

# Ontologies and Databases

Enrico Franconi

KRDB Research Centre - Free University of Bozen-Bolzano, Italy

http://www.inf.unibz.it/~franconi

### **Summary**

- ▶ What is an Ontology
- ► (Description Logics for Conceptual Modelling)
- Queries via a Conceptual Schema

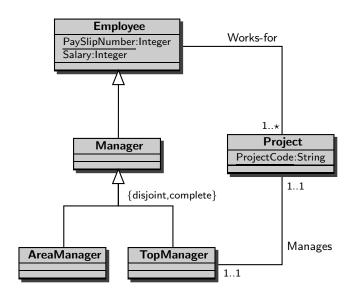
### What is an Ontology

- ► An ontology is a formal conceptualisation of the world: a conceptual schema.
- ► An ontology specifies a set of constraints, which declare what should necessarily hold in any possible world.
- Any possible world should conform to the constraints expressed by the ontology.
- ► Given an ontology, a *legal world description* is a finite possible world satisfying the constraints.

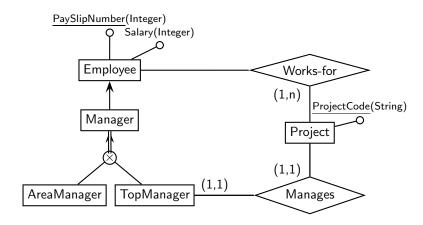
### **Ontologies and Conceptual Data Models**

- An ontology language usually introduces concepts (aka classes, entities), properties of concepts (aka slots, attributes, roles), relationships between concepts (aka associations), and additional constraints.
- ▶ Ontology languages may be simple (e.g., involving only concepts and taxonomies), frame-based (e.g., UML, based on concepts, properties, and binary relationships), or logic-based (e.g. OWL, Description Logics).
- ▶ Ontology languages are typically expressed by means of diagrams.
- Entity-Relationship schemas and UML class diagrams can be considered as ontologies.

#### **UML Class Diagram**



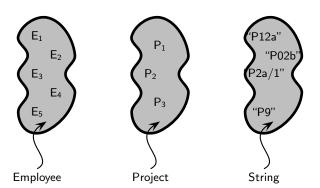
#### **Entity-Relationship Schema**



#### **Semantics**

In a specific world:

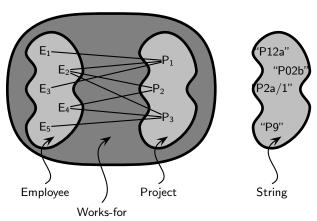
- A class is a set of instances;
- a n-ary relationship is a set of n-tuples of instances;
- ▶ an attribute is a set of pairs of an instance and a domain element.



#### **Semantics**

#### In a specific world:

- A class is a set of instances;
- a n-ary relationship is a set of n-tuples of instances;
- an attribute is a set of pairs of an instance and a domain element.



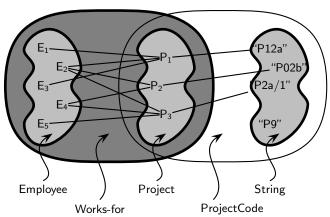
(6/38)

Ontologies and Databases. E. Franconi.

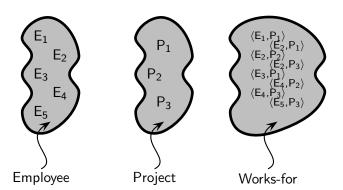
#### **Semantics**

#### In a specific world:

- A class is a set of instances;
- a n-ary relationship is a set of n-tuples of instances;
- an attribute is a set of pairs of an instance and a domain element.



#### A world is described by sets of instances



#### The relational representation of a world

**Employee** 

employe
$E_1$
$E_2$
$E_3$
$E_4$
Es

Project

. 0,000
projectId
$P_1$
$P_2$
$P_3$

String

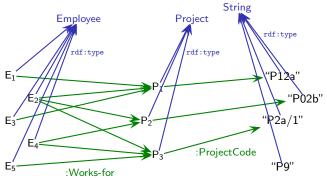
Juling
anystring
"P12a"
"P02b"
"P2a/1"
"Pģ"

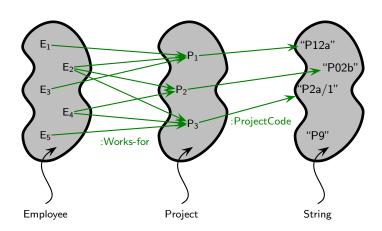
Works-for

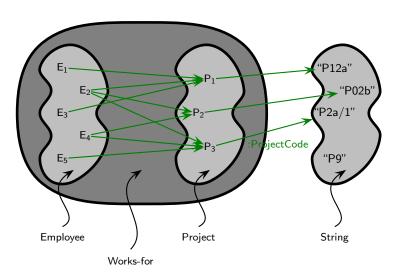
employeeld	projectId
$E_1$	$P_1$
$E_2$	$P_1$
$E_2$	$P_2$
$E_2$	$P_3$
$E_3$	$P_1$
$E_4$	$P_2$
$E_4$	$P_3$
E <sub>5</sub>	P <sub>3</sub>

