Description Logics: a Nice Family of Logics — Reasoning in the Lightweight DL $\mathcal{EL}$ —

Uli Sattler[1] *Thomas Schneider*[2]

[1]School of Computer Science, University of Manchester, UK

[2]Department of Computer Science, University of Bremen, Germany

ESSLLI, 9 August 2012

Plan for today

1. What is $\mathcal{EL}$?
2. Normalisation
3. A simple poly-time reasoning algorithm

And now . . .

1. What is $\mathcal{EL}$?
2. Normalisation
3. A simple poly-time reasoning algorithm

Summary

$\mathcal{EL}$ is a restriction of $\mathcal{ALC}$ ...

- that allows only conjunction and existential restrictions
- that is used to represent, e.g., medical knowledge
- whose standard reasoning problems are tractable

  i.e., there is a worst-case poly-time algorithm for deciding subsumption etc.

- whose extension $\mathcal{EL}^{++}$ with other features, namely:

$\bot$	domain and range restrictions
disjoint concepts	concept and role assertions
role (chain) inclusions	nominals
transitive roles	concrete domains
reflexive roles	

  remains tractable and is a profile of OWL

Syntax and semantics of $\mathcal{EL}$

Concepts

For C, D concepts and R a role name:

Constructor	Syntax	Example	Semantics
top	$\top$		$\Delta^{\mathcal{I}}$
conjunction	$C \sqcap D$	$\mathtt{Human} \sqcap \mathtt{Male}$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
exist. restr.	$\exists r.C$	$\exists\mathtt{hasChild.Human}$	$\{x \mid \exists y.(x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$

Axioms

- $C \sqsubseteq D$
- $C \equiv D$ as a shortcut for "$C \sqsubseteq D$, $D \sqsubseteq C$"

What a tiny logic!?

✓ We **can say** in $\mathcal{EL}$

$\mathsf{Hand} \sqsubseteq \exists \mathsf{hasPart}.\mathsf{Finger}$

✘ but we **can't say**

$\mathsf{Hand} \sqsubseteq {=}5\, \mathsf{hasPart}.\mathsf{Finger}$

$\mathsf{Finger} \sqsubseteq \exists \mathsf{hasPart}^{-}.\mathsf{Hand}$

✘ We'd like to say, but **can't**

$\mathsf{MildFlu} \equiv \mathsf{Flu} \sqcap \forall \mathsf{symptom}.\mathsf{Triv}$

✓ all we **can say** (in $\mathcal{EL}^{++}$) is

$\mathsf{MildFlu} \sqsubseteq \mathsf{Flu}$

$\mathsf{MildFlu} \sqcap \exists \mathsf{symptom}.\mathsf{Fever} \sqsubseteq \bot$

$\mathsf{Fever} \sqcap \mathsf{Triv} \sqsubseteq \bot$

$\mathcal{EL}^{++}$ **is used** in some large-scale ontologies

e.g. bio-medical domain, terminologies:
SNOMED, GALEN, GO (see References)

$\mathcal{EL}^{(+)}$ is not so tiny – an example ontology

$$
\begin{array}{rcl}
\mathsf{Endocardium} & \sqsubseteq & \mathsf{Tissue} \sqcap \exists \mathsf{cont\text{-}in}.\mathsf{HeartWall} \sqcap \\
 & & \qquad \exists \mathsf{cont\text{-}in}.\mathsf{HeartValve} \\
\mathsf{HeartWall} & \sqsubseteq & \mathsf{BodyWall} \sqcap \exists \mathsf{part\text{-}of}.\mathsf{Heart} \\
\mathsf{HeartValve} & \sqsubseteq & \mathsf{BodyValve} \sqcap \exists \mathsf{part\text{-}of}.\mathsf{Heart} \\
\mathsf{Endocarditis} & \sqsubseteq & \mathsf{Inflammation} \sqcap \exists \mathsf{has\text{-}loc}.\mathsf{Endocardium} \\
\mathsf{Inflammation} & \sqsubseteq & \mathsf{Disease} \sqcap \exists \mathsf{acts\text{-}on}.\mathsf{Tissue} \\
\mathsf{Heartdisease} \sqcap \exists \mathsf{has\text{-}loc}.\mathsf{HeartValve} & \sqsubseteq & \mathsf{CriticalDisease} \\
\mathsf{Heartdisease} & \equiv & \mathsf{Disease} \sqcap \exists \mathsf{has\text{-}loc}.\mathsf{Heart}
\end{array}
$$

