Modularity of Ontologies in an Arbitrary Institution

Yazmin Angelica Ibañez[1], Till Mossakowski[⋆2], Donald Sannella[3], and Andrzej Tarlecki[⋆⋆4]

[1] Department of Computer Science, University of Bremen
[2] Faculty of Computer Science, Otto-von-Guericke University of Magdeburg
[3] Laboratory for Foundations of Computer Science, University of Edinburgh
[4] Institute of Informatics, University of Warsaw

Abstract. The notion of module extraction has been studied extensively in the ontology community. The idea is to extract, from a large ontology, those axioms that are relevant to certain concepts of interest (formalised as a subsignature). The technical concept used for the definition of module extraction is that of inseparability, which is related to indistinguishability known from observational specifications.

Module extraction has been studied mainly for description logics and the Web Ontology Language OWL. In this work, we generalise previous definitions and results to an arbitrary inclusive institution. We reveal a small inaccuracy in the formal definition of inseparability, and show that some results hold in an arbitrary inclusive institution, while others require the institution to be weakly union-exact.

This work provides the basis for the treatment of module extraction within the institution-independent semantics of the distributed ontology, modeling and specification language (DOL), which is currently under submission to the Object Management Group (OMG).

1 Introduction

Goguen's and Burstall's invention of the concept of institution to formalise the notion of logical system has stimulated a research programme with the general idea that modular structuring of complex specifications can be studied largely independently of the details of the underlying logical system. José Meseguer has made important contributions to institutions and their translations [1–4, 7, 11–15, 19], and to the study of module systems over arbitrary institutions, see especially [7]. His contributions have been inspiring for our work, and and some of his papers are among those we cite most frequently. In the present work, we study modularity over an arbitrary institution using a concept of *inseparability*, which

⋆ This work has been partially supported by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, FET-Open Grant number: 611553, project COINVENT.

⋆⋆ This work has been partially supported by the National Science Centre, grant 2013/11/B/ST6/01381

has some resemblance to José's notion of *indistinguishability* [22] of protocols, although context and technical details are very different. This paper is dedicated to José on the occasion of his 65th birthday — our congratulations and best wishes, José!

The notion of modularity studied in the specification community is modularity *by construction*: complex specifications are formed from basic specifications (which are simply logical theories in some institution) by means of specification-building operations [7, 20, 21].

In the ontology community, a different notion of modularity has emerged: while even large ontologies with tens of thousands of axioms are often formalised as flat logical theories, the notion of ontological *module extraction* [27] provides an *a posteriori* extraction of relevant parts of ontologies. Module extraction has been studied mainly for description logics, but first attempts for first-order logic have also been made.

In the present paper, we try to cast module extraction in the institution-independent framework and compare it with module notions from specification theory. This work thereby also provides a semantics for certain modularity concepts and constructs in the distributed ontology, modeling and specification language DOL [16, 17], which is currently under submission as a standard to the Object Management Group (OMG).

The problem of module extraction can be phrased as follows: given a subset Σ of the signature of an ontology, find a (minimal) subset of that ontology that is "relevant" for the terms in Σ. For example, the size of well-established ontologies such as SNOWMED CT[5] or GALEN[6] makes it difficult for current tools to navigate through them on a standard computer. Therefore, in an application where only a specific subset of the terms in such huge ontologies is used, it is more practical to reuse only those parts that *cover* all the knowledge about that subset of relevant terms.

The key concept of "relevance" may be formalised in different ways. We will not discuss in any detail here approaches based on syntactic structure of axioms and hierarchy of concepts [6, 25, 26]. Instead, we will focus on *logic-based modules*, for which relevance amounts to entailment (or model) preservation over a signature Σ. That is, given an ontology $\mathcal{O}$, when we say that a module $\mathcal{M}$ (which is a subset of $\mathcal{O}$) "is relevant for" the terms in Σ, we mean that all consequences of $\mathcal{O}$ that can be expressed over Σ are also consequences of $\mathcal{M}$. Then $\mathcal{O}$ is said to be a *conservative extension* (CE) of $\mathcal{M}$. A stronger property is that every model of $\mathcal{M}$ extends to a model of $\mathcal{O}$— we refer to this as *model conservative extension.*

One of the reasons why one might be interested in modularity aspects of an ontology is for reusing information about relevant terms it captures. Reusing a module $\mathcal{M} \subseteq \mathcal{O}$ within another ontology $\mathcal{O}'$ is referred to in the literature as the *module importing scenario.* In this scenario, the signature Σ used to extract $\mathcal{M}$ from $\mathcal{O}$ acts as the *interface* signature between $\mathcal{O}'$ and $\mathcal{O}$ in the sense that

[5] http://ihtsdo.org/snomed-ct/
[6] http://www.opengalen.org/

it contains the set of terms that one is interested in reusing and that might be shared between $\mathcal{O}'$ and $\mathcal{O}$.

Example 1.1. Assume that we have the following OWL-ontology $\mathcal{O}$:

$$\begin{array}{ll} \mathsf{Male} \equiv \mathsf{Human} \sqcap \neg\mathsf{Female}, & \mathsf{Father} \sqsubseteq \mathsf{Human}, \\ \mathsf{Human} \sqsubseteq \forall \mathsf{has\_child}.\mathsf{Human}, & \mathsf{Father} \equiv \mathsf{Male} \sqcap \exists \mathsf{has\_child}.\top \end{array}$$

For readers not familiar with OWL, we provide the translation to first-order logic:

$$\begin{array}{l} \forall x.\mathsf{Male}(x) \leftrightarrow \mathsf{Human}(x) \wedge \neg\mathsf{Female}(x), \\ \qquad \forall x.\mathsf{Father}(x) \rightarrow \mathsf{Human}(x), \\ \forall x.\mathsf{Human}(x) \rightarrow \forall y.\mathsf{has\_child}(x,y) \rightarrow \mathsf{Human}(y) \\ \qquad \forall x.\mathsf{Father}(x) \leftrightarrow \mathsf{Male}(x) \wedge \exists y.\mathsf{has\_child}(x,y) \end{array}$$

Now further assume that we are interested in the terms in $\Sigma = \{\mathsf{Male}, \mathsf{Human}, \mathsf{Female}, \mathsf{has\_child}\}$. Then the subset $\mathcal{M}$ containing only the grey shaded axioms is a Σ-module of $\mathcal{O}$. Indeed, one can show that $\mathcal{O}$ has the same Σ-consequences as $\mathcal{M}$. For example, $\mathsf{Male} \sqcap \exists \mathsf{has\_child}.\top \sqsubseteq \mathsf{Human}$ follows from $\mathcal{O}$, but also from $\mathcal{M}$. □

Ideally, an imported module $\mathcal{M}$ should be as small as possible while still guaranteeing to capture all the relevant knowledge w.r.t. Σ. Importing $\mathcal{M}$ into $\mathcal{O}'$ would have the same observable effect as importing the entire ontology $\mathcal{O}$, e.g., one should get the same answers to a query in both cases.

Observe that the logical view appears to be theoretically sound and elegant and guarantees that by reusing only terms from Σ one is not able to distinguish between importing $\mathcal{M}$ and importing $\mathcal{O}$ into some ontology $\mathcal{O}'$.

This paper contributes the generalization of central notions of ontology module extraction to an arbitrary institution. While doing this, we were also able to correct a small inaccuracy appearing in the definition of inseparability used in the literature. Our paper is organized as follows: in Sect. 2, we recall institutions and inclusion systems (the latter leading to a set-theoretic flavour of signatures, which is generally assumed in the ontology community). Sect. 3 studies conservative extensions and inseparability as a prequisite for module extraction, which is studied in Sect. 4, together with some robustness properties. Sect. 5 concludes the paper.

2 Institutions

The large variety of logical languages in use can be captured at an abstract level using the concept of *institutions* [8]. This allows us to develop results independently of the specific features of a logical system. We use the notions of institution and logical language interchangeably throughout the rest of the paper.

The main idea is to collect the non-logical symbols of the language in signatures and to assign to each signature the set of sentences that can be formed using its symbols. Informally, in typical examples, each signature lists the symbols it consists of, together with their kinds. Signature morphisms are mappings between signatures. We do not assume any details except that signature morphisms can be composed and that there are identity morphisms; this amounts to a category of signatures. Readers unfamiliar with category theory may replace this with a partial order; signature morphisms are then just inclusions. See [18] for details of this simplified foundation.

