)-rewritable** if \( r_{q,T} \) is in \( Q \).
Consider an ontology language that enjoys \( Q\)-rewritability, for a query language \( Q \):

- When \( Q \) is FOL (i.e., the language enjoys FOL-rewritability)
  \( \leadsto \) Query evaluation can be done in SQL, i.e., via an RDBMS
  \( (\text{Note: FOL is in LogSpace}) \).

- When \( Q \) is an NLogSpace-hard language
  \( \leadsto \) Query evaluation requires (at least) linear recursion.

- When \( Q \) is a PTIME-hard language
  \( \leadsto \) Query evaluation requires (at least) recursion (e.g., Datalog).

- When \( Q \) is a coNP-hard language
  \( \leadsto \) Query evaluation requires (at least) power of Disjunctive Datalog.
We now study $Q$-rewritability of query answering over $DL-Lite$ ontologies.

In particular we will show that both $DL-Lite^R$ and $DL-Lite^F$ enjoy FOL-rewritability of conjunctive query answering.
Query answering over unsatisfiable ontologies

- In the case in which an ontology is unsatisfiable, according to the “ex falso quod libet” principle, reasoning is trivialized.

- In particular, query answering is meaningless, since every tuple is in the answer to every query.

- We are not interested in encoding meaningless query answering into the perfect reformulation of the input query. Therefore, before query answering, we will always check ontology satisfiability to single out meaningful cases.

- Thus, in the following, we focus on query answering over satisfiable ontologies.

- We first consider satisfiable $DL$-Lite$_R$ ontologies.
Remark

we call **positive inclusions (PIs)** assertions of the form

\[ Cl \subseteq A \mid \exists Q \]
\[ Q_1 \subseteq Q_2 \]

whereas we call **negative inclusions (NIs)** assertions of the form

\[ Cl \subseteq \neg A \mid \neg \exists Q \]
\[ Q_1 \subseteq \neg Q_2 \]
Query answering in $DL$-$Lite^\mathcal{R}$

Given a CQ $q$ and a satisfiable ontology $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$, we compute $\text{cert}(q, \mathcal{O})$ as follows

1. using $\mathcal{T}$, reformulate $q$ as a union $r_{q,\mathcal{T}}$ of CQs
2. Evaluate $r_{q,\mathcal{T}}$ directly over $\mathcal{A}$ managed in secondary storage via a RDBMS

Correctness of this procedure shows FOL-rewritability of query answering in $DL$-$Lite^\mathcal{R}$

$\leadsto$ Query answering over $DL$-$Lite^\mathcal{R}$ ontologies can be done using RDMBS technology.
Query answering in $DL$-$Lite_R$: Query reformulation

**Intuition:** Use the PIs as basic rewriting rules

\[ q(x) \leftarrow \text{Professor}(x) \]

AssistantProf $\sqsubseteq$ Professor

as a logic rule: \[ \text{Professor}(z) \leftarrow \text{AssistantProf}(z) \]

**Basic rewriting step:**

- **when** the atom unifies with the **head** of the rule.
- **substitute** the atom with the **body** of the rule.

Towards the computation of the perfect reformulation, we add to the input query above the following query

\[ q(x) \leftarrow \text{AssistantProf}(x) \]

We say that the PI AssistantProf $\sqsubseteq$ Professor applies to the atom Professor$(x)$. 
Consider now the query

\[ q(x) \leftarrow \text{teaches}(x, y) \]

as a logic rule:

\[ \text{Professor} \sqsubseteq \exists \text{teaches} \]

We add to the reformulation the query

\[ q(x) \leftarrow \text{Professor}(x) \]
Conversely, for the query

\[ q(x) \leftarrow \text{teaches}(x, \text{databases}) \]

Professor \( \sqsubseteq \exists \text{teaches} \)

as a logic rule: \( \text{teaches}(z_1, z_2) \leftarrow \text{Professor}(z_1) \)

\text{teaches}(x, \text{databases}) \) does not unify with \( \text{teaches}(z_1, z_2) \), since the existentially quantified variable \( z_2 \) in the head of the rule does not unify with the constant \( \text{databases} \).

In this case the PI \textbf{does not apply} to the atom \( \text{teaches}(x, \text{databases}) \).

The same holds for the following query, where \( y \) is \textbf{distinguished}

\[ q(x, y) \leftarrow \text{teaches}(x, y) \]
An analogous behavior with join variables

\[ q(x) \leftarrow \text{teaches}(x, y), \text{Course}(y) \]

\[ \text{Professor} \sqsubseteq \exists \text{teaches} \]

as a logic rule:

\[ \text{teaches}(z_1, z_2) \leftarrow \text{Professor}(z_1) \]

The PI above does not apply to the atom \( \text{teaches}(x, y) \).

Conversely, the PI

\[ \exists \text{teaches}^- \sqsubseteq \text{Course} \]

as a logic rule:

\[ \text{Course}(z_2) \leftarrow \text{teaches}(z_1, z_2) \]

applies to the atom \( \text{Course}(y) \).

We add to the perfect reformulation the query

\[ q(x) \leftarrow \text{teaches}(x, y), \text{teaches}(z, y) \]
Query answering in $DL\text{-}Lite_{R}$: Query reformulation (cont’d)

We now have the query

$$q(x) \leftarrow \text{teaches}(x, y), \text{teaches}(z, y)$$

The PI

$$\text{Professor} \sqsubseteq \exists \text{teaches}$$

as a logic rule:

$$\text{teaches}(z_1, z_2) \leftarrow \text{Professor}(z_1)$$

does not apply to $\text{teaches}(x, y)$ nor $\text{teaches}(z, y)$, since $y$ is in join.

However, we can transform the above query by unifying the atoms $\text{teaches}(x, y), \text{teaches}(z_1, y)$. This rewriting step is called reduce, and produces the following query

$$q(x) \leftarrow \text{teaches}(x, y)$$

We can now apply the PI above, and add to the reformulation the query

$$q(x) \leftarrow \text{Professor}(x)$$
Query answering in $DL$-Lite$^R$: Query reformulation (cont’d)

Reformulate the CQ $q$ into a set of queries: apply to $q$ in all possible ways the PIs in the TBox $T$:

$A_1 \sqsubseteq A_2$ \quad \ldots, $A_2(x), \ldots \leadsto \ldots, A_1(x), \ldots$

$\exists P \sqsubseteq A$ \quad \ldots, $A(x), \ldots \leadsto \ldots, P(x, _), \ldots$

$\exists P^- \sqsubseteq A$ \quad \ldots, $A(x), \ldots \leadsto \ldots, P(_, x), \ldots$

$A \sqsubseteq \exists P$ \quad \ldots, $P(x, _), \ldots \leadsto \ldots, A(x), \ldots$

$A \sqsubseteq \exists P^-$ \quad \ldots, $P(_, x), \ldots \leadsto \ldots, A(x), \ldots$

$\exists P_1 \sqsubseteq \exists P_2$ \quad \ldots, $P_2(x, _), \ldots \leadsto \ldots, P_1(x, _), \ldots$

(_ denotes an unbound variable, i.e., a variable that appears only once)

This corresponds to exploiting ISAs, role typing, and mandatory participation to obtain new queries that could contribute to the answer.

Unifying atoms can make applicable rules that could not be applied otherwise.
Query answering in \( DL-Lite_\mathcal{R} \): Query reformulation (cont’d)

Algorithm PerfectRef \((q, T_P)\)

**Input:** conjunctive query \( q \), set of \( DL-Lite_\mathcal{R} \) PIs \( T_P \)

**Output:** union of conjunctive queries \( PR \)

\[
PR := \{q\};
\]

repeat

\[
PR' := PR;
\]

for each \( q \in PR' \) do

(a) for each \( g \) in \( q \) do

for each PI \( I \) in \( T_P \) do

if \( I \) is applicable to \( g \)

then \( PR := PR \cup \{ q[g/(g,I)] \} \)

(b) for each \( g_1, g_2 \) in \( q \) do

if \( g_1 \) and \( g_2 \) unify

then \( PR := PR \cup \{ \tau(reduce(q, g_1, g_2)) \} \)

until \( PR' = PR \);

return \( PR \)

Notice that NIs do not play any role in the reformulation of the query.
Query answering in *DL-Lite*$_R$: ABox storage

ABox $\mathcal{A}$ stored as a relational database in a standard RDBMS as follows:

- For each atomic concept $A$ used in the ABox:
  - define a unary relational table $\text{tab}_A$
  - populate $\text{tab}_A$ with each $\langle d \rangle$ such that $A(c) \in \mathcal{A}$

- For each atomic role $P$ used in the ABox,
  - define a binary relational table $\text{tab}_P$
  - populate $\text{tab}_P$ with each $\langle a, b \rangle$ such that $P(c_1, c_2) \in \mathcal{A}$

We denote with $\text{DB}(\mathcal{A})$ the database obtained as above.
Query answering in $DL$-$Lite_R$: Query evaluation

Let $r_{q,T}$ be the UCQ returned by the algorithm PerfectRef $(q, T)$

- We denote by $\text{SQL}(r_{q,T})$ the encoding of $r_{q,T}$ into an SQL query over $\text{DB}(\mathcal{A})$.