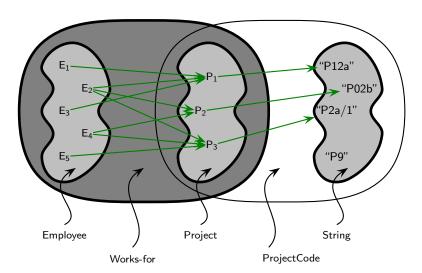
ProjectCode

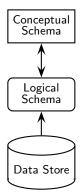
projectId	pcode
$P_1$	"P12a"
$P_2$	"P02b"
P <sub>3</sub>	"P2a/1"

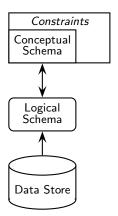


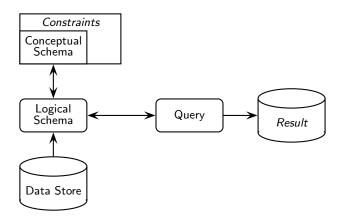


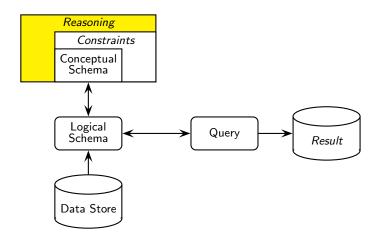










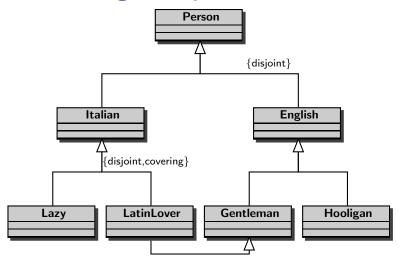


#### Reasoning

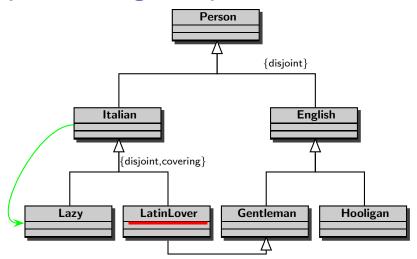
Given an ontology – seen as a collection of constraints – it is possible that additional constraints can be inferred.

- ▶ A class is inconsistent if it denotes the empty set in any legal world description.
- ► A class is a subclass of another class if the former denotes a subset of the set denoted by the latter in any legal world description.
- ► Two classes are equivalent if they denote the same set in any legal world description.
- ▶ A stricter constraint is inferred e.g., a cardinality constraint if it holds in in any legal world description.
- **.**..

#### Simple reasoning example

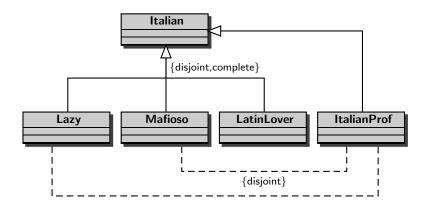


#### Simple reasoning example

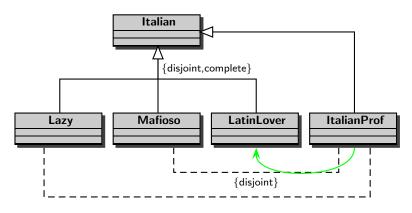


 $\begin{aligned} \mathsf{LatinLover} &= \emptyset \\ \mathsf{Italian} &\subseteq \mathsf{Lazy} \\ \mathsf{Italian} &\equiv \mathsf{Lazy} \end{aligned}$ 

## Reasoning: cute professors

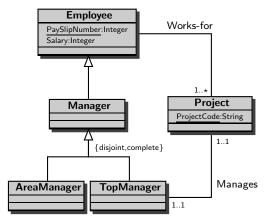


#### Reasoning: cute professors



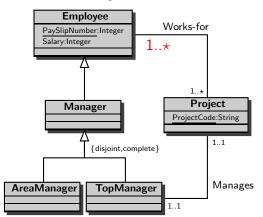
implies Italian Prof  $\subseteq$  Latin Lover

### **Reasoning with Conceptual Schemas**



▶ Managers do not work for a project (she/he just manages it):  $\forall x. \texttt{Manager}(x) \rightarrow \forall y. \neg \texttt{WORKS-FOR}(x, y)$ 

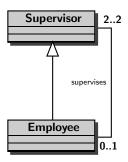
#### **Reasoning with Conceptual Schemas**



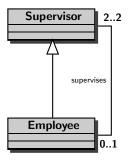
- ▶ Managers do not work for a project (she/he just manages it):  $\forall x$ . Manager(x)  $\rightarrow \forall y$ .  $\neg WORKS$ -FOR(x, y)
- ▶ If the minimum cardinality for the participation of employees to the works-for relationship is increased, then . . .

