Modularity in Ontologies: MEX Modules

Thomas Schneider[1] *Dirk Walther*[2]

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

[2]Center for Advancing Electronics Dresden, TU Dresden, Germany

ESSLLI, 15 August 2013

Modularity for Light-weight DLs

Logic-based modularity in light-weight DLs

- DL-Lite family
  - [Kontchakov, Wolter, Zakharyaschev, 2010]
- $\mathcal{EL}$ family
  - [Lutz, Wolter, 2010]

$\leadsto$ Here we focus on $\mathcal{EL}$.

Description Logic $\mathcal{EL}$

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

$\mathcal{EL}$-syntax:

$$C ::= \top \mid A \mid C \sqcap C \mid \exists r.C$$

$\mathcal{EL}$-TBox $\mathcal{T}$ is a finite set of $\mathcal{EL}$-concept inclusions $C \sqsubseteq D$.

Reasoning tasks:

- Satisfiability of $\mathcal{EL}$-concept C wrt. $\mathcal{EL}$-TBox $\mathcal{T}$
  - trivial: always satisfiable in a one-point model
- Subsumption of $\mathcal{EL}$-concepts C, D wrt. $\mathcal{EL}$-TBox $\mathcal{T}$
  - tractable (decidable in polynomial time)

Modularity reasoning for $\mathcal{EL}$

- Deciding whether two $\mathcal{EL}$-TBoxes are Σ-inseparable wrt. $\mathcal{EL}$ is ExpTime-complete.
- For $\mathcal{EL}$-TBoxes, Σ-inseparability wrt. SO is undecidable.
- For $\mathcal{EL}$-TBoxes, even $\mathcal{T} \equiv_{\Sigma}^{SO} \emptyset$, (equivalently, whether

  $$\{\mathcal{I}_{|\Sigma} \mid \mathcal{I} \models \mathcal{T}\} = \text{ class of all } \Sigma\text{-models})$$

  is undecidable.

We consider $\mathcal{EL}$-TBoxes of a particular form.

$\mathcal{EL}$-terminologies

Definition

An $\mathcal{EL}$-TBox $\mathcal{T}$ is a $\mathcal{EL}$-terminology if

- every axiom is of the form $A \equiv C$, where A is a concept name;
- no concept name A occurs more than once on the left hand side of an axiom.

A $\mathcal{EL}$-terminology $\mathcal{T}$ is acyclic if no concept name refers to itself along definitions:

- let $A \prec_{\mathcal{T}} X$ if there exists an axiom $A \equiv C$ in $\mathcal{T}$ such that X occurs in C.

Then $\mathcal{T}$ is acyclic iff $\prec_{\mathcal{T}}$ is acyclic (equivalently $\prec_{\mathcal{T}}^{+}$ is irreflexive).

In a TBox $\mathcal{T}$, we rewrite $A \sqsubseteq C$ into $A \equiv X \sqcap C$, where X is fresh.

Example

$$\begin{aligned}
\mathsf{Knee} &\equiv \mathsf{Joint} \sqcap \exists \mathsf{hasPart.Patella} \sqcap & (1)\\
&\quad \exists \mathsf{hasFunct.Hinge} \\
\mathsf{Patella} &\sqsubseteq \mathsf{Bone} \sqcap \mathsf{Sesamoid} & (2)\\
\mathsf{Ginglymus} &\equiv \mathsf{Joint} \sqcap \exists \mathsf{hasFunct.Hinge} & (3)\\
\mathsf{Joint} \sqcap \exists \mathsf{hasPart.}(\mathsf{Bone} \sqcap \mathsf{Sesamoid}) &\sqsubseteq \mathsf{Ginglymus} & (4)\\
\mathsf{Ginglymus} &\equiv \mathsf{HingeJoint} & (5)\\
\mathsf{Meniscus} &\equiv \mathsf{FibroCartilage} \sqcap \exists \mathsf{locatedIn.Knee} & (6)
\end{aligned}$$

It is an $\mathcal{EL}$-TBox. But is it an $\mathcal{EL}$-terminology?