- We indicate with $\text{Eval}(\text{SQL}(r_{q,T}), \text{DB}(\mathcal{A}))$ the evaluation of $\text{SQL}(r_{q,T})$ over $\text{DB}(\mathcal{A})$. 
Query answering in $DL\text{-Lite}_R$

**Theorem**

Let $\mathcal{T}$ be a $DL\text{-Lite}_R$ TBox, $\mathcal{T}_P$ the set of PIs in $\mathcal{T}$, $q$ a CQ over $\mathcal{T}$, and let $r_{q,\mathcal{T}} = \text{PerfectRef} (q, \mathcal{T}_P)$. Then, for each ABox $\mathcal{A}$ such that $\langle \mathcal{T}, \mathcal{A} \rangle$ is satisfiable, we have that $\text{cert}(q, \langle \mathcal{T}, \mathcal{A} \rangle) = \text{Eval}(\text{SQL}(r_{q,\mathcal{T}}), \text{DB}(\mathcal{A}))$.

In other words, query answering over a satisfiable $DL\text{-Lite}_R$ ontology is FOL-rewritable.

Notice that we did not mention NIs of $\mathcal{T}$ in the theorem above. Indeed, when the ontology is satisfiable, we can ignore NIs and answer queries as NIs were not specified in $\mathcal{T}$. 
Query answering in $DL$-Lite$^R$: Example

**TBox:**

$\text{Professor} \sqsubseteq \exists \text{teaches} \exists \text{teaches}^{-} \sqsubseteq \text{Course}$

**Query:**

$q(x) \leftarrow \text{teaches}(x, y), \text{Course}(y)$

**Perfect Reformulation:**

$q(x) \leftarrow \text{teaches}(x, y), \text{Course}(y)$
$q(x) \leftarrow \text{teaches}(x, y), \text{teaches}(\_ , y)$
$q(x) \leftarrow \text{teaches}(x, \_)$
$q(x) \leftarrow \text{Professor}(x)$

**ABox:**

$\text{teaches}($John, databases$)$
$\text{Professor}($Mary$)$

It is easy to see that $\text{Eval}(\text{SQL}(r_q,T), \text{DB}(A))$ in this case produces the set $\{\text{John, Mary}\}$. 
Query answering in $\textit{DL-Lite}_R$: An interesting case

TBox: \begin{align*}
\text{Person} & \sqsubseteq \exists \text{hasFather} \\
\exists \text{hasFather}^- & \sqsubseteq \text{Person}
\end{align*}

ABox: \text{Person}(\text{Mary})

Query: \begin{align*}
q(x) & \leftarrow \text{Person}(x), \text{hasFather}(x, y_1), \text{hasFather}(y_1, y_2), \text{hasFather}(y_2, y_3) \\
q(x) & \leftarrow \text{Person}(x), \text{hasFather}(x, y_1), \text{hasFather}(y_1, y_2), \text{hasFather}(y_2, -) \\
& \downarrow \quad \text{Apply } \text{Person} \sqsubseteq \exists \text{hasFather} \text{ to the atom } \text{hasFather}(y_2, -) \\
q(x) & \leftarrow \text{Person}(x), \text{hasFather}(x, y_1), \text{hasFather}(y_1, y_2), \text{Person}(y_2) \\
& \downarrow \quad \text{Apply } \exists \text{hasFather}^- \sqsubseteq \text{Person} \text{ to the atom } \text{Person}(y_2) \\
q(x) & \leftarrow \text{Person}(x), \text{hasFather}(x, y_1), \text{hasFather}(y_1, y_2), \text{hasFather}(-, y_2) \\
& \downarrow \quad \text{Unify atoms } \text{hasFather}(y_1, y_2) \text{ and } \text{hasFather}(-, y_2) \\
q(x) & \leftarrow \text{Person}(x), \text{hasFather}(x, y_1), \text{hasFather}(y_1, y_2) \\
& \downarrow \\
& \ldots \\
q(x) & \leftarrow \text{Person}(x), \text{hasFather}(x, -) \\
& \downarrow \quad \text{Apply } \text{Person} \sqsubseteq \exists \text{hasFather} \text{ to the atom } \text{hasFather}(x, -) \\
q(x) & \leftarrow \text{Person}(x)
\end{align*}
Query answering in $DL-Lite_F$

If we limit our attention to PIs, we can say that $DL-Lite_F$ ontologies are $DL-Lite_R$ ontologies of a special kind (i.e., with no PIs between roles).

As for NIs and functionality assertions, it is possible to show that they can be disregarded in query answering over satisfiable $DL-Lite_F$ ontologies.

The following result is therefore straightforward.

**Theorem**

Let $\mathcal{T}$ be a $DL-Lite_F$ TBox, $\mathcal{T}_P$ the set of PIs in $\mathcal{T}$, $q$ a CQ over $\mathcal{T}$, and let $r_{q,\mathcal{T}} =$PerfectRef $(q, \mathcal{T}_P)$. Then, for each ABox $\mathcal{A}$ such that $\langle \mathcal{T}, \mathcal{A} \rangle$ is satisfiable, we have that $\text{cert}(q, \langle \mathcal{T}, \mathcal{A} \rangle) = \text{Eval}($SQL$(r_{q,\mathcal{T}}), \text{DB}(\mathcal{A}))$.

In other words, query answering over a satisfiable $DL-Lite_F$ ontology is FOL-rewritable.
Satisfiability of ontologies with only PIs

Let us now attack the problem of establishing whether a ontology is satisfiable.

Remember that solving this problem allow us to solve TBox reasoning and identify cases in which query answering is meaningless.

A first notable result says us that PIs alone cannot generate ontology unsatisfiability.

**Theorem**

Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be either a $DL\text{-}Lite_\mathcal{R}$ or a $DL\text{-}Lite_\mathcal{F}$ ontology, where $\mathcal{T}$ contains only PIs. Then, $\mathcal{O}$ is satisfiable.
Unsatisfiability in \( DL-Lite_\mathcal{R} \) ontologies can be however caused by NIs

Example: \( \text{TBox } \mathcal{T} \): Professor \( \sqsubseteq \neg \text{Student} \)
\( \exists \text{teaches } \sqsubseteq \text{Professor} \)

\( \text{ABox } \mathcal{A} \): \( \text{teaches(John, databases)} \)
\( \text{Student(John)} \)

In what follows we provide a mechanism to establish, in an efficient way, whether a \( DL-Lite_\mathcal{R} \) ontology is satisfiable.
Checking satisfiability of $DL$-$Lite_\mathcal{R}$ ontologies

Satisfiability of a $DL$-$Lite_\mathcal{R}$ ontology $O = \langle T, A \rangle$ is reduced to evaluating a FOL-query (in fact a UCQ) over $DB(A)$.

We proceed as follows:

1. Let $T_P$ the set of PIs in $T$.

2. For each NI between concepts (resp. roles) in $T$, we ask $\langle T_P, A \rangle$ if there exists some individual (resp. pair of individuals) that contradicts $N$, i.e., we pose over $\langle T_P, A \rangle$ a boolean CQ $q_N$ such that $\langle T_P, A \rangle \models q_N$ iff $\langle T_P \cup \{N\}, A \rangle$ is unsatisfiable.

3. We exploit PerfectRef to verify if $\langle T_P, A \rangle \models q_N$, i.e., we compute $\text{PerfectRef}(q_N, T_P)$, and evaluate it (in fact its SQL encoding) over $DB(A)$.
Example

\( \text{Pls } T_P: \exists \text{teaches } \sqsubseteq \text{Professor} \)

\( \text{NI } N: \text{Professor } \sqsubseteq \neg \text{Student} \)

\( \text{Query } q_N: q() \leftarrow \text{Student}(x), \text{Professor}(x) \)

\( \text{Perfect Reformulation: } q() \leftarrow \text{Student}(x), \text{Professor}(x) \)

\( q() \leftarrow \text{Student}(x), \text{teaches}(x, _) \)

\( \text{ABox } A: \text{teaches}(\text{John, databases}) \)

\( \text{Student}(\text{John}) \)

It is easy to see that \( \langle T_P, A \rangle \models q_N \), and that \( \langle T_P \cup \{ \text{Professor } \sqsubseteq \neg \text{Student} \}, A \rangle \) is unsatisfiable.
Queries for NIs

For each NI in $\mathcal{T}$ we compute a boolean CQ according to the following rules:

\[
\begin{align*}
A_1 & \sqsubseteq \neg A_2 & \Rightarrow & q() \leftarrow A_1(x), A_2(x) \\
\exists P & \sqsubseteq \neg A \quad \text{or} \quad A & \sqsubseteq \neg \exists P & \Rightarrow & q() \leftarrow P(x, y), A(x) \\
\exists P^- & \sqsubseteq \neg A \quad \text{or} \quad A & \sqsubseteq \neg \exists P^- & \Rightarrow & q() \leftarrow P(y, x), A(x) \\
\exists P_1 & \sqsubseteq \neg \exists P_2 & \Rightarrow & q() \leftarrow P_1(x, y), P_2(x, z) \\
\exists P_1 & \sqsubseteq \neg \exists P_2^- & \Rightarrow & q() \leftarrow P_1(x, y), P_2(z, x) \\
\exists P_1^- & \sqsubseteq \neg \exists P_2 & \Rightarrow & q() \leftarrow P_1(x, y), P_2(y, z) \\
\exists P_1^- & \sqsubseteq \neg \exists P_2^- & \Rightarrow & q() \leftarrow P_1(x, y), P_2(z, y) \\
P_1 & \sqsubseteq \neg P_2 \quad \text{or} \quad P_1^- & \sqsubseteq \neg P_2^- & \Rightarrow & q() \leftarrow P_1(x, y), P_2(x, y) \\
P_1^- & \sqsubseteq \neg P_2 \quad \text{or} \quad P_1 & \sqsubseteq \neg P_2^- & \Rightarrow & q() \leftarrow P_1(x, y), P_2(y, x)
\end{align*}
\]

Given a NI $N \in \mathcal{T}$, we denote with $q_N$ the corresponding CQ.
**Lemma [separation]**

Let $O = \langle T, A \rangle$ be a $DL-Lite_\mathcal{R}$ ontology, and $T_P$ the set of PIs in $T$. Then, $O$ is unsatisfiable iff there exists a NI $N \in T$ such that $\langle T_P, A \rangle \models q_N$.

