

An Overview on Description Logics, and Axiom Pinpointing in \mathcal{EL}

Barış Sertkaya

Theoretical Computer Science
TU Dresden, Germany

June 15, 2009

- 1 Knowledge Representation
- 2 Description Logics
- 3 The DL \mathcal{ALC}
- 4 DL Ontology
- 5 Reasoning and its Complexity
- 6 Web Ontology Language OWL
- 7 Axiom Pinpointing in \mathcal{EL}

- **Description Logics:**
 - logics for representing **conceptual** knowledge.
 - describe the world with a logical language, classify the descriptions
- **Formal Concept Analysis:**
 - field of lattice theory for **conceptual** data analysis
 - derive a classification from the given samples
- combining them, using methods from one field in the other

Develop formalisms that have

- a well-defined **syntax**,
- formal, unambiguous **semantics**,
- **practical** methods for reasoning, and **efficient** implementations

Conceptual knowledge

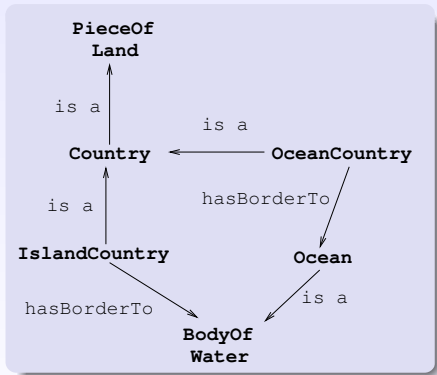
- **Classes**: country, coastal country, ...
- **Relations**: has border to, has neighbour, ...
- **Individuals**: Portugal, Mediterranean, Atlantic, ...

Semantic networks [Quillian(1967)]

- nodes represent classes
- links represent relations
- hasBorderTo: has a border to sth., or only has border to sth.?
- **ambiguous semantics!**

KL-ONE [Brachman & Levesque(1985)]

- logic-based semantics



Description Logics (DLs)

- family of **logic-based knowledge representation formalisms**
- describe an application domain in terms of
 - **concepts** (classes) like Country, Ocean, ...
 - **individuals** like Portugal, Atlantic, ...
 - **roles** (relations) like hasBorderTo, hasNeighbour,...
- well-defined, formal semantics
- decidable fragments of First Order Logic

- $\text{Sea} \sqsubseteq \text{Ocean}$,
- $\text{Country} \sqcap \exists \text{hasBorderTo}.\text{Ocean}$,
- $\text{Ocean}(\textit{Atlantic})$,
- $\text{OceanCountry}(\textit{Portugal})$

\mathcal{ALC} : The smallest propositionally closed DL

- Atomic concepts: A, B, \dots (unary predicates)
- Atomic roles: r, s, \dots (binary predicates)
- Constructors:
 - $\neg C$ (negation)
 - $C \sqcap D$ (conjunction)
 - $C \sqcup D$ (disjunction)
 - $\exists r.C$ (existential restriction)
 - $\forall r.C$ (value restriction)

Examples:

- $\text{Country} \sqcap \neg \text{Island}$
- $\text{Country} \sqcap \forall \text{hasBorderTo}.\text{LandMass}$
- $\text{Country} \sqcap \exists \text{hasNeighbour} . (\exists \text{hasBorderTo}.\text{Ocean})$

based on **interpretations** $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of a **domain** $\Delta^{\mathcal{I}}$, and an **interpretation function** $\cdot^{\mathcal{I}}$

Concept and role names:

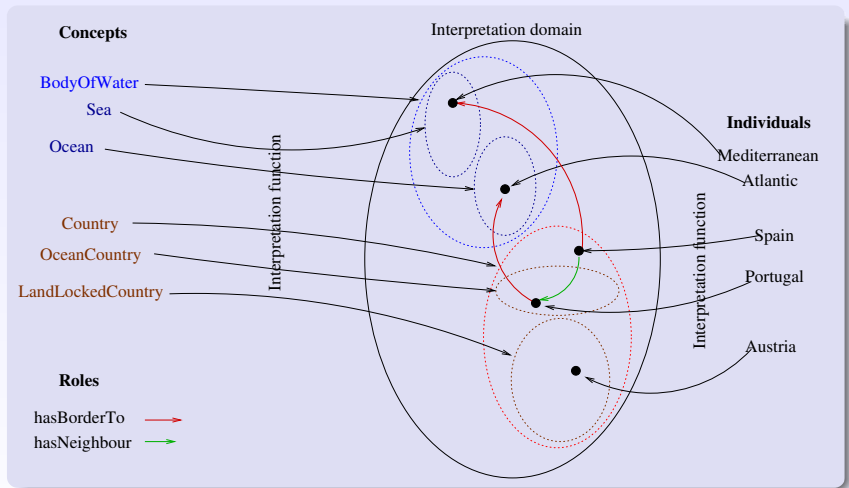
- $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ *(concepts interpreted as sets)*
- $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ *(roles interpreted as binary relations)*

Complex concept descriptions:

- $(\neg C)^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$, $(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$, $(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$
- $(\exists r.C)^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \text{there is } e \in \Delta^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}} \text{ and } e \in C^{\mathcal{I}}\}$
- $(\forall r.C)^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \text{for all } e \in \Delta^{\mathcal{I}} : (d, e) \in r^{\mathcal{I}} \text{ implies } e \in C^{\mathcal{I}}\}$

\mathcal{I} is a **model** of C if $C^{\mathcal{I}} \neq \emptyset$

Example of an interpretation



$$(\text{Country} \sqcap \exists \text{hasNeighbour} . (\exists \text{hasBorderTo} . \text{Ocean}))^{\mathcal{I}} = \{\text{Spain}\}$$

- \mathcal{ALC} is a fragment of FOL
- Concept names are **unary predicates**, role names are **binary predicates**
- Concept descriptions yield formulae with one free variable
- \mathcal{ALC} concept descriptions can be written in the two-variable fragment (\mathcal{L}^2) of FOL

$$\begin{aligned}\forall r.(\exists r.A) &\Rightarrow \forall y.(r(x,y) \rightarrow (\exists z.(r(y,z) \wedge A(z)))) \\ &\Rightarrow \forall y.(r(x,y) \rightarrow (\exists x.(r(y,x) \wedge A(x))))\end{aligned}$$

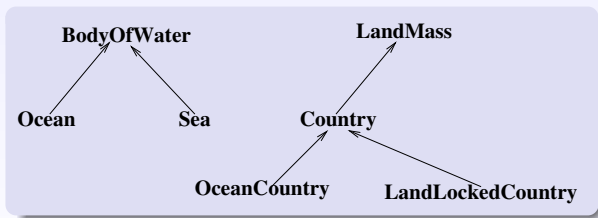
- \mathcal{L}^2 is **decidable** [Mortimer(1975)]

Reasoning tasks

Two main reasoning tasks:

- **concept satisfiability**: Is there a model of C ?
- **concept subsumption**: Does $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ hold for all \mathcal{I} ?
(written $C \sqsubseteq D$?)

Concept subsumption for computing the subsumption hierarchy:



For \mathcal{ALC} , satisfiability and subsumption are mutually reducible:
 $C \sqsubseteq D$ iff $C \sqcap \neg D$ is unsatisfiable

DL Knowledge Bases (Ontologies)

DL Knowledge Base (Ontology) = TBox + ABox

- TBox (\mathcal{T}) defines the terminology of the application domain
- ABox (\mathcal{A}) states facts about a specific “world”

TBox: set of
concept definitions

$\text{LockedCountry} \equiv \text{Country} \sqcap \forall \text{hasBorderTo}.\text{LandMass}$
 $\text{OceanCountry} \equiv \text{Country} \sqcap \exists \text{hasBorderTo}.\text{Ocean}$

ABox: set of concept and
role assertions

$\text{Ocean}(\textit{Atlantic}), \text{Sea}(\textit{Mediterranean}),$
 $\text{hasBorderTo}(\textit{Portugal}, \textit{Atlantic}),$

