Pinpointing in the Description Logic $\mathcal{EL}^+$

Lorenz Leutgeb
lorenz@leutgeb.xyz

2017-01-18

International Center for Computational Logic, TU Dresden

This work is licensed under Creative Commons BY-SA 4.0.

Paper available for download at:

`https://lorenz.leutgeb.xyz/paper/elp.pdf`

Based on:

Pinpointing in the Description Logic $\mathcal{EL}^+$

Baader, Peñaloza, and Suntisrivaraporn

30th Annual German Conference on AI, 2007

(detailed reference in the end)

Agenda

Introduction

Introduction

Syntax and Semantics

Syntax and Semantics [3, Sec. 2, Tbl. 1]

Name	Syntax	Semantics	$\mathcal{HL}$	$\mathcal{EL}$	$\mathcal{EL}^+$
Top	$\top$	Δ^I	•	•	•
Conjunction	$C \sqcap D$	$C^I \cap D^I$	•	•	•
Existential Restr.	$\exists r.C$	$*$		•	•
GCI[1]	$C \sqsubseteq D$	$C^I \subseteq D^I$	•	•	•
Concept Definition	$C \equiv D$	$C^I = D^I$	•	•	•
Role Inclusion	$r_1 \circ \cdots \circ r_n \sqsubseteq s$	$r_1^I \circ \cdots \circ r_n^I \subseteq s^I$			•

$*$: $\{x \in \Delta^I \mid \text{there exists } y \in \Delta^I \text{ s.t. } (x, y) \in r^I \text{ and } y \in C^I\}$

- Concept Descriptions C, D (inductively)
- Role Names $r_1, \ldots, r_n, s$
- Classical Interpretation $I = (\Delta^I, \cdot^I)$

[1]General Concept Inclusion

Example: $\mathcal{HL}$ and Horn Logic Programming (cf. [3, Sec. 2])

$\mathsf{woman} \sqsubseteq \mathsf{person}$	`person :- woman.`
$\mathsf{man} \sqsubseteq \mathsf{person}$	`person :- man.`
$\mathsf{parent} \sqcap \mathsf{woman} \sqsubseteq \mathsf{mother}$	`mother :- parent, woman.`

Introduction

TBoxes and Concept Subsumption

- We consider knowledge bases that are *finite sets of axioms*, called TBoxes, denoted $\mathcal{T}$.
- Key questions wrt. TBoxes are satisfiability and concept subsumption.

Definition (Concept Subsumption, cf. [4, Def. 1])

Given two concept descriptions C, D and a TBox $\mathcal{T}$, C is subsumed by D wrt. $\mathcal{T}$ (written $C \sqsubseteq_{\mathcal{T}} D$) if for every interpretation I that satisfies $\mathcal{T}$ we have $C^I \subseteq D^I$.

Introduction

Pinpointing

Motivation for Pinpointing

Given a TBox ...

$$\mathsf{Human} \sqsubseteq \exists \mathsf{parent}.\mathsf{Human} \qquad \text{Humans have a human parent.} \quad (a_1)$$

$$\mathsf{Human} \sqsubseteq \mathsf{Monkey} \qquad \text{Humans are monkeys.} \quad (a_2)$$

$$\exists \mathsf{parent}.\mathsf{Monkey} \sqsubseteq \mathsf{Animal} \qquad \text{Monkey parent? It's an animal.} \quad (a_3)$$

$$\mathsf{Monkey} \sqsubseteq \mathsf{Animal} \qquad \text{Monkeys are animals.} \quad (a_4)$$

$$\mathsf{Fish} \sqsubseteq \mathsf{Animal} \qquad \text{Fish are animals.} \quad (a_5)$$

Motivation for Pinpointing

Given a TBox ...