The lemma relies on the properties that NIs do not interact with each other, and interaction between NIs and PIs can be managed through PerfectRef. Notably, each NI can be processed individually.
**DL-Lite_\mathcal{R}:** FOL-rewritability of satisfiability

From the lemma above and the theorem on query answering for satisfiable DL-Lite_\mathcal{R}, we get the following result

**Theorem**

Let \( \mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle \) be a DL-Lite_\mathcal{R} ontology, and \( \mathcal{T}_P \) the set of PIs in \( \mathcal{T} \). Then, \( \mathcal{O} \) is unsatisfiable iff there exists a NI \( N \in \mathcal{T} \) such that\( \text{Eval}(\text{SQL}(\text{PerfectRef}(q_N, \mathcal{T}_P)), \text{DB}(\mathcal{A})) \) returns true.

In other words, satisfiability of DL-Lite_\mathcal{R} ontology can be reduced to FOL-query evaluation.
Unsatisfiability in $DL-Lite_{\mathcal{F}}$ ontologies can be caused by NIs and functionality assertions.

Example: \[ \text{TBox } \mathcal{T} : \text{Professor } \sqsubseteq \neg \text{Student} \]
\[ \exists \text{teaches } \sqsubseteq \text{Professor} \]
\[ (\text{funct} \ \text{teaches}^{-}) \]

\[ \text{ABox } \mathcal{A} : \text{teaches}(\text{John}, \text{databases}) \]
\[ \text{teaches}(\text{Michael}, \text{databases}) \]

In what follows we extend to $DL-Lite_{\mathcal{F}}$ ontologies the technique for $DL-Lite_{\mathcal{R}}$ ontology satisfiability given before.
Checking satisfiability of $DL$-$Lite\mathcal{F}$ ontologies

Satisfiability of a $DL$-$Lite\mathcal{F}$ ontology $\mathcal{O} = \langle T, A \rangle$ is reduced to evaluating a FOL-query over $DB(A)$.

We deal with NIs exactly as done in $DL$-$Lite\mathcal{R}$ ontologies (indeed, limited to NIs, $DL$-$Lite\mathcal{F}$ ontologies are $DL$-$Lite\mathcal{R}$ ontologies of a special kind).

As for functionality assertions we proceed as follows:

1. For each functionality assertion $F \in T$ we ask if there exists two pairs of individuals in $A$ that contradict $F$, i.e., we pose over $A$ a boolean FOL query $q_F$ such that $A \models q_F$ iff $\langle \{F\}, A \rangle$ is unsatisfiable.

2. To verify if $A \models q_F$, we evaluate $SQL(q_F)$ over $DB(A)$.
Example

**Functionality** $F$: \((\text{funct} \ telescaches^-)\)

**Query** $q_F$: $q() \leftarrow \text{TeachesTo}(x, y), \text{TeachesTo}(z, y), x \neq z$

**ABox** $\mathcal{A}$: $\text{teaches}(\text{John}, \text{databases})$
          $\text{teaches}(\text{Michael}, \text{databases})$

It is easy to see that $\mathcal{A} \models q_F$, and that \(\langle \{(\text{funct} \ telescaches^-)\}, \mathcal{A}\rangle\), is unsatisfiable.
Queries for functionality assertions

For each functionality assertion in $T$ we compute a FOL query according to the following rules:

$$\begin{align*}
\text{(funct } P) & \rightsquigarrow q() \leftarrow P(x, y), P(x, z), y \neq z \\
\text{(funct } P^-) & \rightsquigarrow q() \leftarrow P(x, y), P(z, y), x \neq z
\end{align*}$$

Given a functionality assertion $F \in T$, we denote with $q_F$ the corresponding FOL query.
**DL-Lite\(_R\): From satisfiability to query answering**

**Lemma**

Let \( O = \langle T, A \rangle \) be a DL-Lite\(_F\) ontology, and \( T_P \) the set of PIs in \( T \). Then, \( O \) is unsatisfiable iff one of the following condition holds

1. (a) there exists a NI \( N \in T \) such that \( \langle T_P, A \rangle \models q_N \)
2. (b) there exists a functionality assertion \( F \in T \) such that \( A \models q_F \).

The lemma relies on the properties that NIs do not interact with each other, and interaction between NIs and PIs can be managed through PerfectRef.

It also exploits the properties that NIs and PIs do not interact with functionalities: indeed, no functionality assertions are contradicted in a DL-Lite\(_F\) ontology \( O \), beyond those explicitly contradicted by the ABox.

Notably, the lemma asserts that to check ontology satisfiability, each NI and each functionality can be processed individually.
By the lemma above and the theorem on query answering for satisfiable $DL\text{-}Lite_{\mathcal{F}}$, the following result follows:

**Theorem**

Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a $DL\text{-}Lite_{\mathcal{F}}$ ontology, and $\mathcal{T}_P$ the set of PIs in $\mathcal{O}$. Then, $\mathcal{O}$ is unsatisfiable iff one of the following condition holds.

(a) there exists a NI $N \in \mathcal{T}$ such that 
$\text{Eval}(\text{SQL}(\text{PerfectRef}(q_N, \mathcal{T}_P)), \text{DB(}\mathcal{A}))$ returns *true*.

(b) there exists a functionality assertion $F \in \mathcal{T}$ such that 
$\text{Eval}(\text{SQL}(q_F), \text{DB(}\mathcal{A}))$ returns *true*.

In other words, satisfiability of a $DL\text{-}Lite_{\mathcal{F}}$ ontology can be reduced to FOL-query evaluation.
Outline

7 TBox reasoning

8 TBox & ABox Reasoning

9 Complexity of reasoning in Description Logics
   - Complexity of reasoning in DL-Lite
   - Data complexity of query answering in DLs
   - NLOGSPACE-hard DLs
   - PTIME-hard DLs
   - coNP-hard DLs
Complexity of query answering over satisfiable ontologies

**Theorem**

Query answering over both $DL-Lite_R$ and $DL-Lite_F$ ontologies is

1. **NP-complete** in the size of query and ontology (combined comp.).
2. **PTime** in the size of the ontology.
3. **LogSpace** in the size of the ABox (data complexity).

**Proof (sketch)**

1. We **guess** the derivation of one of the CQs of the perfect rewriting, and an assignment to its existential variables. Checking the derivation and evaluating the guessed CQ over the ABox is then polynomial in combined complexity. NP-hardness follows from combined complexity of evaluating CQs over a database.

2. The number of CQs in the perfect reformulation is polynomial in the size of the TBox, and we can get them in PTime.

3. Is the data complexity of evaluating FOL queries over a database.
Complexity of reasoning in \( DL-Lite \)

### Complexity of ontology satisfiability

**Theorem**

Checking satisfiability of both \( DL-Lite_R \) and \( DL-Lite_F \) ontologies is

1. \( \text{PTIME} \) in the size of the ontology (combined complexity).
2. \( \text{LogSpace} \) in the size of the ABox (data complexity).

**Proof (sketch)**

Follows directly from the algorithm for ontology satisfiability and the complexity of query answering over satisfiable ontologies.
Theorem

TBox reasoning over both $DL-Lite_{\mathcal{R}}$ and $DL-Lite_{\mathcal{F}}$ ontologies is \textbf{PTime} in the size of the TBox (schema complexity).

Proof (sketch)

Follows from the previous theorem, and from the reduction of TBox reasoning to ontology satisfiability. Indeed, the size of the ontology constructed in the reduction is polynomial in the size of the input TBox.
Can we further extend these results to more expressive ontology languages?