Ontologies and Databases. E. Franconi. (14/38)

## The democratic company



#### The democratic company

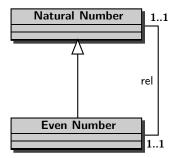


#### implies

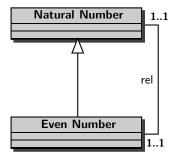
"the classes Employee and Supervisor necessarily contain an infinite number of instances".

Since legal world descriptions are *finite* possible worlds satisfying the constraints imposed by the conceptual schema, the schema is inconsistent.

## How many numbers?



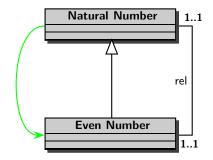
## How many numbers?



#### implies

"the classes Natural Number and Even Number contain the same number of instances".

## How many numbers?



#### implies

"the classes Natural Number and Even Number contain the same number of instances".

Only if the domain is finite: Natural Number ≡ Even Number

#### **Summary**

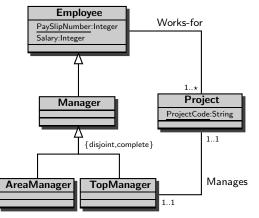
- ► Logic and Conceptual Modelling
- ► Description Logics for Conceptual Modelling
- ▶ Queries via a Conceptual Schema

# **Encoding Conceptual Schemas in** (Description) Logics

- ▶ Object-oriented data models (e.g., UML and ODMG)
- ► Semantic data models (e.g., EER and ORM)
- ► Frame-based and web ontology languages (e.g., RDFS, OWL)

# **Encoding Conceptual Schemas in** (Description) Logics

- Object-oriented data models (e.g., UML and ODMG)
- ► Semantic data models (e.g., EER and ORM)
- ► Frame-based and web ontology languages (e.g., RDFS, OWL)
- ▶ Theorems **prove** that a conceptual schema and its encoding as DL knowledge bases constrain every world description in the same way i.e., the models of the DL theory correspond to the legal world descriptions of the conceptual schema, and vice-versa.

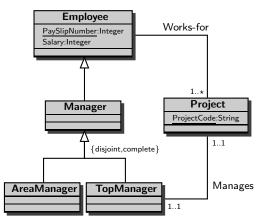


```
emp/2: Employee \sqcap act/2: Project
Works-for
                         man/2: TopManager □ prj/2: Project
Manages
                         \exists^{=1}[worker](PaySlipNumber \sqcap num/2 : Integer)\sqcap
Employee
                          \exists^{=1}[payee](Salary \cap amount/2 : Integer)
                          \exists \leq 1 [\text{num}] (\text{PaySlipNumber} \sqcap \text{worker}/2 : \text{Employee})
Manager
                         \texttt{Employee} \sqcap (\texttt{AreaManager} \sqcup \texttt{TopManager})
                         Manager \sqcap \neg TopManager
AreaManager
TopManager
                         Manager \sqcap \exists^{=1}[man]Manages
                          \exists^{\geq 1}[act]Works-for \sqcap \exists^{=1}[prj]Manages
Project
```

#### **Set-based Constraints**

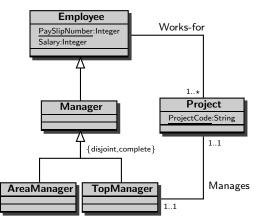
```
Works-for \subseteq Employee \times Project
Manages \subseteq TopManager \times Project
Employee \subseteq \{e \mid \sharp (\mathsf{PaySlipNumber} \cap (\{e\} \times \mathsf{Integer})) \geq 1\}
Employee \subseteq \{e \mid \sharp(\mathsf{Salary} \cap (\{e\} \times \mathsf{Integer})) \geq 1\}
Project \subseteq \{p \mid \sharp(ProjectCode \cap (\{p\} \times String)) \ge 1\}
TopManager \subseteq \{m \mid 1 \geq \sharp (\mathsf{Manages} \cap (\{m\} \times \Omega)) \geq 1\}
Project \subseteq \{p \mid 1 \ge \sharp (\mathsf{Manages} \cap (\Omega \times \{p\})) \ge 1\}
Project \subseteq \{p \mid \sharp (\mathsf{Works\text{-}for} \cap (\Omega \times \{p\})) > 1\}
Manager \subseteq Employee
AreaManager ⊂ Manager
\mathsf{TopManager} \subseteq \mathsf{Manager}
AreaManager \cap TopManager = \emptyset
Manager \subseteq AreaManager \cup TopManager
```

