

Safe and Economic Re-Use of Ontologies: A Logic-Based Methodology and Tool Support

Ernesto Jiménez-Ruiz¹, Bernardo Cuenca Grau², Ulrike Sattler³,
Thomas Schneider³, and Rafael Berlanga¹

¹ Universitat Jaume I, Spain
{berlanga,ejimenez}@uji.es

² University of Oxford, UK
berg@comlab.ox.ac.uk

³ University of Manchester, UK
{sattler,schneider}@cs.man.ac.uk

Abstract. Driven by application requirements and using well-understood theoretical results, we describe a novel methodology and a tool for modular ontology design. We support the user in the *safe* use of imported symbols and in the *economic* import of the relevant part of the imported ontology. Both features are supported in a well-understood way: safety guarantees that the semantics of imported concepts is not changed, and economic import guarantees that no difference can be observed between importing the whole ontology and importing the relevant part.

1 Motivation

Ontology design and maintenance require an expertise in both the domain of application and the ontology language. Realistic ontologies typically model different aspects of an application domain at various levels of granularity; prominent examples are the National Cancer Institute Ontology (NCI)¹ [1], which describes diseases, drugs, proteins, etc., and GALEN², which represents knowledge mainly about the human anatomy, but also about other domains such as drugs.

Ontologies such as NCI and GALEN are used in bio-medical applications as *reference ontologies*, i.e., ontology developers reuse these ontologies and customise them for their specific needs. For example, ontology designers use concepts³ from NCI or GALEN and refine them (e.g., add new sub-concepts), generalise them (e.g., add new super-concepts), or refer to them when expressing a property of some other concept (e.g., define the concept `Polyarticular_JRA` by referring to the concept `Joint` from GALEN).

One of such use cases is the development within the Health-e-Child project of an ontology, called JRAO, to describe a kind of arthritis called JRA (Juvenile

¹ Online browser: <http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do>, latest version: ftp://ftp1.nci.nih.gov/pub/cacore/EVS/NCI_Thesaurus

² <http://www.co-ode.org/galen>

³ We use the Description Logic terms “concept” and “role” instead of the OWL terms “class” and “property”.

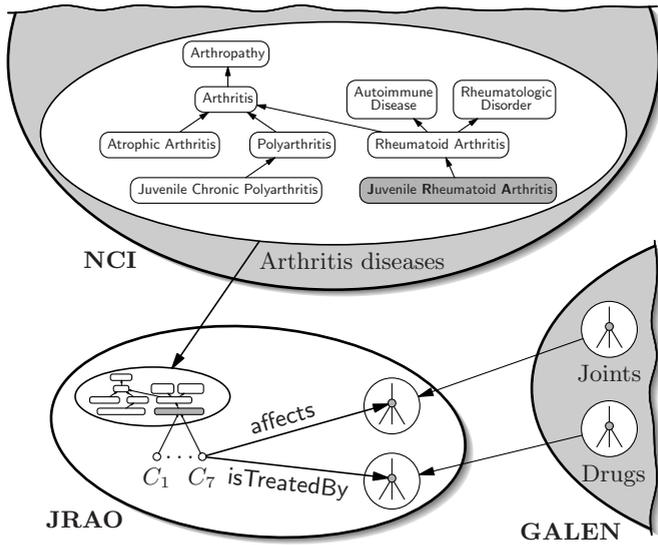


Fig. 1. Constructing the ontology JRAO reusing fragments of GALEN and NCI

Rheumatoid Arthritis).⁴ Following the ILAR⁵, JRAO describes the kinds of JRA. Those are distinguished by several factors such as the joints affected or the occurrence of fever, and each type of JRA requires a different treatment. GALEN and NCI contain information that is relevant to JRA, such as detailed descriptions of the human joints as well as diseases and their symptoms. Figure 1 gives a fragment of NCI that defines JRA. It also shows our reuse scenario, where C_1, \dots, C_7 refer to the kinds of JRA to be defined in JRAO.

The JRAO developers want to reuse knowledge from NCI and GALEN for three reasons: (a) they want to save time through reusing existing ontologies rather than writing their own; (b) they value knowledge that is commonly accepted by the community and used in similar applications; (c) they are not experts in all areas covered by NCI and GALEN.

Currently, GALEN, NCI, and JRAO are written in OWL DL [2], and thus they come with a logic-based semantics, which allows for powerful reasoning services for classification and consistency checking. Thus, ontology reuse should take into account the semantics and, more precisely, should provide the following two guarantees. First, when reusing knowledge from NCI and GALEN, the developers of JRAO do not want to change the original meaning of the reused concepts. For example, due to (b) and (c) above, if it follows from the union of JRAO and NCI that JRA is a genetic disorder, then it also follows from NCI alone. Second, only small parts of large ontologies like NCI and GALEN are relevant

⁴ See <http://www.health-e-child.org>. This project aims at creating a repository of ontologies that can be used by clinicians in various applications.

⁵ Int. League of Associations for Rheumatology <http://www.ilarportal.org/>

to the sub-types of JRA. For efficiency and succinctness, the JRAO developers want to import only those axioms from NCI and GALEN that are relevant for JRAO. By importing only fragments of NCI and GALEN, one should not lose important information; for example, if it follows from the union of the JRAO and NCI that JRA is a genetic disorder, then this also follows from the union of JRAO and the chosen fragment of NCI.

Our scenario has two main points in common with other ontology design scenarios: the ontology developer wants to reuse knowledge without damaging it, and also to import only the relevant parts of an existing ontology. To support these scenarios whilst providing the two above guarantees, a logic-based approach to reuse is required. Current tools that support reuse, however, do not implement a logic-based solution and thus do not provide the above guarantees—and neither do existing guidelines and “best practices” for ontology design.

In this paper, we propose a methodology for ontology design in scenarios involving reuse which is based on a well-understood logic-based framework [3]. We describe a tool that implements this methodology and report on experiments.