General TBox: set of
GCIs (general concept
inclusion axiom)

$\text{Ocean} \sqsubseteq \text{BodyOfWater}$
 $\exists \text{hasBorderTo}.\text{BodyOfWater} \sqsubseteq \text{CoastalCountry}$

An interpretation \mathcal{I} is a **model** of

- a TBox \mathcal{T} if it satisfies all its **concept definitions**: $A^{\mathcal{I}} = C^{\mathcal{I}}$ for all $A \equiv C \in \mathcal{T}$
- a **general** TBox \mathcal{T} if it satisfies all its **concept inclusions**: $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all $C \sqsubseteq D \in \mathcal{T}$
- an ABox \mathcal{A} if it satisfies all its **assertions**:
 - $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all $C(a) \in \mathcal{A}$
 - $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all $r(a, b) \in \mathcal{A}$
- a knowledge base $(\mathcal{T}, \mathcal{A})$ if it is a common model of \mathcal{T} and \mathcal{A} .

\mathcal{T} a TBox, \mathcal{A} an ABox, C and D concept descriptions, a an individual name.

TBox Reasoning

- **Satisfiability**: Is there a common model of C and \mathcal{T} ?
- **Subsumption**: Does $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ hold in all models \mathcal{I} of \mathcal{T} ?
(written $C \sqsubseteq_{\mathcal{T}} D$)

ABox Reasoning

- **Consistency**: Is there a common model of \mathcal{A} and \mathcal{T} ?
- **Instance**: Does $a^{\mathcal{I}} \in C^{\mathcal{I}}$ hold in all models \mathcal{I} of \mathcal{A} and \mathcal{T} ?
(written $\mathcal{T}, \mathcal{A} \models C(a)$)

Satisfiability \Leftrightarrow *Subsumption*
Consistency \Leftrightarrow *Instance*

Satisfiability \Rightarrow *Consistency*

Tableau Algorithm:

- decision procedure for **checking ABox consistency**
- tries to generate a **finite model** for the input ABox
- extends the ABox by applying **tableau rules**
- checks for **contradictions**
- if **no rule applies** and **no contradiction** is found, the ABox is **consistent**

Complexity of Reasoning in \mathcal{ALC}

- with empty TBox PSPACE-complete [Schmidt-Schauß & Smolka(1991)]
 - in PSPACE: the tableau algorithm
 - PSPACE-hard: reduction from satisfiability in QBF (quantified boolean formulae)
- with general TBoxes: EXPTIME-complete
 - in EXPTIME: tableau algorithm in [Donini & Massacci(2000)]
 - EXPTIME-hard: [Schild(1991)]

PSPACE: solvable by a **deterministic** machine with **polynomial space**

EXPTIME: solvable by a **deterministic** machine in **exponential time**

$P \subseteq NP \subseteq PSPACE \subseteq EXPTIME$
Horn Boolean \mathcal{ALC} (empty TBox) \mathcal{ALC} (general TBox)

tradeoff between expressivity and computational complexity!

More expressive DLs \mathcal{ALCN} and \mathcal{ALCQ}

\mathcal{ALC} cannot express countries that have at most 3 neighbours

Unqualified number restrictions: Country $\sqcap (\leq 3 \text{ hasNeighbour})$

- at most: $(\leq n r)^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \text{card}(\{e \mid (d, e) \in r^{\mathcal{I}}\}) \leq n\}$
- at least: $(\geq n r)^{\mathcal{I}} = \dots$

\mathcal{ALCN} cannot express countries that have at least 1 neighbour that is an ocean country

Qualified nr. restrns: Country $\sqcap (\geq 1 \text{ hasNeighbour.OceanCountry})$

- at least:
 $(\geq n r.C)^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \text{card}(\{e \mid (d, e) \in r^{\mathcal{I}} \wedge e \in C^{\mathcal{I}}\}) \geq n\}$
- at most: $(\leq n r.C)^{\mathcal{I}} = \dots$

