

Özgür L. Özçep

Ontology-Based Data Access

Lecture 8: Motivation, Description Logics 4 June 2020

> Informationssysteme CS4130 (Summer 2020)

References

► ESSLLI 2010 Course by Calvanese and Zakharyaschev

http://www.inf.unibz.it/~calvanese/teaching/2010-08-ESSLLI-DL-QA/

 Reasoning Web Summer School 2014 course by Kontchakov on Description Logics

http:

//rw2014.di.uoa.gr/sites/default/files/slides/An_Introduction_to_Description_Logics.pdf

 Lecture notes by Calvanese in 2013/2014 course on Ontology and Database Systems

https://www.inf.unibz.it/~calvanese/teaching/14-15-odbs/lecture-notes/

- Course notes by Franz Baader on Description Logics
- Parts of Reasoning Web Summer School 2014 course by Ö. on Ontology-Based Data Access on Temporal and Streaming Data

http://rw2014.di.uoa.gr/sites/default/files/slides/Ontology_Based_Data_Access_on_

Temporal_and_Streaming_Data.pdf

Ontology-Based Data Access as Information Integration

- Data Integration can be considered as information integration purely on DB level
- OBDA can be considered as information integration using an ontology
- Bridges DB world (closes world assumption) and ontology world (open world assumption)

- DB theory: closed-world assumption (CWA)
 - All and only those facts mentioned in DB hold.

- DB theory: closed-world assumption (CWA)
 - ► All and only those facts mentioned in DB hold.
- Simple form of uncertain knowledge expressed by NULLs
 - ► For one incomplete DB there are many completions
 - Nonetheless: Type information on attribute constrains the possible attribute instances

- DB theory: closed-world assumption (CWA)
 - ► All and only those facts mentioned in DB hold.
- Simple form of uncertain knowledge expressed by NULLs
 - ► For one incomplete DB there are many completions
 - Nonetheless: Type information on attribute constrains the possible attribute instances
- In DE incompleteness generated by different schemata
 Flight scenario: Source DB had no flight number, whilst target
 DB has
 - \implies introduction of NULLs for flight number attribute

- DB theory: closed-world assumption (CWA)
 - ► All and only those facts mentioned in DB hold.
- Simple form of uncertain knowledge expressed by NULLs
 - ► For one incomplete DB there are many completions
 - Nonetheless: Type information on attribute constrains the possible attribute instances
- In DE incompleteness generated by different schemata
 Flight scenario: Source DB had no flight number, whilst target
 DB has

 \implies introduction of NULLs for flight number attribute

- Logical theories (ontologies) adhere to open world assumption (OWA)
 - If something is not told, then we do not know
 - Logical theories (ontologies) may have many models

Close-World Assumption (CWA) for DBs

"The world described by DBs is complete"

Exar	nple	
Uni	versity employee	Professor
ID	Name	ID
1	Sokrates	1
2	Platon	2
3	Aristotle	

"3" (= ID of Aristotle) not in table Professor \implies Aristotle is not a professor

Close-World Assumption (CWA) for DBs

"The world described by DBs is complete"

Exar	nple			
Patient		Blood sugar		
ID	Name		ID	value
1	Sokrates		1	90
2	Platon		2	120
3	Aristotle			
"3" n	ot in blood	sugar	tabl	e

"3" not in blood sugar table ⇒? Aristotle has not blood sugar value?

NULLs

- NULLs intended to model incompleteness
- but semantics not clear and hence highly criticized

Lit: L. Libkin. SQL's three-valued logic and certain answers. ACM Trans. Database Syst., 41(1):1:1–1:28, 2016.

Exar	nple		
	Patient		Blood sugar
ID	Name	ID	value [30-600]
1	Sokrates	1	90
2	Platon	2	120
3	Aristotle	3	NULL

Aristotle has a blood sugar value (30 or 31 or ...)

NULLs

- NULLs intended to model incompleteness
- but semantics not clear and hence highly criticized

Lit: L. Libkin. SQL's three-valued logic and certain answers. ACM Trans. Database Syst., 41(1):1:1–1:28, 2016.

Exar	nple		
	Patient	F	Pregnancy
ID	Name	ID	HCG value
1	Sokrates	1	NULL
2	Platon	2	NULL
3	Aristotle	3	NULL
4	Xanthippe	4	NULL
5	Leda	5	130

- Male patient with NULL: no HCG test
- Female patient with NULL: not HCG test (but she has HCG value) or HCG test & not known

OBDA: Motivation and Overview

Ontology-Based Data Access

- Use ontologies as interface
- to access (here: query)
- data stored in some format
- using mappings

. . .



Ontology-Based Data Access

- Use ontologies as interface
- to access (here: query)
- data stored in some format
- using mappings

. . .