2 Preliminaries on Modularity

Based on the application scenario in Section 1, we define the notions of a *conservative extension*, *safety*, and *module* [4, 3]. For simplicity of the presentation, we restrict ourselves to the description logic \mathcal{SHIQ} , which covers most of OWL DL [2]. Therefore the ontologies, entailments and signatures we consider are relative to \mathcal{SHIQ} . The results mentioned in this section, however, can be extended to \mathcal{SHOIQ} and therefore OWL DL [3]. In this and the following section, we have omitted a few technical details, for instance proofs of the propositions. They can be found in a technical report available at <http://www.cs.man.ac.uk/~schneidt/publ/safe-eco-reuse-report.pdf>.

2.1 The Notions of Conservative Extension and Safety

As mentioned in Section 1, when reusing knowledge from NCI and GALEN, the developer of JRAO should not change the original meaning of the reused concepts. This requirement can be formalised using the notion of a *conservative extension* [4, 5]. In the following, we use $\text{Sig}()$ to denote the signature of an ontology or an axiom.

Definition 1 (Conservative Extension). *Let $\mathcal{T}_1 \subseteq \mathcal{T}$ be ontologies, and \mathbf{S} a signature. We say that \mathcal{T} is an \mathbf{S} -conservative extension of \mathcal{T}_1 if, for every axiom α with $\text{Sig}(\alpha) \subseteq \mathbf{S}$, we have $\mathcal{T} \models \alpha$ iff $\mathcal{T}_1 \models \alpha$; \mathcal{T} is a conservative extension of \mathcal{T}_1 if \mathcal{T} is an \mathbf{S} -conservative extension of \mathcal{T}_1 for $\mathbf{S} = \text{Sig}(\mathcal{T}_1)$.*

Definition 1 applies to our example as follows: $\mathcal{T}_1 = \text{NCI}$ is the ontology to be reused, \mathcal{T} is the union of JRAO and NCI, \mathbf{S} represents the symbols reused from NCI, such as JRA and Rheumatologic_Disorder, and α stands for any axiom over the reused symbols only, e.g., $\text{JRA} \sqsubseteq \text{Rheumatologic_Disorder}$.

Definition 1 assumes that the ontology to be reused (e.g. NCI) is static. In practice, however, ontologies such as NCI are under development and may evolve beyond the control of the JRAO developers. Thus, it is convenient to keep NCI separate from the JRAO and make its axioms available on demand via a reference such that the developers of the JRAO need not commit to a particular version of NCI. The notion of *safety* [3] can be seen as a stronger version of conservative extension that abstracts from the particular ontology to be reused and focuses only on the reused *symbols*.

Definition 2 (Safety for a Signature). *Let \mathcal{T} be an ontology and \mathbf{S} a signature. We say that \mathcal{T} is safe for \mathbf{S} if, for every ontology \mathcal{T}' with $\text{Sig}(\mathcal{T}) \cap \text{Sig}(\mathcal{T}') \subseteq \mathbf{S}$, we have that $\mathcal{T} \cup \mathcal{T}'$ is a conservative extension of \mathcal{T}' .*

2.2 The Notion of Module

As mentioned in Section 1, by importing only a fragment of NCI and GALEN, one should not lose important information. This idea can be formalised using the notion of a *module* [3]. Intuitively, when checking an arbitrary entailment over the signature of the JRAO, importing a module of NCI should give exactly the same answers as if the whole NCI had been imported.

Definition 3 (Module for a Signature). *Let $\mathcal{T}'_1 \subseteq \mathcal{T}'$ be ontologies and \mathbf{S} a signature. We say that \mathcal{T}'_1 is a module for \mathbf{S} in \mathcal{T}' (or an \mathbf{S} -module in \mathcal{T}') if, for every ontology \mathcal{T} with $\text{Sig}(\mathcal{T}) \cap \text{Sig}(\mathcal{T}'_1) \subseteq \mathbf{S}$, we have that $\mathcal{T} \cup \mathcal{T}'_1$ is a conservative extension of $\mathcal{T} \cup \mathcal{T}'$ for $\text{Sig}(\mathcal{T})$.*

The notions of safety and module are related as follows:

Proposition 4 ([3], Safety vs. Modules). *If $\mathcal{T}' \setminus \mathcal{T}'_1$ is safe for $\mathbf{S} \cup \text{Sig}(\mathcal{T}'_1)$, then \mathcal{T}'_1 is an \mathbf{S} -module in \mathcal{T}' .*

2.3 Locality Conditions

The decision problems associated with conservative extensions, safety and modules—i.e., whether \mathcal{T} is an \mathbf{S} -conservative extension of \mathcal{T}_1 , whether \mathcal{T} is safe for \mathbf{S} , or whether \mathcal{T}'_1 is an \mathbf{S} -module in \mathcal{T} —are undecidable for *SHOIQ* [6, 3]. Sufficient conditions for safety have been proposed: if an ontology satisfies such conditions, then we can guarantee that it is safe, but the converse does not necessarily hold [3]. By means of Proposition 4, such conditions could be used for extracting modules. A particularly useful condition is *locality* [3]: it is widely applicable in practice and it can be checked syntactically.

As mentioned in Section 1, when using a symbol from NCI or GALEN, the JRAO developers may refine it, extend it, or refer to it for expressing a property of another symbol. The simultaneous refinement and generalisation of a given “external” symbol, however, may compromise safety. For example, JRAO cannot simultaneously contain the following axioms:

$$\text{Polyarticular_JRA} \sqsubseteq \underline{\text{JRA}} \quad (\perp\text{-local}) \quad (1)$$

$$\underline{\text{Juvenile_Chronic_Polyarthritis}} \sqsubseteq \text{Polyarticular_JRA} \quad (\top\text{-local}) \quad (2)$$

where the underlined concepts are reused from NCI, see Figure 1. These axioms imply Juvenile_Chronic_Polyarthritis \sqsubseteq JRA, and therefore an ontology containing axioms (1) and (2) is not safe w.r.t. $\mathbf{S} = \{\text{JRA}, \text{Juvenile_Chronic_Polyarthritis}\}$. Thus, when designing sufficient conditions for safety, we are faced with a fundamental choice depending on whether the ontology designer wants to reuse or generalise the reused concepts. Each choice leads to a different locality condition.

The following definition introduces these conditions and refers to Figure 2. In this figure, A^\dagger and R^\dagger stand for concept and role names not in \mathbf{S} . The letters C and R denote arbitrary concepts and roles.

