Modularity in Ontologies: Formal foundations of modules

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

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

[2]Faculty of Informatics, Technical University of Madrid, Spain

ESSLLI, 2 August 2011

Plan for today

- definition of a module in logical theories
- interface
- robustness properties

credits: based on slides by Frank Wolter

Modules

- 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.
- interface:
  - provides a view on theory (set of observables)
    $\leadsto$ set of observables is a subset of $\mathcal{QL}$ formulated in Σ
  - depends on the application or system

Examples of interfaces

Let T be a logical theory of arithmetic over the signature $\Sigma = \{+, \times, s, <, 0\}$. (We write 1, 2, ... instead of $s(0), s(s(0))$, ..., etc., and n, m, k range over these number.)

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

- Primary school:
  $\mathcal{QL} = \{n + m = k, n \times m = k\}$, $\Sigma = \{0, s, +, \times\}$
- Undergraduate:
  $\mathcal{QL} =$ linear equations, $\Sigma = \{0, s, +, \times\}$
- Mathematician:
  $\mathcal{QL} =$ Diophantine equations, $\Sigma = \{0, s, +, \times\}$
- Logician:
  $\mathcal{QL} = \text{SO}$, $\Sigma = \{0, s, f_1, f_2, \ldots, +, \times\}$

Interfaces of medical ontologies

Let 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 ALC-concept inclusions,
  Σ = predicates relevant to anatomy
- Someone who can ask all relevant questions:
  $\mathcal{QL}$ = Second-order Logic (SO)
  Σ = all predicates in T

Interface for querying instance data

Let T be a TBox defining geopolitical notions. T provides a background theory when querying instance data.

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

Example: Instance data $\mathcal{A}$

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

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

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

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

Module in logical theories

- An ontology M is a module of an ontology O if $\mathsf{M} \subseteq \mathsf{O}$.
- The functionality of an ontology O wrt. an interface $(\mathcal{QL}, \Sigma)$ is the set of $\mathcal{QL}$-formulas φ formulated in Σ that follow from O. Formally,

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

- O is a black box, what matters is its functionality $\mathsf{Th}_\Sigma^{\mathcal{QL}}(\mathsf{O})$

Replacement of a module

Q: When can module M_1 be equivalently replaced by module M_2?

A: Whenever M_1 and M_2 have the same functionality.

(!) Functionality depends on an interface $(\mathcal{QL}, \Sigma)$

A': Whenever $\mathsf{Th}_\Sigma^{\mathcal{QL}}(M_1) = \mathsf{Th}_\Sigma^{\mathcal{QL}}(M_2)$

Replacement of a module

Let SO denote the set of sentences of Second-Order Logic.
$\rightsquigarrow$ expressive enough to describe FO-interpretations (up to isomorphism)

Definition

Let T_1, T_2 be finite sets of SO-sentences, $\mathcal{QL} \subseteq$ SO a query language, and Σ a signature.
Then: T_1 and T_2 are Σ-inseparable wrt. $\mathcal{QL}$, in symbols

$$T_1 \equiv_{\Sigma}^{\mathcal{QL}} T_2,$$

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

$$T_1 \models \varphi \Leftrightarrow T_2 \models \varphi.$$

$T_1 \equiv_{\Sigma}^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_1$ and T_2 have same functionality wrt. $(\mathcal{QL}, \Sigma)$

Inseparability wrt. SO

Theorem

Let T_1 and T_2 be finite sets of SO-sentences and Σ a signature. Then the following are equivalent:

1. $T_1 \equiv_\Sigma^{SO} T_2$
2. $\{M_{|\Sigma} \mid M \models T_1\} = \{M_{|\Sigma} \mid M \models T_2\}$

Proof: Point 2 implies Point 1.

Trivial: When T_1 and T_2 have the same Σ-models, then they have the same Σ-consequences in SO.

