Modularity in Ontologies: Module extraction and its logical foundations (Part A)

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, 13 August 2013

Plan for today

We will discuss **logical foundations** for modules

- modules and interfaces
- inseparability = same functionality w.r.t. interface
  = same answers to queries
- decidability/complexity results

Then, we'll look a bit closer on how to **use these insights** to help ontology engineers re-use ontologies

- in a controlled way
- without (unwanted) side-effects

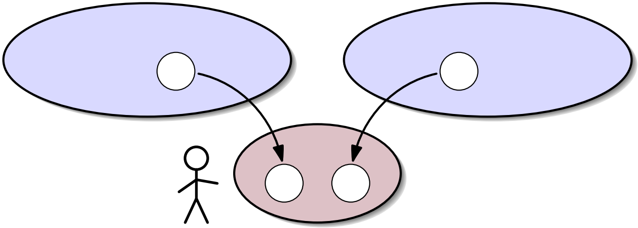
Thanks: partly based on slides by **Uli Sattler** and **Frank Wolter**.

And now . . .

1. Motivation: modular reuse of ontologies
2. Logical foundations of safety and coverage
3. Logical guarantees in detail

Remember: an import/reuse scenario

Take and use knowledge from external ontologies



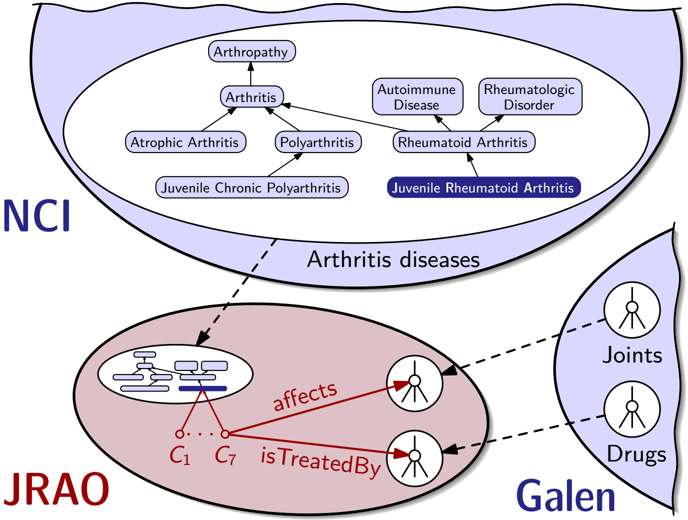
- Provides access to well-established knowledge
- Doesn't require expertise in external disciplines

This scenario is well-understood and implemented.

A *real* example: Health-e-child project

- Build an ontology *JRAO* that describes JRA
  JRA = Juvenile Rheumatoid Arthritis
- Describe JRA subkinds by
  - Joints affected
  - Occurrence of concomitant symptoms, e.g., fever
  - Treatment with certain drugs
- Re-use information provided by biomedical ontologies
  - *NCI*: diseases, drugs, proteins etc.
  - *Galen*: human anatomy

A real example



Why reuse an ontology?

- Saves time and effort
- Provides access to well-established knowledge and terminology
- Doesn't require expertise in drugs, proteins, anatomy etc.

$\rightsquigarrow$ A tool supporting reuse should **guarantee**:

- reusing imported terms doesn't change their meaning **Safety**
- the order of imports doesn't matter **Independence**
- all relevant parts of external ont.s are imported **Coverage**
  in addition, import *only* relevant parts (Economy)

Can we be a bit more specific about

- "doesn't change their meaning"
- "all relevant parts"?

Guarantees by example

Safety

Concerns the **usage of (imported) terms** in the importing ontology:

Let JRA, GeneticDisorder be terms of interest from sig(*NCI*).

$$JRAO \cup NCI \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$$

iff

$$NCI \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$$

Guarantees by example

Independence

Concerns **preservation of safety**:

If *JRAO* is safe for *Galen* and for *NCI*, then

- *JRAO* ∪ *NCI-module* is still safe for *Galen* and
- *JRAO* ∪ *Galen-module* is still safe for *NCI*.