$\mathsf{Human} \sqsubseteq \exists \mathsf{parent}.\mathsf{Human}$	Humans have a human parent.	(a_1)
$\mathsf{Human} \sqsubseteq \mathsf{Monkey}$	Humans are monkeys.	(a_2)
$\exists \mathsf{parent}.\mathsf{Monkey} \sqsubseteq \mathsf{Animal}$	Monkey parent? It's an animal.	(a_3)
$\mathsf{Monkey} \sqsubseteq \mathsf{Animal}$	Monkeys are animals.	(a_4)
$\mathsf{Fish} \sqsubseteq \mathsf{Animal}$	Fish are animals.	(a_5)

We observe that ...

$$\mathsf{Human} \sqsubseteq \mathsf{Animal}$$

Motivation for Pinpointing

Given a TBox ...

$$\mathsf{Human} \sqsubseteq \exists \mathsf{parent}.\mathsf{Human} \qquad \text{Humans have a human parent.} \quad (a_1)$$

$$\mathsf{Human} \sqsubseteq \mathsf{Monkey} \qquad \text{Humans are monkeys.} \quad (a_2)$$

$$\exists \mathsf{parent}.\mathsf{Monkey} \sqsubseteq \mathsf{Animal} \qquad \text{Monkey parent? It's an animal.} \quad (a_3)$$

$$\mathsf{Monkey} \sqsubseteq \mathsf{Animal} \qquad \text{Monkeys are animals.} \quad (a_4)$$

$$\mathsf{Fish} \sqsubseteq \mathsf{Animal} \qquad \text{Fish are animals.} \quad (a_5)$$

We observe that ...

$$\mathsf{Human} \sqsubseteq \mathsf{Animal}$$

And ask ourselves: Why?

Formalization of our question yields Pinpointing, a process that results in *minimal axiom sets*:

Definition (MinA, cf. [4, Def. 2 without partitioning $\mathcal{T}$])

Let $\mathcal{T}$ be a TBox and A, B concept names occurring in it such that $A \sqsubseteq_{\mathcal{T}} B$. Then a minimal axiom set (MinA) for $\mathcal{T}$ w.r.t. $A \sqsubseteq B$ is a subset $\mathcal{S} \subseteq \mathcal{T}$ such that

$$A \sqsubseteq_{\mathcal{S}} B$$

and for all $\mathcal{S}' \subset S$ we have

$$A \not\sqsubseteq_{\mathcal{S}'} B$$

Motivation for Pinpointing

$$\begin{aligned}
\mathsf{Human} &\sqsubseteq \exists \mathsf{parent.Human} && \text{Humans have a human parent.} \quad (a_1)\\
\mathsf{Human} &\sqsubseteq \mathsf{Monkey} && \text{Humans are monkeys.} \quad (a_2)\\
\exists \mathsf{parent.Monkey} &\sqsubseteq \mathsf{Animal} && \text{Monkey parent? It's an animal.} \quad (a_3)\\
\mathsf{Monkey} &\sqsubseteq \mathsf{Animal} && \text{Monkeys are animals.} \quad (a_4)\\
\mathsf{Fish} &\sqsubseteq \mathsf{Animal} && \text{Fish are animals.} \quad (a_5)
\end{aligned}$$

$$\mathsf{Human} \sqsubseteq \mathsf{Animal}$$

Minimal Axiom Sets (MinAs): $\{a_2, a_4\}, \{a_1, a_2, a_3\}$

Example: SNOMED CT [8, Fig. 6.8, p. 128]

$$
\begin{aligned}
\text{direct-procedure-site} &\sqsubseteq \text{procedure-site} \\
\mathit{AmputationOfFinger} &\sqsubseteq \mathit{AmputationOfFingerNotThumb} \\
\text{AmputationOfFingerNotThumb} &\equiv \text{HandExcision} \sqcap \\
&\quad \exists \mathit{roleGroup}.(\\
&\qquad \exists \text{direct-procedure-site}.\text{Finger}_S \sqcap \\
&\qquad \exists \text{method}.\text{Amputation}) \\
\text{AmputationOfHand} &\equiv \text{HandExcision} \sqcap \\
&\quad \exists \mathit{roleGroup}.(\\
&\qquad \exists \text{direct-procedure-site}.\text{Finger}_S \sqcap \\
&\qquad \exists \text{method}.\text{Amputation}) \\
\mathit{Finger}_S &\sqsubseteq \mathit{DigitOfHand}_S \sqcap \mathit{Hand}_P \\
\mathit{Hand}_P &\sqsubseteq \mathit{Hand}_S \sqcap \mathit{UpperExtremity}_P
\end{aligned}
$$

Algorithms

White-Box vs. Black-Box Algorithms

White-Box: Inspects syntax of axioms, more “low-level”.