Institutions also provide a model theory, which introduces semantics for the language and gives a satisfaction relation between the models and the sentences of a signature. The main restriction imposed is the satisfaction condition, which captures the idea that truth is invariant under change of notation (and enlargement of context) along signature morphisms. This relies on two further components of institutions: the translation of sentences along signature morphisms, and the reduction of models against signature morphisms (generalising the notion of model reduct known from logic).

Definition 2.1. *An **institution** [8] is a quadruple $I = (\mathbf{Sign}, \mathbf{Sen}, \mathbf{Mod}, \models)$ consisting of the following:*

- *a category* $\mathbf{Sign}$ *of* signatures *and* signature morphisms*;*
- *a functor* $\mathbf{Sen}\colon \mathbf{Sign} \to \mathbb{S}et$[7] *giving, for each signature Σ, the set of* sentences $\mathbf{Sen}(\Sigma)$*, and for each signature morphism $\sigma\colon \Sigma \to \Sigma'$, the* sentence translation map $\mathbf{Sen}(\sigma)\colon \mathbf{Sen}(\Sigma) \to \mathbf{Sen}(\Sigma')$*, where $\mathbf{Sen}(\sigma)(\varphi)$ is often written as $\sigma(\varphi)$;*
- *a functor* $\mathbf{Mod}\colon \mathbf{Sign}^{op} \to \mathbb{C}at$[8] *giving, for each signature Σ, the category of* models $\mathbf{Mod}(\Sigma)$*, and for each signature morphism $\sigma\colon \Sigma \to \Sigma'$, the* reduct functor $\mathbf{Mod}(\sigma)\colon \mathbf{Mod}(\Sigma') \to \mathbf{Mod}(\Sigma)$*, where $\mathbf{Mod}(\sigma)(M')$ is often written as $M'|_\sigma$. Then $M'|_\sigma$ is called the* σ-reduct *of M', while M' is called a* σ-expansion *of $M'|_\sigma$; and*
- *a satisfaction relation $\models_\Sigma \subseteq |\mathbf{Mod}(\Sigma)| \times \mathbf{Sen}(\Sigma)$ for each $\Sigma \in |\mathbf{Sign}|$,*

such that for each $\sigma\colon \Sigma \to \Sigma'$ in $\mathbf{Sign}$ *the following **satisfaction condition** holds:*

$$(\star) \qquad M' \models_{\Sigma'} \sigma(\varphi) \text{ iff } M'|_\sigma \models_\Sigma \varphi$$

for each $M' \in |\mathbf{Mod}(\Sigma')|$ and $\varphi \in \mathbf{Sen}(\Sigma)$. □

As usual, the satisfaction relation between models and sentences determines a semantic notion of consequence: for any signature $\Sigma \in |\mathbf{Sign}|$, a Σ-sentence $\varphi \in \mathbf{Sen}(\Sigma)$ is a (semantic) *consequence* of a set of Σ-sentences $\Phi \subseteq \mathbf{Sen}(\Sigma)$, written $\Phi \models_\Sigma \varphi$, if for each model $M \in |\mathbf{Mod}(\Sigma)|$, $M \models_\Sigma \varphi$ whenever $M \models_\Sigma \Phi$ (i.e., $M \models_\Sigma \psi$ for all $\psi \in \Phi$). This is an example of how logical notions can be defined in an arbitrary institution. It is easy to see that semantic consequence is preserved under translation w.r.t. signature morphisms: given $\sigma\colon \Sigma \to \Sigma'$, if

[7] $\mathbb{S}et$ is the category having sets as objects and functions as arrows.

[8] $\mathbb{C}at$ is the category of categories and functors. Strictly speaking, $\mathbb{C}at$ is a quasicategory (which is a category that lives in a higher set-theoretic universe).

$\Phi \models_\Sigma \varphi$ then $\sigma(\Phi) \models_{\Sigma'} \sigma(\varphi)$. The opposite implication does not hold in general though.

It is also possible to complement an institution with a proof theory, introducing a derivability or *deductive consequence* relation between sentences, formalised as an *entailment system* [13]. In particular, this can be done for the institutions presented below.

Several institution-independent languages for structured theories have been defined, see e.g. [7, 21]. One of them is the distributed ontology, modeling and specification language DOL [17], which also provides language constructs for module extraction.

Example 2.2. In the institution Prop of propositional logic, signatures are sets of propositional variables and signature morphisms are functions. Models are valuations into $\{T, F\}$ and model reduct is just composition. Sentences are formed inductively from propositional variables by the usual logical connectives. Sentence translation means replacement of propositional variables along the signature morphism. Satisfaction is the usual satisfaction of a propositional sentence under a valuation. □

Example 2.3. OWL signatures consist of sets of atomic classes, individuals, object and data properties. OWL signature morphisms map classes to classes, individuals to individuals, object properties to object properties and data properties to data properties. For an OWL signature Σ, sentences include subsumption relations between classes or properties, membership assertions of individuals in classes and pairs of individuals in properties, and complex role inclusions. Sentence translation along a signature morphism simply replaces non-logical symbols with their image along the morphism. The kinds of symbols are class, individual, object property and data property, respectively, and the set of symbols of a signature is the union of its sets of classes, individuals and properties. Models are (unsorted) first-order structures that interpret concepts as unary and properties as binary predicates, and individuals as elements of the universe of the structure, and satisfaction is the standard satisfaction of description logics. This gives us an institution for OWL.

Strictly speaking, this institution captures OWL 2 DL *without restrictions* in the sense of [23]. The reason is that in an institution, sentences can be used for arbitrary formation of theories. This is related to the presence of union as a specification-building operation, which is also present in DOL. OWL 2 DL's specific restrictions on theory formation can be modelled *inside* this institution, as a constraint on ontologies (theories). This constraint is generally not preserved under unions or extensions. DOL's multi-logic capability allows the clean distinction between ordinary OWL 2 DL and OWL 2 DL without restrictions. □

Example 2.4. In the institution $\mathsf{FOL}^=$ of many-sorted first-order logic with equality, signatures are many-sorted first-order signatures, consisting of a set of sort and sorted operation and predicate symbols. Signature morphisms map sorts, operation and predicate symbols in a compatible way. Models are many-sorted first-order structures. Sentences are closed first-order formulae with atomic formulae including equality between terms of the same sort. Sentence translation means replacement of symbols along the signature morphism. A model reduct interprets a symbol by first translating it along the signature morphism and then interpreting it in the model to be reduced. Satisfaction is the usual satisfaction of a first-order sentence in a first-order structure. □

A *presentation* in an institution $I = (\mathbf{Sign}, \mathbf{Sen}, \mathbf{Mod}, \models)$ is a pair $P = (\Sigma, \Phi)$, where $\Sigma \in |\mathbf{Sign}|$ is a signature and $\Phi \subseteq \mathbf{Sen}(\Sigma)$ is a set of Σ-sentences. Σ is also denoted as $\mathsf{Sig}(P)$, Φ as $\mathsf{Ax}(P)$. We extend the model functor to presentations and write $\mathbf{Mod}(\Sigma, \Phi)$ (or sometimes $\mathbf{Mod}(\Phi)$ if the signature is clear) for the full subcategory of $\mathbf{Mod}(\Sigma)$ that consists of the *models* of (Σ, Φ), i.e., $|\mathbf{Mod}(\Sigma, \Phi)| = \{M \in |\mathbf{Mod}(\Sigma)| \mid M \models_\Sigma \Phi\}$.

A *presentation morphism* $\sigma\colon (\Sigma, \Phi) \to (\Sigma', \Phi')$ is a signature morphism $\sigma\colon \Sigma \to \Sigma'$ such that for all models $M' \in |\mathbf{Mod}(\Sigma', \Phi')|$, $M'|_\sigma \in |\mathbf{Mod}(\Sigma, \Phi)|$. This defines the category $\mathbf{Pres}$ of *presentations* in I. An easy consequence of the satisfaction condition is that presentation morphisms preserve semantic consequence:

Proposition 2.5. *$\sigma\colon (\Sigma, \Phi) \to (\Sigma', \Phi')$ is a presentation morphism iff for all Σ-sentences φ, if $\Phi \models_\Sigma \varphi$ then $\Phi' \models_{\Sigma'} \sigma(\varphi)$.* □

Each presentation (Σ, Φ) *generates* a *theory* $(\Sigma, \mathsf{cl}_\models(\Phi))$, where $\mathsf{cl}_\models(\Phi) = \{\varphi \in \mathbf{Sen}(\Sigma) \mid \Phi \models_\Sigma \varphi\}$ is the closure of Φ under semantic consequence. The category $\mathbf{Th}$ of *theories* in I is the full subcategory of $\mathbf{Pres}$ with objects (Σ, Φ) such that Φ is closed under semantic consequence. The closure under semantic consequence extends to the functor $\mathsf{cl}_\models\colon \mathbf{Pres} \to \mathbf{Th}$, which together with the inclusion $\mathbf{Th} \hookrightarrow \mathbf{Pres}$ establishes the equivalence between $\mathbf{Pres}$ and $\mathbf{Th}$.