Essentially NO!
(unless we take particular care)
Beyond **DL-Lite**

We now consider DL languages that allow for constructs not present in **DL-Lite** or for combinations of constructs that are not legal in **DL-Lite**. We recall here syntax and semantics of constructs used in what follows.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Syntax</th>
<th>Example</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>conjunction</td>
<td>$C_1 \sqcap C_2$</td>
<td>Doctor $\sqcap$ Male</td>
<td>$C_1^T \sqcap C_2^T$</td>
</tr>
<tr>
<td>disjunction</td>
<td>$C_1 \sqcup C_2$</td>
<td>Doctor $\sqcup$ Lawyer</td>
<td>$C_1^T \sqcup C_2^T$</td>
</tr>
<tr>
<td>qual. exist. restr.</td>
<td>$\exists Q.C$</td>
<td>$\exists$child$.$Male</td>
<td>${a \mid \exists b. (a, b) \in Q^T \land b \in C^T }$</td>
</tr>
<tr>
<td>qual. univ. restr.</td>
<td>$\forall Q.C$</td>
<td>$\forall$child$.$Male</td>
<td>${a \mid \forall b. (a, b) \in Q^T \rightarrow b \in C^T }$</td>
</tr>
</tbody>
</table>
### Summary of results on data complexity

<table>
<thead>
<tr>
<th>Cl</th>
<th>Cr</th>
<th>$\mathcal{F}$</th>
<th>$\mathcal{R}$</th>
<th>Data complexity of query answering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$DL$-Lite$_{\mathcal{F}}$</td>
<td>$\sqrt{-}$</td>
<td>$-$</td>
<td>in LogSpace</td>
</tr>
<tr>
<td>2</td>
<td>$DL$-Lite$_{\mathcal{R}}$</td>
<td>$-$</td>
<td>$\sqrt{-}$</td>
<td>in LogSpace</td>
</tr>
<tr>
<td>3</td>
<td>$A \mid \exists P.A$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>4</td>
<td>$A$</td>
<td>$A \mid \forall P.A$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>5</td>
<td>$A$</td>
<td>$A \mid \exists P.A$</td>
<td>$\sqrt{-}$</td>
<td>$-$</td>
</tr>
<tr>
<td>6</td>
<td>$A \mid \exists P.A \mid A_1 \sqcap A_2$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>7</td>
<td>$A \mid A_1 \sqcap A_2$</td>
<td>$A \mid \forall P.A$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>8</td>
<td>$A \mid A_1 \sqcap A_2$</td>
<td>$A \mid \exists P.A$</td>
<td>$\sqrt{-}$</td>
<td>$-$</td>
</tr>
<tr>
<td>9</td>
<td>$A \mid \exists P.A \mid \exists P^- .A$</td>
<td>$A \mid \exists P$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>10</td>
<td>$A$</td>
<td>$A \mid \exists P .A \mid \exists P^- .A$</td>
<td>$\sqrt{-}$</td>
<td>$-$</td>
</tr>
<tr>
<td>11</td>
<td>$A \mid \exists P .A$</td>
<td>$A \mid \exists P .A$</td>
<td>$\sqrt{-}$</td>
<td>$-$</td>
</tr>
<tr>
<td>12</td>
<td>$A \mid \neg A$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>13</td>
<td>$A$</td>
<td>$A \mid A_1 \sqcup A_2$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>14</td>
<td>$A \mid \forall P.A$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

All \textbf{NLogSpace} and \textbf{PTIME} hardness results hold already for atomic queries.

D. Calvanese, D. Lembo
Ontology-Based Data Access
ISWC’07 – Nov. 12, 2007
Observations

- **DL-Lite-family** is FOL-rewritable, hence $\text{LogSpace}$ – holds also with $n$-ary relations $\rightsquigarrow DLR-Lite_\mathcal{F}$ and $DLR-Lite_\mathcal{R}$.
- **RDFS** is a subset of $DL-Lite_\mathcal{R} \rightsquigarrow$ is FOL-rewritable, hence $\text{LogSpace}$.
- **Horn-SHIQ** [HMS05] is $\text{PTIME-hard}$ even for instance checking (line 11).
- **DLP** [GHVD03] is $\text{PTIME-hard}$ (line 6)
- **$\mathcal{EL}$** [BBL05] is $\text{PTIME-hard}$ (line 6).
Qualified existential quantification in the lhs of inclusions

Adding **qualified existential on the lhs** of inclusions makes instance checking (and hence query answering) **NLogSpace-hard**:

<table>
<thead>
<tr>
<th>Cl</th>
<th>Cr</th>
<th>$\mathcal{F}$</th>
<th>$\mathcal{R}$</th>
<th>Data complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$A</td>
<td>\exists P . A$</td>
<td>$A$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

Hardness proof is by a reduction from reachability in directed graphs:

Ontology $\mathcal{O}$: a single inclusion assertion $\exists P . A \sqsubseteq A$

Database $\mathcal{D}$: encodes graph using $P$ and asserts $A(d)$

Result:

$(\mathcal{O}, \mathcal{D}) \models A(s)$ iff $d$ is reachable from $s$ in the graph.
Instance checking (and hence query answering) is $\text{NLogSpace}$-hard in data complexity for:

<table>
<thead>
<tr>
<th></th>
<th>$Cl$</th>
<th>$Cr$</th>
<th>$F$</th>
<th>$R$</th>
<th>Data complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$A \mid \exists P.A$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
<td>$\text{NLogSpace}$-hard</td>
</tr>
</tbody>
</table>

By reduction from reachability in directed graphs

4

|   | $A$ | $A \mid \forall P.A$ | $-$ | $-$ | $\text{NLogSpace}$-hard |

Follows from 3 by replacing $\exists P.A_1 \sqsubseteq A_2$ with $A_1 \sqsubseteq \forall P^- . A_2$

5

|   | $A$ | $A \mid \exists P.A$ | $\sqrt{}$ | $-$ | $\text{NLogSpace}$-hard |

Proved by simulating in the reduction $\exists P.A_1 \sqsubseteq A_2$

via $A_1 \sqsubseteq \exists P^- . A_2$ and (funct $P^-$)
Path System Accessibility

Instance of Path System Accessibility: $PS = (N, E, S, t)$ with

- $N$ a set of nodes
- $E \subseteq N \times N \times N$ an accessibility relation
- $S \subseteq N$ a set of source nodes
- $t \in N$ a terminal node

Accessibility of nodes is defined inductively:

- each $n \in S$ is accessible
- if $(n, n_1, n_2) \in E$ and $n_1, n_2$ are accessible, then also $n$ is accessible

Given $PS$, checking whether $t$ is accessible, is $\text{PTIME}$-complete.
Reduction from Path System Accessibility

Given an instance $PS = (N, E, S, t)$, we construct

- **TBox $T$** consisting of the inclusion assertions
  
  $\exists P_1 . A \sqsubseteq B_1$
  
  $\exists P_2 . A \sqsubseteq B_2$
  
  $B_1 \sqcap B_2 \sqsubseteq A$
  
  $\exists P_3 . A \sqsubseteq A$

- **ABox $A$** encoding the accessibility relation using $P_1$, $P_2$, and $P_3$, and asserting $A(s)$ for each source node $s \in S$

  $e_1 = (n, \ldots, \cdot)$
  
  $e_2 = (n, s_1, s_2)$
  
  $e_3 = (n, \ldots, \cdot)$

Result:

$\langle T, A \rangle \models A(t)$ iff $t$ is accessible in $PS$. 
Are obtained when we can use in the query two concepts that cover another concept. This forces reasoning by cases on the data.

Query answering is coNP-hard in data complexity for:

<table>
<thead>
<tr>
<th></th>
<th>$Cl$</th>
<th>$Cr$</th>
<th>$\mathcal{F}$</th>
<th>$\mathcal{R}$</th>
<th>Data complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>$A \mid \neg A$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
<td>coNP-hard</td>
</tr>
<tr>
<td>15</td>
<td>$A$</td>
<td>$A \mid A_1 \sqcup A_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>coNP-hard</td>
</tr>
<tr>
<td>16</td>
<td>$A \mid \forall P.A$</td>
<td>$A$</td>
<td>$-$</td>
<td>$-$</td>
<td>coNP-hard</td>
</tr>
</tbody>
</table>

All three cases are proved by adapting the proof of coNP-hardness of instance checking for $\mathcal{ALC}$ by [DLNS94].
2+2-SAT: satisfiability of a 2+2-CNF formula, i.e., a CNF formula where each clause has exactly 2 positive and 2 negative literals.

Example: \( \varphi = c_1 \land c_2 \land c_3 \), with
\[
\begin{align*}
c_1 & = \ell_1 \lor \ell_2 \lor \neg \ell_3 \lor \neg \ell_4 \\
c_2 & = false \lor false \lor \neg \ell_1 \lor \neg \ell_4 \\
c_3 & = false \lor \ell_4 \lor \neg true \lor \neg \ell_2 
\end{align*}
\]

2+2-SAT is NP-complete [DLNS94].
Reduction from 2+2-SAT

2+2-CNF formula $\varphi = c_1 \land \cdots \land c_k$ over letters $\ell_1, \ldots, \ell_n, true, false$

- **ABox** $A_\varphi$ constructed from $\varphi$ (concepts $L$, $T$, $F$, roles $P_1$, $P_2$, $N_1$, $N_2$):
  - for each letter $\ell_i$: $L(\ell_i)$
  - for each clause $c = \ell_1 \lor \ell_2 \lor \neg \ell_3 \lor \neg \ell_4$:
    - $P_1(c, \ell_1)$, $P_2(c, \ell_2)$, $N_1(c, \ell_3)$, $N_2(c, \ell_4)$
  - $T(true)$, $F(false)$

- **TBox** $T = \{ L \sqsubseteq T \sqcup F \}$

- $q() \leftarrow P_1(c, \ell_1), P_2(c, \ell_2), N_1(c, \ell_3), N_2(c, \ell_4), F(\ell_1), F(\ell_2), T(\ell_3), T(\ell_4)$

We have: $\langle T, A_\varphi \rangle \models q$ iff $\varphi$ is not satisfiable.