#### **Deducing constraints**



Managers are employees who do not work for a project (she/he just manages it):  $\texttt{Employee} \sqcap \neg (\exists^{\geq 1} [\texttt{emp}] \texttt{Works-for}) \sqsubseteq \texttt{Manager}, \quad \texttt{Manager} \sqsubseteq \neg (\exists^{\geq 1} [\texttt{emp}] \texttt{Works-for})$ 

#### **Deducing constraints**



Managers are employees who do not work for a project (she/he just manages it):  $\texttt{Employee} \sqcap \neg (\exists^{\geq 1} [\texttt{emp}] \texttt{Works-for}) \sqsubseteq \texttt{Manager}, \quad \texttt{Manager} \sqsubseteq \neg (\exists^{\geq 1} [\texttt{emp}] \texttt{Works-for})$ 

For every project, there is at least one employee who is not a manager:  $Project \sqsubseteq \exists^{\geq 1}[act](Works-for \sqcap emp : \neg Manager)$ 

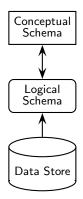
Ontologies and Databases. E. Franconi. (21/38)

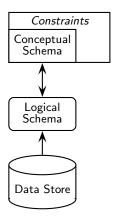
#### iocom: Intelligent Conceptual Modelling

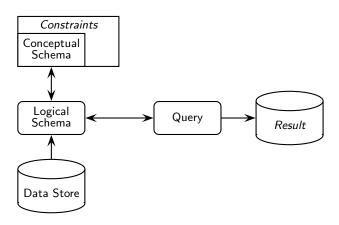
- ▶ i•com allows for the specification of multiple EER (or UML) diagrams and inter- and intra-schema constraints;
- ightharpoonup Complete logical reasoning is employed by the tool using a hidden underlying  $\mathcal{DLR}$  inference engine;
- ▶ i•com verifies the specification, infers implicit facts and stricter constraints, and manifests any inconsistencies during the conceptual modelling phase.

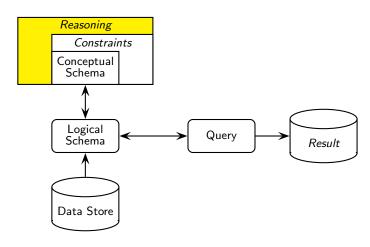
#### **Summary**

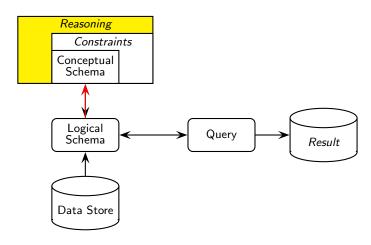
- ► Logic and Conceptual Modelling
- Description Logics for Conceptual Modelling
- ► Queries via a Conceptual Schema

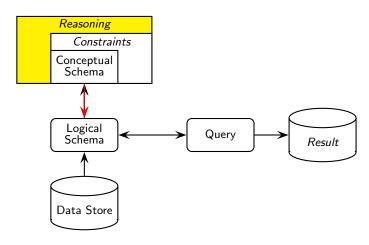


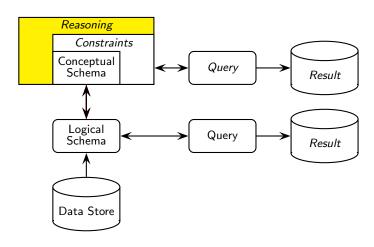


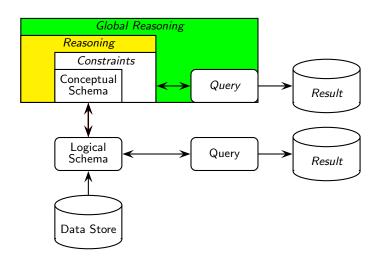


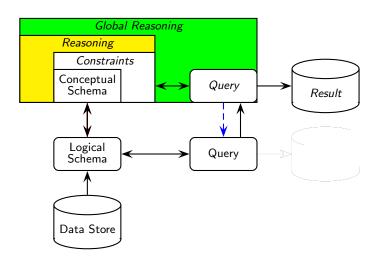


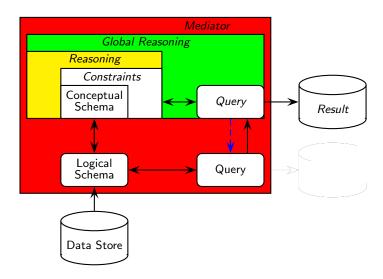


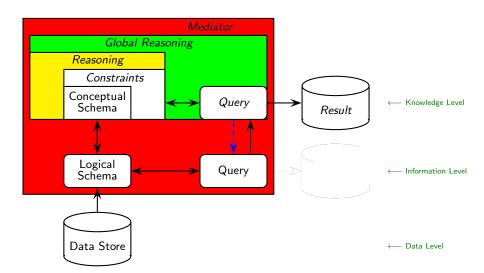












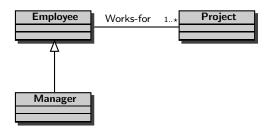
# Queries via Conceptual Schemas: the DB assumption

