A Logical Framework for Modularity of Ontologies*

Bernardo Cuenca Grau, Ian Horrocks, Yevgeny Kazakov and **Ulrike Sattler**
The University of Manchester
School of Computer Science
Manchester, M13 9PL, UK
{bcg, horrocks, ykazakov, sattler }@cs.man.ac.uk

Abstract

Modularity is a key requirement for collaborative ontology engineering and for distributed ontology reuse on the Web. Modern ontology languages, such as OWL, are logic-based, and thus a useful notion of modularity needs to take the semantics of ontologies and their implications into account. We propose a logic-based notion of modularity that allows the modeler to specify the *external* signature of their ontology, whose symbols are assumed to be defined in some other ontology. We define two restrictions on the usage of the external signature, a syntactic and a slightly less restrictive, semantic one, each of which is decidable and guarantees a certain kind of "black-box" behavior, which enables the controlled merging of ontologies. Analysis of real-world ontologies suggests that these restrictions are not too onerous.

1 Motivation

Modularity is a key requirement for many tasks concerning ontology design, maintenance and integration, such as the collaborative development of ontologies and the merging of independently developed ontologies into a single, reconciled ontology. Modular representations are easier to understand, reason with, extend and reuse. Unfortunately, in contrast to other disciplines such as software engineering—in which modularity is a well established notion—ontology engineering is still lacking a useful, well-defined notion of modularity.

Modern ontology languages, such as OWL [Patel-Schneider *et al.*, 2004], are logic-based; consequently, a notion of modularity needs to take into account the semantics of the ontologies and their implications.

In this paper, we propose a logic-based framework for modularity of ontologies in which we distinguish between *external* and *local* symbols of ontologies. Intuitively, the external symbols of an ontology are those that are assumed to be defined somewhere externally in other ontologies; the remaining, local symbols, are assumed to be defined within the ontology itself by possibly reusing the meaning of external symbols. Hence by merging an ontology with external ontologies we import the meaning of its external symbols to define the meaning of its local symbols.

To make this idea work, we need to impose certain constraints on the usage of the external signature: in particular, merging ontologies should be "safe" in the sense that they do not produce unexpected results such as new inconsistencies or subsumptions between imported symbols. To achieve this kind of safety, we use the notion of *conservative extensions* to define modularity of ontologies, and then prove that a property of some ontologies, called *locality*, can be used to achieve modularity. More precisely, we define two notions of locality for $\mathcal{SHIQ}$ TBoxes: (i) a tractable syntactic one which can be used to provide guidance in ontology editing tools, and (ii) a more general semantic one which can be checked using a DL-reasoner. Additionally, we present an extension of locality to the more expressive logic $\mathcal{SHOIQ}$ [Horrocks and Sattler, 2005]. Finally, we analyse existing ontologies and conclude that our restrictions to local TBoxes are quite natural.

When integrating independently developed ontologies, we often have to identify different symbols in the ontologies having the same intended meaning. This is a problem known as ontology matching or mapping,[1] which we are not concerned with here: we consider ontologies sharing some part of their signature, and how we can make sure that that their merge (i.e. the union of their axioms) is "well-behaved".

2 Preliminaries

We introduce the description logic $\mathcal{SHOIQ}$, which provides the foundation for OWL.

A $\mathcal{SHOIQ}$-signature is the disjoint union $\mathbf{S} = \mathbf{R} \uplus \mathbf{C} \uplus \mathbf{I}$ of sets of *role names* (denoted by $R, S, \cdots$), *concept names* (denoted by $A, B, \cdots$) and *nominals* (denoted by $i, j, k, \cdots$). A $\mathcal{SHOIQ}$*-role* is either $R \in \mathbf{R}$ or an *inverse role* R^- with $R \in \mathbf{R}$. We denote by $\mathsf{Rol}(\mathbf{S})$ the set of $\mathcal{SHOIQ}$-roles for the signature $\mathbf{S}$. The set $\mathsf{Con}(\mathbf{S})$ of $\mathcal{SHOIQ}$*-concepts* for the signature $\mathbf{S}$ is defined by the grammar

$$\mathsf{Con}(\mathbf{S}) ::= A \mid j \mid (\neg C) \mid (C_1 \sqcap C_2) \mid (\exists R.C) \mid (\geqslant n\, S.C)$$

*This work is supported by the EU Project TONES (Thinking ONtologiES) ref: IST-007603 and by the EPSRC Project REOL (Reasoning in Expressive Ontology Languages) ref: EP/C537211/1.