Intuition: each model of $T$ partitions $L$ into $T$ and $F$, and corresponds to a truth assignment to $\ell_1, \ldots, \ell_n$. $q$ asks for a false clause.
Part IV

Linking data to ontologies
Outline

10 The Description Logic $DL$-$Lite_A$

11 Connecting ontologies to relational data
The Description Logic $DL$-Lite$_A$

10 The Description Logic $DL$-Lite$_A$
- Missing features in $DL$-Lite
- Combining functionality and role inclusions
- Syntax and semantics of $DL$-Lite$_A$
- Reasoning in $DL$-Lite$_A$

11 Connecting ontologies to relational data
What is missing in \textit{DL-Lite} wrt popular data models?

Let us consider UML class diagrams that have the following features:
- functionality of associations (i.e., \textit{roles})
- inclusion (i.e., ISA) between associations
- attributes of concepts and associations, possibly functional
- covering constraints in hierarchies

What can we capture of these while maintaining FOL-rewritability?
1. We can \textit{forget about covering constraints}, since they make query answering $\text{coNP}$-hard in data complexity (see Part 3).
2. Attributes of concepts are “syntactic sugar” (they could be modeled by means of \textit{roles}), but their functionality is an issue.
3. We could also add \textit{attributes of roles} (we won’t discuss this here).
4. \textit{Functionality and role inclusions} are present separately (in $\text{DL-Lite}_F$ and $\text{DL-Lite}_R$), but \textit{were not allowed to be used together}.

Let us first analyze this last point.
Combining functionalities and role inclusions

We have seen till now that:

- By including in DL-Lite both functionality of roles and qualified existential quantification (i.e., $\exists P.A$), query answering becomes \textbf{NLogSpace}-hard (and \textbf{PTIME}-hard with also inverse roles) in data complexity (see Part 3).
- Qualified existential quantification can be simulated by using role inclusion assertions (see Part 2).
- When the data complexity of query answering is \textbf{NLogSpace} or above, the DL does not enjoy FOL-rewritability.

As a consequence of these results, we get:

To preserve FOL-rewritability, we need to restrict the interaction of functionality and role inclusions.

Let us analyze on an example the effect of an unrestricted interaction.
Combining functionalities and role inclusions – Example

\( TBox \ T: \quad A \sqsubseteq \exists P \quad P \sqsubseteq S \\
\exists P^- \sqsubseteq A \quad (\text{funct } S) \)

\( ABox \ A: \quad A(c_1), \ S(c_1, c_2), \ S(c_2, c_3), \ldots, \ S(c_{n-1}, c_n) \)

\[
\begin{align*}
A(c_1), & \quad A \sqsubseteq \exists P \quad \models P(c_1, x), \text{ for some } x \\
P(c_1, x), & \quad P \sqsubseteq S \quad \models S(c_1, x) \\
S(c_1, x), & \quad S(c_1, c_2), \quad (\text{funct } S) \quad \models x = c_2 \\
P(c_1, c_2), & \quad \exists P^- \sqsubseteq A \quad \models A(c_2) \\
A(c_2), & \quad A \sqsubseteq \exists P \quad \models A(c_2) \\
\vdots & \quad \models A(c_n) 
\end{align*}
\]

Hence, we get:

- If we add \( B(c_n) \) and \( B \sqsubseteq \neg A \), the ontology becomes inconsistent.
- Similarly, the answer to the following query over \( \langle T, A \rangle \) is true:

\[
q() \leftarrow A(z_1), S(z_1, z_2), S(z_2, z_3), \ldots, S(z_{n-1}, z_n), A(z_n)
\]
Restrictions on combining functionalities and role inclusions

Note: The number of unification steps above depends on the data. Hence this kind of deduction cannot be mimicked by a FOL (or SQL) query, since it requires a form of recursion. As a consequence, we get:

Combining functionality and role inclusions is problematic.

It breaks separability, i.e., functionality assertions may force existentially quantified objects to be unified with existing objects.

Note: the problems are caused by the interaction among:

- an inclusion $P \sqsubseteq S$ between roles,
- a functionality assertion $(\text{funct } S)$ on the super-role, and
- a cycle of concept inclusion assertions $A \sqsubseteq \exists P$ and $\exists P^- \sqsubseteq A$.

Since we do not want to limit cycles of ISA, we pose suitable restrictions on the combination of functionality and role inclusions.
Features of $DL$-Lite$_A$

$DL$-Lite$_A$ is a Description Logic designed to capture as much features as possible of conceptual data models, while preserving nice computational properties for query answering.

- **Enjoys FOL-rewritability**, and hence is $\text{LogSpace}$ in data complexity.
- **Allows for both functionality assertions and role inclusion assertions**, but restricts in a suitable way their interaction.
- **Takes into account the distinction between objects and values:**
  - Objects are elements of an abstract interpretation domain.
  - Values are elements of concrete data types, such as integers, strings, etc.
- **Values are connected to objects through attributes**, rather than roles (we consider here only concept attributes and not role attributes $[\text{CDGL}^+06a]$).
Syntax of the DL-Lite_\text{A} description language

- **Concept expressions:**
  \[ B \quad \rightarrow \quad A \mid \exists Q \mid \delta(U) \]
  \[ C \quad \rightarrow \quad \top_C \mid B \mid \neg B \mid \exists Q.C \]

- **Role expressions:**
  \[ Q \quad \rightarrow \quad P \mid P^- \]
  \[ R \quad \rightarrow \quad Q \mid \neg Q \]

- **Value-domain expressions:** (each \( T_i \) is one of the RDF datatypes)
  \[ E \quad \rightarrow \quad \rho(U) \]
  \[ F \quad \rightarrow \quad \top_D \mid T_1 \mid \cdots \mid T_n \]

- **Attribute expressions:**
  \[ V \quad \rightarrow \quad U \mid \neg U \]
Semantics of $DL\text{-}Lite_A$ – Objects vs. values

We make use of an alphabet $\Gamma$ of constants, partitioned into:

- an alphabet $\Gamma_O$ of object constants.
- an alphabet $\Gamma_V$ of value constants, in turn partitioned into alphabets $\Gamma_{V_i}$, one for each RDF datatype $T_i$.

The interpretation domain $\Delta^I$ is partitioned into:

- a domain of objects $\Delta^I_O$
- a domain of values $\Delta^I_V$

The semantics of $DL\text{-}Lite_A$ descriptions is determined as usual, considering the following:

- The interpretation $C^I$ of a concept $C$ is a subset of $\Delta^I_O$.
- The interpretation $R^I$ of a role $R$ is a subset of $\Delta^I_O \times \Delta^I_O$.
- The interpretation $val(v)$ of each value constant $v$ in $\Gamma_V$ and RDF datatype $T_i$ is given a priori (e.g., all strings for $xsd:\text{string}$).
- The interpretation $V^I$ of an attribute $V$ is a subset of $\Delta^I_O \times \Delta^I_V$. 
## Semantics of the DL-Lite<sub>A</sub> constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Syntax</th>
<th>Example</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>top concept</td>
<td>⊤&lt;sub&gt;C&lt;/sub&gt;</td>
<td></td>
<td>⊤&lt;sub&gt;C&lt;/sub&gt; = Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>atomic concept</td>
<td>A</td>
<td>Doctor</td>
<td>A&lt;sup&gt;T&lt;/sup&gt; ⊆ Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>existential restriction</td>
<td>⊤Q</td>
<td>⊤child^-</td>
<td>{o</td>
</tr>
<tr>
<td>qualified exist. restriction</td>
<td>⊤Q.C</td>
<td>⊤child.Male</td>
<td>{o</td>
</tr>
<tr>
<td>concept negation</td>
<td>⊤¬B</td>
<td>¬∃child</td>
<td>Δ&lt;sup&gt;T&lt;/sup&gt; \ B&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>attribute domain</td>
<td>δ(U)</td>
<td>δ(salary)</td>
<td>{o</td>
</tr>
<tr>
<td>atomic role</td>
<td>⊤P</td>
<td>child</td>
<td>⊤P&lt;sup&gt;T&lt;/sup&gt; ⊆ Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt; × Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>inverse role</td>
<td>⊤P^-</td>
<td>child^-</td>
<td>{(o, o')</td>
</tr>
<tr>
<td>role negation</td>
<td>⊤¬Q</td>
<td>¬manages</td>
<td>(Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt; × Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;) \ Q&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>top domain</td>
<td>⊤&lt;sub&gt;D&lt;/sub&gt;</td>
<td></td>
<td>⊤&lt;sub&gt;D&lt;/sub&gt; = Δ&lt;sub&gt;V&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>datatype</td>
<td>⊤Ti</td>
<td>xsd:int</td>
<td>val(T&lt;sub&gt;i&lt;/sub&gt;) ⊆ Δ&lt;sub&gt;V&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>attribute range</td>
<td>⊤ρ(U)</td>
<td>ρ(salary)</td>
<td>{v</td>
</tr>
<tr>
<td>atomic attribute</td>
<td>⊤U</td>
<td>salary</td>
<td>⊤U&lt;sup&gt;T&lt;/sup&gt; ⊆ Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt; × Δ&lt;sub&gt;V&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>attribute negation</td>
<td>⊤¬U</td>
<td>¬salary</td>
<td>(Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt; × Δ&lt;sub&gt;V&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;) \ U&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>object constant</td>
<td>⊤c</td>
<td>john</td>
<td>c&lt;sup&gt;T&lt;/sup&gt; ∈ Δ&lt;sub&gt;O&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
<tr>
<td>value constant</td>
<td>⊤v</td>
<td>'john'</td>
<td>val(v) ∈ Δ&lt;sub&gt;V&lt;/sub&gt;&lt;sup&gt;T&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
DL-Lite assertions