Definition 5 (Syntactic \perp -Locality and \top -Locality). *Let \mathbf{S} be a signature. An axiom α is \perp -local w.r.t. \mathbf{S} (\top -local w.r.t. \mathbf{S}) if $\alpha \in \text{Ax}(S)$, as defined in Figure 2 (a) ((b)). An ontology \mathcal{T} is \perp -local (\top -local) w.r.t. \mathbf{S} if α is \perp -local (\top -local) w.r.t. \mathbf{S} for all $\alpha \in \mathcal{T}$.*

(a) \perp -Locality	Let $C^\dagger \in \text{Con}(\bar{\mathbf{S}})$, $C_{(i)}^s \in \text{Con}(\mathbf{S})$
$\text{Con}(\bar{\mathbf{S}}) ::= A^\dagger \mid \neg C^s \mid C \sqcap C^\dagger \mid C^\dagger \sqcap C \mid \exists R.C^\dagger \mid \geq n R.C^\dagger \mid (\exists R^\dagger.C) \mid (\geq n R^\dagger.C)$	
$\text{Con}(\mathbf{S}) ::= \neg C^\dagger \mid C_1^s \sqcap C_2^s$	
$\text{Ax}(\mathbf{S}) ::= C^\dagger \sqsubseteq C \mid C \sqsubseteq C^s \mid R^\dagger \sqsubseteq R \mid \text{Trans}(R^\dagger)$	
(b) \top -Locality	Let $C^s \in \text{Con}(\mathbf{S})$, $C_{(i)}^\dagger \in \text{Con}(\bar{\mathbf{S}})$
$\text{Con}(\mathbf{S}) ::= \neg C^\dagger \mid C \sqcap C^s \mid C^s \sqcap C \mid \exists R.C^s \mid \geq n R.C^s$	
$\text{Con}(\bar{\mathbf{S}}) ::= A^\dagger \mid \neg C^s \mid C_1^\dagger \sqcap C_2^\dagger \mid \exists R^\dagger.C^\dagger \mid \geq n R^\dagger.C^\dagger$	
$\text{Ax}(\mathbf{S}) ::= C^s \sqsubseteq C \mid C \sqsubseteq C^\dagger \mid R \sqsubseteq R^\dagger \mid \text{Trans}(R^\dagger)$	

Fig. 2. Syntactic locality conditions

Axiom (2) is \top -local w.r.t. $\mathbf{S} = \{\text{Juvenile_Chronic_Polyarthritis}\}$, and Axiom (1) is \perp -local w.r.t. $\mathbf{S} = \{\text{JRA}\}$. Note that the locality conditions allow us to refer to a reused concept for expressing a property of some other concept; for example, the axiom Polyarticular_JRA $\sqsubseteq \geq 5$ affects Joint is \perp -local w.r.t. $\mathbf{S} = \{\text{Joint}\}$.

Both \top -locality and \perp -locality are sufficient for safety:

Proposition 6 ([3], Locality Implies Safety). *If an ontology \mathcal{T} is \perp -local or \top -local w.r.t. \mathbf{S} , then \mathcal{T} is safe for \mathbf{S} .*

Propositions 4 and 6 suggest the following definition of modules in terms of locality.

Definition 7. *Let $\mathcal{T}_1 \subseteq \mathcal{T}$ be ontologies, and \mathbf{S} a signature. We say that \mathcal{T}_1 is a \perp -module (\top -module) for \mathbf{S} in \mathcal{T} if $\mathcal{T} \setminus \mathcal{T}_1$ is \perp -local (\top -local) w.r.t. $\mathbf{S} \cup \text{Sig}(\mathcal{T}_1)$.*

We illustrate these notions by an example. Figure 3 (a) shows an example ontology (TBox). The set of external symbols is $\mathbf{S}_0 = \{A, t_1, t_2\}$. In order to extract

$\mathcal{T} = \{ \begin{array}{lll} A \sqsubseteq A_2, & A_2 \sqsubseteq \forall s_2.E_2, & A_1 \sqsubseteq A, \quad \forall s_1.E_1 \sqsubseteq A_1, \\ A_2 \sqsubseteq \exists r_2.C_2, & A_2 \sqsubseteq \forall t_2.F_2, & \exists r_1.C_1 \sqsubseteq A_1, \quad \forall t_1.F_1 \sqsubseteq A_1 \\ A_2 \sqsubseteq \forall r_2.D_2, & & \forall r_1.D_1 \sqsubseteq A_1 \end{array} \}$
--

(a) The TBox

Consideration	Consequence
(1) $A \in \mathbf{S}_0, A_2 \notin \mathbf{S}_0 \Rightarrow A \sqsubseteq A_2 \notin \text{Ax}(\mathbf{S}_0)$	$\mathbf{S}_1 = \mathbf{S}_0 \cup \{A_2\}$
(2) $A_2 \in \mathbf{S}_1, C_2 \notin \mathbf{S}_1 \Rightarrow \exists r_2.C_2 \in \text{Con}(\bar{\mathbf{S}}_1)$ $\Rightarrow A_2 \sqsubseteq \exists r_2.C_2 \notin \text{Ax}(\mathbf{S}_1)$	$\mathbf{S}_2 = \mathbf{S}_1 \cup \{r_2, C_2\}$
(3) $A_2 \in \mathbf{S}_2, r_2 \in \mathbf{S}_2, D_2 \notin \mathbf{S}_2 \Rightarrow \exists r_2.\neg D_2 \notin \text{Con}(\bar{\mathbf{S}}_2)$ $\Rightarrow \neg\exists r_2.\neg D_2 \notin \text{Con}(\bar{\mathbf{S}}_2) \Rightarrow A_2 \sqsubseteq \forall r_2.D_2 \notin \text{Ax}(\mathbf{S}_2)$	$\mathbf{S}_3 = \mathbf{S}_2 \cup \{D_2\}$
(4) $A_2 \in \mathbf{S}_3, s_2 \notin \mathbf{S}_3, E_2 \notin \mathbf{S}_3 \Rightarrow \exists s_2.\neg E_2 \in \text{Con}(\bar{\mathbf{S}}_3)$ $\Rightarrow \neg\exists s_2.\neg E_2 \in \text{Con}(\bar{\mathbf{S}}_3) \Rightarrow A_2 \sqsubseteq \forall s_2.E_2 \in \text{Ax}(\mathbf{S}_3)$	$\mathbf{S}_4 = \mathbf{S}_3$
(5) analogous to (3)	$\mathbf{S}_5 = \mathbf{S}_4 \cup \{F_2\}$
$\mathbf{S}_5 = \{A, A_2, C_2, D_2, F_2, r_2, t_1, t_2\}$	
The \perp -module consists of \mathbf{S}_5 and all axioms $\alpha \in \mathcal{T}$ with $\text{Sig}(\alpha) \subseteq \mathbf{S}_5$.	