- ▶ Basic assumption: consistent information with respect to the constraints introduced by the conceptual schema
- ▶ DB assumption: complete information about each term appearing in the conceptual schema
- ▶ Problem: answer a query over the conceptual schema vocabulary

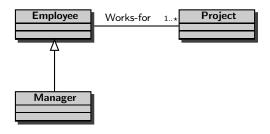
# Queries via Conceptual Schemas: the DB assumption

- ▶ Basic assumption: consistent information with respect to the constraints introduced by the conceptual schema
- ▶ DB assumption: complete information about each term appearing in the conceptual schema
- ▶ Problem: answer a query over the conceptual schema vocabulary
- Solution: use a standard DB technology (e.g., SQL, datalog, etc)

#### **Example with DB assumption**

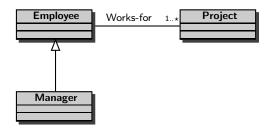


#### **Example with DB assumption**



```
Employee = { John, Mary, Paul }
Manager = { John, Paul }
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-B \rangle }
Project = { Prj-A, Prj-B }
```

#### **Example with DB assumption**



Employee = { John, Mary, Paul }

```
Manager = { John, Paul }
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-B \rangle }
Project = { Prj-A, Prj-B }

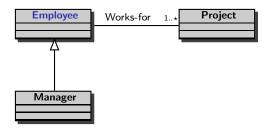
Q(X) :- Manager(X), Works-for(X,Y), Project(Y)
\Rightarrow { John }
```

#### Weakening the DB assumption

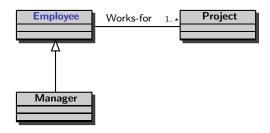
▶ The DB assumption is against the principle that a conceptual schema presents a richer vocabulary than the data stores (i.e., it plays the role of an ontology).

#### Weakening the DB assumption

- ▶ The DB assumption is against the principle that a conceptual schema presents a richer vocabulary than the data stores (i.e., it plays the role of an ontology).
- ► Partial DB assumption (or conceptual schema with *DBox*): complete information about *some* term appearing in the conceptual schema
- Standard DB technologies do not apply
- ▶ The query answering problem in this context is inherently complex

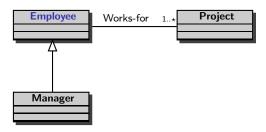


```
Manager = { John, Paul }
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-B \rangle }
Project = { Prj-A, Prj-B }
```



```
Manager = { John, Paul }
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-B \rangle }
Project = { Prj-A, Prj-B }
```

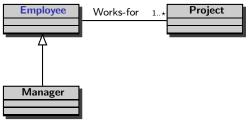
Q(X) := Employee(X)



```
Manager = { John, Paul }
Works-for = { ⟨John,Prj-A⟩, ⟨Mary,Prj-B⟩ }
Project = { Prj-A, Prj-B }

Q(X) :- Employee(X)

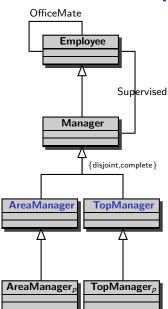
⇒ { John, Paul, Mary }
```



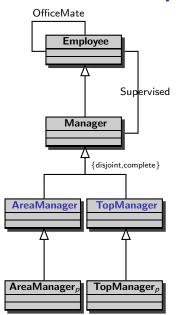
```
Manager = { John, Paul }
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-B \rangle }
Project = { Pri-A, Pri-B }
Q(X) := Employee(X)
\Longrightarrow { John, Paul, Mary }
```

Q'(X) :- Manager(X) ∪ Works-for(X,Y)