[1] The website http://www.ontologymatching.org provides extensive information about this area.

where $A \in \mathbf{C}$, $j \in \mathbf{I}$, $C_{(i)} \in \mathsf{Con}(\mathbf{S})$, $R, S \in \mathsf{Rol}(\mathbf{S})$, with S a *simple* role,[2] and n a positive integer. We use the following abbreviations: $C \sqcup D$ stands for $\neg(\neg C \sqcap \neg D)$; $\top$ and $\bot$ stand for $A \sqcup \neg A$ and $A \sqcap \neg A$, respectively; $\forall R.C$ and $\leqslant n\, S.C$ stand for $\neg\exists R.\neg C$ and $\neg(\geqslant n{+}1\, S.C)$, respectively.

A $\mathcal{SHOIQ}$-TBox $\mathcal{T}$, or ontology, is a finite set of *role inclusion axioms* (RIs) $R_1 \sqsubseteq R_2$ with $R_i \in \mathsf{Rol}(\mathbf{S})$, *transitivity axioms* $\mathsf{Trans}(R)$ with $R \in \mathbf{R}$ and *general concept inclusion axioms* (GCIs) $C_1 \sqsubseteq C_2$ with $C_i \in \mathsf{Con}(\mathbf{S})$. We use $A \equiv C$ as an abbreviation for the two GCIs $A \sqsubseteq C$ and $C \sqsubseteq A$. The signature $\mathsf{Sig}(\alpha)$ (respectively $\mathsf{Sig}(\mathcal{T})$) of an axiom α (respectively of a TBox $\mathcal{T}$) is the set of symbols occurring in α (respectively in $\mathcal{T}$). A $\mathcal{SHIQ}$-TBox is a $\mathcal{SHOIQ}$-TBox that does not contain nominals.

Given a signature $\mathbf{S} = \mathbf{R} \uplus \mathbf{C} \uplus \mathbf{I}$, an $\mathbf{S}$-*interpretation* $\mathcal{I}$ is a pair $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set, called the *domain* of the interpretation, and $\cdot^{\mathcal{I}}$ is the *interpretation function* that assigns to each $R \in \mathbf{R}$ a binary relation $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, to each $A \in \mathbf{C}$ a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, and to every $j \in \mathbf{I}$ a singleton set $j^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$. The interpretation function is extended to complex roles and concepts as follows:

$$
\begin{aligned}
(R^-)^{\mathcal{I}} &= \{\langle x,y\rangle \mid \langle y,x\rangle \in R^{\mathcal{I}}\}\\
(\neg C)^{\mathcal{I}} &= \Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}\\
(C \sqcap D)^{\mathcal{I}} &= C^{\mathcal{I}} \cap D^{\mathcal{I}}\\
(\exists R.C)^{\mathcal{I}} &= \{x \in \Delta^{\mathcal{I}} \mid \exists y.\langle x,y\rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}\\
(\geqslant n\, R.C)^{\mathcal{I}} &= \{x \in \Delta^{\mathcal{I}} \mid\\
&\qquad \sharp\{y \in \Delta^{\mathcal{I}} \mid \langle x,y\rangle \in R^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\} \geq n\}
\end{aligned}
$$

The *satisfaction* relation $\mathcal{I} \models \alpha$ between an interpretation $\mathcal{I}$ and a $\mathcal{SHOIQ}$-axiom α (read as $\mathcal{I}$ *satisfies* α) is defined as follows: $\mathcal{I} \models (R_1 \sqsubseteq R_2)$ iff $R_1^{\mathcal{I}} \subseteq R_2^{\mathcal{I}}$; $\mathcal{I} \models \mathsf{Trans}(R)$ iff $R^{\mathcal{I}}$ is transitive; $\mathcal{I} \models (C \sqsubseteq D)$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all axioms in $\mathcal{T}$. A TBox $\mathcal{T}$ *implies* an axiom α (written $\mathcal{T} \models \alpha$) if $\mathcal{I} \models \alpha$ for every model $\mathcal{I}$ of $\mathcal{T}$. An axiom α is a *tautology* if it is satisfied in every interpretation.