Ontologies

- Ontologies are triples of the form $\mathcal{O} = (\sigma, \mathcal{T}, \mathcal{A})$
 - Signature σ : Non-logical vocabulary $\sigma = N_i \cup N_C \cup N_R$
 - Tbox *T*: set of *σ*-axioms in some logic to capture terminological knowledge This lecture: ontologies represented in Description Logics (DLs)
 - Abox A: set of σ-axioms in (same logic) to capture assertional/contingential knowledge
- Note: Sometimes only tbox termed ontology

Ontologies

- Ontologies are triples of the form $\mathcal{O} = (\sigma, \mathcal{T}, \mathcal{A})$
 - Signature σ : Non-logical vocabulary $\sigma = N_i \cup N_C \cup N_R$
 - Tbox *T*: set of *σ*-axioms in some logic to capture terminological knowledge This lecture: ontologies represented in Description Logics (DLs)
 - Abox A: set of σ-axioms in (same logic) to capture assertional/contingential knowledge
- Note: Sometimes only tbox termed ontology
- Semantics defined on the basis of σ -interpretations $\mathcal I$
 - $\mathcal{I} \models Ax$ iff \mathcal{I} makes all axioms in Ax true
 - $Mod(Ax) = \{\mathcal{I} \models Ax\}$

General Idea

- A: Represents facts in domain of interest
- Open world assumption: *Mod*(A) is not a singleton
- *T*: Constrains *Mod*(*A*) with intended *σ* readings
- Usually one has only approximations of intended models *IM*



General Idea

- A: Represents facts in domain of interest
- Open world assumption: *Mod*(A) is not a singleton
- *T*: Constrains *Mod*(*A*) with intended *σ* readings
- Usually one has only approximations of intended models *IM*
- Realize inference services on the basis of the constrained interpretations



WARNING: A Misconception

With ontologies one does not declare data structures

- Abox data in most cases show pattern of data structures
- One does not have to re-model patterns/constraints in the abox data
 - ► Knowing "All A are B" in the abox is different from stipulating A ⊆ B (the former is known as integrity constraint)
 - Add A ⊆ B, if you need to handle this relation for objects not mentioned in the abox
- Motto: Keep the tbox simple

Ontology-Based Data Access

- Use ontologies as interface
- to access (here: query)
- data stored in some format
- using mappings

. . .



Reasoning Services

- Different standard and nonstandard reasoning services exist
- May be reducible to each other

Example (Reasoning Services)

consistency check, subsumption check, taxonomy calculations, most specific subsumer, most specific concept, matching, ...

- In classical OBDA focus on
 - Consistency checking: $Mod(\mathcal{A} \cup \mathcal{T}) \neq \emptyset$.
 - Query answering

Reasoning Services

- Different standard and nonstandard reasoning services exist
- May be reducible to each other

Example (Reasoning Services)

consistency check, subsumption check, taxonomy calculations, most specific subsumer, most specific concept, matching, ...

- In classical OBDA focus on
 - Consistency checking: $Mod(\mathcal{A} \cup \mathcal{T}) \neq \emptyset$.
 - Query answering
- Next to abox and tbox language query language QL over σ is a relevant factor for OBDA
- Certain query answering

 $cert(\psi(\vec{x}), \mathcal{T} \cup \mathcal{A}) = \{ \vec{a} \in (N_i)^n \mid \mathcal{T} \cup \mathcal{A} \models \psi[\vec{x}/\vec{a}] \}$

Ontology-Based Data Access

- Use ontologies as interface
- to access (here: query)
- data stored in some format
- using mappings



Backend Data Sources

- Classically: relational SQL DBs with static data
- Possible extensions: non-SQL DBs
 - datawarehouse repositories for statistical applications
 - pure logfiles
 - RDF repositories
- Non-static data
 - historical data (stored in temporal DB)
 - dynamic data coming in streams
- Originally intended for multiple DBs but ...

Ontology-Based Data Access

- Use ontologies as interface
- to access (here: query)
- data stored in some format
- using mappings

. . .



Mappings

- Mappings have an important crucial role in OBDA
- Lift data to the ontology level
 - Data level: (nearly) close world
 - Ontology Level: open world

Definition (Schema of Mappings)

 $m:\psi(\vec{f}(\vec{x})) \longleftarrow Q(\vec{x},\vec{y})$

- $\psi(\vec{f}(\vec{x}))$: Template (query) for generating abox axioms
- $Q(\vec{x}, \vec{y})$: Query over the backend sources
- Function \vec{f} translates backend instantiations of \vec{x} to constants

• Mappings *M* over backend sources generates abox $\mathcal{A}(M, DB)$.

Example Scenario: Measurements

Example schema for measurement and event data in DB SENSOR(<u>SID</u>, CID, Sname, TID, description) SENSORTYPE(<u>TID</u>, Tname)

COMPONENT(<u>CID</u>, superCID, AID, Cname) ASSEMBLY(<u>AID</u>, AName, ALocation) MEASUREMENT(<u>MID</u>, MtimeStamp, SID, Mval) MESSAGE(<u>MesID</u>, MesTimeStamp, MesAssemblyID, catID, MesEventText) CATEGORY(catID, catName)

Example Scenario: Measurements

Example schema for measurement and event data in DB SENSOR(<u>SID</u>, CID, Sname, TID, description) SENSORTYPE(<u>TID</u>, Tname)