#### **Andrea's Example**

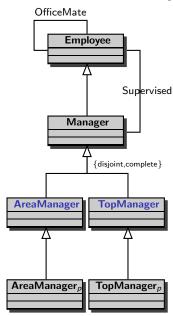


#### **Andrea's Example**



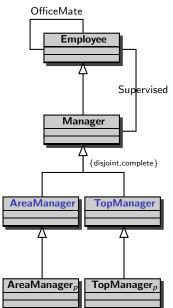
```
\begin{split} & \text{Employee} = \left\{ \text{ Andrea, Paul, Mary, John } \right\} \\ & \text{Manager} = \left\{ \text{ Andrea, Paul, Mary} \right\} \\ & \text{AreaManager}_{\rho} = \left\{ \text{ Paul } \right\} \\ & \text{TopManager}_{\rho} = \left\{ \text{ Mary } \right\} \\ & \text{Supervised} = \left\{ \left\langle \text{John,Andrea} \right\rangle, \left\langle \text{John,Mary} \right\rangle, \\ & \text{OfficeMate} = \left\{ \left\langle \text{Mary,Andrea} \right\rangle, \left\langle \text{Andrea,Paul} \right\rangle, \right\} \\ & \text{OfficeMate} = \left\{ \left\langle \text{Mary,Andrea} \right\rangle, \left\langle \text{Andrea,Paul} \right\rangle, \\ & \text{Mary, Andrea} \right\} \\ & \text{Mary, Andrea} \\ & \text{Mary, Andrea, Andrea
```

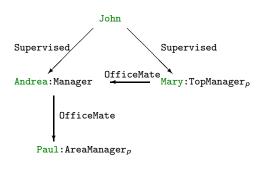
#### **Andrea's Example**

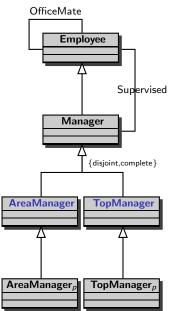


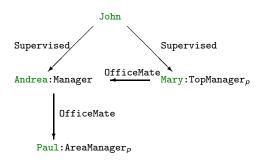
```
Employee = { Andrea, Paul, Mary, John }
Manager = { Andrea, Paul, Mary}
AreaManager_p = \{ Paul \}
TopManager_p = \{ Mary \}
Supervised = { \langle John, Andrea \rangle, \langle John, Mary \rangle }
OfficeMate = { \langle Mary, Andrea \rangle, \langle Andrea, Paul \rangle }
                     John
                                   Supervised
Supervised
                     \underbrace{\texttt{OfficeMate}}_{\texttt{Mary:TopManager}_p}
Andrea: Manager
           OfficeMate
     Paul: AreaManager,
```



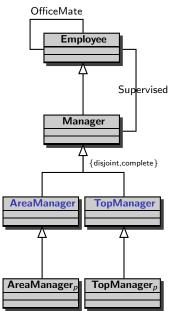


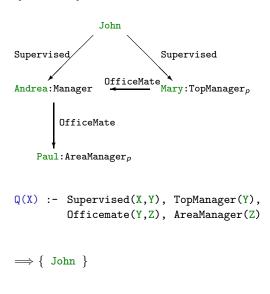






Q(X) :- Supervised(X,Y), TopManager(Y), Officemate(Y,Z), AreaManager(Z)

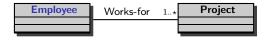




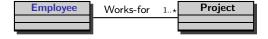
#### Partial incomplete DB assumption

- ► Partial DB assumption (or conceptual schema with <u>DBox</u>): complete information about *some* term appearing in the conceptual schema
- Partial incomplete DB assumption (or conceptual schema with <u>ABox</u>): incomplete information about some term appearing in the conceptual schema
- ► The partial incomplete DB assumption (conceptual schema with ABox) is crucial in data integration scenarios.

#### Partial incomplete DB assumption

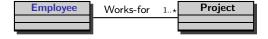


#### Partial incomplete DB assumption



#### DBox:

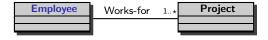
```
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-A \rangle }
Project = { Prj-A, Prj-B }
```



DBox:

 $\Longrightarrow$ 

INCONSISTENT

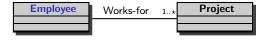


#### DBox:

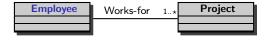
```
Works-for = { \langle John, Prj-A \rangle, \langle Mary, Prj-A \rangle }
Project = { Prj-A, Prj-B }
```

#### → INCONSISTENT

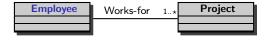
```
Works-for \supseteq { \langle John, Prj-A \rangle, \langle Mary, Prj-A \rangle }
Project \supseteq { Prj-A, Prj-B }
```



```
Works-for \supseteq { \langle John, Prj-A \rangle, \langle Mary, Prj-A \rangle } Project \supseteq { Prj-A, Prj-B }
```



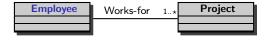
```
Works-for ⊇ { ⟨John,Prj-A⟩, ⟨Mary,Prj-A⟩ }
Project ⊇ { Prj-A, Prj-B }
Q(X) :- Works-for(Y,X)
```



```
Works-for ⊇ { ⟨John,Prj-A⟩, ⟨Mary,Prj-A⟩ }
Project ⊇ { Prj-A, Prj-B }

Q(X) :- Works-for(Y,X)

⇒ { Prj-A, Prj-B }
```



```
Works-for ⊇ { ⟨John,Prj-A⟩, ⟨Mary,Prj-A⟩ }
Project ⊇ { Prj-A, Prj-B }

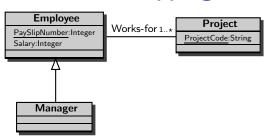
Q(X) :- Works-for(Y,X)

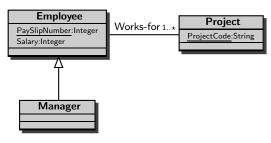
⇒ { Prj-A, Prj-B }

⇒ Q'(X) :- Project(X) ∪ Works-for(Y,X)
```