Example

$$
\begin{aligned}
\mathsf{Knee} &\equiv \mathsf{Joint} \sqcap \exists \mathsf{hasPart.Patella} \sqcap && (1)\\
&\quad \exists \mathsf{hasFunct.Hinge} \\
\mathsf{Patella} &\sqsubseteq \mathsf{Bone} \sqcap \mathsf{Sesamoid} && (2)\\
\mathsf{Ginglymus} &\equiv \mathsf{Joint} \sqcap \exists \mathsf{hasFunct.Hinge} && (3)\\
\mathsf{Joint} \sqcap \exists \mathsf{hasPart.(Bone} \sqcap \mathsf{Sesamoid)} &\sqsubseteq \mathsf{Ginglymus} && (4)\\
\mathsf{Ginglymus} &\equiv \mathsf{HingeJoint} && (5)\\
\mathsf{Meniscus} &\equiv \mathsf{FibroCartilage} \sqcap \exists \mathsf{locatedIn.Knee} && (6)
\end{aligned}
$$

no complex LHSs allowed

Example

$$\text{Knee} \equiv \text{Joint} \sqcap \exists \text{hasPart.Patella} \sqcap \exists \text{hasFunct.Hinge} \quad (1)$$

$$\text{Patella} \sqsubseteq \text{Bone} \sqcap \text{Sesamoid} \quad (2)$$

$$\text{Ginglymus} \equiv \text{Joint} \sqcap \exists \text{hasFunct.Hinge} \quad (3)$$

$$\text{Joint} \sqcap \exists \text{hasPart.}(\text{Bone} \sqcap \text{Sesamoid}) \sqsubseteq \text{Ginglymus} \quad (4)$$

$$\text{Ginglymus} \equiv \text{HingeJoint} \quad (5)$$

$$\text{Meniscus} \equiv \text{FibroCartilage} \sqcap \exists \text{locatedIn.Knee} \quad (6)$$

no multiple occurrences of a concept name on LHSs of axioms

Prominent Example: SNOMED CT

- Systematised Nomenclature of Medicine (Clinical Terms)
- $\sim 400,000$ terms
- used in health care in the US, UK, Australia, etc.
- acyclic $\mathcal{EL}$-terminology (+ role box)

Plan for $\mathcal{EL}$-terminologies

- deciding '$\mathcal{T} \equiv_{\Sigma}^{SO} \emptyset$' in polynomial time,
  then $\mathcal{T}$ is safe ➻ *Tuesday's lecture*
- extract modules

Deciding '$\mathcal{T} \equiv_{\Sigma}^{\mathsf{SO}} \emptyset$'

Theorem

The following problem can be solved in polynomial time:
given an acyclic $\mathcal{EL}$-terminology $\mathcal{T}$, decide whether

$$\mathcal{T} \equiv_{\Sigma}^{SO} \emptyset.$$

For the proof, we distinguish two types of syntactic dependencies between Σ-symbols in $\mathcal{T}$:

(a) direct: 'definition' of a Σ-symbol uses another Σ-symbol

(b) indirect: two Σ-symbols are 'defined' using common non-Σ-symbol

Direct Σ-dependencies

Let $\mathcal{T}$ be an acyclic $\mathcal{EL}$-terminology.

(a) $\mathcal{T}$ contains a direct Σ-dependency if there exist $A, X \in \Sigma$ such that $A \prec_{\mathcal{T}}^{+} X$.

Theorem

If an acyclic $\mathcal{EL}$-terminology $\mathcal{T}$ contains a direct Σ-dependency, then $\mathcal{T} \not\equiv_{\Sigma}^{SO} \emptyset$.

Proof. Suppose $\mathcal{T}$ contains a syntactic Σ-dependency $A \prec_{\Sigma}^{+} X$. Take an interpretation $\mathcal{I}$ with $A^{\mathcal{I}} = \Delta^{\mathcal{I}}$ and $X^{\mathcal{I}} = \emptyset$. Then $\mathcal{I}$ can't be expanded to a model of $\mathcal{T}$.

- Does not work for acyclic ALC-terminologies!
- From now on, we assume $\mathcal{T}$ does not contain direct Σ-dependencies.

Indirect Σ-dependencies

Decomposing an acyclic $\mathcal{EL}$-terminology

- Let $\mathcal{T}$ be an acyclic $\mathcal{EL}$-terminology and Σ a signature.
- Take partition
$$\mathcal{T} = \mathcal{T}_\Sigma \cup \mathcal{T}',$$
where
$$\mathcal{T}_\Sigma = \{A \equiv C \mid A \in \Sigma \text{ or } \exists B \in \Sigma,\ B \prec^+_\mathcal{T} A\}$$
- $\mathcal{T}_\Sigma$ does not contain Σ-role names
(as there are no direct Σ-dependencies in $\mathcal{T}$)

Theorem

If $\mathcal{I} \models \mathcal{T}_\Sigma$, then there exists $\mathcal{J} \models \mathcal{T}$ such that $\mathcal{J}_{|\Sigma} = \mathcal{I}_{|\Sigma}$.