Guarantees by example

Coverage

Concerns what we would consider a **module**:

Let JRA, GeneticDisorder be terms of interest from sig(NCI).

$$JRAO \cup NCI \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$$

iff

$$JRAO \cup \textbf{\textit{NCI-module}} \models \texttt{JRA} \sqsubseteq \texttt{GeneticDisorder}$$

Big, hopeful questions

1. How do we formalise these guarantees?
2. How do we define module notions and import methodologies that provide these guarantees?

And now . . .

1. Motivation: modular reuse of ontologies
2. Logical foundations of safety and coverage
3. Logical guarantees in detail

What is a module?

General definition, e.g., from systems theory

Definition

A module is a part of a system which **functions independently** from the system. The connection between the module and the system is provided by an **interface**.

- an interface enables interoperability between systems
- a system functions through the boundaries of an interface
- what matters is the functionality
  (we can treat the system itself as **black box**)

In logical theories

An interface

- is a tuple $(\mathcal{QL}, \Sigma)$ of a **query language** $\mathcal{QL}$ and a **signature** Σ
- provides a view on a theory (set of observables)
  $\rightsquigarrow$ set of observables is a subset of $\mathcal{QL}$ formulated in Σ
- depends on the application or system

Examples of interfaces

Let $\mathcal{T}$ be a logical theory of arithmetic over the signature $\Sigma = \{+, \times, s, <, 0\}$.

Let $n, m, k \in \{0, 1, 2, \dots\} = \{0, s(0), s(s(0)), \dots\}$.

Interfaces $(\mathcal{QL}, \Sigma)$:

- *Primary school:*

  $\mathcal{QL} = \{n + m = k, n \times m = k\}$ $\quad\Sigma = \{0, s, +, \times\}$
- *Undergraduate:*

  $\mathcal{QL} =$ linear equations $\quad\Sigma = \{0, s, +, \times\}$
- *Mathematician:*

  $\mathcal{QL} =$ Diophantine equations $\quad\Sigma = \{0, s, +, \times\}$
- *Logician:*

  $\mathcal{QL} = \mathrm{SO}$ $\quad\Sigma = \{0, s, f_1, f_2, \dots, +, \times\}$

Interfaces of medical ontologies

Let $\mathcal{T}$ be a TBox defining terms of some medical domain.

Interfaces $(\mathcal{QL}, \Sigma)$:

- *Hospital clerk:*

  $\mathcal{QL}$ = all inclusions $A \sqsubseteq B$, where A, B are concept names
  Σ = predicates relevant to hospital administration

- *Researcher (oncologist):*

  $\mathcal{QL}$ = all inclusions $A \sqsubseteq B$, where A, B are concept names
  Σ = predicates relevant to cancer research

- *Terminologist (expert in anatomy):*

  $\mathcal{QL}$ = all inclusions $C \sqsubseteq D$, where C, D are $\mathcal{ALC}$-concepts
  Σ = predicates relevant to anatomy

- *Someone who can ask all relevant questions:*

  $\mathcal{QL}$ = second-order logic (SO)
  Σ = all predicates in $\mathcal{T}$

Interface for querying instance data

Let $\mathcal{T}$ be a TBox defining geopolitical notions.
$\mathcal{T}$ provides a background theory when querying instance data.

Query language $\mathcal{QL}$:
$\mathcal{A} \rightarrow q$, where $\mathcal{A}$ represents instance data and q is a query.