\mathcal{ALCQI} : \mathcal{ALCQ} + inverse roles

Inverse roles: $(r^-)^{\mathcal{I}} = \{(e, d) \mid (d, e) \in r^{\mathcal{I}}\}$

countries recognized by
an EU member

Country $\sqcap \exists \text{recognizes}^- . \text{EUmember}$

\mathcal{ALCQO} : \mathcal{ALCQ} + nominals

Nominals: $\{a_1, \dots, a_n\}^{\mathcal{I}} = \{a_1^{\mathcal{I}}, \dots, a_n^{\mathcal{I}}\}$

countries that have
territories in Europe
or Asia

Country $\sqcap \exists \text{hasTerritoryIn} . \{Asia, Europe\}$

Very expressive DLs *SHIF*, *SHOIN*

- transitive roles (R_+): hasGoodRelationsWith₊
- role hierarchy (\mathcal{H}): hasNeighbour \sqsubseteq hasBorderTo
- functional restriction (\mathcal{F}): hasCapitalCity
- Concrete domains (D): natural numbers, reals, time, ...
- S : *ALC* with transitive roles

OWL: the standard ontology language for **The Semantic Web!**

- **OWL Lite** = *SHIF* (D) (reasoning EXPTIME)
- **OWL DL** = *SHOIN* (D) (reasoning NEXPTIME)
- **OWL Full** = beyond DLs (reasoning undecidable!)

($\text{EXPTIME} \subseteq \text{NEXPTIME}$: solvable in exp. time by a nond. mach.)

- very high worst-case complexity: EXPTIME , NEXPTIME
- almost 20 years of research
- highly optimized tableau algorithms
- efficient implementations
- Reasoners: **FaCT++**, **RACERPRO**, **Pellet**, **KAON2**, **Hermit** ...
- can classify large real world ontologies written in very expressive DLs!

- DLs with **polynomial** time reasoning!
- still **expressive enough** for widely used **bio-medial ontologies**
 - **SNOMED** (Systematized Nomenclature of Medicine): medical terminology for diseases, treatments, ...
 - **Gene Ontology**: controlled vocabulary to describe gene and gene product attributes
- \mathcal{EL}^+ constructors:
 - \top (top)
 - $C \sqcap D$ (conjunction)
 - $\exists r.C$ (existential restriction)
 - $C \sqsubseteq D$ (GCI)
 - $r_1 \circ \dots \circ r_n \sqsubseteq s$ (role inclusion)

- SNOMED contains more than 350.000 axioms
- ontology development is an error prone task
- in SNOMED Amputation-of-Finger is classified as a subconcept of Amputation-of-Arm
- obviously a modelling error
- what is the reason, which axioms out of 350.000 are responsible?
- **Axiom pinpointing**: finding **small** subsets of the knowledge base that have given consequence
- **Minimal explanation** of $C \sqsubseteq D$ in \mathcal{T} is a $\mathcal{T}' \subseteq \mathcal{T}$ s.t. $\mathcal{T}' \models C \sqsubseteq D$ and \mathcal{T}' minimal

Complexity of Axiom pinpointing in \mathcal{EL}^+

- finding a minimum cardinality explanation is NP-complete
[Baader *et al.*(2007)Baader, Peñaloza, & Suntisrivaraporn]
- finding one minimal explanation is polynomial: remove axioms one by one and test
- what about **finding all minimal explanations**?
- there can be **exponentially many** of them!
- clearly **not** computable in polynomial time

Enumeration complexity

- **polynomial delay**: polynomial time **between each consecutive solution**
- **output-polynomial**: polynomial in the size of the input **and the total number of solutions**

Complexity of Axiom pinpointing in \mathcal{EL}^+

- finding a minimum cardinality explanation is NP-complete
[Baader *et al.*(2007)Baader, Peñaloza, & Suntisrivaraporn]
- finding one minimal explanation is polynomial: remove axioms one by one and test
- what about **finding all minimal explanations**?
- there can be **exponentially many** of them!
- clearly **not** computable in polynomial time

Enumeration complexity

- **polynomial delay**: polynomial time **between each consecutive solution**
- **output-polynomial**: polynomial in the size of the input **and the total number of solutions**

Complexity of Axiom pinpointing in \mathcal{EL}^+

We have considered the following:

Can we enumerate min. exps. in a given order with poly. delay?