$$
\mathcal{EL}^{+} \left\{
\begin{array}{rcl}
\mathsf{part\text{-}of} \circ \mathsf{part\text{-}of} & \sqsubseteq & \mathsf{part\text{-}of} \\
\mathsf{part\text{-}of} & \sqsubseteq & \mathsf{cont\text{-}in} \\
\mathsf{has\text{-}loc} \circ \mathsf{cont\text{-}in} & \sqsubseteq & \mathsf{has\text{-}loc}
\end{array}
\right.
$$

Taken from [Baader et al. 2006]

Satisfiability and subsumption

Satisfiability + coherence are trivial: every $\mathcal{EL}$-TBox is coherent

because ?

Satisfiability and subsumption

Satisfiability + coherence are trivial: every $\mathcal{EL}$-TBox is coherent

- $\mathcal{I}$ with $A^\mathcal{I} = \Delta^\mathcal{I}$ and $r^\mathcal{I} = \Delta^\mathcal{I} \times \Delta^\mathcal{I}$,
  for all concept names A and role names r,
  satisfies every $\mathcal{EL}$ axiom
- ($\mathcal{I}$ with $A^\mathcal{I} = r^\mathcal{I} = \emptyset$ doesn't – **why?**)

Subsumption ?

Satisfiability and subsumption

Satisfiability + coherence are trivial: every $\mathcal{EL}$-TBox is coherent

- $\mathcal{I}$ with $A^\mathcal{I} = \Delta^\mathcal{I}$ and $r^\mathcal{I} = \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, for all concept names A and role names r, satisfies every $\mathcal{EL}$ axiom
- ($\mathcal{I}$ with $A^\mathcal{I} = r^\mathcal{I} = \emptyset$ doesn't – **why?**)

Subsumption isn't:
does the following TBox entail $A \sqsubseteq B$? $A' \sqsubseteq B'$?

$$\begin{aligned} \exists r.A &\sqsubseteq \exists r.B \\ A' &\equiv \exists r.\exists r.A \\ B' &\equiv \exists r.\exists r.B \end{aligned}$$

(Without negation, they are no longer interreducible.)

Roadmap

Goal: present a decision procedure for subsumption in $\mathcal{EL}$

Outline:

- Normalisation procedure
- Decision procedure

  (simple, naïve, without optimisations)

And now . . .

Normal form

...keeps the reasoning procedure simple

Definition

An $\mathcal{EL}$ ontology is in **normal form** if all axioms have these forms:

$$A_1 \sqcap A_2 \sqsubseteq B$$
$$A \sqsubseteq B$$
$$A \sqsubseteq \exists r.B$$
$$\exists r.A \sqsubseteq B$$

$A_{(i)}, B$: **atomic** concepts or $\top$ $\qquad$ r: role

The normalisation procedure

- . . . applies **normalisation rules** to axioms in a given TBox $\mathcal{T}$
- each rule transforms an axiom into one or several shorter ones
- old axiom is removed from $\mathcal{T}$; new axioms are added