COMPONENT(<u>CID</u>, superCID, AID, Cname) ASSEMBLY(<u>AID</u>, AName, ALocation) MEASUREMENT(<u>MID</u>, MtimeStamp, SID, Mval) MESSAGE(<u>MesID</u>, MesTimeStamp, MesAssemblyID, catID, MesEventText) CATEGORY(catID, catName)

For mapping

m: $Sens(x) \wedge name(x, y) \leftarrow$

SELECT f(SID) as x, Sname as y FROM SENSOR

the row data in SENSOR table

SENSOR

(123, comp45, TempSens, TC255, 'A temperature sensor')

generates facts

 $Sens(f(123)), name(f(123), TempSens) \in \mathcal{A}(m, DB)$

OBDA in the Classical Sense

- Keep the data where they are because of large volume
- Abox is virtual (no materialization)
- ► First-order logic (FOL) perfect rewriting + unfolding



Ontologies and Description Logics

Description Logics

Definition

Description logics (DLs) are logics for use in knowledge representation with special attention on a good balance of expressibility and feasibility of reasoning services

Can be mapped to fragments of FOL

Description Logics

Definition

Description logics (DLs) are logics for use in knowledge representation with special attention on a good balance of expressibility and feasibility of reasoning services

- Can be mapped to fragments of FOL
- ► Use
 - ► as ontology representation language for conceptual modeling
 - in particular in the semantic web
 - ► Formal counterpart of standard web ontology language (OWL)
 - ▶ and in particular for ontology-based data access (OBDA)

Description Logics

Definition

Description logics (DLs) are logics for use in knowledge representation with special attention on a good balance of expressibility and feasibility of reasoning services

- Can be mapped to fragments of FOL
- Use
 - ► as ontology representation language for conceptual modeling
 - in particular in the semantic web
 - ► Formal counterpart of standard web ontology language (OWL)
 - ▶ and in particular for ontology-based data access (OBDA)
- Have been investigated for ca. 30 years now
 - Many theoretical insights on various different purpose DLs
 - General-purpose reasoners (RacerPro, Fact++, ...) and specific reasoners (Quest,...)
 - Various editing tools (most notably Protege)

Family of DLs

- Variable-free logics centered around concepts
- concepts = one-ary predicates in FOL = classes in OWL



An (Semi-)Expressive Logic: ALC

• Vocabulary: constants N_i , atomic concepts N_C , roles N_R

- An (Semi-)Expressive Logic: ALC
 - Vocabulary: constants N_i , atomic concepts N_C , roles N_R
 - Concept(description)s: syntax

$$C ::= A \quad (\text{for } A \in N_C) \mid C \sqcap C \mid C \sqcup C \mid \neg C \mid \\ \forall r.C \mid \exists r.C \quad (\text{for } r \in N_R) \mid \bot \mid \top$$
- An (Semi-)Expressive Logic: ALC
 - ► Vocabulary: constants N_i , atomic concepts N_C , roles N_R
 - Concept(description)s: syntax

$$C ::= A \quad (\text{for } A \in N_C) \mid C \sqcap C \mid C \sqcup C \mid \neg C \mid \\ \forall r.C \mid \exists r.C \quad (\text{for } r \in N_R) \mid \bot \mid \top$$

Concept(description)s: semantics

An (Semi-)Expressive Logic: ALC

▶ Vocabulary: constants N_i , atomic concepts N_C , roles N_R

Concept(description)s: syntax

$$C ::= A \quad (\text{for } A \in N_C) \mid C \sqcap C \mid C \sqcup C \mid \neg C \mid \\ \forall r.C \mid \exists r.C \quad (\text{for } r \in N_R) \mid \bot \mid \top$$

Concept(description)s: semantics



- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ for all $A \in N_C$
- $c^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ for all $c \in N_i$

$$\ \, \mathbf{r}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \\ \text{ for all } r \in N_r$$

An (Semi-)Expressive Logic: ALC

- ▶ Vocabulary: constants N_i , atomic concepts N_C , roles N_R
- Concept(description)s: syntax

$$C ::= A \quad (\text{for } A \in N_C) \mid C \sqcap C \mid C \sqcup C \mid \neg C \mid \\ \forall r.C \mid \exists r.C \quad (\text{for } r \in N_R) \mid \bot \mid \top$$

Concept(description)s: semantics



- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ for all $A \in N_C$
- $c^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ for all $c \in N_i$
- $\mathsf{r}^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ for all $r \in N_r$

$$\bullet \ (C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$$

 $\bullet \ (C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$

$$\bullet \neg C = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$$

- ► $(\forall r.C)^{\mathcal{I}} = \{ d \in \Delta^{\mathcal{I}} \mid \text{ for all } e \in \Delta^{\mathcal{I}} :$ If $(d, e) \in r^{\mathcal{I}}$ then $e \in C^{\mathcal{I}} \}$
- ► $(\exists r.C)^{\mathcal{I}} = \{ d \in \Delta^{\mathcal{I}} \mid \text{ there is } e \in \Delta^{\mathcal{I}} \text{ s.t. } (d, e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}} \}$