Inseparability wrt. SO

Theorem

Let T_1 and T_2 be finite sets of SO-sentences and Σ a signature. Then the following are equivalent:

1. $T_1 \equiv_{\Sigma}^{SO} T_2$
2. $\{M_{|\Sigma} \mid M \models T_1\} = \{M_{|\Sigma} \mid M \models T_2\}$

Proof: Point 1 implies Point 2.
Suppose $M \models T_1$, but there does not exist $M' \models T_2$ with $M'_{|\Sigma} = M_{|\Sigma}$. Then

$$M \not\models \exists P_1 \cdots \exists P_n . \bigwedge T_2,$$

where $\{P_1, \ldots, P_n\} = \mathrm{sig}(T_2) \setminus \Sigma$. Hence

- $T_2 \models \exists P_1 \cdots \exists P_n . \bigwedge T_2$;
- $T_1 \not\models \exists P_1 \cdots \exists P_n . \bigwedge T_2$.

Conservativity and inseparability: FO

Definition

Let $T_1 \subseteq T_2$ be finite sets FO-sentences. Then

1. T_2 is a deductive Σ-conservative extension of T_1 in FO iff T_1 and T_2 are Σ-inseparable wrt. FO
2. T_2 is a model Σ-conservative extension of T_1 iff $\{M_{|\Sigma} \mid M \models T_1\} = \{M_{|\Sigma} \mid M \models T_2\}$

Corollary

By previous theorem:

(2)' T_2 is a model Σ-conservative extension of T_1 iff T_1 and T_2 are Σ-inseparable wrt. SO

Conservativity and inseparability: DLs

- Let L be a description logic.
- Denote with $\mathcal{QL}_{\mathsf{L}}$ the set of concept inclusions $C \sqsubseteq D$, where C, D are L-concepts.

Definition

Let $T_1 \subseteq T_2$ be TBoxes.
Then:

- T_2 is a deductive Σ-conservative extension of T_1 in L iff T_1 and T_2 are Σ-inseparable wrt. $\mathcal{QL}_{\mathsf{L}}$
- T_2 is a model Σ-conservative extension of T_1 iff T_1 and T_2 are Σ-inseparable wrt. SO

Reasoning task: deciding inseparability

Can we automatically decide whether two theories have the same functionality (i.e. whether they are inseparable)?

Theorem

Deciding Σ-inseparability of EL-TBoxes wrt. $\mathcal{QL}_{\mathsf{EL}}$ is ExpTime-complete [Lutz, Wolter, 2010]

Theorem

Deciding Σ-inseparability of ALC-TBoxes wrt. $\mathcal{QL}_{\mathsf{ALC}}$ is 2ExpTime-complete [Konev, Lutz, Walther, Wolter, 2008]

Theorem

Deciding Σ-inseparability of EL-TBoxes wrt. SO is undecidable [Lutz, Wolter, 2010]

Programme:

- properties of inseparability relation $\equiv_{\Sigma}^{\mathcal{QL}}$
- connection to the interpolation property

Simple properties of $\equiv_\Sigma^{\mathcal{QL}}$

Let $\mathcal{QL} \subseteq$ SO, Σ a signature and T_1, T_2 be finite sets of SO sentences.
Then:

- $\equiv_\Sigma^{\mathcal{QL}}$ is an equivalence relation
- robustness under vocabulary reduction:
  for all $\Sigma' \subseteq \Sigma$,

  $$T_1 \equiv_\Sigma^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_1 \equiv_{\Sigma'}^{\mathcal{QL}} T_2$$

- robustness under query language reduction:
  for all $\mathcal{QL}' \subseteq \mathcal{QL}$,

  $$T_1 \equiv_\Sigma^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_1 \equiv_\Sigma^{\mathcal{QL}'} T_2$$

Robustness properties: vocabulary extension

Q: When can we extend the vocabulary of the interface without losing inseparability?