(b) Extracting the \perp -module

Fig. 3. An example illustrating \perp - and \top -modules

the \perp -module, we extend \mathbf{S}_0 stepwise as in Figure 3 (b). The \top -module is obtained analogously and consists of $\mathbf{S}'_5 = \{A, A_1, C_1, D_1, F_1, r_1, t_1, t_2\}$ and all axioms $\alpha \in \mathcal{T}$ with $\text{Sig}(\alpha) \subseteq \mathbf{S}'_5$.

It is clear that \perp -modules and \top -modules are modules as in Definition 3:

Proposition 8 ([3], Locality-based Modules are Modules). *Let \mathcal{T}_1 be either a \perp -module or a \top -module for \mathbf{S} in \mathcal{T} and let $\mathbf{S}' = \mathbf{S} \cup \text{Sig}(\mathcal{T}_1)$. Then \mathcal{T}_1 is an \mathbf{S}' -module in \mathcal{T} .*

These modules enjoy an important property which determines their scope: suppose that \mathcal{T}_1 (\mathcal{T}_2) is a \perp -module (\top -module) for \mathbf{S} in \mathcal{T} , then \mathcal{T}_1 (\mathcal{T}_2) will contain all super-concepts (sub-concepts) in \mathcal{T} of all concepts in \mathbf{S} :

Proposition 9 ([3], Module Scope). *Let \mathcal{T} be an ontology, X, Y be concept names in $\mathcal{T} \cup \{\top\} \cup \{\perp\}$, $\alpha := (X \sqsubseteq Y)$, $\beta := (Y \sqsubseteq X)$, and $\mathcal{T}_X \subseteq \mathcal{T}$ with $X \in \mathbf{S}$. Then the following statements hold.*

- (i) *If \mathcal{T}_X is a \perp -module in \mathcal{T} for \mathbf{S} , then $\mathcal{T}_X \models \alpha$ iff $\mathcal{T} \models \alpha$.*
- (ii) *If \mathcal{T}_X is a \top -module in \mathcal{T} for \mathbf{S} , then $\mathcal{T}_X \models \beta$ iff $\mathcal{T} \models \beta$.*

For example, if we were to reuse the concept JRA from NCI as shown in Figure 1 by extracting a \perp -module for a signature that contains JRA, such a module would contain all the super-concepts of JRA in NCI, namely Rheumatoid_Arthritis, Autoimmune_Disease, Rheumatologic_Disorder, Arthritis, and Arthropathy. Since such a fragment is a module, it will contain the axioms necessary for entailing those subsumption relations between the listed concepts that hold in NCI.

Finally, given \mathcal{T} and \mathbf{S} , there is a unique minimal \perp -module and a unique minimal \top -module for \mathbf{S} in \mathcal{T} [3]. We denote these modules by $\text{UpMod}(\mathcal{T}, \mathbf{S})$ and $\text{LoMod}(\mathcal{T}, \mathbf{S})$. This is motivated by the alternative terms “upper/lower module” that refer to the property from Proposition 9. Following a similar approach as exemplified in Figure 3 (b), the modules can be computed efficiently [3].

3 A Novel Methodology for Ontology Reuse

Based on our scenario in Section 1 and the theory of modularity summarised in Section 2, we propose a novel methodology for designing an ontology when knowledge is to be borrowed from several external ontologies. This methodology provides precise guidelines for ontology developers to follow, and ensures that a set of logical guarantees will hold at certain stages of the design process.

3.1 The Methodology

We propose the working cycle given in Figure 4. This cycle consists of an *offline phase*—which is performed independently from the current contents of the external ontologies—and an *online phase*—where knowledge from the external ontologies is extracted and transferred into the current ontology. Note that this separation is not strict: The first phase is called “offline” simply because it does not need to be performed online. However, the user may still choose to do so.

The Offline Phase starts with the ontology \mathcal{T} being developed, e.g., JRAO. The ontology engineer specifies the set \mathbf{S} of symbols to be reused from external ontologies, and associates to each symbol the external ontology from which it will be borrowed. In Figure 4 this signature selection is represented in the *Repeat* loop: each $\mathbf{S}_i \subseteq \mathbf{S}$ represents the external symbols to be borrowed from a particular ontology \mathcal{T}'_i ; in our example, we have $\mathbf{S} = \mathbf{S}_1 \uplus \mathbf{S}_2$, where \mathbf{S}_1 is associated with NCI and contains JRA, and \mathbf{S}_2 is associated with GALEN and contains symbols related to joints and drugs. This part of the offline phase may involve an “online” component since the developer may browse through the external ontologies to choose the symbols she wants to import.

Next, the ontology developer decides, for each \mathbf{S}_i , whether she wants to refine or generalise the symbols from this set. For instance, in the reuse example shown in Figure 1, the concept JRA from NCI is refined by the sub-concepts abbreviated C_1, \dots, C_7 . In both cases, the user may also reference the external symbols via roles; in our example, certain types of JRA are defined by referencing concepts in GALEN (e.g., joints) via the roles *affects* and *isTreatedBy*. As argued in Section 1, refinement and generalisation, combined with reference, constitute the main possible intentions when reusing external knowledge. Therefore it is reasonable for the user, both from the modelling and tool design perspectives, to declare her intention. This step is represented by the *For* loop in Figure 4.