A presentation morphism $\sigma\colon (\Sigma, \Phi) \to (\Sigma', \Phi')$ is *model-conservative* if for each model $M \in |\mathbf{Mod}(\Sigma, \Phi)|$ there is a model $M' \in |\mathbf{Mod}(\Sigma', \Phi')|$ that is a σ-*expansion* of M, i.e., $M'|_\sigma = M$. A presentation morphism $\sigma\colon (\Sigma, \Phi) \to (\Sigma', \Phi')$ is *consequence-conservative* if for all Σ-sentences $\varphi \in \mathbf{Sen}(\Sigma)$, $\Phi \models_\Sigma \varphi$ whenever $\Phi' \models_{\Sigma'} \sigma(\varphi)$ (the opposite implication always holds).

Proposition 2.6. *If a presentation morphism is model-conservative then it is consequence-conservative as well.* □

The opposite implication does not hold in general: model-conservativity is a strictly stronger notion than consequence-conservativity. However, in some logics, the two notions may coincide:

Example 2.7. In the institution Prop of propositional logic (see Example 2.2), a presentation morphism is model-conservative iff it is consequence-conservative. Consider a presentation morphism $\sigma\colon (V, \Phi) \to (\mathsf{V}', \Phi')$ in Prop. Assume that σ is not model-conservative, and let $m\colon V \to \{T, F\}$ be such that $m \models \Phi$ and m has no σ-expansion to a model of Φ'. For each propositional variable $p \in V$, let $\varphi_{m,p}$ be p if $m(p) = T$ and $\neg p$ if $m(p) = F$. Consider $\Psi' = \Phi' \cup \{\sigma(\varphi_{m,p}) \mid p \in V\}$. Since there is no model $m'\colon V' \to \{T, F\}$ such that $m' \models \Phi'$ and $m'|_\sigma = m$, Ψ'

has no model, and so *false* is a semantic consequence of Ψ'. By compactness of propositional logic, for some finite set of variables $p_1, \ldots, p_n \in V$, the implication $\sigma(\varphi_{m,p_1}) \wedge \ldots \wedge \sigma(\varphi_{m,p_n}) \Rightarrow \mathit{false}$ is a consequence of Φ'. However, the implication $\varphi_{m,p_1} \wedge \ldots \wedge \varphi_{m,p_n} \Rightarrow \mathit{false}$ is not a consequence of Φ, and hence σ is not consequence-conservative. □

The signatures of the standard institutions presented above come naturally equipped with a notion of subsignature, hence signature inclusion, and a well-defined way of forming a union of signatures. These concepts can be captured in a categorical setting using *inclusion systems* [5, 9]. However, we will work with a slightly different version of this notion:

Definition 2.8. *An* ***inclusive category*** *is a category with a broad subcategory*[9] *which is a partially ordered class with a least element (denoted $\emptyset$), non-empty products (denoted $\cap$) and finite coproducts (denoted $\cup$), such that for each pair of objects A, B, the following is a pushout in the category:*

$$\begin{array}{ccc} A \cap B & \hookrightarrow & A \\ \downarrow & & \downarrow \\ B & \hookrightarrow & A \cup B \end{array}$$

□

For any objects A and B of an inclusive category, we write $A \subseteq B$ if there is an inclusion from A to B; the unique such inclusion will then be denoted by $\iota_{A \subseteq B} \colon A \hookrightarrow B$, or simply $A \hookrightarrow B$.

A functor between two inclusive categories is inclusive if it takes inclusions in the source category to inclusions in the target category.

Definition 2.9. *An institution $I = (\mathbf{Sign}, \mathbf{Sen}, \mathbf{Mod}, \models)$ is* ***inclusive***[10] *if*

- **Sign** *is an inclusive category,*
- **Sen** *is inclusive and preserves intersections,*[11] *and*
- *each model category is inclusive, and reduct functors are inclusive.*[12]

Moreover, we asume that reducts w.r.t. signature inclusions are surjective on objects. □

---

[9] That is, with the same objects as the original category.

[10] Even though we use the same term as in [9], since the overall idea is the same, on one hand, some of our assumptions here are weaker than in [9], and on the other hand, we require a bit more structure on the category of signatures.

[11] That is, for any family of signatures $\mathbb{S} \subseteq |\mathbf{Sign}|$, $\mathbf{Sen}(\bigcap \mathbb{S}) = \bigcap_{\Sigma \in \mathbb{S}} \mathbf{Sen}(\Sigma)$.

[12] That is, we have a model functor $\mathbf{Mod} \colon \mathbf{Sign}^{op} \to \mathbb{IC}at$, where $\mathbb{IC}at$ is the (quasi)category of inclusive categories and inclusive functors.

The empty object in the category of signatures will be referred to as the empty signature (indeed, in typical signature categories it is empty) and will be written as $\Sigma_\emptyset$.

Since in any inclusive institution the category of signatures has arbitrary intersections, for any set of sentences $\Phi \subseteq \bigcup_{\Sigma \in |\mathbf{Sign}|} \mathbf{Sen}(\Sigma)$, there exists the least signature $\mathsf{Sig}(\Phi)$ such that $\Phi \subseteq \mathbf{Sen}(\mathsf{Sig}(\Phi))$.

The assumption that reducts are surjective on models is rather mild and ensures that semantic consequence is not only preserved but also reflected under signature extension. Then, given $\Phi \subseteq \mathbf{Sen}(\Sigma)$ and $\varphi \in \mathbf{Sen}(\Sigma)$ (or, equivalently, $\mathsf{Sig}(\Phi) \cup \mathsf{Sig}(\varphi) \subseteq \Sigma$), we have that $\Phi \models_\Sigma \varphi$ if and only if $\Phi \models_{\mathsf{Sig}(\Phi) \cup \mathsf{Sig}(\varphi)} \varphi$. In particular, this justifies use of the notation $\Phi \models \varphi$ without any explicit reference to the signature over which sentences and consequence between them are considered. Moreover, $\Phi \models \varphi$ if and only if $|\mathbf{Mod}(\Sigma, \Phi)| \subseteq |\mathbf{Mod}(\Sigma, \{\varphi\})|$ for every signature $\Sigma \supseteq \mathsf{Sig}(\Phi) \cup \mathsf{Sig}(\varphi)$.

The institutions Prop, OWL and $\mathsf{FOL}^=$ sketched above in Examples 2.2, 2.3 and 2.4 can be equipped with the obvious inclusion system on their signatures and models, and then become inclusive institutions.

In inclusive institutions, if $\Sigma_1 \subseteq \Sigma_2$ via an inclusion $\iota \colon \Sigma_1 \hookrightarrow \Sigma_2$ and $M \in \mathbf{Mod}(\Sigma_2)$, we write $M|_{\Sigma_1}$ for $M|_\iota$. Note that $\mathbf{Sen}(\iota) \colon \mathbf{Sen}(\Sigma_1) \to \mathbf{Sen}(\Sigma_2)$ is the usual set-theoretic inclusion, hence its application may be omitted.

For some results, we need an amalgamation property on models. An inclusive institution I is called *(weakly) union-exact* if all intersection-union signature pushouts in **Sign** are (weakly) amalgamable. More specifically, the latter means that for any pushout

$$\begin{array}{ccc} \Sigma_1 \cap \Sigma_2 & \longrightarrow & \Sigma_1 \\ \downarrow & & \downarrow \\ \Sigma_2 & \longrightarrow & \Sigma_1 \cup \Sigma_2 \end{array}$$

in **Sign**, any pair $(M_1, M_2) \in \mathbf{Mod}(\Sigma_1) \times \mathbf{Mod}(\Sigma_2)$ that is *compatible* in the sense that M_1 and M_2 reduce to the same $(\Sigma_1 \cap \Sigma_2)$-model can be *amalgamated* to a unique (or weakly amalgamated to a not necessarily unique) $(\Sigma_1 \cup \Sigma_2)$-model: there exists a (unique) $M \in \mathbf{Mod}(\Sigma_1 \cup \Sigma_2)$ that reduces to M_1 and M_2, respectively.