Example:

- Instance data $\mathcal{A}$

$$\{ \quad \begin{array}{l} \text{Country}(\text{France}), \text{Country}(\text{Columbia}), \ldots, \\ \text{LocatedinEurope}(\text{France}), \ldots \end{array} \quad \}$$

- Query: $q = \text{EuropeanCountry}(\text{France})$
- Then

$$\mathcal{T} \models \mathcal{A} \rightarrow q$$

if $\mathcal{T} \models \text{Country} \sqcap \text{LocatedinEurope} \sqsubseteq \text{EuropeanCountry}$

Functionality and modules of logical theories

Functionality of an ontology $\mathcal{O}$ w.r.t. an interface $(\mathcal{QL}, \Sigma)$:

- the set of $\mathcal{QL}$-formulas φ formulated in Σ that follow from $\mathcal{O}$.

  Formally: $\mathsf{Th}_{\Sigma}^{\mathcal{QL}}(\mathcal{O}) = \{\varphi \in \mathcal{QL} \mid \mathcal{O} \models \varphi,\ \mathsf{sig}(\varphi) \subseteq \Sigma\}$

- $\mathcal{O}$ is considered as a **black box**:
  we're only interested in its functionality $\mathsf{Th}_{\Sigma}^{\mathcal{QL}}(\mathcal{O})$

Question:
When can $\mathcal{O}$ be equivalently replaced by a module $\mathcal{M} \subseteq \mathcal{O}$?

Answer:
Whenever $\mathcal{M}$ has **the same functionality** as $\mathcal{O}$ (w.r.t. interface $(\mathcal{QL}, \Sigma)$)

$\mathcal{M}$ is a **module** of $\mathcal{O}$ if $\mathcal{M} \subseteq \mathcal{O}$ and $\mathsf{Th}_{\Sigma}^{\mathcal{QL}}(\mathcal{M}) = \mathsf{Th}_{\Sigma}^{\mathcal{QL}}(\mathcal{O})$

Inseparability

Definition

Let $\mathcal{T}_1, \mathcal{T}_2$ be finite sets of sentences,[1] $\mathcal{QL}$ a query language,[1] and Σ a signature.

$\mathcal{T}_1$ and $\mathcal{T}_2$ are **Σ-inseparable w.r.t. $\mathcal{QL}$**, in symbols

$$\mathcal{T}_1 \equiv_\Sigma^{\mathcal{QL}} \mathcal{T}_2,$$

if for all $\varphi \in \mathcal{QL}$ with $\mathrm{sig}(\varphi) \subseteq \Sigma$:

$$\mathcal{T}_1 \models \varphi \;\Leftrightarrow\; \mathcal{T}_2 \models \varphi.$$

Example: Let $\mathcal{T}_1 = \{A \sqsubseteq \exists r.B', \exists r.B' \sqsubseteq B\}$, $\mathcal{T}_2 = \{A \sqsubseteq B\}$, $\Sigma = \{A, B\}$. Then $\mathcal{T}_1 \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}_2$ but $\mathcal{T}_1 \not\equiv_{\Sigma\cup\{r\}}^{\mathcal{EL}} \mathcal{T}_2$. Example: Let $\mathcal{T}_1 = \{A \sqsubseteq \neg B\}$, $\mathcal{T}_2 = \emptyset$, $\Sigma = \{A, B\}$. Then $\mathcal{T}_1 \equiv_\Sigma^{\mathcal{EL}} \mathcal{T}_2$ but $\mathcal{T}_1 \not\equiv_\Sigma^{\mathcal{ALC}} \mathcal{T}_2$.