The normalisation rules

NF1	Input	$C \equiv D$	
	Output	$C \sqsubseteq D$	$D \sqsubseteq C$
NF2	Input	$\mathbf{C} \sqsubseteq \mathbf{D}$	
	Output	$\mathbf{C} \sqsubseteq A$	$A \sqsubseteq \mathbf{D}$
NF3	Input	$\exists r.\mathbf{C} \sqsubseteq D$	
	Output	$\mathbf{C} \sqsubseteq A$	$\exists r.A \sqsubseteq D$
NF4	Input	$\mathbf{C} \sqcap D \sqsubseteq E$	
	Output	$\mathbf{C} \sqsubseteq A$	$A \sqcap D \sqsubseteq E$
NF5	Input	$B \sqsubseteq \exists r.\mathbf{C}$	
	Output	$B \sqsubseteq \exists r.A$	$A \sqsubseteq \mathbf{C}$
NF6	Input	$B \sqsubseteq C \sqcap D$	
	Output	$B \sqsubseteq C$	$B \sqsubseteq D$

C D E	arbitrary concepts
$\mathbf{C}$ $\mathbf{D}$	complex concepts
B	atomic concept
A	fresh atomic concept

The normalisation procedure

Given TBox $\mathcal{T}$, apply **NF1–NF7** axiom-wise until none can be applied

The result $\mathcal{T}'$

- contains new atomic concepts $A_1, \ldots, A_k$
- is of size linear in the size of $\mathcal{T}$

Lemma

- For every model $\mathcal{I} \models \mathcal{T}$, there is a model $\mathcal{J} \models \mathcal{T}'$ such that $X^{\mathcal{J}} = X^{\mathcal{I}}$ for all $X \notin \{A_1, \ldots, A_k\}$.
- For every model $\mathcal{I} \models \mathcal{T}'$, it holds that $\mathcal{I} \models \mathcal{T}$.

Consequence: $\mathcal{T}'$ is equivalent to $\mathcal{T}$ w.r.t. subsumption:

$$\mathcal{T} \models C \sqsubseteq D \text{ iff } \mathcal{T}' \models C \sqsubseteq D$$

for all C, D that don't use the A_i

Details and Example: see [Suntisrivaraporn 2005, pg. 37–39]

And now . . .

Initial assumptions

Input: TBox $\mathcal{T}$, **atomic** concepts A, B

Question: does $\mathcal{T} \models A \sqsubseteq B$ hold?

Assumption of A, B being atomic is no real restriction:

$$\begin{array}{rcl} \mathcal{T} & \models & C \sqsubseteq D \\ & \Updownarrow & \\ \mathcal{T} \cup \{A \equiv C,\ B \equiv D\} & \models & A \sqsubseteq B \end{array}$$

Shorter notation: $A \sqsubseteq_{\mathcal{T}} B$ abbreviates $\mathcal{T} \models A \sqsubseteq B$

Deciding subsumptions via subsumer sets

Subsumer of A: a concept name B (including $\top$) with $A \sqsubseteq_{\mathcal{T}} B$

Subsumer set $S(A)$: set that contains subsumers of A

Representation of subsumer sets: in a labelled graph $G(\mathcal{T})$

- Nodes of $G(\mathcal{T})$ = concept names (including $\top$) in $\mathcal{T}$
- Label of node A: $S(A)$

  B in label $S(A)$ means $A \sqsubseteq_{\mathcal{T}} B$
- Label of edge (A, B): set $R(A, B)$ of roles

  $r \in R(A, B)$ means $A \sqsubseteq_{\mathcal{T}} \exists r.B$

Outline of the procedure:

1. Set $S(A) = \{A, \top\}$ for every A
2. Monotonically build $G(\mathcal{T})$
   by exhaustively applying completion rules
3. Check whether $B \in S(A)$

The completion rules (1)

Completion rule R1 for node X

If $A_1 \sqcap A_2 \sqsubseteq B$ in $\mathcal{T}$
and $\{A_1, A_2\} \subseteq S(X)$ and $B \notin S(X)$

then $S(X) := S(X) \cup \{B\}$

Completion rule R2 for node X

If $A \sqsubseteq \exists r.B$ in $\mathcal{T}$
and $A \in S(X)$ and $r \notin R(X, B)$

then $R(X, B) := R(X, B) \cup \{r\}$

The completion rules (2)