Tbox and Abox

• Terminological Box (tbox) T

- Finite set of general concept inclusions (GCIs)
- GCI: axioms of form C ⊆ D (for arbitrary concept descriptions)
 C ≡ D abbreviates {C ⊆ D, D ⊆ C}
- Semantics: $\mathcal{I} \models C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

Tbox and Abox

• Terminological Box (tbox) \mathcal{T}

- Finite set of general concept inclusions (GCIs)
- GCI: axioms of form C ⊆ D (for arbitrary concept descriptions)
 C ≡ D abbreviates {C ⊆ D, D ⊆ C}
- Semantics: $\mathcal{I} \models C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

• Assertional Box (abox) A

- Finite set of assertions
- Assertion: C(a) (concept assertion), r(a, b) (role assertion)
- Semantics:

 $\mathcal{I} \models C(a) \text{ iff } a^{\mathcal{I}} \in C^{\mathcal{I}} \\ \mathcal{I} \models r(a, b) \text{ iff } (a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}.$

Tbox and Abox

• Terminological Box (tbox) \mathcal{T}

- Finite set of general concept inclusions (GCIs)
- GCI: axioms of form C ⊑ D (for arbitrary concept descriptions)
 C ≡ D abbreviates {C ⊑ D, D ⊑ C}
- Semantics: $\mathcal{I} \models C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$.

• Assertional Box (abox) A

- Finite set of assertions
- Assertion: C(a) (concept assertion), r(a, b) (role assertion)
- Semantics:

 $\mathcal{I} \models C(a) \text{ iff } a^{\mathcal{I}} \in C^{\mathcal{I}} \\ \mathcal{I} \models r(a, b) \text{ iff } (a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}.$

• Ontology: $(\sigma, \mathcal{T}, \mathcal{A})$ or just $\mathcal{T} \cup \mathcal{A}$

Example (University)

 $\mathcal{T} = \{ GradStudent \sqsubseteq Student, \\ GradStudent \sqsubseteq \exists takesCourse.GradCourse \} \\ \mathcal{A} = \{ GradStudent(john) \} \end{cases}$

Consider the following interpretations

Example (University)

 $\mathcal{T} = \{ GradStudent \sqsubseteq Student, \\ GradStudent \sqsubseteq \exists takesCourse.GradCourse \} \\ \mathcal{A} = \{ GradStudent(john) \} \end{cases}$

Consider the following interpretations

 $\blacktriangleright \mathcal{I}_1$:

- $john^{\mathcal{I}_1} = j$
- GradStudent^{I_1} = {*j*}
- Student^{\mathcal{I}_1} = {*j*}
- ▶ GradCourse^I = {s}
- takesCourse^{\mathcal{I}_1} = {(j, s)}

• $\mathcal{I}_1 \models \mathcal{T} \cup \mathcal{A}$

• \mathcal{I}_2 :

- ► $john^{\mathcal{I}_2} = j$
- GradStudent^{I_2} = {*j*}
- ► Student^{*I*₂} = {*j*}
- GradCourse^{\mathcal{I}_2} = {*j*}
- takesCourse^{I_2} = {(j, j)}

 $\blacktriangleright \ \mathcal{I}_2 \models \mathcal{T} \cup \mathcal{A}$

Example (University)

 $\mathcal{T} = \{ GradStudent \sqsubseteq Student, \\ GradStudent \sqsubseteq \exists takesCourse.GradCourse \} \\ \mathcal{A} = \{ GradStudent(john) \} \end{cases}$

Consider the following interpretations

► *I*₃ :

- GradStudent^{I_1} = {*j*}
- Student^{\mathcal{I}_1} = {*j*}
- GradCourse^{\mathcal{I}_1} = \emptyset
- ▶ takesCourse^{I1} = Ø
- $\blacktriangleright \ \mathcal{I}_3 \not\models \mathcal{T} \cup \mathcal{A}$

Stricter notion of Tbox

- Above definition of tbox very general
 - "Meanings" of concept names determined only implicitly in the whole ontology
 - No guarantee for unique extensions
- Early notion of tbox more related to idea of explicitly defining concept names
- $C \equiv D$ used as abbreviation for $C \sqsubseteq D$ and $D \sqsubseteq C$
- Concept definition: $A \equiv D$ (where A atomic)

Definition

A TBox in a strict sense is a finite set of concept definitions not defining a concept multiple times or in a cyclic manner. Defined concepts occur on the lhs, primitive concept on the rhs of definitions.

Digression: Implicit vs. Explicit Definability

- Sometimes a general tbox may fix the denotation of a concept name w.r.t. denotations of the others =>> implicit definability
- Maybe then it can also be defined explicitly?

Definition

Given an FOL theory Ψ over signature σ and a predicate symbol R.

- *R* is implicitly defined in Ψ iff for any two models 𝔅 ⊨ Ψ and 𝔅 ⊨ Ψ agreeing on σ \ {*R*} one has *R*^𝔅 = *R*^𝔅.
- ► *R* is explicitly defined in Ψ by a formula $\phi(\vec{x})$ not containing *R* iff $\Psi \models \forall \vec{x} R(\vec{x}) \leftrightarrow \phi(\vec{x})$

Digression: Beth Definability Theorem

For FOL both implicit and explicit definability coincide

Theorem

An FOL theory defines a predicate implicitly iff it defines it explicitly

 Though DLs can be embedded into FOL, the equivalence of implicit and explicit definability does not transfer necessarily to DLs

► At least it does for *ALC* theories

Lit: B. ten Cate, E. Franconi, and I. Seylan. Beth definability in expressive description logics. J. Artif. Int. Res., 48(1): 347–414, Oct. 2013.