The institutions Prop, OWL and $\mathsf{FOL}^=$ sketched above are all union-exact.

3 Conservative Extensions and Inseparability

An ontology is typically presented as a collection of concepts/objects, relations, properties and axioms — thus a presentation of a theory in some suitable logic, with OWL being a typical example. The goal of this paper is to study some concepts used in the research on ontologies and their modularisation independently of the logic in use. We make this precise by presenting these concepts in the context of an arbitrary (but fixed for now) inclusive institution $I = (\mathbf{Sign}, \mathbf{Sen}, \mathbf{Mod}, \models)$. The presentation below is based on the general concepts and facts conerning inclusive institutions, as spelled out in Sect. 2. To stay in tune with the literature and concerns of the domain we consider, we will adjust the terminology and notation appropriately.

An *ontology* $\mathcal{O}$ in a logic given as the institution I is just a set of sentences $\mathcal{O} \subseteq \bigcup_{\Sigma \in |\mathbf{Sign}|} \mathbf{Sen}(\Sigma)$ in I. As explained in Sect. 2, for each ontology $\mathcal{O}$ we have its signature $\mathsf{Sig}(\mathcal{O})$, which is the least signature over which all the sentences in $\mathcal{O}$ may be considered.

Note that if we want an ontology to be always considered over a larger signature with some extra symbols without changing its intended meaning, we need to add trivially true sentences that involve the additional symbols. In many typical institutions such sentences always exists (for instance, $p \vee \neg p$ in Prop, etc.); if this is not the case, we may want to expand our institution by some trivial sentences to "declare" that some extra symbols are to be considered.

Ontology inclusions give a starting notion to study relationships between ontologies. If $\mathcal{O} \subseteq \mathcal{O}'$ then we say that $\mathcal{O}'$ is an extension of $\mathcal{O}$. As in Sect. 2, conservativity of such an extension may be defined in two variants: based on models and based on semantic consequence (deduction), respectively. However, we are often interested in further nuances, where conservativity is considered up to an indicated signature of current interest.

Consider ontologies $\mathcal{O} \subseteq \mathcal{O}'$ and a signature $\Sigma \in |\mathbf{Sign}|$.

1. $\mathcal{O}'$ is a *model Σ-conservative extension (Σ-mCE)* of $\mathcal{O}$ if for every $(\mathsf{Sig}(\mathcal{O}) \cup \Sigma)$-model $\mathcal{I}$ of $\mathcal{O}$ there exists a $(\mathsf{Sig}(\mathcal{O}') \cup \Sigma)$-model $\mathcal{I}$' of $\mathcal{O}$' such that $\mathcal{I}|_\Sigma = \mathcal{I}'|_\Sigma$.
2. $\mathcal{O}'$ is a *consequence Σ-conservative extension (Σ-cCE)* of $\mathcal{O}$ if for every Σ-sentence α, we have $\mathcal{O}' \models \alpha$ iff $\mathcal{O} \models \alpha$.

Proposition 2.6 essentially applies here as well, so that the notion of model Σ-conservative extension is strictly stronger than that of consequence Σ-conservative extension, and it clearly does not depend on the expressiveness of the institution. Thus if $\mathcal{O}'$ is a Σ-mCE of $\mathcal{O}$ then $\mathcal{O}'$ is a Σ-cCE of $\mathcal{O}$ as well, while the converse does not hold. However, for propositional logic, the two concepts are equivalent, see Example 2.7.

We have parameterised above both concepts of conservative extension by a specific signature to indicate the focus of current interest. Further concepts will be developed in a similar fashion, taking the signature of interest into account. As this signature may vary here rather arbitrarily, we will need to adjust any ontology to cover it explicitly, turning the ontology into a presentation in I: for an ontology $\mathcal{O}$ and a signature Σ, we define $\mathcal{O}{\uparrow}\Sigma = (\mathsf{Sig}(\mathcal{O}) \cup \Sigma, \mathsf{Ax}(\mathcal{O}))$.

Now, when a signature of interest is indicated, the notion of ontology extension may be refined as follows. Again, this comes in two flavours: one based on models, the other on sentences (consequence).

Given ontologies $\mathcal{O}'$ and $\mathcal{O}$ and a signature Σ:

1. $\mathcal{O}'$ is a model Σ-extension of $\mathcal{O}$ if for all models $\mathcal{I}' \in |\mathbf{Mod}(\mathcal{O}'{\uparrow}\Sigma)|$ there is $\mathcal{I} \in |\mathbf{Mod}(\mathcal{O}{\uparrow}\Sigma)|$ such that $\mathcal{I}'|_\Sigma = \mathcal{I}|_\Sigma$.
2. $\mathcal{O}'$ is a consequence Σ-extension of $\mathcal{O}$ if for all Σ-sentences α, we have $\mathcal{O}'{\uparrow}\Sigma \models \alpha$ if $\mathcal{O}{\uparrow}\Sigma \models \alpha$.

Clearly, if $\mathcal{O} \subseteq \mathcal{O}'$ then $\mathcal{O}'$ is a model Σ-extension of $\mathcal{O}$, for any signature Σ. Moreover, essentially by Prop. 2.5, if $\mathcal{O}'$ is a model Σ-extension of $\mathcal{O}$ then it is a consequence Σ-extension of $\mathcal{O}$ as well.

The model Σ-extension condition may be rewritten as follows:

$$\{\mathcal{I}'|_\Sigma \mid \mathcal{I}' \in |\mathbf{Mod}(\mathcal{O}'{\uparrow}\Sigma)|\} \subseteq \{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}(\mathcal{O}{\uparrow}\Sigma)|\}$$

One may feel tempted to simplify this and instead write

$$\{\mathcal{I}'|_\Sigma \mid \mathcal{I}' \models \mathcal{O}'\} \subseteq \{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{O}\}$$

However, this formally makes little sense unless we assume $\Sigma \subseteq \mathsf{Sig}(\mathcal{O}') \cap \mathsf{Sig}(\mathcal{O})$. This would be a strong assumption concerning the signature of interest (even if $\mathcal{O} \subseteq \mathcal{O}'$), especially when we come to discussing robustness properties below. If this condition does not hold, it is not entirely clear what $\mathcal{I}|_\Sigma$ should mean. One possible interpretation might be "remove all model components whose names do not occur in Σ" (consider reducts to $\Sigma \cap \mathsf{Sig}(\mathcal{O}')$ and $\Sigma \cap \mathsf{Sig}(\mathcal{O})$, respectively). But even then, the two definitions depart: consider (in OWL) $\mathcal{O}' = \{C \sqsubseteq C\}$, $\mathcal{O} = \{C' \sqsubseteq C'\}$, and $\Sigma = \{C, C'\}$. Then according to our definition, $\mathcal{O}'$ is a model Σ-extension of $\mathcal{O}$, but this would not be the case if the apparently simplified condition was used.[13] In fact, the simpler condition cannot be met in a non-trivial way unless $\Sigma \cap \mathsf{Sig}(\mathcal{O}') = \Sigma \cap \mathsf{Sig}(\mathcal{O})$, another strong assumption we rather avoid.

One may now want to define a module in an ontology $\mathcal{O}$ to be another ontology $\mathcal{M}$ such that $\mathcal{M} \subseteq \mathcal{O}$ and the inclusion is conservative (either in the model-based sense, or in the consequence-based sense). However, we want this concept to work for an arbitrary signature of interest. The appropriate requirement is formulated in terms of *inseparability*. One intuition is that inseparability is a proper generalisation of conservative extension to a more symmetric situation.

Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be ontologies and Σ a signature. Then $\mathcal{O}_1$ and $\mathcal{O}_2$ are *model Σ-inseparable*, written $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$ if

$$\{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}(\mathcal{O}_1{\uparrow}\Sigma)|\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}(\mathcal{O}_2{\uparrow}\Sigma)|\}$$

Note that in the literature, a simpler condition is commonly used:

$$\{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{O}_1\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{O}_2\}$$

However, this "simplification" is dubious — all the comments concerning the definition of model Σ-extension above apply here as well.