Black-Box: Relies on reasoning services (subsumption) only.

Algorithms

Pinpointing via Labeling

Subsumption Algorithm for $\mathcal{EL}^+$

Based on following completion rules[2] wrt. a TBox $\mathcal{T}$. Rule i is applicable if $a_i \in \mathcal{T}$ and $P_i \subseteq \mathcal{T}' \setminus \mathcal{T}$. If rule i is applied, then q_i is added to $\mathcal{T}'$.

	Applicability				**Result**
i	a_i (axiom)			P_i (set of $\mathcal{T}$-seq)	q_i ($\mathcal{T}$-seq)
1	$A_1 \sqcap \cdots \sqcap A_n$	$\sqsubseteq$	B	$X \sqsubseteq A_1, \ldots, X \sqsubseteq A_n$	$X \sqsubseteq B$
2	A	$\sqsubseteq$	$\exists r.B$	$X \sqsubseteq A$	$X \sqsubseteq \exists r.B$
3	$\exists r.A$	$\sqsubseteq$	B	$X \sqsubseteq \exists r.Y, Y \sqsubseteq A$	$X \sqsubseteq B$
4	r	$\sqsubseteq$	s	$X \sqsubseteq \exists r.Y$	$X \sqsubseteq \exists s.Y$
5	$r \circ r'$	$\sqsubseteq$	s	$X \sqsubseteq \exists r.Y, Y \sqsubseteq \exists r'.Z$	$X \sqsubseteq \exists s.Z$

[2] adapted from [4, 3, Fig. 1], see also [8, Fig. 5.2, p. 104]

Algorithm 1: $\textsc{Subsumption}(\mathcal{T}, A, B)$

Input: An $\mathcal{EL}^+$ TBox $\mathcal{T}$ in normal form over $\mathsf{N_C}$ and $A, B \in \mathsf{N_C}$.

Output: "yes" if $A \sqsubseteq_{\mathcal{T}} B$ holds, "no" otherwise.

$\mathcal{T}' := \{A \sqsubseteq A, A \sqsubseteq \top \mid A \in \mathsf{N_C}\}$

while *there is a rule i s.t.* $1 \leq i \leq 5$, $a_i \in \mathcal{T}$ *and* $P_i \subseteq \mathcal{T}'$ **do**

    $\mathcal{T}' := \mathcal{T}' \cup \{q_i\}$

end

return "yes" **if** $A \sqsubseteq B \in \mathcal{T}'$ **otherwise** "no"

Subsumption Algorithm for $\mathcal{EL}^+$

The algorithm ...

- requires a normalized TBox. Normalization is always possible and can be computed in linear time[3].

[3] proof for $\mathcal{EL}$ in [2, Lemma 6.2] and $\mathcal{EL}^{++}$ in [1, Lemma 1]
[4] [3, Thm. 1]

The algorithm ...

- requires a normalized TBox. Normalization is always possible and can be computed in linear time[3].
- restricts to concept names. We may introduce new concept names $A \sqsubseteq C, D \sqsubseteq B$ for any concepts C, D.

[3] proof for $\mathcal{EL}$ in [2, Lemma 6.2] and $\mathcal{EL}^{++}$ in [1, Lemma 1]
[4] [3, Thm. 1]

The algorithm ...

- requires a normalized TBox. Normalization is always possible and can be computed in linear time[3].
- restricts to concept names. We may introduce new concept names $A \sqsubseteq C, D \sqsubseteq B$ for any concepts C, D.
- is correct[4]: $A \sqsubseteq_{\mathcal{T}} B$ iff $A \sqsubseteq B \in \mathcal{T}'$

[3] proof for $\mathcal{EL}$ in [2, Lemma 6.2] and $\mathcal{EL}^{++}$ in [1, Lemma 1]
[4] [3, Thm. 1]

Subsumption Algorithm for $\mathcal{EL}^+$

The algorithm ...