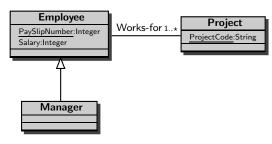
## View based Query Processing

- ▶ Mappings between the conceptual schema terms and the information source terms are not necessarily atomic.
- Mappings can be given in terms of a set of ABox (or DBox) facts:
  - ► GAV (global-as-view): ABox (or DBox) facts over the information source vocabulary are associated to terms in the conceptual schema
    - both the DB and the partial DB assumptions are special cases of GAV
    - an ER schema can be easily mapped to its corresponding relational schema in some normal form via a GAV mapping
  - ► LAV (*local-as-view*): a ABox or DBox over the conceptual schema vocabulary is associated to each term in the information source;
  - GLAV: mix of the above.
- ▶ It is non-trivial, even in the pure GAV setting which is wrongly believed to be computable by simple view unfolding.



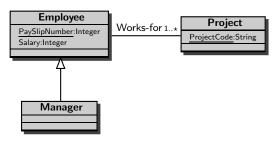


- 1-Employee(PaySlipNumber,Salary,ManagerP)
- ${\tt 2-Works-for}\overline{({\tt PaySlipNumber}\,,{\tt ProjectCode})}$



```
1-Employee(PaySlipNumber, Salary, ManagerP)
2-Works-for(PaySlipNumber, ProjectCode)
```

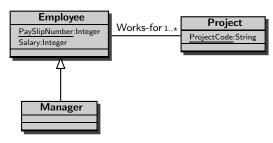
Project(Y) :- 2-Works-for(X,Y)



```
1-Employee(PaySlipNumber, Salary, ManagerP)
2-Works-for(PaySlipNumber, ProjectCode)
```

```
 \begin{split} & \text{Employee}(X) & :- & 1-\text{Employee}(X,Y,\text{false}) & \text{Works-for}(X,Y) & :- & 2-\text{Works-for}(X,Y) \\ & \text{Manager}(X) & :- & 1-\text{Employee}(X,Y,\text{true}) & \text{Salary}(X,Y) & :- & 1-\text{Employee}(X,Y,Z) \\ & \text{Project}(Y) & :- & 2-\text{Works-for}(X,Y) & :- & 1-\text{Employee}(X,Y,Z) \\ \end{split}
```

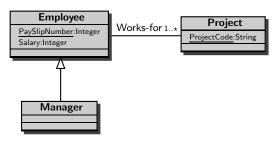
Q(X) :- Employee(X)



```
1-Employee(PaySlipNumber, Salary, ManagerP)
2-Works-for(PaySlipNumber, ProjectCode)
```

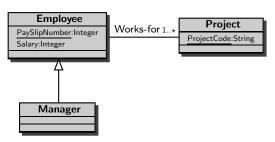
```
 \begin{split} & \text{Employee}(\textbf{X}) & :- & 1\text{-Employee}(\textbf{X},\textbf{Y},\text{false}) & \text{Works-for}(\textbf{X},\textbf{Y}) & :- & 2\text{-Works-for}(\textbf{X},\textbf{Y}) \\ & \text{Manager}(\textbf{X}) & :- & 1\text{-Employee}(\textbf{X},\textbf{Y},\text{true}) & \text{Salary}(\textbf{X},\textbf{Y}) & :- & 1\text{-Employee}(\textbf{X},\textbf{Y},\textbf{Z}) \\ & \text{Project}(\textbf{Y}) & :- & 2\text{-Works-for}(\textbf{X},\textbf{Y}) & :- & 1\text{-Employee}(\textbf{X},\textbf{Y},\textbf{Z}) \\ \end{split}
```

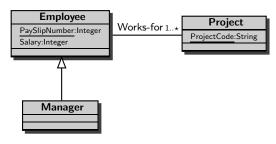
```
\begin{array}{lll} \mathbb{Q}(\mathbb{X}) & := & \text{Employee}(\mathbb{X}) \\ \Longrightarrow & \mathbb{Q}^{?}(\mathbb{X}) & := & 1-\text{Employee}(\mathbb{X},\mathbb{Y},\mathbb{Z}) \ \cup \ 2-\text{Works-for}(\mathbb{X},\mathbb{W}) \end{array}
```