An $\mathbf{S}$-interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is an *expansion* of an $\mathbf{S}'$-interpretation $\mathcal{I}' = (\Delta^{\mathcal{I}'}, \cdot^{\mathcal{I}'})$ if $\mathbf{S} \supseteq \mathbf{S}'$, $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}'}$, and $X^{\mathcal{I}} = X^{\mathcal{I}'}$ for every $X \in \mathbf{S}'$. The *trivial expansion of* $\mathcal{I}'$ *to* $\mathbf{S}$ is an expansion $\mathcal{I} = (\Delta^{\mathcal{I}'}, \cdot^{\mathcal{I}})$ of $\mathcal{I}'$ such that $X^{\mathcal{I}} = \varnothing$ for every $X \in \mathbf{S} \setminus \mathbf{S}'$.

3 Modularity of Ontologies

In this section, we propose a logical formalization for the notion of modularity for ontologies. In analogy to software engineering, we should be able to compose complex ontologies from simpler (modular) ontologies in a consistent and well-defined way, in particular without unintended interactions between the component ontologies. This notion of modularity would be useful for both the collaborative development of a single ontology by different domain experts, and the integration of independently developed ontologies, including the reuse of existing third-party ontologies.

In order to formulate our notion of modularity, we will distinguish between the local and external symbols of an ontology. We assume that the signature $\mathsf{Sig}(\mathcal{T})$ of a TBox $\mathcal{T}$ is partitioned into two parts: the *local signature* $\mathsf{Loc}(\mathcal{T})$ of $\mathcal{T}$ and the *external signature* $\mathsf{Ext}(\mathcal{T})$ of $\mathcal{T}$. To distinguish between the elements of these signatures, we will underline the external signature elements whenever used within the context of $\mathcal{T}$. Intuitively, $\mathsf{Ext}(\mathcal{T})$ specifies the concept and role names that are (or can be) imported from other ontologies, while $\mathsf{Loc}(\mathcal{T})$ specifies those that are defined in $\mathcal{T}$.

As a motivating example, imagine a set of bio-medical ontologies that is being developed collaboratively by a team of experts. Suppose that one group of experts designs an ontology $\mathcal{G}$ about genes and another group designs an ontology $\mathcal{D}$ about diseases. Now certain genes are defined in terms of the diseases they cause. For example, the gene ErbB2 is described in $\mathcal{G}$ as an Oncogene that is found in humans and is associated with a disease called Adrenocarcinoma[3]

$$
\begin{aligned}
\mathsf{ErbB2} \;\equiv\;& \mathsf{Oncogene} \sqcap \exists \mathsf{foundIn.Human}\\
& \sqcap \exists \mathsf{associatedWith}.\underline{\mathsf{Adrenocarcinoma}}
\end{aligned}
$$

The concept Adrenocarcinoma is described in $\mathcal{D}$, which is under the control of a different group of modelers. So, this concept is external for $\mathcal{G}$, whereas the remaining concept and role names are local for $\mathcal{G}$. Now one consequence of these ontologies being modularly "well-behaved" would be that the gene experts building $\mathcal{G}$ should not change the knowledge about diseases, even if they are using them in their axioms.

Another example is the integration of a *foundational* (or "upper") ontology $\mathcal{U}$ and a *domain* ontology $\mathcal{O}$. Foundational ontologies, such as CYC, SUMO,[4] and DOLCE, provide a structure upon which ontologies for specific subject matters can be constructed and are assumed to be the result of an agreement between experts. Suppose that an ontology developer wants to reuse the generic concept of a Substance from $\mathcal{U}$ in their ontology $\mathcal{O}$ about *Chemicals*. For such a purpose, they state that the concept Organic_Chemical in their chemical ontology $\mathcal{O}$ is more specific than Substance in $\mathcal{U}$ by using the axiom: $\mathsf{Organic\_Chemical} \sqsubseteq \underline{\mathsf{Substance}}$, where $\mathsf{Substance} \in \mathsf{Ext}(\mathcal{O})$. Since foundational ontologies are well-established ontologies that one does not control and, typically, does not have complete knowledge about, it is especially important that the merge $\mathcal{O} \cup \mathcal{U}$ does not produce new logical consequences w.r.t. $\mathcal{U}$—even if $\mathcal{U}$ changes.