Clearly, $\mathcal{O}_1$ and $\mathcal{O}_2$ are model Σ-inseparable iff $\mathcal{O}_1$ is a model Σ-extension of $\mathcal{O}_2$ and $\mathcal{O}_2$ is a model Σ-extension of $\mathcal{O}_1$. Moreover if $\mathcal{O}_1 \subseteq \mathcal{O}_2$ and $\Sigma \subseteq$

[13] This remains true even if $\mathcal{I}$ varies over models of arbitrary signatures, which seems to be a widespread understanding in the ontology modularity community. Note that $\mathcal{I} \models \mathcal{O}$ still entails that $\mathcal{I}$ interprets at least the symbols occurring in $\mathcal{O}$.

$\mathsf{Sig}(\mathcal{O}_1)$ then $\mathcal{O}_1$ and $\mathcal{O}_2$ are model Σ-inseparable iff $\mathcal{O}_2$ is a model Σ-conservative extension of $\mathcal{O}_1$.

Model Σ-inseparability provides a very strong form of equivalence between ontologies considered from the perspective given by Σ: $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$ guarantees that $\mathcal{O}_1$ can be replaced by $\mathcal{O}_2$ in any application that refers only to symbols from Σ. Moreover, since this notion does not depend on the expressibility of the underlying institution, we may arbitrarily strengthen the logic without affecting this equivalence.

Weaker versions of inseparability relations can be defined. To begin with, as usual, we consider a deductive variant: $\mathcal{O}_1$ and $\mathcal{O}_2$ are *consequence Σ-inseparable*, written $\mathcal{O}_1 \equiv^s_\Sigma \mathcal{O}_2$, if for all Σ-sentences φ

$$\mathcal{O}_1 \models \varphi \text{ iff } \mathcal{O}_2 \models \varphi$$

Let us recall here again that the semantic consequences might be taken over any signatures that encompass all symbols used either in the ontology or in the sentences considered.

Given a signature Σ, in many contexts we are not interested in preservation of all Σ-sentences, but it is sufficient to consider only sentences of some specific form that express the properties we really care about. This extra twist may be captured by considering a set of Σ-sentences $\Lambda \subseteq \mathbf{Sen}(\Sigma)$, and weakening consequence Σ-inseparability as follows: $\mathcal{O}_1$ and $\mathcal{O}_2$ are *Λ-consequence Σ-inseparable*, written $\mathcal{O}_1 \equiv^\Lambda_\Sigma \mathcal{O}_2$, if for all Σ-sentences $\varphi \in \Lambda$

$$\mathcal{O}_1 \models \varphi \text{ iff } \mathcal{O}_2 \models \varphi$$

For instance, in OWL, one relevant choice of the set Λ is to consider all subsumptions between atomic concepts.

It is easy to see now that indeed, the above three equivalences are gradually coarser:

Proposition 3.1. *For any signature Σ and set of Σ-sentences $\Lambda \subseteq \mathbf{Sen}(\Sigma)$, we have $\equiv^m_\Sigma \subseteq \equiv^s_\Sigma \subseteq \equiv^\Lambda_\Sigma$.* □

In particular, this means that if two ontologies are model Σ-inseparable then they are Σ-inseparable by any set of sentences, even if we strengthen the logic in use. Whatever sentences we add to our institution, no matter how strong they would be, two ontologies that are model Σ-inseparable will have the same consequences among them.

We mentioned above that one may want to consider various signatures Σ, changing the focus of interest through which ontologies are considered. In particular, this means that to use Λ-consequence Σ-inseparability, we have to provide the set of sentences over each such signature Σ. What one wants then is an inclusive functor $\Lambda\colon \mathbf{Sign} \to \mathbb{S}et$ with $\Lambda(\Sigma) \subseteq \mathbf{Sen}(\Sigma)$ for all signatures Σ. This implies that for $\Sigma' \subseteq \Sigma$, $\Lambda(\Sigma') \subseteq \Lambda(\Sigma)$, capturing the intuition that the sentences to be preserved cannot be disregarded when signature is enlarged. For any signature Σ, slightly abusing the notation, we write $\equiv^\Lambda_\Sigma$ for $\equiv^{\Lambda(\Sigma)}_\Sigma$.

Given the above arrangements, the inseparability relations defined are preserved when the signature considered is narrowed:

Proposition 3.2. *Given any signatures* $\Sigma' \subseteq \Sigma$*, we have* $\equiv^m_\Sigma \subseteq \equiv^m_{\Sigma'}$*,* $\equiv^s_\Sigma \subseteq \equiv^s_{\Sigma'}$*, and* $\equiv^A_\Sigma \subseteq \equiv^A_{\Sigma'}$*.* □

For a given institution, an *inseparability relation* is a family $\mathcal{S} = \langle \equiv^{\mathcal{S}}_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$ of equivalence relations on the family of presentations. The informal intuition we want to capture is that for any two ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$, $\mathcal{O}_1 \equiv^{\mathcal{S}}_\Sigma \mathcal{O}_2$ means that $\mathcal{O}_1$ and $\mathcal{O}_2$ are indistinguishable w.r.t. Σ, i.e., they represent the same knowledge of interest about the topics expressible in the signature Σ. Any specific definition of the inseparability relation determines the exact meaning of the terms "indistinguishable" and "the knowledge of interest". However, since "the knowledge of interest" relevant for a signature should not be disregarded when the signature is enlarged, it is desirable that the inseparability relations are monotone in the following sense:

Definition 3.3 ([10]). *An inseparability relation* $\mathcal{S} = \langle \equiv^{\mathcal{S}}_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$ *is* monotone *if*

1. *for any signatures* $\Sigma' \subseteq \Sigma$*,* $\equiv^{\mathcal{S}}_\Sigma \subseteq \equiv^{\mathcal{S}}_{\Sigma'}$ *(the inseparability relation gets finer when the signature gets larger), and*
2. *if* $\mathcal{O}_1 \subseteq \mathcal{O}_2 \subseteq \mathcal{O}_3$ *and* $\mathcal{O}_1 \equiv^{\mathcal{S}}_\Sigma \mathcal{O}_3$ *then* $\mathcal{O}_1 \equiv^{\mathcal{S}}_\Sigma \mathcal{O}_2$ *and* $\mathcal{O}_2 \equiv^{\mathcal{S}}_\Sigma \mathcal{O}_3$ *(the intuition here is: since larger ontologies capture more of "the knowledge of interest", we also require that any ontology squeezed between an ontology and its inseparable extension is inseparable from both of them).*

□

The inseparability relations defined above ($\langle \equiv^m_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$, $\langle \equiv^s_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$, and $\langle \equiv^A_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$) are typical examples we will use in the following. It is easy to show that all are monotone.

Monotonicity can be reformulated as robustness under signature restrictions. We now recall further robustness properties from the literature [27, 10].

Definition 3.4. *An inseparability relation* $\mathcal{S} = \langle \equiv^{\mathcal{S}}_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$ *is*

– robust under signature extensions *if for all ontologies* $\mathcal{O}_1$ *and* $\mathcal{O}_2$ *and all signatures* Σ*,* Σ' *with* $\Sigma' \cap (\mathsf{Sig}(\mathcal{O}_1) \cup \mathsf{Sig}(\mathcal{O}_2)) \subseteq \Sigma$

$$\mathcal{O}_1 \equiv_\Sigma \mathcal{O}_2 \text{ implies } \mathcal{O}_1 \equiv_{\Sigma'} \mathcal{O}_2$$

– robust under replacement *if for all ontologies* $\mathcal{O}$*,* $\mathcal{O}_1$ *and* $\mathcal{O}_2$ *and all signatures* Σ *with* $\mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$*, we have*

$$\mathcal{O}_1 \equiv_\Sigma \mathcal{O}_2 \text{ implies } \mathcal{O}_1 \cup \mathcal{O} \equiv_\Sigma \mathcal{O}_2 \cup \mathcal{O}$$

– robust under joins *if for all ontologies* $\mathcal{O}_1$ *and* $\mathcal{O}_2$ *and all signatures* Σ *with* $\mathsf{Sig}(\mathcal{O}_1) \cap \mathsf{Sig}(\mathcal{O}_2) \subseteq \Sigma$*, we have for* $i = 1, 2$

$$\mathcal{O}_1 \equiv_\Sigma \mathcal{O}_2 \text{ implies } \mathcal{O}_i \equiv_\Sigma \mathcal{O}_1 \cup \mathcal{O}_2$$

□

We have the following result on robustness:

Theorem 3.5. *Model inseparability is robust under replacement. In a union-exact inclusive institution, model inseparability is also robust under signature extensions and joins.*