TBox assertions can have the following forms:

- $B \sqsubseteq C$: concept inclusion assertion
- $Q \sqsubseteq R$: role inclusion assertion
- $E \sqsubseteq F$: value-domain inclusion assertion
- $U \sqsubseteq V$: attribute inclusion assertion
- $(\text{funct } Q)$: role functionality assertion
- $(\text{funct } U)$: attribute functionality assertion

ABox assertions: $A(c), \ P(c, c'), \ U(c, d)$,

where $c, c'$ are object constants
$d$ is a value constant
## Semantics of the $DL\text{-}Lite_\mathcal{A}$ assertions

<table>
<thead>
<tr>
<th>Assertion</th>
<th>Syntax</th>
<th>Example</th>
<th>Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td>conc. incl.</td>
<td>$B \sqsubseteq C$</td>
<td>Father $\sqsubseteq \exists \text{child}$</td>
<td>$B^I \subseteq C^I$</td>
</tr>
<tr>
<td>role incl.</td>
<td>$Q \sqsubseteq R$</td>
<td>father $\sqsubseteq \text{anc}$</td>
<td>$Q^I \subseteq R^I$</td>
</tr>
<tr>
<td>v.dom. incl.</td>
<td>$E \sqsubseteq F$</td>
<td>$\rho(\text{age}) \sqsubseteq \text{xsd:int}$</td>
<td>$E^I \subseteq F^I$</td>
</tr>
<tr>
<td>attr. incl.</td>
<td>$U \sqsubseteq V$</td>
<td>offPhone $\sqsubseteq \text{phone}$</td>
<td>$U^I \subseteq V^I$</td>
</tr>
<tr>
<td>role funct.</td>
<td>$\text{funct} \ Q$</td>
<td>$\text{funct} \ \text{father}$</td>
<td>$\forall o, o', o''. (o, o') \in Q^I \land (o, o'') \in Q^I \rightarrow o' = o''$</td>
</tr>
<tr>
<td>att. funct.</td>
<td>$\text{funct} \ U$</td>
<td>$\text{funct} \ \text{ssn}$</td>
<td>$\forall o, v, v'. (o, v) \in U^I \land (o, v') \in U^I \rightarrow v = v'$</td>
</tr>
<tr>
<td>mem. asser.</td>
<td>$A(c)$</td>
<td>Father(bob)</td>
<td>$c^I \in A^I$</td>
</tr>
<tr>
<td>mem. asser.</td>
<td>$P(c_1, c_2)$</td>
<td>child(bob, ann)</td>
<td>$(c_1^I, c_2^I) \in P^I$</td>
</tr>
<tr>
<td>mem. asser.</td>
<td>$U(c, d)$</td>
<td>phone(bob, ’2345’)</td>
<td>$(c^I, \text{val}(d)) \in U^I$</td>
</tr>
</tbody>
</table>
Restriction on TBox assertions in $DL-Lite_A$ ontologies

As shown, to ensure FOL-rewritability, we have to impose a restriction on the use of functionality and role/attribute inclusions.

Restriction on $DL-Lite$ TBoxes

No functional role or attribute can be specialized by using it in the right-hand side of a role or attribute inclusion assertions.

Formally:

- If $\exists P.C$ or $\exists P^-.C$ appears in $\mathcal{T}$, then $(\text{funct } P)$ and $(\text{funct } P^-)$ are not in $\mathcal{T}$.

- If $Q \sqsubseteq P$ or $Q \sqsubseteq P^-$ is in $\mathcal{T}$, then $(\text{funct } P)$ and $(\text{funct } P^-)$ are not in $\mathcal{T}$.

- If $U_1 \sqsubseteq U_2$ is in $\mathcal{T}$, then $(\text{funct } U_2)$ is not in $\mathcal{T}$.
DL-Lite\(_A\) – Example

**Employee**
- empCode: Integer
- salary: Integer

**Manager**
- boss

**AreaManager** ⊑ Employee

**TopManager** ⊑ Employee

**Project**
- projectName: String

**Manager** ⊑ Employee

**AreaManager** ⊑ Manager

**TopManager** ⊑ Manager

**Manager** ⊑ ¬TopManager

Employee ⊑ δ(salary)

δ(salary) ⊑ Employee

ρ(salary) ⊑ xsd:int

(funct salary)

∃worksFor ⊑ Employee

∃worksFor⁻ ⊑ Project

∃worksFor⁻ ⊑ Employee

∃worksFor⁻ ⊑ Project

(funct manages)

(funct manages⁻)

manages ⊑ worksFor

Note: in DL-Lite\(_A\) we still cannot capture:
- completeness of the hierarchy
- number restrictions
Reasoning in $DL\text{-}Lite_\mathcal{A}$ – Separation

It is possible to show that, by virtue of the restriction on the use of role inclusion and functionality assertions, all nice properties of $DL\text{-}Lite_\mathcal{F}$ and $DL\text{-}Lite_\mathcal{R}$ continue to hold also for $DL\text{-}Lite_\mathcal{A}$.

In particular, w.r.t. satisfiability of a $DL\text{-}Lite_\mathcal{A}$ ontology $\mathcal{O}$, we have:

- NIs do not interact with each other.
- NIs and PIs do not interact with functionality assertions.

We obtain that for $DL\text{-}Lite_\mathcal{A}$ a separation result holds:

- Each NI and each functionality can be checked independently from the others.
- A functionality assertion is contradicted in an ontology $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ only if it is explicitly contradicted by its ABox $\mathcal{A}$. 
Ontology satisfiability in $DL$-Lite$_A$

Due to the separation property, we can associate
- to each NI $N$ a boolean CQ $q_N$, and
- to each functionality assertion $F$ a boolean CQ $q_F$.
and check satisfiability of $\mathcal{O}$ by suitably evaluating $q_N$ and $q_F$.

**Theorem**

Let $\mathcal{O} = \langle \mathcal{T}, \mathcal{A} \rangle$ be a $DL$-Lite$_A$ ontology, and $\mathcal{T}_P$ the set of PIs in $\mathcal{O}$. Then, $\mathcal{O}$ is unsatisfiable iff one of the following condition holds:

- There exists a NI $N \in \mathcal{T}$ such that $\text{Eval}(\text{SQL}(\text{PerfectRef}(q_N, \mathcal{T}_P)), \text{DB}(\mathcal{A}))$ returns true.
- There exists a functionality assertion $F \in \mathcal{T}$ such that $\text{Eval}(\text{SQL}(q_F), \text{DB}(\mathcal{A}))$ returns true.
Query answering in $DL-Lite_\mathcal{A}$

- Queries over $DL-Lite_\mathcal{A}$ ontologies are analogous to those over $DL-Lite_\mathcal{R}$ and $DL-Lite_\mathcal{F}$ ontologies, except that they can also make use of attribute and domain atoms.

- Exploiting the previous result, the query answering algorithm of $DL-Lite_\mathcal{R}$ can be easily extended to deal with $DL-Lite_\mathcal{A}$ ontologies:
  - Assertions involving attribute domain and range can be dealt with as for role domain and range assertions.
  - $\exists Q.C$ in the right hand-side of concept inclusion assertions can be eliminated by making use of role inclusion assertions.
  - Disjointness of roles and attributes can be checked similarly as for disjointness of concepts, and does not interact further with the other assertions.
Complexity of reasoning in $DL$-$Lite_A$

As for ontology satisfiability, $DL$-$Lite_A$ maintains the nice computational properties of $DL$-$Lite_R$ and $DL$-$Lite_F$ also w.r.t. query answering. Hence, we get the same characterization of computational complexity.

**Theorem**

For $DL$-$Lite_A$ ontologies:

- Checking **satisfiability of the ontology** is
  - $\text{PTime}$ in the size of the **ontology** (combined complexity).
  - $\text{LogSpace}$ in the size of the **ABox** (data complexity).

- **TBox reasoning** is $\text{PTime}$ in the size of the **TBox**.