Reasoning services

- Semantical notions as in FOL but additional notions due to focus on concepts
- Let $\mathcal{O} = (\sigma, \mathcal{T}, \mathcal{A})$

Definition (Basic Semantical Notions)

- Model: $\mathcal{I} \models \mathcal{O}$ iff $\mathcal{I} \models \mathcal{T} \cup \mathcal{A}$
- Satisfiability: \mathcal{O} is satisfiable iff $\mathcal{T} \cup \mathcal{A}$ is satisfiable
- Coherence: O is coherent iff T ∪ A has a model I s.t. for all concept names A^I ≠ Ø
- Concept satisfiability: *C* is satisfiable w.r.t. \mathcal{O} iff there is $\mathcal{I} \models \mathcal{O}$ s.t. $C^{\mathcal{I}} \neq \emptyset$
- ► Subsumption: *C* is subsumed by *D* w.r.t. \mathcal{O} iff $\mathcal{O} \models C \sqsubseteq D$ iff $\mathcal{T} \cup \mathcal{A} \models C \sqsubseteq D$
- ▶ Instance check: *a* is an instance of *C* w.r.t. \mathcal{O} iff $\mathcal{O} \models C(a)$

Reduction Examples

- Many of the semantical notions are reducible to each other
- We give only one example

Exercise

Show that subsumption can be reduced to satisfiability tests (allowing the introduction of new constants). More concretely:

 $C \sqsubseteq D$ w.r.t. \mathcal{O} iff $(\sigma \cup \{b\}, \mathcal{T}, \mathcal{A} \cup \{C(b), \neg D(b)\})$ is not satisfiable (where *b* is a fresh constant).

Extended Reasoning Services

Definition

- ▶ Instance retrieval: Find all constants x s.t. $\mathcal{O} \models C(x)$
- Query answering: Certain answers cert(φ(x), O) = { a ∈ N_i | O ⊨ φ[x/a]}
- Classification: Compute the subsumption hierarchy of all concept names
- Realization: Compute the most specific concept name to which a given constant belongs
- Pinpointing, matching,

Example (Certain Answers for Conjunctive Queries)

 $\mathcal{T} = \{ \top \sqsubseteq \mathit{Male} \sqcup \mathit{Female}, \mathit{Male} \sqcap \mathit{Female} \sqsubseteq \bot \}$

 $\mathcal{A} = \{ \textit{friend(john, susan), friend(john, andrea), female(susan), } \\ \textit{likes(susan, andrea), likes(andrea, bill), Male(bill) } \}$

 $Q(x) = \exists y, z(friend(x, y) \land Female(y) \land likes(y, z) \land Male(z))$

•
$$cert(Q(x), \mathcal{O}) = ?$$

Example (Certain Answers for Conjunctive Queries)

 $\mathcal{T} = \{ \top \sqsubseteq \textit{Male} \sqcup \textit{Female}, \textit{Male} \sqcap \textit{Female} \sqsubseteq \bot \}$

 $\mathcal{A} = \{ \textit{friend(john, susan), friend(john, andrea), female(susan), } \\ \textit{likes(susan, andrea), likes(andrea, bill), Male(bill) } \}$

 $Q(x) = \exists y, z(friend(x, y) \land Female(y) \land likes(y, z) \land Male(z))$

• $cert(Q(x), \mathcal{O}) = ?$

- We have to consider all possible models of the ontology
- But here there are actually two classes: Andrea is male vs. Andrea is not male.

Example (Certain Answers for Conjunctive Queries)

$$\mathcal{T} = \{ \top \sqsubseteq Male \sqcup Female, Male \sqcap Female \sqsubseteq \bot \}$$

 $\mathcal{A} = \{ \textit{friend(john, susan)}, \textit{friend(john, andrea)}, \textit{female(susan)}, \\ \textit{likes(susan, andrea)}, \textit{likes(andrea, bill)}, \textit{Male(bill)} \}$

$$Q(x) = \exists y, z(friend(x, y) \land Female(y) \land likes(y, z) \land Male(z))$$



- Most DLs (such as \mathcal{ALC}) can be embedded into FOL
- Notion of embedding is well-defined as FOL structures are used for semantics of DLs.