- requires a normalized TBox. Normalization is always possible and can be computed in linear time[3].
- restricts to concept names. We may introduce new concept names $A \sqsubseteq C, D \sqsubseteq B$ for any concepts C, D.
- is correct[4]: $A \sqsubseteq_{\mathcal{T}} B$ iff $A \sqsubseteq B \in \mathcal{T}'$
- runs in time polynomial in the size of the input TBox[4].

[3] proof for $\mathcal{EL}$ in [2, Lemma 6.2] and $\mathcal{EL}^{++}$ in [1, Lemma 1]
[4] [3, Thm. 1]

The algorithm ...

- requires a normalized TBox. Normalization is always possible and can be computed in linear time[3].
- restricts to concept names. We may introduce new concept names $A \sqsubseteq C, D \sqsubseteq B$ for any concepts C, D.
- is correct[4]: $A \sqsubseteq_{\mathcal{T}} B$ iff $A \sqsubseteq B \in \mathcal{T}'$
- runs in time polynomial in the size of the input TBox[4].
- actually computes all concept subsumptions. This is easily extended to a *classification algorithm*.

[3] proof for $\mathcal{EL}$ in [2, Lemma 6.2] and $\mathcal{EL}^{++}$ in [1, Lemma 1]
[4] [3, Thm. 1]

Pinpointing Formula

$$\begin{aligned} \mathsf{Human} &\sqsubseteq \exists \mathsf{parent}.\mathsf{Human} && \text{Humans have a human parent.} \quad (a_1) \\ \mathsf{Human} &\sqsubseteq \mathsf{Monkey} && \text{Humans are monkeys.} \quad (a_2) \\ \exists \mathsf{parent}.\mathsf{Monkey} &\sqsubseteq \mathsf{Animal} && \text{Monkey parent? It's an animal.} \quad (a_3) \\ \mathsf{Monkey} &\sqsubseteq \mathsf{Animal} && \text{Monkeys are animals.} \quad (a_4) \\ \mathsf{Fish} &\sqsubseteq \mathsf{Animal} && \text{Fish are animals.} \quad (a_5) \end{aligned}$$

$$\mathsf{Human} \sqsubseteq \mathsf{Animal}$$

Minimal Axiom Sets (MinAs): $\{a_2, a_4\}, \{a_1, a_2, a_3\}$

Pinpointing Formula

$\text{Human} \sqsubseteq \exists \text{parent}.\text{Human}$	Humans have a human parent.	(a_1)
$\text{Human} \sqsubseteq \text{Monkey}$	Humans are monkeys.	(a_2)
$\exists \text{parent}.\text{Monkey} \sqsubseteq \text{Animal}$	Monkey parent? It's an animal.	(a_3)
$\text{Monkey} \sqsubseteq \text{Animal}$	Monkeys are animals.	(a_4)
$\text{Fish} \sqsubseteq \text{Animal}$	Fish are animals.	(a_5)

$$\text{Human} \sqsubseteq \text{Animal}$$

Pinpointing Formula: $a_2 \wedge (a_4 \vee (a_1 \wedge a_3))$

Pinpointing Formula

- Let $\mathit{lab}(\mathcal{T})$ be the set of labels of all axioms in $\mathcal{T}$.
- Let $\mathcal{V} \subseteq \mathit{lab}(\mathcal{T})$ be a valuation wrt. $\mathcal{T}$.
- Let $\mathcal{T}_{\mathcal{V}} = \{a \in \mathcal{T} \mid \mathit{lab}(a) \in \mathcal{V}\}$ be the selection of axioms with a label that is true under $\mathcal{V}$.

Definition (Pinpointing Formula [4, Def. 3])

Given an $\mathcal{EL}^{+}$ TBox $\mathcal{T}$ and concept names A, B occurring in it, a monotone Boolean formula ψ over $\mathit{lab}(\mathcal{T})$ is a pinpointing formula for $\mathcal{T}$ wrt. $A \sqsubseteq B$ if for every valuation $\mathcal{V} \subseteq \mathit{lab}(\mathcal{T})$ it holds that $A \sqsubseteq_{\mathcal{T}_{\mathcal{V}}} B$ iff $\mathcal{V}$ satisfies ψ.