Definition

Let L $\subseteq$ SO and $\mathcal{QL} \subseteq$ SO.
Then $(L, \mathcal{QL})$ is *robust* under vocabulary extensions if, for all finite sets T_1 and T_2 of L-sentences and signatures Σ, Σ' with $\Sigma' \cap \mathrm{sig}(T_1 \cup T_2) \subseteq \Sigma$, the following holds:

$$T_1 \equiv_{\Sigma}^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_1 \equiv_{\Sigma'}^{\mathcal{QL}} T_2.$$

A: If $(L, \mathcal{QL})$ is robust under vocabulary extensions, then we can add vocabulary not in T_1 and T_2

Intuition: if we have two Σ-indistinguishable ontologies, then they are still Σ'-indistinguishable

Robustness properties: joins

Definition

Let L $\subseteq$ SO and $\mathcal{QL} \subseteq$ SO.
Then $(L, \mathcal{QL})$ is *robust* under joins if, for all finite sets T_1 and T_2 of L-sentences and signatures Σ with $\mathrm{sig}(T_1) \cap \mathrm{sig}(T_2) \subseteq \Sigma$, the following holds for $i = 1, 2$:

$$T_1 \equiv_\Sigma^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_i \equiv_\Sigma^{\mathcal{QL}} T_1 \cup T_2.$$

Application: if we have two indistinguishable ontologies, it suffices to import just one of them

Robustness properties: replacement

Definition

Let L $\subseteq$ SO and $\mathcal{QL} \subseteq$ SO.
Then $(L, \mathcal{QL})$ is *robust* under replacement if, for all all finite sets T, T_1 and T_2 of L-sentences and signatures Σ with $\text{sig}(T) \cap \text{sig}(T_1 \cup T_2) \subseteq \Sigma$, the following holds:

$$T_1 \equiv_\Sigma^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_1 \cup T \equiv_\Sigma^{\mathcal{QL}} T_2 \cup T.$$

Intuition: indistinguishability is insensitive towards adding certain contexts

Application: important property for re-use
➻ *Wednesday's lecture*

Concept hierarchy

Let $\mathcal{QL}_C$ denote the set of implications $A \sqsubseteq B$ such that A, B are concept names.

Theorem

(ALC, $\mathcal{QL}_C$) is robust under vocabulary extensions, but not under joins nor replacement.

Counterexample:

$$\mathcal{T}_1 = \{A \sqsubseteq \exists r.B\}, \quad \mathcal{T}_2 = \{\exists r.B \sqsubseteq E\}, \quad \Sigma = \{r, A, B, E\}.$$

Then:

- $\mathcal{T}_1 \equiv_\Sigma^{\mathcal{QL}_C} \mathcal{T}_2$
- $\mathcal{T}_1 \cup \mathcal{T}_2 \not\equiv_\Sigma^{\mathcal{QL}_C} \mathcal{T}_2$
- $\mathcal{T}_1 \cup \{B \equiv \bot\} \not\equiv_\Sigma^{\mathcal{QL}_C} \mathcal{T}_2 \cup \{B \equiv \bot\}$

Basic observations

Theorem

Let $\mathsf{L} \subseteq \mathcal{QL}$.
If $(\mathcal{QL}, \mathcal{QL})$ is robust under vocabulary extensions (joins, replacement), then $(\mathsf{L}, \mathcal{QL})$ is robust under vocabulary extensions (joins, replacement).

Robustness under joins: Suppose

$$T_1 \equiv_{\Sigma}^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_i \equiv_{\Sigma}^{\mathcal{QL}} T_1 \cup T_2,$$

for all finite sets T_1 and T_2 of $\mathcal{QL}$-sentences. Then, because $\mathsf{L} \subseteq \mathcal{QL}$,

$$T_1 \equiv_{\Sigma}^{\mathcal{QL}} T_2 \quad \Rightarrow \quad T_i \equiv_{\Sigma}^{\mathcal{QL}} T_1 \cup T_2,$$

for all finite sets T_1 and T_2 of L-sentences.

Basic observations

Theorem

(SO,SO) is robust under vocabulary extensions, joins and replacement.