► Correspondence idea Concept names = unary predicates, roles = binary predicates, GCI = ∀ rules

- ▶ Most DLs (such as *ALC*) can be embedded into FOL
- Notion of embedding is well-defined as FOL structures are used for semantics of DLs.
- ► Correspondence idea Concept names = unary predicates, roles = binary predicates, GCI = ∀ rules
- ► Define for any concept description and variable x its corresponding x-open formula \(\tau_x(C)\)
 - $\tau_x(A) = A(x)$
 - $\tau_x(C \sqcap D) = \tau_x(C) \land \tau_x(D)$
 - $\tau_x(C \sqcup D) = \tau_x(C) \lor \tau_x(D)$
 - $\tau_x(\neg C) = \neg \tau_x(C)$
 - $\tau_x(\forall r.C) = \forall y(r(x,y) \rightarrow \tau_y(C))$

- Most DLs (such as ALC) can be embedded into FOL
- Notion of embedding is well-defined as FOL structures are used for semantics of DLs.
- ► Correspondence idea Concept names = unary predicates, roles = binary predicates, GCI = ∀ rules
- ► Define for any concept description and variable x its corresponding x-open formula \(\tau_x(C)\)
 - $\tau_x(A) = A(x)$
 - $\tau_x(C \sqcap D) = \tau_x(C) \land \tau_x(D)$
 - $\tau_x(C \sqcup D) = \tau_x(C) \lor \tau_x(D)$
 - $\tau_x(\neg C) = \neg \tau_x(C)$
 - $\tau_x(\forall r.C) = \forall y(r(x,y) \rightarrow \tau_y(C))$
- Abox axioms not changed

- ▶ Most DLs (such as *ALC*) can be embedded into FOL
- Notion of embedding is well-defined as FOL structures are used for semantics of DLs.
- ► Correspondence idea Concept names = unary predicates, roles = binary predicates, GCI = ∀ rules
- ► Define for any concept description and variable x its corresponding x-open formula \(\tau_x(C)\)
 - $\tau_x(A) = A(x)$
 - $\tau_x(C \sqcap D) = \tau_x(C) \land \tau_x(D)$
 - $\tau_x(C \sqcup D) = \tau_x(C) \lor \tau_x(D)$
 - $\tau_x(\neg C) = \neg \tau_x(C)$
 - $\tau_x(\forall r.C) = \forall y(r(x,y) \rightarrow \tau_y(C))$
- Abox axioms not changed
- ▶ Tbox axioms: $C \sqsubseteq D$ becomes $\forall x(\tau_x(C) \rightarrow \tau_x(D))$

- For translation two variables are sufficient ("2 finger movement")
- Hence: DLs embeddable into known 2-variable fragment of FOL
- Also the fragment is a guarded fragment: one quantifies over variables fixed within atom.

DL Family

- Different DLs for different purposes
 - What is more important: Expressivity or feasibility?
 - Which kinds of reasoning services does one have to provide?
- Differences regarding
 - the allowed set of concept constructors
 - the allowed set of role constructors
 - the allowed types of tbox axioms
 - the allowed types of abox axioms
 - the allowance of concrete domains and attributes (such as hasAge with range the domain of integers)

Family of DLs and their Namings

- ► AL: attributive language
- \triangleright C: (full) complement/negation
- \blacktriangleright *I*: inverse roles $((r^{-1})^{\mathcal{I}} = \{(d, e) \in \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}} \mid (e, d) \in r^{\mathcal{I}}\})$
- \blacktriangleright \mathcal{H} : role inclusions

(hasFather \Box hasParent)

(trans isReachable)

- \blacktriangleright S: ACC + transitive roles
- \blacktriangleright N: ungualified number restrictions

 $((> n r)^{\mathcal{I}} = \{ d \in \Delta^{\mathcal{I}} \mid \#(\{e \mid (d, e) \in r^{\mathcal{I}}\}) > n) \}$ $\{b\}^{\mathcal{I}} = \{b^{\mathcal{I}}\}$

- O: nominals
- Q: qualified number restrictions

 $((\geq n r.C)^{\mathcal{I}} = \{ d \in \Delta^{\mathcal{I}} \mid \#(\{e \mid (d, e) \in r^{\mathcal{I}}\} \text{ and } e \in C^{\mathcal{I}}) \geq n) \}$

- F: functionality constraints

▶ \mathcal{R} : role chains and $\exists R.Self$ (hasFather \circ hasMother \sqsubseteq hasgrandMa) $(narcist \equiv \exists likes.Self)$

 $\mathcal{I} \models (func R)$ iff $R^{\mathcal{I}}$ is a function

► OWL 2 is SROID

- Lightweight DLs favor feasibility over expressibility by, roughly, disallowing disjunction
- In principle three lightweight logics that have corresponding OWL 2 profiles (https://www.w3.org/TR/owl2-profiles/)

- Lightweight DLs favor feasibility over expressibility by, roughly, disallowing disjunction
- In principle three lightweight logics that have corresponding OWL 2 profiles (https://www.w3.org/TR/owl2-profiles/)
- ► *EL* (OWL 2 EL)
 - \blacktriangleright No inverses, no negation, no \forall
 - polynomial time algorithms for all the standard reasoning tasks with large ontologies

- Lightweight DLs favor feasibility over expressibility by, roughly, disallowing disjunction
- In principle three lightweight logics that have corresponding OWL 2 profiles (https://www.w3.org/TR/owl2-profiles/)
- ► *EL* (OWL 2 EL)
 - \blacktriangleright No inverses, no negation, no \forall
 - polynomial time algorithms for all the standard reasoning tasks with large ontologies
- DL-Lite (OWL 2 QL)
 - Tbox: No qualified existentials on lhs
 - Feasible CQ answering using rewriting and unfolding leveraging RDBS technology