At this stage, we want to ensure that the designer of \mathcal{T} does not change the original meaning of the reused concepts, independently of what their particular meaning is in the external ontologies. This requirement can be formalised using the notion of safety introduced in Section 2:

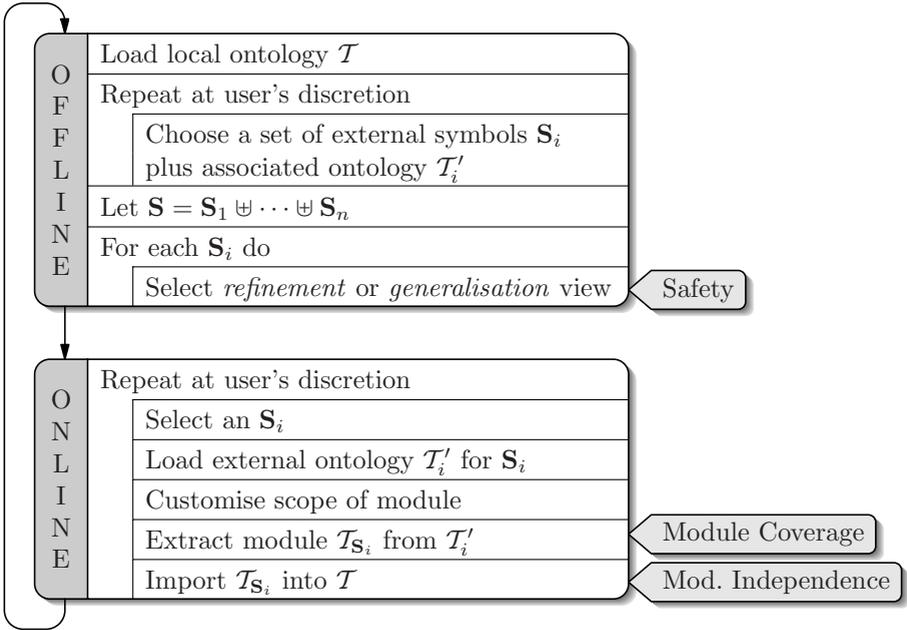


Fig. 4. The two phases of import with the required guarantees

Definition 10 (Safety Guarantee). *Given an ontology \mathcal{T} and signatures $\mathbf{S}_1, \dots, \mathbf{S}_n$, \mathcal{T} guarantees safety if \mathcal{T} is safe for \mathbf{S}_i for all $1 \leq i \leq n$.*

In the next subsection, we will show how to guarantee safety.

In the Online Phase, the ontology engineer imports the relevant knowledge from each of the external ontologies. As argued in Section 1 and 2, we aim at extracting only those fragments from the external ontologies that are relevant to the reused symbols, and therefore the extracted fragments should be modules in the sense of Definition 3.

As shown in Figure 4, the import for each external ontology \mathcal{T}'_i is performed in four steps. First, \mathcal{T}'_i is loaded; by doing so, the ontology engineer commits to a particular version of it. Second, the scope of the module to be extracted from \mathcal{T}'_i is customised; in practice, this means that the ontology engineer is given a view of \mathcal{T}'_i and enabled to extend \mathbf{S}_i by specifying requirements such as: “The module has to contain the concept ‘joint’, all its direct super-concepts and two levels of its sub-concepts”. In the third step, the actual fragment of \mathcal{T}'_i is extracted. At this stage, we should ensure that the extracted fragment is a module for the customised signature according to Definition 3. Therefore, the following guarantee should be provided for each external ontology and customised signature:

Definition 11 (Module Coverage Guarantee). *Let \mathbf{S} be a signature and $\mathcal{T}'_{\mathbf{S}}, \mathcal{T}'$ be ontologies with $\mathcal{T}'_{\mathbf{S}} \subseteq \mathcal{T}'$ such that $\mathbf{S} \subseteq \text{Sig}(\mathcal{T}'_{\mathbf{S}})$. Then, $\mathcal{T}'_{\mathbf{S}}$ guarantees coverage of \mathbf{S} if $\mathcal{T}'_{\mathbf{S}}$ is a module for \mathbf{S} in \mathcal{T}' .*

Finally, the actual module $\mathcal{T}_{\mathbf{S}_i}$ is imported. The effect of this import is that the ontology \mathcal{T} being developed evolves to $\mathcal{T} \cup \mathcal{T}_{\mathbf{S}_i}$; as a consequence, the safety guarantee with respect to the remaining external ontologies might be compromised. Such an effect is obviously undesirable and hence the following guarantee should be provided after the import:

Definition 12 (Module Independence Guarantee). *Given an ontology \mathcal{T} and signatures $\mathbf{S}_1, \mathbf{S}_2$, we say that \mathcal{T} guarantees module independence if, for all \mathcal{T}_1 with $\text{Sig}(\mathcal{T}) \cap \text{Sig}(\mathcal{T}_1) \subseteq \mathbf{S}_1$, it holds that $\mathcal{T} \cup \mathcal{T}_1$ is safe for \mathbf{S}_2 .*

3.2 Achieving the Logical Guarantees

In order to provide the necessary guarantees of our methodology, we will now make use of the locality conditions introduced in Section 2.3 and some general properties of conservative extensions, safety and modules.

The Safety Guarantee. In Section 2.3, we argue that the simultaneous refinement and generalisation of an external concept may compromise safety. To preserve safety, we propose to use two locality conditions: \perp -locality, suitable for refinement, and \top -locality, suitable for generalisation. These conditions can be checked syntactically using the grammars defined in Figure 2, and therefore they can be easily implemented in a tool. In order to achieve the safety guarantee at the end of the offline phase, we propose to follow the procedure sketched in Figure 5, where “locality of \mathcal{T} for \mathbf{S}_i according to the choice for \mathbf{S}_i ” means \perp -locality if \mathbf{S}_i is to be refined and \top -locality if \mathbf{S}_i is to be generalised.