Robustness under joins: suppose

$$\{M_{|\Sigma} \mid M \models T_1\} = \{M_{|\Sigma} \mid M \models T_2\}$$

and $\text{sig}(T_1) \cap \text{sig}(T_2) \subseteq \Sigma$. Then every $M_{|\Sigma}$ with $M \models T_1$ can be expanded to a model M' of T_2. Moreover, we may assume that the interpretation of $\text{sig}(T_1)$ in M and M' is the same. But then $M' \models T_1 \cup T_2$ and $M'_{|\Sigma} = M_\Sigma$. Hence

$$\{M_{|\Sigma} \mid M \models T_1\} = \{M_{|\Sigma} \mid M \models T_1 \cup T_2\}.$$

Interpolation

Definition

A logic L has weak interpolation iff for every finite set T of L-sentences and L-sentence φ such that $T \models \varphi$, there exists a set $I(T, \varphi)$ of L-sentence such that

- $\text{sig}(I(T, \varphi)) \subseteq \text{sig}(T) \cap \text{sig}(\varphi)$;
- $T \models I(T, \varphi)$;
- $I(T, \varphi) \models \varphi$.

L has interpolation if there always exists a finite set $I(T, \varphi)$ with these properties.

There are many results on variants of interpolation in logic and software specification!

Interpolation

Details for logicians:

- For any compact logic L, weak interpolation implies interpolation.
- L is compact if $T \models \varphi$ implies that there exists a finite subset T' of T such that $T' \models \varphi$.

Robustness under Vocabulary Extensions and Interpolation

Weak interpolation implies robustness under vocabulary extensions.

Theorem

If $\mathcal{QL}$ has weak interpolation, then $(\mathcal{QL}, \mathcal{QL})$ is robust under vocabulary extensions.

Proof. Suppose $T_1 \equiv_\Sigma^{\mathcal{QL}} T_2$. Let $\varphi \in \mathcal{QL}$ with $\mathrm{sig}(\varphi) \cap \mathrm{sig}(T_1 \cup T_2) \subseteq \Sigma$ such that $T_1 \models \varphi$. We show $T_2 \models \varphi$. By weak interpolation,

$$T_1 \models I(T_1, \varphi) \models \varphi.$$

From $\mathrm{sig}(I(T_1, \varphi)) \subseteq \Sigma$

$$T_2 \models I(T_1, \varphi).$$

Hence $T_2 \models \varphi$.

Robustness under Vocabulary Extensions and Interpolation

Robustness under vocabulary extensions implies weak interpolation.

Theorem

Suppose $(\mathcal{QL}, \mathcal{QL})$ is robust under vocabulary extensions for possibly infinite sets of sentences. Then $\mathcal{QL}$ has weak interpolation.

Proof. Assume $T \models \varphi$. Set $\Sigma = \text{sig}(T) \cap \text{sig}(\varphi)$ and

$$T_\Sigma = \{\psi \in \mathcal{QL} \mid T \models \psi, \text{sig}(\psi) \subseteq \Sigma\}.$$

Then

- T and T_Σ are Σ-inseparable wrt. $\mathcal{QL}$.

By robustness under vocabulary extensions,

- T and T_Σ are Σ'-inseparable wrt. $\mathcal{QL}$, for $\Sigma' = \text{sig}(\varphi)$.

From $T \models \varphi$ we obtain $T_\Sigma \models \varphi$. Hence $I(T, \varphi) = T_\Sigma$ is a weak interpolant.

Interpolation

The following logics L have interpolation (and thus (L,L) is robust under vocabulary extensions):

- propositional logic
- FO [standard textbooks on logic]
- EL [Sofronie-Stokkermans, 2006]
- ALC, ALCQ, ALCI, ALCQI, etc. [Konev, Lutz, Walther, Wolter, 2008]
- SO [Goranko, Otto, 2007]

Counterexamples: nominals

ALCO extends ALC by nominals, which are interpreted as singleton sets.

Theorem

ALCO does not have weak interpolation.
(ALC, ALCO) is not robust under vocabulary extensions.