- **Query answering** is
  - $\text{NP-complete}$ in the size of the query and the ontology (comb. com.).
  - $\text{PTime}$ in the size of the **ontology**.
  - $\text{LogSpace}$ in the size of the **ABox** (data complexity).
Outline

10 The Description Logic $DL-Lite_A$

11 Connecting ontologies to relational data
   - The impedance mismatch problem
   - Ontology-Based Data Access System
   - Query answering in Ontology-Based Data Access Systems
Managing ABoxes

In all the previous discussion, we have assumed that the data is maintained in the ABox of the ontology:

- The ABox is perfectly compatible with the TBox:
  - the vocabulary of concepts, roles, and attributes is the one used in the TBox.
  - An ABox “stores” abstract objects, and these objects and their properties are those returned by queries over the ontology.

- There may be different ways to manage the ABox from a physical point of view:
  - Description Logics reasoners maintain the ABox is main-memory data structures.
  - When ABoxes become large, managing them in secondary storage may be required, but this is again handled directly by the reasoner.
There are several situations where the assumptions of having the data in an ABox managed directly by the ontology system (e.g., a Description Logics reasoner) is not feasible or realistic:

- When the ABox is very large, so that it requires relational database technology.
- When have no direct control over the data since it belongs to some external organization, which controls the access to it.
- When multiple data sources need to be accessed, such as in Information Integration.

We would like to deal with such situation by keeping the data in the external (relational) storage, and performing query answering by leveraging the capabilities of the relational engine.
The impedance mismatch problem

We have to deal with the **impedance mismatch problem**:

- Sources store data, which is constituted by values taken from concrete domains, such as strings, integers, codes, . . .
- Instead, instances of concepts and relations in an ontology are (abstract) objects.

The solution is to define a **mapping language** that allows for specifying how to transform data into objects:

- **Basic idea**: use Skolem functions in the head of the mapping to “generate” the objects.
- **Semantics**: objects are denoted by terms (of exactly one level of nesting), and different terms denote different objects (unique name assumption on terms).
Impedance mismatch – Example

Actual data is stored in a DB:
- An Employee is identified by her SSN.
- A Project is identified by its name.

D₁[SSN: String, PrName: String]
Employees and Projects they work for

D₂[Code: String, Salary: Int]
Employee’s Code with salary

D₃[Code: String, SSN: String]
Employee’s Code with SSN

Intuitively:

- An employee should be created from her SSN: \text{pers}(SSN)
- A project should be created from its Name: \text{proj}(PrName)
Creating object identifiers

We need to associate to the data in the tables objects in the ontology.

- We introduce an alphabet $\Lambda$ of function symbols, each with an associated arity.
- To denote values, we use value constants in $\Gamma_V$ as before.
- To denote objects, we use object terms instead of object constants. An object term has the form $f(d_1, \ldots, d_n)$, with $f \in \Lambda$, and each $d_i$ a value constant in $\Gamma_V$.

Example

- If a person is identified by its SSN, we can introduce a function symbol $\text{pers}/1$. If $\text{VRD56B25}$ is a SSN, then $\text{pers}(\text{VRD56B25})$ denotes a person.
- If a person is identified by its name and dateOfBirth, we can introduce a function symbol $\text{pers}/2$. Then $\text{pers}(\text{Vardi}, 25/2/56)$ denotes a person.
Mapping assertions

Mapping assertions are used to extract the data from the DB to populate the ontology.

We make use of variable terms, which are as object terms, but with variables instead of values as arguments of the functions.

A mapping assertion between a database $\mathcal{D}$ and a TBox $\mathcal{T}$ has the form

$$\Phi \rightsquigarrow \Psi$$

where

- $\Phi$ is an arbitrary SQL query of arity $n > 0$ over $\mathcal{D}$.
- $\Psi$ is a conjunctive query over $\mathcal{T}$ of arity $n' > 0$ without non-distinguished variables, possibly involving variable terms.
Mapping assertions – Example

\[ D_1[SSN: \text{String}, PrName: \text{String}] \]
Employees and Projects they work for

\[ D_2[Code: \text{String}, Salary: \text{Int}] \]
Employee’s Code with salary

\[ D_3[Code: \text{String}, SSN: \text{String}] \]
Employee’s Code with SSN

\[ M_1: \text{SELECT SSN, PrName FROM } D_1 \]
\[ \leadsto \text{Employee(pers(SSN))}, \text{Project(proj(PrName))}, \text{projectName(proj(PrName), PrName)}, \text{workFor(pers(SSN), proj(PrName))} \]

\[ M_2: \text{SELECT SSN, Salary FROM } D_2, D_3 \text{ WHERE } D_2.CODEx = D_3.CODEx \]
\[ \leadsto \text{Employee(pers(SSN))}, \text{salary(pers(SSN), Salary)} \]
The mapping assertions are a crucial part of an Ontology-Based Data Access System. 

**Ontology-Based Data Access System**

is a triple $\mathcal{O} = \langle T, M, D \rangle$, where

- $T$ is a TBox.
- $D$ is a relational database.
- $M$ is a set of mapping assertions between $T$ and $D$.

*Note:* we could consider also mapping assertions between the datatypes of the database and those of the ontology.
Semantics of mappings

We first need to define the semantics of mappings.

**Definition**

An interpretation $\mathcal{I}$ satisfies a mapping assertion $\Phi(\vec{x}) \leadsto \Psi(\vec{t}, \vec{y})$ in $\mathcal{M}$ with respect to a database $\mathcal{D}$, if for each tuple of values $\vec{v} \in \text{Eval}(\Phi, \mathcal{D})$, and for each ground atom in $\Psi[\vec{x}/\vec{v}]$, we have that:

- if the ground atom is $A(s)$, then $s^{\mathcal{I}} \in A^{\mathcal{I}}$.
- if the ground atom is $T(s)$, then $s^{\mathcal{I}} \in T^{\mathcal{I}}$.
- if the ground atom is $P(s_1, s_2)$, then $(s_1^{\mathcal{I}}, s_2^{\mathcal{I}}) \in P^{\mathcal{I}}$.
- if the ground atom is $U(s_1, s_2)$, then $(s_1^{\mathcal{I}}, s_2^{\mathcal{I}}) \in U^{\mathcal{I}}$.

Intuitively, $\mathcal{I}$ satisfies $\Phi \leadsto \Psi$ w.r.t. $\mathcal{D}$ if all facts obtained by evaluating $\Phi$ over $\mathcal{D}$ and then propagating the answers to $\Psi$, hold in $\mathcal{I}$.

*Note:* $\Psi[\vec{x}/\vec{v}]$ denotes $\Psi$ where each $x_i$ has been substituted with $v_i$. 
Semantics of an OBDA system

Model of an OBDA system

An interpretation $\mathcal{I}$ is a model of $\mathcal{O} = \langle \mathcal{T}, \mathcal{M}, \mathcal{D} \rangle$ if:

- $\mathcal{I}$ is a model of $\mathcal{T}$;
- $\mathcal{I}$ satisfies $\mathcal{M}$ w.r.t. $\mathcal{D}$, i.e., satisfies every assertion in $\mathcal{M}$ w.r.t. $\mathcal{D}$.

An OBDA system $\mathcal{O}$ is satisfiable if it admits at least one model.
Answering queries over an OBDA system

In an OBDA system $\mathcal{O} = \langle \mathcal{T}, \mathcal{M}, \mathcal{D} \rangle$

- Queries are posed over the TBox $\mathcal{T}$.
- The data needed to answer queries is stored in the database $\mathcal{D}$.
- The mapping $\mathcal{M}$ is used to bridge the gap between $\mathcal{T}$ and $\mathcal{D}$.

Two approaches to exploit the mapping:

- bottom-up approach: simpler, but less efficient
- top-down approach: more sophisticated, but also more efficient

**Note:** Both approaches require to first split the TBox queries in the mapping assertions into their constituent atoms. This is possible, since all variables in such queries are distinguished.
Bottom-up approach to query answering

Consists in a straightforward application of the mappings:

1. Propagate the data from $D$ through $M$, materializing an ABox $A_M,D$ (the constants in such an ABox are values and object terms).

2. Apply to $A_M,D$ and to the TBox $T$, the satisfiability and query answering algorithms developed for $DL-Lite_A$.

This approach has several drawbacks (hence is only theoretical):

- The technique is no more $\text{LogSpace}$ in the data, since the ABox $A_M,D$ to materialize is in general polynomial in the size of the data.
- $A_M,D$ may be very large, and thus it may be infeasible to actually materialize it.
- Freshness of $A_M,D$ with respect to the underlying data source(s) may be an issue, and one would need to propagate updates (cf. Data Warehousing).
Top-down approach to query answering

Consists of three steps:

1. **Reformulation:** Compute the perfect reformulation $q' = \text{PerfectRef}(q, T_P)$ of the original query $q$, using the PIs $T_P$ of the TBox $T$.

2. **Unfolding:** Compute from $q'$ a new query $q''$ by unfolding $q'$ using (the split version of) the mappings $\mathcal{M}$.
   - Essentially, each atom in $q'$ that unifies with an atom in $\Psi$ is substituted with the corresponding query $\Phi$ over the database.
   - The unfolded query $q''$ is such that $\text{Eval}(q'', \mathcal{D}) = \text{Eval}(q', A_M, \mathcal{D})$.