- **NO!** cannot be done with polynomial delay unless $P = NP$
- checking whether a given minimal explanation is the first one is **coNP-complete**

Can we compute all min. exps. in output-polynomial time?

- We do **not** know :(
- the corresponding decision problem is in coNP
- it is **TRANS-HYP-hard** (*hypergraph problem open for 20 years*)

Complexity of Axiom pinpointing in \mathcal{EL}^+

We have considered the following:

Can we enumerate min. exps. in a given order with poly. delay?

- **NO!** cannot be done with polynomial delay unless $P = NP$
- checking whether a given minimal explanation is the first one is **coNP-complete**

Can we compute all min. exps. in output-polynomial time?

- We do **not** know :(
- the corresponding decision problem is in coNP
- it is **TRANS-HYP-hard** (*hypergraph problem open for 20 years*)

We have considered the following:

Can we enumerate min. exps. in a given order with poly. delay?

- **NO!** cannot be done with polynomial delay unless $P = NP$
- checking whether a given minimal explanation is the first one is **coNP-complete**

Can we compute all min. exps. in output-polynomial time?

- We do **not** know :(
- the corresponding decision problem is in coNP
- it is **TRANS-HYP-hard** (*hypergraph problem open for 20 years*)

We have considered the following:

Can we enumerate min. exps. in a given order with poly. delay?

- **NO!** cannot be done with polynomial delay unless $P = NP$
- checking whether a given minimal explanation is the first one is **coNP-complete**

Can we compute all min. exps. in output-polynomial time?

- We do **not** know :(
- the corresponding decision problem is in coNP
- it is **TRANS-HYP-hard** (*hypergraph problem open for 20 years*)

We have considered the following:

- Is there a minimal explanation that contains a given axiom?
 - NP-complete
- Is there a min. exp. that does not contain any of the given subsets of the original KB?
 - NP-complete
- Determine the number of minimal explanations
 - #P-complete

- DLs are **logic-based** KR formalisms
- **decidable** fragments of FOL
- theoretical background of **OWL**: the W3C standard ontology language for **The Semantic Web**
- Lightweight DLs, **polynomial** time reasoning
- **Axiom pinpointing**: finding explanations



Baader, F., Peñaloza, R., & Suntisrivaraporn, B. (2007).

Pinpointing in the description logic \mathcal{EL}^+ .

In J. Hertzberg, M. Beetz, & R. Englert (Eds.), *Proceedings of the 30th German Conference on Artificial Intelligence (KI2007)*. Springer-Verlag, vol. 4667 of *Lecture Notes in Artificial Intelligence*, 52–67.



Brachman, R. J. & Levesque, H. J. (Eds.) (1985).

Readings in Knowledge Representation.

Los Altos (CA), USA: Morgan-Kaufmann.



Donini, F. M. & Massacci, F. (2000).

EXPTIME tableaux for \mathcal{ALC} .

Artificial Intelligence, 124(1), 87–138.



Mortimer, M. (1975).

On languages with two variables.

Zeitschrift für mathematische Logik und Grundlagen der Mathematik, 21, 135–140.



Quillian, M. R. (1967).

Word concepts: A theory and simulation of some basic capabilities.

Behavioral Science, 12, 410–430.

Republished in [Brachman & Levesque(1985)].



Schild, K. (1991).

A correspondence theory for terminological logics: preliminary report.

In J. Mylopoulos & R. Reiter (Eds.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI'91)*. Morgan Kaufmann, 466–471.



Schmidt-Schauß, M. & Smolka, G. (1991).

Attributive concept descriptions with complements.

Artificial Intelligence, 48(1), 1–26.