In both examples, we have argued that ontology integration should be carried out in such a way that consequences of a TBox $\mathcal{T}'$ are not changed when elements of $\mathcal{T}'$ are reused in another TBox $\mathcal{T}$. This property can be formalized using the notion of a *conservative extension* [Ghilardi *et al.*, 2006].

Definition 1 (Conservative Extension). Let $\mathcal{T}$ and $\mathcal{T}'$ be TBoxes. Then $\mathcal{T} \cup \mathcal{T}'$ is a *conservative extension* of $\mathcal{T}'$ if, for every axiom α with $\mathsf{Sig}(\alpha) \subseteq \mathsf{Sig}(\mathcal{T}')$ we have $\mathcal{T} \cup \mathcal{T}' \models \alpha$ iff $\mathcal{T}' \models \alpha$. ◊

Thus, given $\mathcal{T}$ and $\mathcal{T}'$, their union $\mathcal{T} \cup \mathcal{T}'$ does not yield new consequences in the language of $\mathcal{T}'$ if $\mathcal{T} \cup \mathcal{T}'$ is a conser-

[2] See [Horrocks and Sattler, 2005] for a precise definition of simple roles.

[3] Example from the National Cancer Institute Ontology `http://www.mindswap.org/2003/CancerOntology`.

[4] See `http://ontology.teknowledge.com`.

vative extension of $\mathcal{T}'$. A useful notion of modularity, however, should abstract from the particular $\mathcal{T}'$ under consideration. In fact, the external signature should be the core notion in a modular representation as opposed to its particular definition in a particular ontology $\mathcal{T}'$. This is especially important when $\mathcal{T}'$ may evolve, and where this evolution is beyond our control—which, for example, could well be the case when using the "imports" construct provided by OWL. Consequently, in order for $\mathcal{T}$ to use $\mathsf{Ext}(\mathcal{T})$ in a modular way, $\mathcal{T} \cup \mathcal{T}'$ should be a conservative extension of *any* $\mathcal{T}'$ over $\mathsf{Ext}(\mathcal{T})$.

Furthermore, it is important to ensure that, whenever two independent parts $\mathcal{T}_1$ and $\mathcal{T}_2$ of an ontology $\mathcal{T}$ under the control of different modelers are developed in a modular way, then $\mathcal{T}$ remains modular as well.

These requirements can be formalized as follows:

Definition 2 (Modularity). A set $\boldsymbol{\mathcal{M}}$ of TBoxes $\mathcal{T}$ with $\mathsf{Sig}(\mathcal{T}) = \mathsf{Loc}(\mathcal{T}) \uplus \mathsf{Ext}(\mathcal{T})$ is a *modularity class* if the following conditions hold:

M1. If $\mathcal{T} \in \boldsymbol{\mathcal{M}}$, then $\mathcal{T} \cup \mathcal{T}'$ is a conservative extension of every $\mathcal{T}'$ such that $\mathsf{Sig}(\mathcal{T}') \cap \mathsf{Loc}(\mathcal{T}) = \varnothing$;

M2. If $\mathcal{T}_1, \mathcal{T}_2 \in \boldsymbol{\mathcal{M}}$, then $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \in \boldsymbol{\mathcal{M}}$ with $\mathsf{Loc}(\mathcal{T}) = \mathsf{Loc}(\mathcal{T}_1) \cup \mathsf{Loc}(\mathcal{T}_2)$. ♢

Please note that our framework is independent of the DL under consideration. Also, Definition 2 does not define a modularity class uniquely, but just states conditions for being one. When the modularity class is clear from the context, we will call its elements *modular ontologies*.

In the next section, we focus our attention on the logic $\mathcal{SHIQ}$, and show that it is possible to define a reasonable modularity class such that (1) checking its membership can be done using standard reasoning tools, (2) it has an inexpensive syntactic approximation that can be used to guide the modeling of ontologies in a modular way, and (3) our analysis of existing ontologies shows that they seem to conform "naturally" with its restrictions.