- Given a pinpointing formula ψ, we can construct corresponding MinAs [3, Prop. 1]:

$$\{\mathcal{T}_{\mathcal{V}} \mid \mathcal{V} \models \psi \text{ and } \mathcal{V} \text{ is } \subseteq\text{-minimal}\}$$

Algorithm 2: ALLMINAS($\mathcal{T}$)

Input: An $\mathcal{EL}^+$ TBox $\mathcal{T}$ in normal form over $\mathsf{N_C}$.

Output: A TBox $\mathcal{T}'$ and a labeling function.

Assign $\mathcal{T}' := \{A \sqsubseteq A, A \sqsubseteq \top \mid A \in \mathsf{N_C}\}$ and $lab(a) := \texttt{true}$ f. a. $a \in \mathcal{T}'$
while *there is a rule* i *s.t.* $1 \leq i \leq 5$, $a_i \in \mathcal{T}$ *and* $P_i \subseteq \mathcal{T}'$ **do**
    $\phi = lab(a_i) \wedge \bigwedge_{p \in P_i} lab(p)$
    if $q_i \notin \mathcal{T}'$ **then**
        $\mathcal{T}' := \mathcal{T}' \cup \{q_i\}$
        $lab(q_i) := \phi$
    else
        $\psi = lab(q_i)$
        if $\psi \vee \phi \not\equiv \psi$ **then**
            $lab(q_i) := \psi \vee \phi$
        end
    end
end
return $(\mathcal{T}', lab)$

- We obtain $\mathcal{T}'$ like before, but additionally a labeling *lab*.
- All $\subseteq$-minimal valuations that satisfy $lab(A \sqsubseteq B)$ correspond to a MinA for $\mathcal{T}$ wrt. $A \sqsubseteq B$ [3, Thm. 2].
- The algorithm runs in time exponential in the size of the input TBox[5], exhibited by

$$\mathcal{T}_n := \left\{B_{i-1} \sqsubseteq P_i \sqcap Q_i, \quad P_i \sqsubseteq B_i, \quad Q_i \sqsubseteq B_i \quad \middle| \quad 1 \geq i \geq n\right\}$$

which yields 2^n MinAs for $\mathcal{T}_n$ wrt. $B_0 \sqsubseteq B_n$ [6].

[5] direct argumentation in [3, Sec. 3]
[6] cf. [3, Example 1]

Algorithms

Pinpointing via Subsumption as Black-Box

Algorithm 3: LinOneMinA($\mathcal{T}$, A, B) (cf. [3, 5, Alg. 1], [8, Alg. 7, p. 97])

Input: A TBox $\mathcal{T} = \{a_1, \ldots, a_n\}$ over $\mathsf{N_C}$ and $A, B \in \mathsf{N_C}$.

Output: If $A \sqsubseteq_{\mathcal{T}} B$ holds, one MinA for $\mathcal{T}$ wrt. $A \sqsubseteq B$, else $\emptyset$.

if $A \not\sqsubseteq_{\mathcal{T}} B$ **then**
    return $\emptyset$
end
$\mathcal{S} := \mathcal{T}$
foreach $a_i \in \mathcal{T}$ **do**
    if $A \sqsubseteq_{\mathcal{S} \setminus \{a_i\}} B$ **then**
        $\mathcal{S} := \mathcal{S} \setminus \{a_i\}$
    end
end
return $\mathcal{S}$

Pinpointing via Subsumption as Black-Box

- Linearily scans $\mathcal{T}$ with one call to subsumption per axiom. Thus, runs in polynomial time overall. [3, Thm. 6]
- “… did not terminate on SNOMED CT in 48hrs …” [8, p. 97].
- Can be improved by using a “sliding window” approach or binary search.
- Black-Box algorithms that compute all MinAs are certainly possible.