Proof. Robustness under replacement: Consider ontologies $\mathcal{O}$, $\mathcal{O}_1$ and $\mathcal{O}_2$ and a signature Σ such that $\mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$ and $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$. We need to show that $\mathcal{O}_1 \cup \mathcal{O} \equiv^m_\Sigma \mathcal{O}_2 \cup \mathcal{O}$, which amounts to showing

$$\{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}((\mathcal{O}_1 \cup \mathcal{O})\uparrow\Sigma)|\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}((\mathcal{O}_2 \cup \mathcal{O})\uparrow\Sigma)|\}$$

By symmetry, it suffices to prove one inclusion. Let $\mathcal{I} \in |\mathbf{Mod}((\mathcal{O}_1 \cup \mathcal{O})\uparrow\Sigma)|$. Define $\mathcal{I}' = \mathcal{I}|_{\mathsf{Sig}(\mathcal{O}_1)\cup\Sigma}$. By $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$, we know that $\mathcal{I}'|_\Sigma$ has an expansion to an $\mathcal{O}_2\uparrow\Sigma$-model $\mathcal{I}''$. But since $\mathcal{I} \models \mathcal{O}$ and $\mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$, also $\mathcal{I}'' \models \mathcal{O}$. Hence $\mathcal{I}'' \in |\mathbf{Mod}((\mathcal{O}_2 \cup \mathcal{O})\uparrow\Sigma)|$, and obviously $\mathcal{I}|_\Sigma = \mathcal{I}|''_\Sigma$.

Robustness under signature extensions: Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be ontologies Σ, Σ' be signatures with $\Sigma' \cap (\mathsf{Sig}(\mathcal{O}_1) \cup \mathsf{Sig}(\mathcal{O}_2)) \subseteq \Sigma$. Assume $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$. We need to show that $\mathcal{O}_1 \equiv^m_{\Sigma'} \mathcal{O}_2$, which amounts to showing

$$\{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}(\mathcal{O}_1\uparrow\Sigma')|\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}(\mathcal{O}_2\uparrow\Sigma')|\}$$

By symmetry, it suffices to prove one inclusion. Let $\mathcal{I}'_1 \in \mathbf{Mod}(\mathcal{O}_1\uparrow\Sigma')$. Since $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$, $\mathcal{I}|_\Sigma$ has an expansion to an $\mathcal{O}_1\uparrow\Sigma$-model $\mathcal{I}_2$. From $\Sigma \subseteq \Sigma'$ and $\Sigma' \cap \mathsf{Sig}(\mathcal{O}_2) \subseteq \Sigma$ we get $\Sigma' \cap \mathsf{Sig}(\mathcal{O}_2\uparrow\Sigma) = \Sigma$. Since also $\Sigma' \cup \mathsf{Sig}(\mathcal{O}_2\uparrow\Sigma) = \mathsf{Sig}(\mathcal{O}_2\uparrow\Sigma')$ the following diagram

$$\begin{array}{ccc}
(\Sigma', \emptyset) & \hookrightarrow & \mathcal{O}_2\uparrow\Sigma' \\
\uparrow & & \uparrow \\
(\Sigma, \emptyset) & \hookrightarrow & \mathcal{O}_2\uparrow\Sigma
\end{array}$$

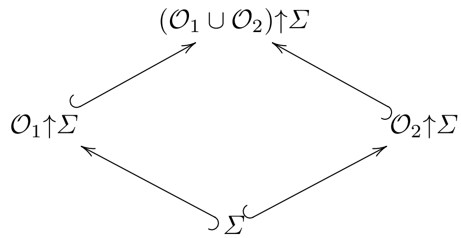
is an intersection-union-pushout in **Pres**. Hence, by weak union-exactness, we can amalgamate $\mathcal{I}'_1|_{\Sigma'}$ and $\mathcal{I}_2$ to $\mathcal{I}'_2 \in \mathbf{Mod}(\mathcal{O}_2\uparrow\Sigma')$, which gives us the desired expansion of $\mathcal{I}'_1|_{\Sigma'}$.

Robustness under joins: Consider ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$ and a signature Σ such that $\mathsf{Sig}(\mathcal{O}_1 \cap \mathcal{O}_2) \subseteq \Sigma$ and $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$. Then we need to show $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_1 \cup \mathcal{O}_2$ and $\mathcal{O}_2 \equiv^m_\Sigma \mathcal{O}_1 \cup \mathcal{O}_2$. We only prove the former; the latter follows by symmetry. We need to show

$$\{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}(\mathcal{O}_1\uparrow\Sigma)|\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \in |\mathbf{Mod}((\mathcal{O}_1 \cup \mathcal{O}_2)\uparrow\Sigma)|\}$$

The inclusion from right to left is obvious. For the converse inclusion, let $\mathcal{I}_1 \in |\mathbf{Mod}(\mathcal{O}_1\uparrow\Sigma)|$. Since $\mathcal{O}_1 \equiv^m_\Sigma \mathcal{O}_2$, $\mathcal{I}_1|_\Sigma$ has an expansion $\mathcal{I}_1 \in |\mathbf{Mod}(\mathcal{O}_2\uparrow\Sigma)|$. From $\mathsf{Sig}(\mathcal{O}_1 \cap \mathcal{O}_2) \subseteq \Sigma$ we get $\mathsf{Sig}(\mathcal{O}_1\uparrow\Sigma) \cap \mathsf{Sig}(\mathcal{O}_2\uparrow\Sigma) = \Sigma$. Moreover, we have

$\mathsf{Sig}(\mathcal{O}_1{\uparrow}\Sigma) \cup \mathsf{Sig}(\mathcal{O}_2{\uparrow}\Sigma) = (\mathcal{O}_1 \cup \mathcal{O}_2){\uparrow}\Sigma$. This implies that



is an intersection-union-pushout in **Pres**. Hence, by weak union-exactness, we can amalgamate $\mathcal{I}_1$ and $\mathcal{I}_2$ to $\mathcal{I}'' \in \mathbf{Mod}((\mathcal{O}_1 \cup \mathcal{O}_2){\uparrow}\Sigma)$, which gives us the desired expansion of $\mathcal{I}_1|_\Sigma$. ❏

4 Module Notions

Equipped with the concepts introduced in the previous sections, we are now ready to introduce the notion of an ontology module. In fact, following the literature, we will put forward a number of concepts, and will study their properties and their mutual relationships. As in Sect. 3, we work in the framework of a logical system formalised as an inclusive institution $I = (\mathbf{Sign}, \mathbf{Sen}, \mathbf{Mod}, \models)$ (Sect. 2).

The notions of a module we present below may be parameterised by an arbitrary inseparability relation $\mathcal{S} = \langle \equiv^{\mathcal{S}}_{\Sigma} \rangle_{\Sigma \in |\mathbf{Sign}|}$.

Definition 4.1 ([10]). *Let $\mathcal{O}$ be an ontology, $\mathcal{M} \subseteq \mathcal{O}$ and Σ a signature. We call $\mathcal{M}$*

- *a* (plain) Σ-module *of $\mathcal{O}$ induced by $\mathcal{S}$ if $\mathcal{M} \equiv^{\mathcal{S}}_{\Sigma} \mathcal{O}$;*
- *a* self-contained Σ-module *of $\mathcal{O}$ induced by $\mathcal{S}$ if $\mathcal{M} \equiv^{\mathcal{S}}_{\Sigma \cup \mathsf{Sig}(\mathcal{M})} \mathcal{O}$;*
- *a* depleting Σ-module *of $\mathcal{O}$ induced by $\mathcal{S}$ if $\mathcal{O} \setminus \mathcal{M} \equiv^{\mathcal{S}}_{\Sigma \cup \mathsf{Sig}(\mathcal{M})} \emptyset$.* □

Example 1.1 shows a plain ontology module. The intuition is that the module $\mathcal{M}$ already contains all the relevant information from $\mathcal{O}$ if attention is restricted to the concepts (symbols) in signature Σ. Note however that the module in Example 1.1 is not depleting: this follows from the fact that $\mathcal{O} \setminus \mathcal{M}$ has still some non-trivial consequences relevant w.r.t. Σ, e.g., $\mathcal{O} \setminus \mathcal{M} \models \mathsf{Male} \sqcap \exists \mathsf{has\_child}.\top \sqsubseteq \mathsf{Human}$.

The main advantage of depleting over plain modules is that minimal depleting modules exist, see Thm. 4.6 below. Therefore, DOL uses the minimal depleting module as semantics of the module extraction operator. It is unclear how one could give a definite semantics to this operator in terms of plain modules, because there may be multiple pairwise incomparable minimal plain modules.