Completion rule R3 for nodes X, Y

If	$\exists r.A \sqsubseteq B$ in $\mathcal{T}$ and $r \in R(X, Y)$ and $A \in S(Y)$ and $B \notin S(X)$
then	$S(X) := S(X) \cup \{B\}$

The "naïve" subsumption algorithm [Baader et al. 2006]

Algorithm 1

Input: $\mathcal{EL}$ ontology $\mathcal{T}$, atomic classes A, B
Output: *yes* if $A \sqsubseteq_{\mathcal{T}} B$, *no* otherwise

---

For each atomic concept A in $\mathcal{T}$ plus $\top$ {
    create a node with label $\{A, \top\}$
}

while some rule is applicable to some axiom in $\mathcal{T}$ {
    choose axiom α and rule **R***i* applicable to α
    apply **R***i* to α
}

if $B \in S(A)$ **then output** *yes*
else output *no*

Summary

Algorithm 1 ...

- terminates in time polynomial in the size of $\mathcal{T}$

Corollary

Subsumption in $\mathcal{EL}$ can be decided in polynomial time.

- constructs a **canonical model** of $\mathcal{T}$
- is **sound** and **complete**: outputs *yes* iff $A \sqsubseteq_{\mathcal{T}} B$
- works "one-pass": computes all $A \sqsubseteq_{\mathcal{T}} B$ at once
- is still slow for big ontologies:
  crux = search for applicable rules

Extensions

Smarter versions of Algorithm 1 ...

- are goal-oriented:
  only apply rules that are necessary for (dis)proving $A \sqsubseteq_{\mathcal{T}} B$
- are implemented in the reasoner CEL for the extension $\mathcal{EL}^{++}$
- can be extended even to the Horn fragment of $\mathcal{SHIQ}$

For details see [Baader et al. 2005, Baader et al. 2006, Kazakov 2009].

Homework

You're cordially invited to
apply the normalisation procedure to the TBox

$$\mathcal{T} = \{ \begin{array}{rcl} A & \sqsubseteq & B \sqcap \exists r.C\,, \\ C & \sqsubseteq & \exists s.D\,, \\ \exists r.\exists s.\top \sqcap B & \sqsubseteq & D \end{array} \}$$

and then check whether it entails

$$A \sqsubseteq D.$$

That's all for today. Thanks!

References: links

Bio-medical ontologies

- SNOMED, the systematized nomenclature of human and veterinary medicine
  http://en.wikipedia.org/wiki/SNOMED_CT
- GALEN http://www.opengalen.org
- GO, the Gene Ontology http://www.geneontology.org

References: articles (1)

- F. Baader.
  Terminological cycles in a description logic with existential restrictions.
  In *Proc. IJCAI*, pages 325–330, 2003.
  http://lat.inf.tu-dresden.de/research/papers.html#2003

- F. Baader, S. Brandt, and C. Lutz.
  Pushing the $\mathcal{EL}$ envelope.
  In *Proc. IJCAI*, pages 364–369, 2005.
  http://www.ijcai.org/papers/0372.pdf

- F. Baader, C. Lutz, and B. Suntisrivaraporn.
  Efficient reasoning in $\mathcal{EL}^+$.
  In *Description Logics*, volume 189 of *CEUR Workshop Proc.*, 2006.
  http://www.ceur-ws.org/Vol-189/submission_8.pdf

References: articles (2)

S. Brandt.

Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and – what else?

In *Proc. ECAI*, pages 298–302, 2004.

http://www.cs.man.ac.uk/~sbrandt/papers.html

Y. Kazakov:

Consequence-Driven Reasoning for Horn $\mathcal{SHIQ}$ Ontologies.

In *Proc. IJCAI*, pages 2040–2045, 2009.

B. Suntisrivaraporn.

Optimization and Implementation of Subsumption Algorithms for the Description Logic $\mathcal{EL}$ with Cyclic TBoxes and General Concept Inclusion Axioms.

Masters thesis, Technische Universität Dresden, Germany, 2005.

http://lat.inf.tu-dresden.de/research/papers.html#2005