Complexity and Tradeoffs in Practice

Complexity

1. Computing all MinAs takes exponential time.
   Is there an output polynomial algorithm?
2. Computing one MinA takes polynomial time.
   This is still bad for large knowledge bases.
   Can we make trade-offs to be faster in practice?

All MinAs

An output polynomial algorithm?

- [3, Thm. 5] shows this is not possible for the case of a TBox with non-refutable part (unless P = NP).
- In [7, Thm. 2] computing all MinAs is established to be as least as hard as computing the set of all minimal transversals of a hypergraph, which is in coNP and no output polynomial alorithm is known (cf. [6]).

Also, computing properties wrt. all MinAs cannot be achieved in polynomial time (unless P = NP) [3, Sec. 4].

One MinA

Polynomial, but still too slow in practice.

1. Let's take some $\mathcal{T}' \subset \mathcal{T}$ with $A \sqsubseteq_{\mathcal{T}'} B$ and run the algorithm on that!
2. To get $\mathcal{T}'$, take the labeling algorithm and drop the re-labeling branch.
3. Other greedy algorithms might perform well/better.

Results: 10min vs. 7hrs with just 2.59% difference in size.[7]

[7] simplified, cf. [3, Sec. 5]

Recap

- Introduction
  - Syntax and Semantics
  - TBoxes and Concept Subsumption
  - Pinpointing
- Algorithms
  - Pinpointing via Labeling
  - Pinpointing via Subsumption as Black-Box
- Complexity and Tradeoffs in Practice

Questions, please!

References I

F. Baader, S. Brandt, and C. Lutz.
Pushing the $\mathcal{EL}$ envelope.
In L. P. Kaelbling and A. Saffiotti, editors, *IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005*, pages 364–369. Professional Book Center, 2005.

F. Baader, I. Horrocks, C. Lutz, and U. Sattler.
An Introduction to Description Logic.
Cambridge University Press, 2017.

References II

F. Baader, R. Peñaloza, and B. Suntisrivaraporn.
Pinpointing in the description logic $\mathcal{EL}^+$.
In J. Hertzberg, M. Beetz, and R. Englert, editors, *KI 2007: Advances in Artificial Intelligence, 30th Annual German Conference on AI, KI 2007, Osnabrück, Germany, September 10-13, 2007, Proceedings*, volume 4667 of *Lecture Notes in Computer Science*, pages 52–67. Springer, 2007.

F. Baader, R. Peñaloza, and B. Suntisrivaraporn.
Pinpointing in the description logic $\mathcal{EL}^+$.
In D. Calvanese, E. Franconi, V. Haarslev, D. Lembo, B. Motik, A. Turhan, and S. Tessaris, editors, *Proceedings of the 2007 International Workshop on Description Logics (DL2007),*

Brixen-Bressanone, near Bozen-Bolzano, Italy, 8-10 June, 2007, volume 250 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2007.

F. Baader and B. Suntisrivaraporn.
Debugging SNOMED CT using axiom pinpointing in the description logic $\mathcal{EL}^+$.
In R. Cornet and K. A. Spackman, editors, *Proceedings of the Third International Conference on Knowledge Representation in Medicine, Phoenix, Arizona, USA, May 31st - June 2nd, 2008*, volume 410 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.

T. Eiter and G. Gottlob.
Hypergraph transversal computation and related problems in logic and AI.
In S. Flesca, S. Greco, N. Leone, and G. Ianni, editors, *Logics in Artificial Intelligence, European Conference, JELIA 2002, Cosenza, Italy, September, 23-26, Proceedings*, volume 2424 of *Lecture Notes in Computer Science*, pages 549–564. Springer, 2002.

References V

R. Peñaloza and B. Sertkaya.
Axiom pinpointing is hard.
In B. C. Grau, I. Horrocks, B. Motik, and U. Sattler, editors, *Proceedings of the 22nd International Workshop on Description Logics (DL 2009), Oxford, UK, July 27-30, 2009*, volume 477 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2009.

B. Suntisrivaraporn.
Polynomial time reasoning support for design and maintenance of large-scale biomedical ontologies.
PhD thesis, Dresden University of Technology, Germany, 2009.