Let

$$T_1 = \{\top \sqsubseteq \exists r.\top\}, \quad T_2 = T_1 \cup \{A \sqsubseteq \forall r.\neg A, \neg A \sqsubseteq \forall r.A\}.$$

$T_1 \equiv_{\Sigma}^{\mathsf{ALCO}} T_2$ for $\Sigma = \{r\}$.
Observe $\{a\} \sqsubseteq \forall r.\neg\{a\}$ separates the two TBoxes wrt. $\mathcal{QL}_{\mathsf{ALCO}}$, for any nominal $\{a\}$.
Thus, $T_1 \not\equiv_{\Sigma'}^{\mathsf{ALCO}} T_2$ for $\Sigma' = \Sigma \cup \{a\}$.

Counterexample: role hierarchies

ALCH extends ALC by axioms of the form $r \sqsubseteq s$, for roles r, s.

Theorem

(ALCH, ALCH) is not robust under vocabulary extensions.

Let

$$T_1 = \{\top \sqsubseteq \forall r_i \forall r_j.\bot \mid i, j = 1, 2\} \cup \{\exists r_1.\top \equiv \exists r_2.\top\},$$

$$T_2 = T_1 \cup \{s \sqsubseteq r_1, s \sqsubseteq r_2, \exists r_1.\top \sqsubseteq \exists s.\top\}.$$

Then $T_1 \equiv_\Sigma^{\text{ALCH}} T_2$ for $\Sigma = \{r_1, r_2\}$.
$\exists r_1.\top \sqcap \forall r_1.A \sqsubseteq \exists r_2.A$ separates the two ontologies, where A is a fresh concept name.

Robustness under joins

Theorem

$(\mathcal{QL},\mathcal{QL})$ is robust under joins for any $\mathcal{QL}$ from ALC, ALCQ, ALCI and ALCQI.

Proof in [Konev, Lutz, Walther, Wolter, 2008].

Theorem

(ALCH, ALCH) and (ALCO, ALCO) are not robust under joins.

Robustness under Replacement

$\mathcal{QL}$ is closed under Boolean operators if
$\varphi, \psi \in \mathcal{QL}$ implies $\neg\varphi$, $\varphi \wedge \psi \in \mathcal{QL}$.

Theorem

If $\mathcal{QL}$ is closed under Boolean operators and
$(\mathcal{QL}, \mathcal{QL})$ is robust under vocabulary extensions,
then $(\mathcal{QL}, \mathcal{QL})$ is robust under replacement.

Let $T, \varphi \subseteq \mathcal{QL}$ with $\text{sig}(T, \varphi) \cap \text{sig}(T_1 \cup T_2) \subseteq \Sigma$ and

$$T_1 \equiv_{\Sigma}^{\mathcal{QL}} T_2.$$

Then

$$T_1 \models \bigwedge T \rightarrow \varphi \Leftrightarrow T_2 \models \bigwedge T \rightarrow \varphi.$$

Hence

$$T_1 \cup T \models \varphi \Leftrightarrow T_2 \cup T \models \varphi.$$

Robustness under Replacement

Theorem

(ALC,ALC) is not robust under replacement.

Let

$$T_1 = \emptyset, \quad T_2 = \{A \sqsubseteq \exists r.B\}, \quad \Sigma = \{A, B\}.$$

The class of Σ-reducts of models of T_2 is axiomatised by

$$\exists x A(x) \rightarrow \exists x B(x).$$

Hence $T_1 \equiv_{\Sigma}^{\mathrm{ALC}} T_2$. Let $T = \{A \equiv \top, B \equiv \bot\}$. Then

$$T_1 \cup T \not\models \top \sqsubseteq \bot \quad T_2 \cup T \models \top \sqsubseteq \bot.$$

Conclusion

We have shown:

- conditions under which ontologies can be replaced with one another (inseparability)
- preservation of inseparability under change of certain parameters (robustness properties)
- investigation of which logics imply robustness
- characterisation of robustness via interpolation
- relevance to application scenarios, e.g., import/re-use
  ➺ *Wednesday's lecture*

Course overview

3. Locality
   - Locality classes and locality-based modules
   - Module extraction algorithms and experiments
4. Versioning and Forgetting
   - Logical difference
   - Forgetting/uniform interpolants
5. Recent Advances/Current Work
   - Atomic decomposition
   - Signature decomposition, relevance of terms