The intuition of depleting modules is as follows: In addition to the properties of plain Σ-modules, for a depleting Σ-module $\mathcal{M}$ of $\mathcal{O}$, the difference $\mathcal{O} \setminus \mathcal{M}$ has no knowledge about $\Sigma \cup \mathsf{Sig}(\mathcal{M})$. This means that the difference of $\mathcal{O}$ and its module $\mathcal{M}$ does not entail any axioms over $\Sigma \cup \mathsf{Sig}(\mathcal{M})$ other than tautologies.

A different formulation of this observation involves the notion of safety. We say that $\mathcal{O}$ is *safe* for Σ if, for every ontology $\mathcal{O}'$ with $\mathsf{Sig}(\mathcal{O}) \cap \mathsf{Sig}(\mathcal{O}') \subseteq \Sigma$, we have that $\mathcal{O} \cup \mathcal{O}'$ is a model Σ-conservative extension of $\mathcal{O}'$. Alternatively, this notion can be formulated in terms of inseparability: an ontology $\mathcal{O}$ is safe for a signature Σ if and only if $\mathcal{O} \equiv^m_\Sigma \emptyset$.

Now if $\mathcal{M}$ is a depleting Σ-module, then $\mathcal{O} \setminus \mathcal{M}$ is safe for $\mathsf{Sig}(\mathcal{M})$ and so the module can be maintained separately outside of $\mathcal{O}$ without the risk of unintended interaction with the rest of $\mathcal{O}$. Also note that checking depleting Σ modules is exactly the same problem as deciding Σ-inseparability from the empty ontology.

In the rest of this paper we will focus on modules induced by model inseparability $\langle \equiv^m_\Sigma \rangle_{\Sigma \in |\mathbf{Sign}|}$, leaving similar developments for other inseparability relations introduced in Sect. 3 for future study. We therefore drop all qualifications "induced by $\mathcal{S}$" in the terminology below.

Modules induced by model inseparability are essentially based on model conservative extensions:

Proposition 4.2. *For any ontology $\mathcal{O}$, $\mathcal{M} \subseteq \mathcal{O}$ and signature Σ, $\mathcal{M}$ is a Σ-module of $\mathcal{O}$ if and only if $\mathcal{O}$ is a model Σ-conservative extension of $\mathcal{M}$.* □

We say that a subontology $\mathcal{M} \subseteq \mathcal{O}$ *covers* all the knowledge that $\mathcal{O}$ has about Σ if $\mathcal{O}$ is a consequence Σ-conservative extension of $\mathcal{M}$, that is, if for every sentence $\alpha \in \mathbf{Sen}(\Sigma)$, we have that $\mathcal{O} \models \alpha$ if and only if $\mathcal{M} \models \alpha$.

A "plain" Σ-module $\mathcal{M}$ of $\mathcal{O}$ covers all knowledge that $\mathcal{O}$ has about Σ. In fact, this claim holds also when any extension of the institution I with arbitrarily strong sentences (but the same signatures and models) is allowed.

The notion of self-contained module is stronger than the plain Σ-module notion in that it requires the module to preserve entailments that can be formulated in the interface signature *plus* the signature of the module. That is, it covers all the knowledge that $\mathcal{O}$ has about $\Sigma \cup \mathsf{Sig}(\mathcal{M})$. Formally, monotonicity of the model inseparability relations, see Prop. 3.2, easily implies:

Proposition 4.3. *If $\mathcal{M}$ is a self-contained Σ-module of $\mathcal{O}$, then $\mathcal{M}$ is a (plain) Σ-module of $\mathcal{O}$ as well.* □

Since $\equiv^m_\Sigma$ enjoys robustness under replacement (Thm. 3.5), we get as in [10]:

Proposition 4.4. *If $\mathcal{M}$ is a depleting Σ-module of $\mathcal{O}$, then it is a self-contained Σ-module.* □

Comparison of the various module notions can be carried out examining properties relevant for ontology reuse. The robustness properties for inseparability (see Def. 3.4) can be transferred to modules as follows:

Robustness under signature restrictions. This property means that a module of an ontology w.r.t. a signature Σ is also a module of this ontology w.r.t. any subsignature of Σ. This property is important because it means that we do not need to import a different module when we restrict the set of terms that we are interested in.

Robustness under signature extensions. This means that a module of an ontology $\mathcal{O}$ w.r.t. a signature Σ is also a module of $\mathcal{O}$ w.r.t. any $\Sigma' \supseteq \Sigma$ as long as $\Sigma' \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$. This means that we do not need to import a different module when extending the set of relevant terms with terms not from $\mathcal{O}$.

Robustness under replacement. This property means that if $\mathcal{M}$ is a module of $\mathcal{O}$ w.r.t. Σ, then the result of importing $\mathcal{M}$ into another ontology $\mathcal{O}'$ is a module of the result of importing $\mathcal{O}$ into $\mathcal{O}'$. Formally, for any ontology $\mathcal{O}'$, if $\mathcal{M}$ is a Σ-module of $\mathcal{O}$, then $\mathcal{O}' \cup \mathcal{M}$ is a Σ-module of $\mathcal{O}' \cup \mathcal{O}$. (The precise restrictions to signatures that are needed to ensure this property can vary.) This is called *module coverage* in the literature: importing a module does not affect its property of being a module.

Robustness under joins. It seems that this property of inseparability relations cannot be usefully transferred to ontology modules. However, together with robustness under replacement, it implies that it is not necessary to import two indistinguishable versions of the same ontology. This shows that it is still useful to have the property.

We have summarized the relevant properties of the modules of each kind in Table 1, which follow from the properties of inseparability relations stated in Sect. 3:

Theorem 4.5. *The module notions appearing as column heads in Table 1 have the properties appearing as row head, if marked with a ✓ or some additional assumptions that are needed. If marked with a ✗, there is a counterexample showing that that the property does not hold. It is assumed that all module notions are based on model inseparability.*

Indeed, the condition needed for robustness under replacement is very limited for plain modules, since the importing ontology $\mathcal{O}'$ must have a signature contained in the signature of interest Σ. This seems to be unrealistic in practice. The other module notions have a more liberal condition: $\mathsf{Sig}(\mathcal{O}') \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma \cup \mathsf{Sig}(\mathcal{M})$, which means that the importing ontology $\mathcal{O}'$ may overlap with the imported ontology $\mathcal{O}$ only w.r.t. the signature of interest plus that of the module. This is more realistic.

Given an ontology $\mathcal{O}$ and a signature of interest Σ, the crucial task is to determine a module $\mathcal{M}$ of $\mathcal{O}$ w.r.t. Σ. Clearly, such a module always exists: the entire ontology $\mathcal{O}$ is one example. However, what we are really interested in is *small* modules of $\mathcal{O}$ w.r.t. Σ. The following theorem establishes existence of such modules (see Thm. 72 in [10]), and so is a starting point for various methods of module extraction.

Theorem 4.6. *Let $\mathcal{O}$ be an ontology and Σ be a signature. Then there is a* unique *minimal depleting Σ-module of $\mathcal{O}$.* □

Properties	Module Notions		
	plain	self-contained	depleting
inseparability	$\mathcal{O} \equiv^m_\Sigma \mathcal{M}$	$\mathcal{O} \equiv^m_{\Sigma \cup \mathsf{Sig}(\mathcal{M})} \mathcal{M}$	$\mathcal{O} \setminus \mathcal{M} \equiv^m_{\Sigma \cup \mathsf{Sig}(\mathcal{M})} \emptyset$
mCE (cCE)	✓	✓	✓
self-contained	✗	✓	✓
depleting	✗	✗	✓
robustness under signature restrictions	✓	✓	✓
robustness under signature extensions	$\Sigma' \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$ plus weak union-exactness	$\Sigma' \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$ plus weak union-exactness	$\Sigma' \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma$ plus weak union-exactness
robustness under replacement	$\mathsf{Sig}(\mathcal{O}') \subseteq \Sigma$	$\mathsf{Sig}(\mathcal{O}') \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma \cup \mathsf{Sig}(\mathcal{M})$	$\mathsf{Sig}(\mathcal{O}') \cap \mathsf{Sig}(\mathcal{O}) \subseteq \Sigma \cup \mathsf{Sig}(\mathcal{M})$

Table 1. Properties of Σ-modules

5 Conclusions

We have generalised the basic notions of ontology module extraction to an arbitrary institution. They can now be applied to logics other than OWL, most notably first-order logic, but also modal logics and more exotic logics. For some nice properties of modules, union-exactness of the institution is needed. While many institutions enjoy this property, some do not, e.g. CASL [24].