4 Modularity of $\mathcal{SHIQ}$ ontologies

In this section we define a particular modularity class, the class of *local ontologies*, which captures many practical examples of modularly developed ontologies. We first give a syntactic definition of local ontologies and then generalize it to a semantic one. Finally, we prove that our semantic definition leads to a *maximal* class of modular TBoxes.

Definition 2 excludes already many $\mathcal{SHIQ}$-TBoxes. Property M1, in particular, implies that no modular TBox $\mathcal{T}$ can contain the two axioms below at the same time:

$$A \sqsubseteq C_1' \qquad \textit{(local)} \tag{1}$$

$$C_2' \sqsubseteq A \qquad \textit{(non-local)} \tag{2}$$

where A is a local concept name and C_1', C_2' are constructed using $\mathsf{Ext}(\mathcal{T})$. These axioms imply $C_2' \sqsubseteq C_1'$, which indeed changes the meaning of the external concepts C_i'. At this point, we are faced with a fundamental choice as to the type of axioms to disallow. Each choice leads to a different modularity class. We argue that, analogously to software engineering, where refinement is the main application of modularity, axioms of type (1) fit better with ontology integration scenarios, such as those sketched in Section 3, than axioms of type (2).

As discussed in Section 3, the external names are "imported" in order to *reuse* them in the definition of other concepts, and not to further constrain their meaning. Intuitively, axioms of type (1) are consistent with this idea, whereas axioms of type (2) are not.

The principal difference between these two axioms is that (2) forces the external concept C_2' to contain *only* instances of the local concept name A, thus bounding the *size* of possible interpretations of C_2' once the meaning of A is established. In contrast, (1) still (in principle) allows for interpretations of C_1' of unbounded size. Note that this argument does not prohibit all inclusion axioms between external concepts and local ones. For example, in contrast to (2), the axiom

$$C_2' \sqsubseteq \neg A \qquad \textit{(local)} \tag{3}$$

still leaves sufficient "freedom" for the interpretation of C_2', even if the interpretation for A is fixed. In fact, this axiom is equivalent to $A \sqsubseteq \neg C_2'$, and thus is of type (1).

Our choice of the types of simple axioms to disallow can be generalized to more complex axioms; for example, all axioms below should be forbidden for the reasons given above:

$$C_1' \sqsubseteq A_1 \sqcup A_2 \quad \text{and} \quad A \equiv C_2' \sqcup \exists R.B \quad \textit{(non-local)} \tag{4}$$

The last axiom is disallowed because it implies (2).

Even if an axiom does not explicitly involve the external symbols, it may still constrain their meaning. In fact, certain GCIs have a global effect and impose constraints on all elements of the models of an ontology, and thereby on the interpretation of external concepts. For example, it is easy to see that the axioms

$$\top \sqsubseteq A \quad \text{and} \quad \neg A_1 \sqsubseteq A_2 \quad \textit{(non-local)} \tag{5}$$

imply (2) and the first axiom in (4) respectively.[5] These observations lead to the following definition:

Definition 3 (Locality). Let $\mathbf{S}$ be a $\mathcal{SHIQ}$-signature and let $\mathbf{E} \subseteq \mathbf{S}$ be the *external signature*. The following grammar defines the two sets $\mathcal{C}_{\mathbf{E}}^{+}$ and $\mathcal{C}_{\mathbf{E}}^{-}$ of *positively* and *negatively local* concepts w.r.t. $\mathbf{E}$:

$$\begin{aligned} \mathcal{C}_{\mathbf{E}}^{+} &::= A \mid (\neg C^{-}) \mid (C \sqcap C^{+}) \mid (\exists R^{+}.C) \mid (\exists R.C^{+}) \mid \\ &\qquad \mid (\geqslant n\, R^{+}.C) \mid (\geqslant n\, R.C^{+})\,. \\ \mathcal{C}_{\mathbf{E}}^{-} &::= (\neg C^{+}) \mid (C_1^{-} \sqcap C_2^{-})\,. \end{aligned}$$

where A is a concept name from $\mathbf{S} \setminus \mathbf{E}$, $R \in \mathsf{Rol}(\mathbf{S})$, $C \in \mathsf{Con}(\mathbf{S})$, $C^{+} \in \mathcal{C}_{\mathbf{E}}^{+}$, $C_{(i)}^{-} \in \mathcal{C}_{\mathbf{E}}^{-}$, $i = 1, 2$, and $R^{+} \notin \mathsf{Rol}(\mathbf{E})$.[6]