- Lightweight DLs favor feasibility over expressibility by, roughly, disallowing disjunction
- In principle three lightweight logics that have corresponding OWL 2 profiles (https://www.w3.org/TR/owl2-profiles/)
- ► *EL* (OWL 2 EL)
 - \blacktriangleright No inverses, no negation, no \forall
 - polynomial time algorithms for all the standard reasoning tasks with large ontologies
- DL-Lite (OWL 2 QL)
 - Tbox: No qualified existentials on lhs
 - Feasible CQ answering using rewriting and unfolding leveraging RDBS technology
- RL (OWL 2 RL)
 - Tbox restriction: "Only concept names on the rhs"
 - Polynomial time algorithms leveraging rule-extended database technologies operating directly on RDF triples

Comparison

	RL	EL	QL	
inverse roles	+	-	+	
rhs qual. exist	-	+	+	
lhs qual. exist.	+	+	-	

Complexity

- ► A nearly complete picture of reasoning services for DLs
- Research in DL community as of now resembles complexity farming
- DL complexity navigator:

http://www.cs.man.ac.uk/~ezolin/dl (Last update 2013)

Tableaux Calculus for \mathcal{ALC}

- ► Efficient calculi are at the core of DL reasoners
- ► Tableaux calculi have been implemented successfully
- Refutation calculus based on disjunctive normal form

Tableaux Calculus for \mathcal{ALC}

- Efficient calculi are at the core of DL reasoners
- ► Tableaux calculi have been implemented successfully
- Refutation calculus based on disjunctive normal form
- ► We demonstrate it here at an example for *ALC* tboxes
- For a full description and proofs see handbook article by Baader

Lit: F. Baader and W. Nutt. Basic description logics. In F. Baader et al., editors, The Description Logic Handbook, pages 43–95. Cambridge University Press, 2003.

Tableaux Example

- ► *ALC* tableau gives tests for satisfiability of abox
- by checking whether obvious contradictions (clashes with complementary literals) are contained
- An abox that is complete (no rules applicable anymore) and open (no clashes) describes a model
- Algorithm applies tableau rules to extend abox

Rules

- Starts with an abox A₀ which is in negation normal form (NNF): all ¬ appear only in front of concept names)
- ► Apply rules to construct new aboxes; indeterminism due to ⊔ rule

Rule	Condition	$\sim \rightarrow$	Effect
\sim_{\Box}	$(C \sqcap D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x), \mathcal{D}(x)\}$
\sim_{\sqcup}	$(C \sqcup D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x)\}$ or $\mathcal{A} \cup \{D(x)\}$
\rightsquigarrow_\exists	$(\exists r.C)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{r(x, y), C(y)\}$ for fresh y
$\rightsquigarrow_{\forall}$	$(\forall r.C)(x), r(x,y) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(\mathbf{y})\}$

Rules

- Starts with an abox A₀ which is in negation normal form (NNF): all ¬ appear only in front of concept names)
- ► Apply rules to construct new aboxes; indeterminism due to ⊔ rule

Rule	Condition	$\sim \rightarrow$	Effect
\sim_{\Box}	$(C \sqcap D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x), \mathcal{D}(x)\}$
\sim_{\sqcup}	$(C \sqcup D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x)\}$ or $\mathcal{A} \cup \{D(x)\}$
\rightsquigarrow_\exists	$(\exists r.C)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{r(x, y), C(y)\}$ for fresh y
$\rightsquigarrow_{\forall}$	$(\forall r.C)(x), r(x,y) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(y)\}$

Rules applicable only if they lead to an addition of assertions
Rules

- Starts with an abox A₀ which is in negation normal form (NNF): all ¬ appear only in front of concept names)
- ► Apply rules to construct new aboxes; indeterminism due to ⊔ rule

Rule	Condition	$\sim \rightarrow$	Effect
\sim_{\Box}	$(C \sqcap D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x), \mathcal{D}(x)\}$
\sim_{\sqcup}	$(C \sqcup D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x)\}$ or $\mathcal{A} \cup \{D(x)\}$
\rightsquigarrow_\exists	$(\exists r.C)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{r(x, y), C(y)\}$ for fresh y
$\rightsquigarrow_{\forall}$	$(\forall r.C)(x), r(x,y) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(y)\}$

- Rules applicable only if they lead to an addition of assertions
- One obtains a tree with aboxes (due to indeterminism)

Rules

- Starts with an abox A₀ which is in negation normal form (NNF): all ¬ appear only in front of concept names)
- ► Apply rules to construct new aboxes; indeterminism due to ⊔ rule

Rule	Condition	$\sim \rightarrow$	Effect
\sim_{\Box}	$(C \sqcap D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x), \mathcal{D}(x)\}$
\sim_{\sqcup}	$(C \sqcup D)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(x)\}$ or $\mathcal{A} \cup \{D(x)\}$
\rightsquigarrow_\exists	$(\exists r.C)(x) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{r(x, y), C(y)\}$ for fresh y
$\rightsquigarrow \forall$	$(\forall r.C)(x), r(x,y) \in \mathcal{A}$	\sim	$\mathcal{A} \cup \{\mathcal{C}(y)\}$