Input:

\mathcal{T} : Ontology

$\mathbf{S}_1, \dots, \mathbf{S}_n$: disjoint signatures

a choice among refinement and generalisation for each \mathbf{S}_i

Output:

ontology \mathcal{T}_1 that guarantees safety

- 1: $\mathcal{T}_1 := \mathcal{T}$
 - 2: **while** exists \mathbf{S}_i such that \mathcal{T} **not** local according to the selection for \mathbf{S}_i **do**
 - 3: check locality of \mathcal{T}_1 w.r.t. \mathbf{S}_i according to the choice for \mathbf{S}_i
 - 4: **if** non-local **then**
 - 5: $\mathcal{T}_1 :=$ repair \mathcal{T}_1 until it is local for \mathbf{S}_i according to the choice for \mathbf{S}_i
 - 6: **end if**
 - 7: **end while**
 - 8: **return** \mathcal{T}_1
-

Fig. 5. A procedure for checking safety

It is immediate to see that the following holds upon completion of the procedure: for each \mathbf{S}_i , if the user selected the refinement (generalisation) view, then \mathcal{T} is \perp -local (\top -local) w.r.t. \mathbf{S}_i . This is sufficient to guarantee safety, as given by the following proposition:

Proposition 13. *Let \mathcal{T} be an ontology and $\mathbf{S} = \mathbf{S}_1 \uplus \dots \uplus \mathbf{S}_n$ be the union of disjoint signatures. If, for each \mathbf{S}_i , either \mathcal{T} is \perp -local or \top -local w.r.t. \mathbf{S}_i , then \mathcal{T} guarantees safety.*

The Module Coverage Guarantee. The fragment extracted for each customised signature in the online phase must satisfy the module coverage guarantee. As seen in Section 2.3, \perp -locality and \top -locality can also be used for extracting modules in the sense of Definition 3. Given an external ontology \mathcal{T}' and customised signature \mathbf{S}_i , the scope of the \perp -module and \top -module is determined by Proposition 9: as shown in Figure 3, the \perp -module will contain all the super-concepts in \mathcal{T}' of the concepts in \mathbf{S}_i , whereas the \top -module will contain all the sub-concepts.

The construction in Figure 3 also shows that the extraction of \perp -modules or \top -modules may introduce symbols not in \mathbf{S}_i , and potentially unnecessary. To make the module as small as possible, we proceed as follows, given \mathcal{T}'_i and \mathbf{S}_i : first, extract the minimal \perp -module M for \mathbf{S} in \mathcal{T}'_i ; then, extract the minimal \top -module for \mathbf{S} in M . The fragment obtained at the end of this process satisfies the module coverage guarantee as given in the following proposition:

Proposition 14. *Let $\mathcal{T}'_2 = \text{LoMod}(\text{UpMod}(\mathcal{T}', \mathbf{S}), \mathbf{S})$. Then \mathcal{T}'_2 guarantees coverage of \mathbf{S} in \mathcal{T}' .*

The Module Independence Guarantee. When a module is imported in the online phase (see Figure 4), module independence should be guaranteed—that is, importing a module for a signature \mathbf{S}_i from an external ontology \mathcal{T}'_i into \mathcal{T} should not affect the safety of \mathcal{T} w.r.t. to the other external ontologies. The following proposition establishes that this guarantee always holds provided that the imported module in \mathcal{T}'_i does not share symbols with the remaining external ontologies \mathcal{T}'_j . In practice, this is always the default situation since different reference ontologies have different namespaces and therefore their signatures are disjoint. The following proposition is an immediate consequence of the syntactic definition of \perp -locality and \top -locality.

Proposition 15. *Let \mathcal{T} be an ontology and $\mathbf{S}_1, \mathbf{S}_2$ disjoint signatures. If, for each $i = 1, 2$, \mathcal{T} is \perp -local or \top -local w.r.t. \mathbf{S}_i , then \mathcal{T} guarantees module independence.*

4 The Ontology Reuse Tool

We have developed a Protégé 4⁶ plugin that supports the methodology presented in Section 3. The plugin and user manual can be downloaded from <http://krono.act.uji.es/people/Ernesto/safety-ontology-reuse>.

⁶ Ontology Editor Protégé 4: <http://www.co-ode.org/downloads/protege-x/>

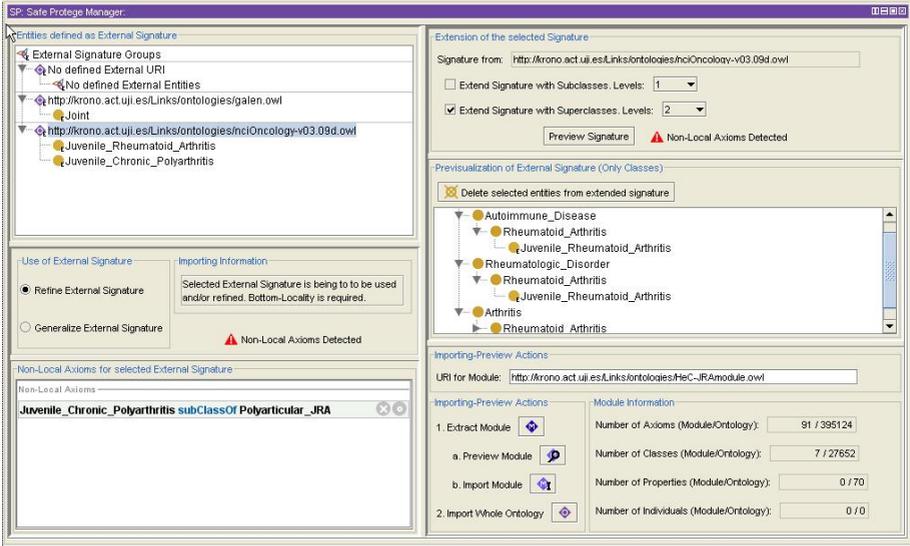


Fig. 6. Protégé Safe Manager Interface

The Offline Phase. The first step of the offline phase involves the selection of the external entities. Our plugin provides functionality for declaring entities as external as well as for defining the external ontology URI (or signature subgroup) for the selected entities; this information is stored in the ontology using OWL 1.1 annotations [7] as follows: we use an ontology annotation axiom per external ontology, an entity annotation axiom to declare an entity external, and an entity annotation axiom per external entity to indicate its external ontology. The set of external entities with the same external ontology URI can be viewed as one of the S_i . Finally, the UI of the plugin also allows for the specification, for each external ontology, whether it will be refined or generalised. Once the external entities have been declared and divided into groups, the tool allows for safety checking of the ontology under development w.r.t. each group of external symbols separately. The safety check uses \perp -locality (\top -locality) for signature groups that adopt the refinement (generalisation) view. The non-local axioms to be repaired are appropriately displayed.