We have entirely neglected questions of decidability or efficient computability of modules. While Thm. 4.6 provides a general method for computing the minimum depleting module, it is based on an oracle for inseparability. Future work should hence study computationally interesting approaches to module extraction, like different versions of locality, and generalize these to an arbitrary institution as well.

Acknowledgement

We thank Thomas Schneider for discussions and feedback.

References

1. David A. Basin, Manuel Clavel, and José Meseguer. Reflective metalogical frameworks. *ACM Trans. Comput. Log.*, 5(3):528–576, 2004.
2. Artur Boronat, Alexander Knapp, José Meseguer, and Martin Wirsing. What is a multi-modeling language? In A. Corradini and U. Montanari, editors, *Recent Trends in Algebraic Development Techniques, 19th International Workshop,*

WADT 2008, Pisa, Italy, June 13-16, 2008, Revised Selected Papers, volume 5486 of *Lecture Notes in Computer Science*, pages 71–87. Springer, 2008.
3. M. Cerioli and J. Meseguer. May I borrow your logic? (transporting logical structures along maps). *Theoretical Computer Science*, 173:311–347, 1997.
4. Manuel Clavel and José Meseguer. Reflection in conditional rewriting logic. *Theor. Comput. Sci.*, 285(2):245–288, 2002.
5. R. Diaconescu, J.A. Goguen, and P. Stefaneas. Logical Support for Modularisation. In *2nd Workshop on Logical Environments*, pages 83–130. CUP, New York, 1993.
6. Paul Doran, Valentina Tamma, and Luigi Iannone. Ontology module extraction for ontology reuse: An ontology engineering perspective. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, CIKM '07, pages 61–70, New York, NY, USA, 2007. ACM.
7. Francisco Durán and José Meseguer. Structured theories and institutions. *Theor. Comput. Sci.*, 309(1-3):357–380, 2003.
8. J. A. Goguen and R. M. Burstall. Institutions: Abstract model theory for specification and programming. *Journal of the Association for Computing Machinery*, 39:95–146, 1992. Predecessor in: LNCS 164, 221–256, 1984.
9. Joseph A. Goguen and Grigore Rosu. Composing hidden information modules over inclusive institutions. In O. Owe, S. Krogdahl, and T. Lyche, editors, *From Object-Orientation to Formal Methods, Essays in Memory of Ole-Johan Dahl*, volume 2635 of *Lecture Notes in Computer Science*, pages 96–123. Springer, 2004.
10. Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev. Logic-based ontology comparison and module extraction, with an application to DL-Lite. *Artif. Intell.*, 174(15):1093–1141, 2010.
11. Salvador Lucas and José Meseguer. Localized operational termination in general logics. In R. De Nicola and R. Hennicker, editors, *Software, Services, and Systems - Essays Dedicated to Martin Wirsing on the Occasion of His Retirement from the Chair of Programming and Software Engineering*, volume 8950 of *Lecture Notes in Computer Science*, pages 91–114. Springer, 2015.
12. Narciso Martí-Oliet, José Meseguer, and Miguel Palomino. Theoroidal maps as algebraic simulations. In J.L. Fiadeiro, P.D. Mosses, and F. Orejas, editors, *Recent Trends in Algebraic Development Techniques, 17th International Workshop, WADT 2004, Barcelona, Spain, March 27-29, 2004, Revised Selected Papers*, volume 3423 of *Lecture Notes in Computer Science*, pages 126–143. Springer, 2004.
13. J. Meseguer. General logics. In *Logic Colloquium 87*, pages 275–329. North Holland, 1989.
14. José Meseguer. Membership algebra as a logical framework for equational specification. In F. Parisi-Presicce, editor, *Recent Trends in Algebraic Development Techniques, 12th International Workshop, WADT'97, Tarquinia, Italy, June 1997, Selected Papers*, volume 1376 of *Lecture Notes in Computer Science*, pages 18–61. Springer, 1997.
15. José Meseguer and Narciso Martí-Oliet. From abstract data types to logical frameworks. In E. Astesiano, G. Reggio, and A. Tarlecki, editors, *Recent Trends in Data Type Specification, 10th Workshop on Specification of Abstract Data Types Joint with the 5th COMPASS Workshop, S. Margherita, Italy, May 30 - June 3, 1994, Selected Papers*, volume 906 of *Lecture Notes in Computer Science*, pages 48–80. Springer, 1994.
16. Till Mossakowski, Mihai Codescu, Fabian Neuhaus, and Oliver Kutz. The distributed ontology, modeling, and specification language - DOL. In A. Koslow and A. Buchsbaum, editors, *The Road to Universal Logic*, volume II of *Studies in Universal Logic*. Springer, 2015.

17. Till Mossakowski, Oliver Kutz, Mihai Codescu, and Christoph Lange. The distributed ontology, modeling and specification language. In C. Del Vescovo, T. Hahmann, D. Pearce, and D. Walther, editors, *WoMo 2013*, volume 1081 of *CEUR-WS online proceedings*, 2013.
18. Till Mossakowski, Oliver Kutz, and Christoph Lange. Semantics of the distributed ontology language: Institutes and institutions. In N. Martí-Oliet and M. Palomino, editors, *Recent Trends in Algebraic Development Techniques, 21th International Workshop, WADT 2012*, volume 7841 of *Lecture Notes in Computer Science*, pages 212–230. Springer, 2013.
19. Miguel Palomino, José Meseguer, and Narciso Martí-Oliet. A categorical approach to simulations. In J.L. Fiadeiro, N. Harman, M. Roggenbach, and J.J.M.M. Rutten, editors, *Algebra and Coalgebra in Computer Science: First International Conference, CALCO 2005, Swansea, UK, September 3-6, 2005, Proceedings*, volume 3629 of *Lecture Notes in Computer Science*, pages 313–330. Springer, 2005.
20. D. Sannella and A. Tarlecki. Specifications in an arbitrary institution. *Information and Control*, 76:165–210, 1988. Earlier version in *Proceedings, International Symposium on the Semantics of Data Types*, LNCS, Volume 173, Springer, 1984.
21. Donald Sannella and Andrzej Tarlecki. *Foundations of Algebraic Specification and Formal Software Development.* Monographs in Theoretical Computer Science. Springer, Berlin, 2012.
22. Sonia Santiago, Santiago Escobar, Catherine Meadows, and José Meseguer. A formal definition of protocol indistinguishability and its verification using Maude-NPA. In S. Mauw and C.D. Jensen, editors, *Security and Trust Management - 10th International Workshop, STM 2014, Wroclaw, Poland, September 10-11, 2014. Proceedings*, volume 8743 of *Lecture Notes in Computer Science*, pages 162–177. Springer, 2014.
23. Michael Schneider, Sebastian Rudolph, and Geoff Sutcliffe. Modeling in OWL 2 without restrictions. In M. Rodriguez-Muro, S. Jupp, and K. Srinivas, editors, *Proceedings of the 10th International Workshop on OWL: Experiences and Directions (OWLED 2013) co-located with 10th Extended Semantic Web Conference (ESWC 2013), Montpellier, France, May 26-27, 2013.*, volume 1080 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.
24. Lutz Schröder, Till Mossakowski, Andrzej Tarlecki, Bartek Klin, and Piotr Hoffman. Amalgamation in the semantics of CASL. *Theor. Comput. Sci.*, 331(1):215–247, 2005.
25. Julian Seidenberg and Alan L. Rector. Web ontology segmentation: analysis, classification and use. In L. Carr, D. De Roure, A. Iyengar, C.A. Goble, and M. Dahlin, editors, *Proceedings of the 15th international conference on World Wide Web, WWW 2006, Edinburgh, Scotland, UK, May 23-26, 2006*, pages 13–22. ACM, 2006.
26. Heiner Stuckenschmidt and Michel C. A. Klein. Structure-based partitioning of large concept hierarchies. In S.A. McIlraith, D. Plexousakis, and F. van Harmelen, editors, *The Semantic Web - ISWC 2004: Third International Semantic Web Conference, Hiroshima, Japan, November 7-11, 2004. Proceedings*, volume 3298 of *Lecture Notes in Computer Science*, pages 289–303. Springer, 2004.
27. Heiner Stuckenschmidt, Christine Parent, and Stefano Spaccapietra, editors. *Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization*, volume 5445 of *Lecture Notes in Computer Science*. Springer, 2009.