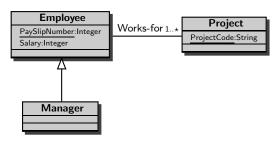
```
1-Employee(PaySlipNumber, Salary, ManagerP)
2-Works-for(PaySlipNumber, ProjectCode)
```

```
 \begin{split} & \text{Employee}(\textbf{X}) & :- & 1-\text{Employee}(\textbf{X},\textbf{Y},\text{false}) & \text{Works-for}(\textbf{X},\textbf{Y}) & :- & 2-\text{Works-for}(\textbf{X},\textbf{Y}) \\ & \text{Manager}(\textbf{X}) & :- & 1-\text{Employee}(\textbf{X},\textbf{Y},\text{true}) & \text{Salary}(\textbf{X},\textbf{Y}) & :- & 1-\text{Employee}(\textbf{X},\textbf{Y},\textbf{Z}) \\ & \text{Project}(\textbf{Y}) & :- & 2-\text{Works-for}(\textbf{X},\textbf{Y}) & :- & 1-\text{Employee}(\textbf{X},\textbf{Y},\textbf{Z}) \\ \end{split}
```





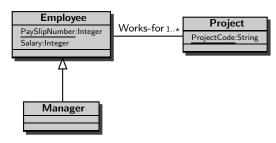
- 1-Employee(PaySlipNumber,Salary,ManagerP)
- ${\tt 2-Works-for}\overline{({\tt PaySlipNumber}}\,, {\tt ProjectCode})$



```
{\tt 1-Employee}(\underline{{\tt PaySlipNumber}}, {\tt Salary}, {\tt ManagerP})
```

 ${\tt 2-Works-for}\overline{({\tt PaySlipNumber}\,,{\tt ProjectCode})}$ 

```
1-Employee(X,Y,Z) :- Manager(X), Salary(X,Y), Z=true
1-Employee(X,Y,Z) :- Employee(X), ¬Manager(X), Salary(X,Y), Z=false
2-Works-for(X,Y) :- Works-for(X,Y)
```

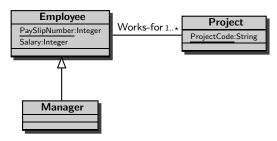


```
{\tt 1-Employee}(\underline{{\tt PaySlipNumber}}, {\tt Salary}, {\tt ManagerP})
```

2-Works-for(PaySlipNumber, ProjectCode)

```
1-Employee(X,Y,Z) :- Manager(X), Salary(X,Y), Z=true
1-Employee(X,Y,Z) :- Employee(X), ¬Manager(X), Salary(X,Y), Z=false
2-Works-for(X,Y) :- Works-for(X,Y)
```

```
Q(X) :- Manager(X), Works-for(X,Y), Project(Y)
```



```
1-Employee(PaySlipNumber, Salary, ManagerP)
2-Works-for(PaySlipNumber, ProjectCode)
```

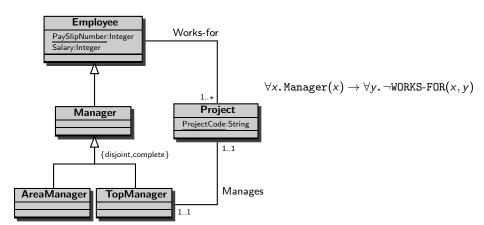
1-Employee(X,Y,Z) :- Manager(X), Salary(X,Y), Z=true
1-Employee(X,Y,Z) :- Employee(X), ¬Manager(X), Salary(X,Y), Z=false
2-Works-for(X,Y) :- Works-for(X,Y)

```
Q(X) :- Manager(X), Works-for(X,Y), Project(Y)

⇒ Q'(X) :- 1-Employee(X,Y,true), 2-Works-for(X,Z)
```

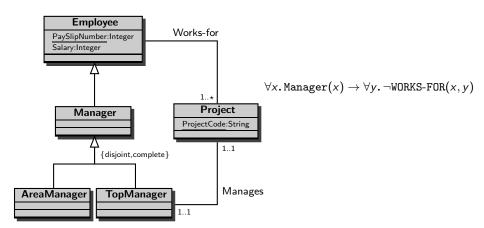
### Reasoning over queries

Q(X,Y) :- Employee(X), Works-for(X,Y), Manages(X,Y)



### Reasoning over queries

Q(X,Y) :- Employee(X), Works-for(X,Y), Manages(X,Y)



→ INCONSISTENT QUERY!

#### **Conclusions**

#### **Conclusions**

Do you want to exploit conceptual schema knowledge (i.e., an ontology) in your data intensive application?

#### **Conclusions**

Do you want to exploit conceptual schema knowledge (i.e., an ontology) in your data intensive application?

Pay attention!



Made with LATEX2e