Figure 6 shows the *Safe Protégé Manager* component with the set of signature subgroups in the top left corner, and the non-local axioms in the bottom left corner. Note that, in this phase, our tool does allow the user to work completely offline, without the need of extracting and importing external knowledge, and even without knowing exactly from which ontology the reused entities will come from. Indeed, the specification of the URI of the external ontologies is optional at this stage, and, even if indicated, such URI may not refer to a real ontology, but it may simply act as a temporary name.

The Online Phase. In the online phase, the user chooses external ontologies and imports axioms from them. At this stage, the groups of external symbols to

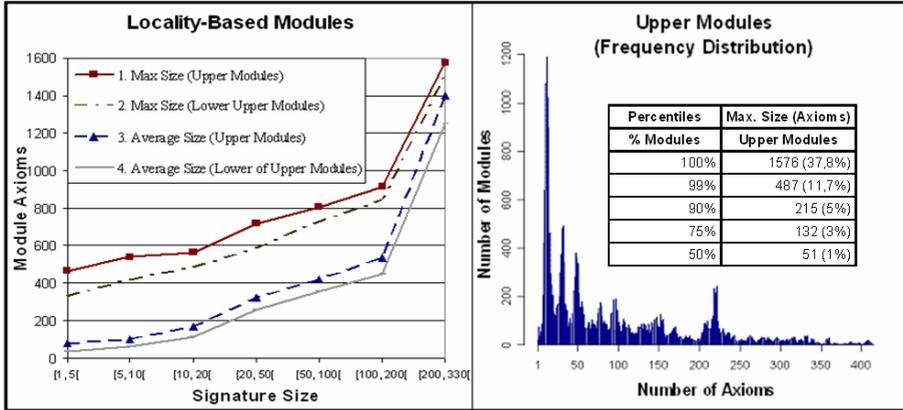


Fig. 7. Module information obtained from Galen. (left) Module average size against max size. (right) Frequency distributions for module sizes.

be imported should refer to the location of a “real” external ontology. Once an external signature group S_i has been selected for import, the selected signature can be customised by adding super-concepts and sub-concepts of the selected symbols. The tool provides the functionality for previewing the concept hierarchy of the corresponding external ontology for this purpose. Once the specific signature group under consideration has been customised, a module for it can be extracted. The module is computed using the procedure in Proposition 14; the user can compute the module, preview it in a separate frame, and either import it or cancel the process and come back to the signature customisation stage. The user is also given the option to import the whole external ontology instead of importing a module. Note that currently the import of a module is done “by value”, in the sense that the module becomes independent from the original ontology: if the external ontology on the Web evolves, the previously extracted module will not change.

The right hand side of the *Safe Protégé Manager* component (see Figure 6) includes the set of necessary components to support the proposed steps for the online phase.

5 Evaluation

So far, we have demonstrated our tool to various ontology developers⁷ who have expressed great interest, and we are currently working on a proper user study. In the following, we describe the experiments we have performed to prove that locality-based modules are reasonably sized compared to the whole ontology.

⁷ Thanks to Elena Beißwanger, Sebastian Brandt, Alan Rector, and Holger Stenzhorn for valuable comments and feedback.

For each concept name A in NCI or Galen⁸, we have proceeded as follows. (i) For each pair (u, ℓ) between $(0, 0)$ and $(3, 3)$, construct the signature $\mathbf{S}(A, u, \ell)$ by taking A , its super-concepts in the next u levels and its sub-concepts in the next ℓ levels. (ii) Extract $\mathcal{T}'_1 = \text{UpMod}(\mathcal{T}', \mathbf{S}_A)$ and $\mathcal{T}'_2 = \text{LoMod}(\mathcal{T}'_1, \mathbf{S}_A)$, see Proposition 14. From now on, we will refer to the modules \mathcal{T}'_1 and \mathcal{T}'_2 mentioned above as “Upper Module (UM)” and “Lower of Upper Module (LUM)”.

We have grouped the extracted modules according to the size of the input signature in order to evaluate its impact on the size of the modules. Figure 7 shows the obtained results for Galen (3161 entities and 4170 axioms), where the size of a module is the number of its axioms. The following conclusions can be drawn from the empirical results: first, the modules obtained are small on average since 99% of UMs have at most 487 axioms ($\sim 12\%$ of the size of Galen), and 99% of the LUMs contain at most and 386 axioms ($\sim 9\%$ of the size of Galen) for initial signatures $\mathbf{S}(A, u, \ell)$ containing between 1 and 330 entities; second, the growth in the size of the modules w.r.t. the size of the initial signature is smooth and linear up to initial signatures containing 100 entities. We have similar findings for NCI v3 (27,722 entities and 395,124 axioms, most of them annotations), NCI v7 (63,754 entities and 867,374 axioms, most of them annotations) and SNOMED (entities and 389,541 axioms and 389,544 entities).

In addition to the “synthetic” experiments, we have undertaken some real experiments in the context of the Health-e-Child project’s user scenario, see Section 1. The experiments focus on JRA and Cardiomyopathies (CMP)—a group of diseases that are central to the project. Using our tool, members of the project have manually selected signatures that are relevant to JRA and Cardiomyopathies from both Galen and NCI, expanded these signatures as in the case of the synthetic tests, and extracted the corresponding modules. For example, in the case of JRA in Galen, the initial signature and expanded signature consisted of 40 and 131 entities. The following tables show the sizes of all signatures and modules extracted from (a) Galen and (b) NCI.

(a) Disease	JRA	JRA	CMP	(b) Disease	JRA	JRA	CMP
Signature size	40	131	77	Signature size	48	356	124
# axioms	490	1151	614	# axioms	300	1258	537
# concepts	296	663	334	# concepts	193	613	283
# roles	69	116	56	# roles	17	21	10

6 Related Work

Ontology Engineering Methodologies: Several ontology engineering methodologies can be found in the literature; prominent examples are Methontology [8], On-To-Knowledge (OTK) [9], and ONTOCLEAN [10]. These methodologies, however, do not address ontology development scenarios involving reuse. Our proposed methodology is complementary and can be used in combination with them.