3. **Evaluation:** Delegate the evaluation of $q''$ to the relational DBMS managing $\mathcal{D}$.

For details, see [PLC+07].
Computational complexity of query answering

Theorem

Query answering in a $DL$-$Lite_A$ OBDM system $\mathcal{O} = \langle T, M, D \rangle$ is

1. **NP-complete** in the size of the query.
2. **PTime** in the size of the TBox $T$ and the mappings $M$.
3. **LogSpace** in the size of the database $D$.

Moreover, the **LogSpace** result is actually a consequence of the fact that query answering in such a setting can be reduced to evaluating an SQL query over the relational database.
Part V

Hands-on session
Outline

12 The MASTRO system

13 XML MASTRO input files
Outline

12 The MASTRO system

13 XML MASTRO input files
The system **MASTRO**

- **MASTRO** is a Java-based tool for Ontology-based Data Access.
- It allows for the specification of Ontologies in the DL $DL\text{-}Lite_\mathcal{A}$.
- In **MASTRO**, $DL\text{-}Lite_\mathcal{A}$ TBoxes are connected to an external RDBMS through suitable mappings.
- In each mapping a generic SQL query over the external RDBMS is put in correspondence with a CQ without existential variables expressed over the $DL\text{-}Lite_\mathcal{A}$ TBox.
At its core, Mastro uses QuOnto (http://www.dis.uniroma1.it/~quonto/) a reasoner for $DL-Lite_\mathcal{A}$, which provides query reformulation services (QuOnto implements the algorithm PerfectRef).

Notice that QuOnto is not designed to support ontology-based data access, and therefore it is not able to deal with mappings to external RDBMs $\leadsto$ Mastro provides this support.
Reasoning in MASTRO

The basic services provided by MASTRO are:

- Specification of a $DL$-$Lite_A$ OBDA System
- Query answering, for computing certain answer for (unions of) conjunctive queries over $DL$-$Lite_A$ OBDA Systems
- Consistency check, for verifying satisfiability of $DL$-$Lite_A$ OBDA Systems

_These are the only services supported by the version of MASTRO demonstrated at the tutorial._

However, the full version of MASTRO also allows for TBox reasoning, meta-level reasoning, and ontology updates.

We are currently working also on query answering of complex (i.e., FOL) queries, introduction on new $DL$-$Lite$ constructs (e.g., identification assertions).
Input formats

- **Mastro** has its own Java-based interface, and accepts inputs in a proprietary XML format.
- That is, to give as input a TBox, an ABox, i.e., a set of mapping assertions to an external RDBMS, and a query, we must specify them into XML, according to a specific DTD.
- Nonetheless, the XML syntax to be used is very simple.
Outline

12 The MASTRO system

13 XML MASTRO input files
<alphabet>
  <atomicC>professor</atomicC>
  <atomicC>assistantProf</atomicC>
  ....
  <atomicCA>name</atomicCA>
  ....
  <atomicR>WORKS_FOR</atomicR>
  ....
  <atomicRA>date</atomicRA>
</alphabet>
**XML TBox: inclusion assertions**

\[
\text{assistantProfessor} \sqsubseteq \neg \text{fullProf}
\]

```xml
<inclusionAssertion>
  <basicC>
    <atomicC>assistantProf</atomicC>
  </basicC>
  <generalC>
    <signedC sign="negative">
      <basicC>
        <atomicC>fullProf</atomicC>
      </basicC>
    </signedC>
  </generalC>
</inclusionAssertion>
```

\[
\exists \text{TAKES\_COURSE}\sqsubseteq \text{course}
\]

```xml
<inclusionAssertion>
  <basicC>
    <exists>
      <basicR dir="inverse">
        <atomicR>TAKES\_COURSE</atomicR>
      </basicR>
    </exists>
  </basicC>
  <generalC>
    <signedC sign="positive">
      <basicC>
        <atomicC>course</atomicC>
      </basicC>
    </signedC>
  </generalC>
</inclusionAssertion>
```
XML TBox: functionality assertions

(funct \textit{ENROLLED})

\begin{verbatim}
<funct>
  <basicR dir="direct">
    <atomicR>ENROLLED</atomicR>
  </basicR>
</funct>
\end{verbatim}

(funct \textit{term})

\begin{verbatim}
<funct>
  <atomicCA>term</atomicCA>
</funct>
\end{verbatim}
XML ABox: mapping specification

It is probably better to see the mapping as follows

\[
\text{SELECT C\_Name, Term, Prof\_Id} \sim \begin{align*}
\text{course}(\text{course}(\text{C\_Name})), \\
\text{name}(\text{course}(\text{C\_Name}), \text{C\_Name}), \\
\text{term}(\text{course}(\text{C\_Name}), \text{Term}), \\
\text{professor}(\text{prof}(\text{Prof\_Id})), \\
\text{TEACHES}(\text{prof}(\text{Prof\_Id}), \text{course}(\text{C\_Name}))
\end{align*}
\]

\[
\text{FROM Course\_Tab}
\]

\[
\text{SELECT C\_Name, Term, Prof\_Id} \sim \begin{align*}
\text{course}(X), X = \text{course}(\text{C\_Name}), \\
\text{name}(X, Y), Y = \text{C\_Name}, \\
\text{term}(X, Z), Z = \text{Term}, \\
\text{professor}(W), W = \text{prof}(\text{Prof\_Id}), \\
\text{TEACHES}(W, X)
\end{align*}
\]

\[
\text{FROM Course\_Tab}
\]
XML ABox: mapping specification

SELECT C_Name, Term, Prof_Id
FROM Course_Tab

\[ course(X), X = \text{course}(C\_Name), \]
\[ \text{name}(X, Y), Y = C\_Name, \]
\[ \text{term}(X, Z), Z = \text{Term}, \]
\[ \text{professor}(W), W = \text{prof}(Prof\_Id), \text{TEACHES}(W, X) \]

\(<\text{mapping}>\)
\(<\text{head}>\)
\(<\text{CQBody}>\)
\hspace{1cm} \ldots \ldots \]
\(<\text{atom}>\)
\hspace{1cm} \langle\text{AtomicConceptAttribute name="term">\]
\hspace{2cm} \langle\text{term}>\]
\hspace{3cm} \langle\text{var name="X"/>\]
\hspace{2cm} \langle\text{term}>\]
\hspace{3cm} \langle\text{var name="Z"/>\]
\hspace{2cm} \langle\text{/term}>\]
\hspace{2cm} \langle\text{/AtomicConceptAttribute>\]
\hspace{2cm} \langle\text{/term}>\]
\hspace{1cm} \langle\text{/atom}>\]
\hspace{1cm} \ldots \ldots \]
\(<\text{/CQBody}>\]
\(<\text{/head}>\]
\hspace{1cm} \ldots \ldots \]
XML ABox: mapping specification

```
SELECT C_Name, Term, Prof_Id
FROM Course_Tab
```

\[ \sim course(X), X = course(C_Name), \]

\[ \text{name}(X, Y), Y = C_Name, \]

\[ \text{term}(X, Z), Z = \text{Term}, \]

\[ \text{professor}(W), W = \text{prof}(\text{Prof_Id}), \text{TEACHES}(W, X) \]

```
<mapping>

........

<map>

<objMap>

<dtVar>X</dtVar>

<sqlObjVar funct="course">C_Name</sqlObjVar>

</objMap>

</map>

........

<map>

<valueMap>

<dtVar>Z</dtVar>

<sqlValueVar type="xs:integer">Term</sqlValueVar>

</valueMap>

</map>

<body>SELECT C_Name, Term, Prof_Id FROM Course_Tab</body>

........
```

D. Calvanese, D. Lembo
Ontology-Based Data Access
ISWC'07 – Nov. 12, 2007 (209/217)
Acknowledgements

- Giuseppe De Giacomo
- Enrico Franconi
- Maurizio Lenzerini
- Marco Ruzzi
- Antonella Poggi
- Riccardo Rosati
- Sergio Tessaris
References I

Reasoning over extended ER models.
In Proc. of the 26th Int. Conf. on Conceptual Modeling (ER 2007),
volume 4801 of Lecture Notes in Computer Science, pages 277–292.

Pushing the $\mathcal{EL}$ envelope.
In Proc. of the 19th Int. Joint Conf. on Artificial Intelligence

Reasoning on UML class diagrams.
References II


*The Description Logic Handbook: Theory, Implementation and Applications.*


On the decidability of query containment under constraints.


DL-Lite: Tractable description logics for ontologies.

References III

Tailoring OWL for data intensive ontologies.

Linking data to ontologies: The description logic DL-LiteA.

Data complexity of query answering in description logics.
References IV


References V


References VI

Combining Horn rules and description logics in CARIN.

Inverse roles make conjunctive queries hard.

Description logic systems.
In Baader et al. [BCM+03], chapter 8, pages 282–305.

Characterizing data complexity for conjunctive query answering in expressive description logics.

Linking data to ontologies.

*J. on Data Semantics, 2007.*

To appear.