- Rules applicable only if they lead to an addition of assertions
- One obtains a tree with aboxes (due to indeterminism)
- Within each abox a tree-like structure is established (tree-model property)

• Given: $\mathcal{T} = \{GoodStudent \equiv Smart \sqcap Studious\}$

- Given: $\mathcal{T} = \{GoodStudent \equiv Smart \sqcap Studious\}$
- Subsumption test: $\mathcal{T} \vDash \exists knows.Smart \sqcap \exists knows.Studious \sqsubseteq \exists knows.GoodStudent$

- Given: $T = \{GoodStudent \equiv Smart \sqcap Studious\}$
- Subsumption test: $\mathcal{T} \vDash \exists knows.Smart \sqcap \exists knows.Studious \sqsubseteq \exists knows.GoodStudent$
- Reduction to abox satisfiability: { 3knows.Smart ¬ 3knows.Studious ¬ ¬(3knows.GoodStudent)(a)} satisfiable?

- Given: $T = \{GoodStudent \equiv Smart \sqcap Studious\}$
- Subsumption test: $\mathcal{T} \vDash \exists knows.Smart \sqcap \exists knows.Studious \sqsubseteq \exists knows.GoodStudent$
- ▶ Reduction to abox satisfiability: {∃knows.Smart □ ∃knows.Studious □ ¬(∃knows.GoodStudent)(a)} satisfiable?
- Expansions of definition {∃knows.Smart □ ∃knows.Studious □ ¬(∃knows.(Smart □ Studious))(a)} satisfiable?

- Given: $T = \{GoodStudent \equiv Smart \sqcap Studious\}$
- Subsumption test: $\mathcal{T} \vDash \exists knows.Smart \sqcap \exists knows.Studious \sqsubseteq \exists knows.GoodStudent$
- Reduction to abox satisfiability: { 3knows.Smart ¬ 3knows.Studious ¬ (3knows.GoodStudent)(a)} satisfiable?
- Expansions of definition {∃knows.Smart □ ∃knows.Studious □ ¬(∃knows.(Smart □ Studious))(a)} satisfiable?
- Transform to NNF {∃knows.Smart □ ∃knows.Studious □ ∀knows.(¬Smart □ ¬Studious)(a)} satisfiable?

- Given: $T = \{GoodStudent \equiv Smart \sqcap Studious\}$
- Subsumption test: $\mathcal{T} \vDash \exists knows.Smart \sqcap \exists knows.Studious \sqsubseteq \exists knows.GoodStudent$
- Reduction to abox satisfiability: { 3knows.Smart ¬ 3knows.Studious ¬ (3knows.GoodStudent)(a)} satisfiable?
- Expansions of definition {∃knows.Smart □ ∃knows.Studious □ ¬(∃knows.(Smart □ Studious))(a)} satisfiable?
- Transform to NNF {∃knows.Smart □ ∃knows.Studious □ ∀knows.(¬Smart □ ¬Studious)(a)} satisfiable?
- ► Expansion step assumes a tbox in a strict sense, but GCIs can be transformed to definitions (i.e., axioms of the form A ≡ C) using additional symbols

Example (A Tableau Derivation)

- {∃knows.Smart □ ∃knows.Studious □ ∀knows.(¬Smart □ ¬Studious)(a)}
- Abbreviation: $\{\exists r.A \sqcap \exists r.B \sqcap \forall r.(\neg A \sqcup \neg B)(a)\}$



Example (The partial tree model in the aboxes)

- {∃knows.Smart □ ∃knows.Studious □ ∀knows.(¬Smart □ ¬Studious)(a)}
- Abbreviation: $\{\exists r.A \sqcap \exists r.B \sqcap \forall r.(\neg A \sqcup \neg B)(a)\}$



Canonical tree model(s) can be directly read off (here from abox A_{5.21} which contains no clash):

 $\mathcal{I} = \left(\{a, b, c\}, \cdot^{\mathcal{I}} \right) \text{ with}$ $r^{\mathcal{I}} = \left\{ (a, b), (a, c) \right\} \quad A^{\mathcal{I}} = \{b\} \quad B^{\mathcal{I}} = \{c\}$

Tableaux Calculus

- ► The tableau calculus for *ALC* is complete, correct, and terminates.
- Hence, the following properties hold

Theorem

- ALC abox satisfiability (concept satisfiability, subsumption...) is decidable
- ALC has the finite model property, i.e. if an ALC ontology has a model, then it has a finite model.
- ► *ALC* has the tree model property

Tableaux Calculus

- ► The tableau calculus for *ALC* is complete, correct, and terminates.
- Hence, the following properties hold

Theorem

- ALC abox satisfiability (concept satisfiability, subsumption...) is decidable
- ALC has the finite model property, i.e. if an ALC ontology has a model, then it has a finite model.
- ► *ALC* has the tree model property