⁸ We have used a fragment of GALEN expressible in OWL:

<http://krono.act.uji.es/Links/ontologies/galen.owl/view>

Ontology Segmentation and Ontology Integration Techniques: In the last few years, a growing body of work has been developed addressing Ontology Modularisation, Ontology Mapping and Alignment, Ontology Merging, Ontology Integration and Ontology Segmentation, see [11, 12, 13] for surveys. This field is diverse and has originated from different communities.

In particular, numerous techniques for extracting fragments of ontologies. Most of them, such as [14, 15, 16], rely on syntactic heuristics for detecting relevant axioms. These techniques do not attempt to formally specify the intended outputs and do not provide any guarantees.

Ontology Reuse techniques: There are various proposals for “safely” combining modules; most of these proposals, such as \mathcal{E} -connections, Distributed Description Logics and Package-based Description Logics propose a specialised semantics for controlling the interaction between the importing and the imported modules to avoid side-effects, for an overview see [17]. In contrast, in our paper we assume that reuse is performed by simply building the logical union of the axioms in the modules under the standard semantics; instead, we provide the user with a collection of reasoning services, such as safety testing, to check for side-effects. Our paper is based on other work on modular reuse of ontologies [18, 19, 6, 5] which enables us to provide the necessary guarantees. We extend this work with a methodology and tool support.

7 Lessons Learned and Future Work

We have described a logic-based approach to the reuse of ontologies that is both *safe* (i.e., we guarantee that the meaning of the imported symbols is not changed) and *economic* (i.e., we import only the module relevant for a given set of symbol and we guarantee that we do not lose any entailments compared to the import of the whole ontology). We have described a methodology that makes use of this approach, have implemented tool support for it in Protégé, and report on experiments that indicate that our modules are indeed of acceptable size.

In the future, we will extend the tool support so that the user can “shop” for symbols to reuse: it will allow to browse an ontology for symbols to reuse and provide a simple mechanism to pick them and, on “check-out”, will compute the relevant module. Moreover, we are working on more efficient ways of module extraction that make use of already existing computations. Next, we plan to carry out a user study to learn more about the usefulness of the interface and how to further improve it. Finally, our current tool support implements a “by value” mechanism: modules are extracted and added at the user’s request. In addition, we would like to support import “by reference”: a feature that checks whether the imported ontology has changed and thus a new import is necessary.

Acknowledgements

This work was partially supported by the PhD Fellowship Program of the *Generalitat Valenciana*, by the *Fundació Caixa Castelló-Bancaixa*, and by the UK EPSRC grant no. EP/E065155/1.

References

- [1] Golbeck, J., Fragoso, G., Hartel, F.W., Hendler, J.A., Oberthaler, J., Parsia, B.: The National Cancer Institute's Thésaurus and Ontology. *JWS* 1(1), 75–80 (2003)
- [2] Horrocks, I., Patel-Schneider, P.F., van Harmelen, F.: From *SHIQ* and RDF to OWL: The making of a web ontology language. *JWS* 1(1), 7–26 (2003)
- [3] Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. *J. of Artificial Intelligence Research* 31, 273–318 (2008)
- [4] Ghilardi, S., Lutz, C., Wolter, F.: Did I damage my ontology? A case for conservative extensions in description logics. In: Doherty, P., Mylopoulos, J., Welty, C. (eds.) *KR 2006*, pp. 187–197. AAAI Press, Menlo Park (2006)
- [5] Lutz, C., Walther, D., Wolter, F.: Conservative extensions in expressive description logics. In: *IJCAI 2007*, pp. 453–459. AAAI Press, Menlo Park (2007)
- [6] Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: A logical framework for modularity of ontologies. In: *IJCAI 2007*, pp. 298–304. AAAI Press, Menlo Park (2007)
- [7] Motik, B., Patel-Schneider, P.F., Horrocks, I.: OWL 1.1 Web Ontology Language Structural Specification and Functional-Style Syntax. W3C Member Submission (2007)
- [8] Fernandez, M., Gomez-Perez, A., et al.: Methontology: From ontological art towards ontological engineering. In: AAAI, Stanford, USA (1997)
- [9] Sure, Y., Staab, S., Studer, R.: On-to-knowledge methodology. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, Springer, Heidelberg (2003)
- [10] Guarino, N., Welty, C.: Evaluating ontological decisions with ontoclean. *Commun. ACM* 45(2), 61–65 (2002)
- [11] Kalfoglou, Y., Schorlemmer, M.: Ontology mapping: The state of the art. *The Knowledge Engineering Review* 18, 1–31 (2003)
- [12] Noy, N.F.: Semantic integration: A survey of ontology-based approaches. *SIGMOD Record* 33(4), 65–70 (2004)
- [13] Noy, N.F.: Tools for mapping and merging ontologies. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*. International Handbooks on Information Systems, pp. 365–384. Springer, Heidelberg (2004)
- [14] Noy, N., Musen, M.: The PROMPT suite: Interactive tools for ontology mapping and merging. *Int. J. of Human-Computer Studies* 6(59) (2003)
- [15] Seidenberg, J., Rector, A.L.: Web ontology segmentation: analysis, classification and use. In: *Proc. of WWW 2006*, pp. 13–22. ACM, New York (2006)
- [16] Jiménez-Ruiz, E., Berlanga, R., Nebot, V., Sanz, I.: Ontopath: A language for retrieving ontology fragments. In: *Proc. of ODBASE. LNCS*, pp. 897–914. Springer, Heidelberg (2007)
- [17] Cuenca Grau, B., Kutz, O.: Modular ontology languages revisited. In: *Proc. of the Workshop on Semantic Web for Collaborative Knowledge Acquisition* (2007)
- [18] Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Just the right amount: extracting modules from ontologies. In: Williamson, C.L., Zurko, M.E., Patel-Schneider, P.F., Shenoy, P.J. (eds.) *WWW*, pp. 717–726. ACM, New York (2007)
- [19] Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Ontology reuse: Better safe than sorry. In: *DL 2007. CEUR Workshop Proceedings*, vol. 250 (2007)