$\leadsto$ $\mathcal{T}_1 \equiv_\Sigma^{\mathcal{QL}} \mathcal{T}_2 \;\Leftrightarrow\; \mathcal{T}_1, \mathcal{T}_2$ have same functionality w.r.t. $(\mathcal{QL}, \Sigma)$

Inseparability w.r.t. second-order logic (SO)

Theorem ([Konev, Lutz, Walther, Wolter, 2009])

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be finite sets of SO-sentences and Σ a signature. Then the following are equivalent:

1. $\mathcal{T}_1 \equiv_{\Sigma}^{SO} \mathcal{T}_2$
2. $\{\mathcal{I}_{|\Sigma} \mid \mathcal{I} \models \mathcal{T}_1\} = \{\mathcal{I}_{|\Sigma} \mid \mathcal{I} \models \mathcal{T}_2\}$

- The last condition means: The restrictions to Σ of all models of $\mathcal{T}_1$ and all models of $\mathcal{T}_2$ coincide.
- Proof omitted, rather straightforward
- Why SO?
    - contains all established ontology languages (DLs, FO)
    - expressive enough to describe FO-interpretations (up to isomorphism)

Conservative extensions in DLs

- Restrict ourselves to description logics $\mathcal{L}$ and to $\mathcal{QL}_{\mathcal{L}} = \{C \sqsubseteq D \mid C, D \text{ are } \mathcal{L}\text{-concepts}\}$
- generalisation to arbitrary $\mathcal{L}$ and $\mathcal{QL}_{\mathcal{L}}$ below SO is easy

Definition

Let $\mathcal{M}, \mathcal{O}$ be TBoxes (finite sets of sentences). Then

1. $\mathcal{O}$ is a **deductive Σ-conservative extension** of $\mathcal{M}$ in $\mathcal{QL}_{\mathcal{L}}$ if $\mathcal{M} \subseteq \mathcal{O}$, and $\mathcal{M}$ and $\mathcal{O}$ are Σ-inseparable w.r.t. $\mathcal{QL}_{\mathcal{L}}$
2. $\mathcal{O}$ is a **model Σ-conservative extension** of $\mathcal{M}$ if $\{\mathcal{I}_{|\Sigma} \mid \mathcal{I} \models \mathcal{M}\} = \{\mathcal{I}_{|\Sigma} \mid \mathcal{I} \models \mathcal{O}\}$

Consequence:

(2′) $\mathcal{O}$ is a model Σ-conservative extension of $\mathcal{M}$ iff $\mathcal{M}$ and $\mathcal{O}$ are Σ-inseparable w.r.t. SO

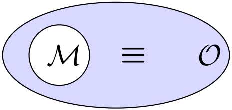
Summary: notion of a module

$\mathcal{M}$ **is a module of** $\mathcal{O}$ for $(\Sigma, \mathcal{QL})$

$\Leftrightarrow$ $\mathcal{M} \subseteq \mathcal{O}$ and $\mathcal{M}, \mathcal{O}$ have the same functionality

$\Leftrightarrow$ $\mathcal{M} \subseteq \mathcal{O}$, and $\mathcal{M}$ and $\mathcal{O}$ are Σ-inseparable w.r.t. $\mathcal{QL}$

$\Leftrightarrow$ $\mathcal{O}$ is a deductive Σ-conservative extension of $\mathcal{M}$



Important questions

- How to compute modules?

  $\rightsquigarrow$ boils down to:

- Can we automatically decide whether two theories are inseparable?

Big, sad theorem ☹

Theorem

- Deciding Σ-inseparability of $\mathcal{EL}$-TBoxes w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ is ExpTime-complete. [Lutz and Wolter 2010]
- Deciding Σ-inseparability of $\mathcal{ALC}$-TBoxes w.r.t. $\mathcal{QL}_{\mathcal{ALC}}$ is 2ExpTime-complete. [Konev, Lutz, Walther, Wolter 2008]
- Deciding Σ-inseparability of $\mathcal{EL}$-TBoxes w.r.t. SO is undecidable. [Lutz and Wolter 2010]

What do we do now?

- Give up? No, modules are too important!
- Drop inseparability? No, safety etc. are too important!
- Approximate? Yes, but from the **right** direction!

Please bear with us ☺
until we've introduced a few more central notions.

And now . . .

1. Motivation: modular reuse of ontologies
2. Logical foundations of safety and coverage
3. Logical guarantees in detail

. . . over to Thomas! . . .