Proof. Expand $\mathcal{I}$ inductively by setting $A^\mathcal{J} := C^\mathcal{J}$ for $A \equiv C \in \mathcal{T}'$.

Checking indirect Σ-dependencies

Theorem

Let $\mathcal{T}$ be an acyclic $\mathcal{EL}$-terminology without direct Σ-dependencies. Then the following conditions are equivalent:

1. $\mathcal{T} \equiv^{SO}_{\Sigma} \emptyset$;
2. Every one-point Σ-interpr. can be expanded to a model of $\mathcal{T}_{\Sigma}$.

Point 2 implies Point 1. Let $\mathcal{I}$ be an interpretation. As $\mathcal{T}_{\Sigma}$ contains no Σ-roles, we may assume that Σ contains no roles. For each d in $\mathcal{I}$, let $\mathcal{J}_{\{d\}} \models \mathcal{T}_{\Sigma}$ be an expansion of $\mathcal{I}_{\{d\}}$. Then

$$\mathcal{J} = \bigcup_{d \in \mathcal{I}} \mathcal{J}_{\{d\}} \models \mathcal{T}_{\Sigma}$$

and $\mathcal{J}$ is an expansion of $\mathcal{I}$.

Polytime algorithm for $\mathcal{T} \equiv_{\Sigma}^{SO} \emptyset$

To decide whether $\mathcal{T} \equiv_{\Sigma}^{SO} \emptyset$, check

1. $\mathcal{T}$ contains no direct Σ-dependencies;
2. every one point Σ-model can be expanded to a model of $\mathcal{T}_\Sigma$.

Point 2 holds iff

For all $A \in \Sigma$,

$$\{X \mid A \prec_{\mathcal{T}}^{+} X\} \not\subseteq \{X \mid \exists B \in \Sigma \setminus \{A\},\ B \prec_{\mathcal{T}}^{+} X\}.$$

Observation: For acyclic ALC-terminologies without Σ-dependencies, one can decide $\mathcal{T} \equiv_{\Sigma}^{SO} \emptyset$ by considering one point-models (then Π_2^p-complete).

Module extraction

From deciding inseparability to module extraction.

- Given acyclic $\mathcal{EL}$-terminology $\mathcal{T}$ and signature Σ, the decision procedure extracts from $\mathcal{T}$ the smallest $M \subseteq \mathcal{T}$ such that

$$\mathcal{T} \setminus M \equiv^{SO}_{\Sigma \cup \mathsf{sig}(M)} \emptyset.$$

➾ *then $\mathcal{T} \setminus M$ is safe for $\Sigma \cup \mathsf{sig}(M)$ wrt. $\mathcal{EL}$ (Tuesday's lecture)*

- Equivalently,

$$M \equiv^{SO}_{\Sigma \cup \mathsf{sig}(M)} \mathcal{T}.$$

➾ *then M is a Σ-module in $\mathcal{T}$ wrt. $\mathcal{EL}$*

Module extraction algorithm

Algorithm

Input: Sig. Σ, acyclic $\mathcal{EL}$-terminology $\mathcal{T}$

$\mathcal{M} \leftarrow \emptyset$, $\Sigma_+ \leftarrow \Sigma$

Repeat $\Sigma_{\text{prev}} \leftarrow \Sigma_+$

For each $\alpha \in \mathcal{O} \setminus \mathcal{M}$

If α **Σ_+-dependent**, then add α to $\mathcal{M}$ and $\text{sig}(\alpha)$ to Σ_+

Until $\Sigma_{\text{prev}} = \Sigma_+$

Return $\mathcal{M}$

Output: smallest $\mathcal{M} \subseteq \mathcal{T}$ such that $\mathcal{T} \setminus \mathcal{M} \equiv^{SO}_{\Sigma \cup \text{sig}(M)} \emptyset$.

Module extraction algorithm

Axiom $\alpha = A \equiv C$ is **Σ_+-dependent** in $\mathcal{T} \setminus \mathcal{M}$ if:

1. direct dependencies
   $A, X \in \Sigma_+$ with $A \prec^+_{\mathcal{T}\setminus\mathcal{M}} X$,

2. indirect dependencies
   $A \in \Sigma_+$ and

$$\{X \mid A \prec^+_{\mathcal{T}\setminus\mathcal{M}} X\} \subseteq \{X \mid \exists B \in \Sigma_+ \setminus \{A\} : B \prec^+_{\mathcal{T}\setminus\mathcal{M}} X\}$$

. . . over to Thomas! . . .