A role inclusion axiom $R^{+} \sqsubseteq R$ or a transitivity axiom $\mathsf{Trans}(R^{+})$ is *local w.r.t.* $\mathbf{E}$. A GCI is *local w.r.t.* $\mathbf{E}$ if it is either of the form $C^{+} \sqsubseteq C$ or $C \sqsubseteq C^{-}$, where $C^{+} \in \mathcal{C}_{\mathbf{E}}^{+}$, $C^{-} \in \mathcal{C}_{\mathbf{E}}^{-}$ and $C \in \mathsf{Con}(\mathbf{S})$. A $\mathcal{SHIQ}$-TBox $\mathcal{T}$ is local if every axiom from $\mathcal{T}$ is local w.r.t. $\mathsf{Ext}(\mathcal{T})$. ♢

Intuitively, the positively local concepts are those whose interpretation is bounded (i.e. its size is limited) when the interpretation of the local symbols is fixed. In this respect,

[5] $\neg A_1 \sqsubseteq A_2$ implies $\top \sqsubseteq A_1 \sqcup A_2$, which implies $C' \sqsubseteq A_1 \sqcup A_2$.

[6] Recall that $\forall R.C$, $(\leqslant n\, R.C)$ and $C_1 \sqcup C_2$ are expressed using the other constructors, so they can be used in local concepts as well.

they behave similarly to local concept names. Negatively local concepts are essentially negations of positively local concepts. Please, note that, given $\mathbf{E} \subseteq \mathbf{S}$, a concept written over $\mathbf{S}$ may be neither in $\mathcal{C}^{+}_{\mathbf{E}}$, nor in $\mathcal{C}^{-}_{\mathbf{E}}$.

Definition 3 can be used to formulate guidelines for constructing modular ontologies, as illustrated by the following example. Moreover, Definition 3 can be used in ontology editors to detect and warn the user of an a priori "dangerous" usage of the external signature—without the need to perform any kind of reasoning.

Example 4 Suppose we are developing $\mathcal{W}$, an ontology about wines, and we want to reuse some concepts and roles from $\mathcal{F}$, an independently developed ontology about food.

$$\mathcal{F}: \quad \mathsf{VealParmesan} \sqsubseteq \mathsf{MeatDish} \sqcap \exists \mathsf{hasIngredient}.\mathsf{Veal}$$
$$\mathsf{DeliciousProduct} \sqsupseteq \exists \mathsf{hasIngredient}.\mathsf{DeliciousProduct}$$
$$\mathsf{Trans}(\mathsf{hasIngredient})$$

$$\mathcal{W}: \quad \mathsf{Chardonnay} \sqsubseteq \mathsf{Wine} \sqcap \exists \mathsf{servedWith}.\underline{\mathsf{VealParmesan}}$$
$$\mathsf{Rioja} \sqsubseteq \mathsf{Wine} \sqcap \exists \underline{\mathsf{hasIngredient}}.\mathsf{Tempranillo}$$
$$\mathsf{RedWine} \sqsupseteq \mathsf{Wine} \sqcap \exists \mathsf{servedWith}.\underline{\mathsf{MeatDish}}$$
$$\mathsf{Tempranillo} \sqsubseteq \underline{\mathsf{DeliciousProduct}}$$

Here $\mathsf{Ext}(\mathcal{W}) = \{\mathsf{hasIngredient}, \mathsf{DeliciousProduct}, \mathsf{VealParmesan}, \mathsf{MeatDish}\}$ and $\mathcal{W}$ is local. ◊

The following Lemma shows that our notion of locality satisfies the desired properties from Definition 2.

Lemma 5 [Locality Implies Modularity]
The set of local $\mathcal{SHIQ}$ TBoxes is a modularity class.

To prove Lemma 5, we use the following property:

Lemma 6 *Let $\mathcal{T}$ be a local $\mathcal{SHIQ}$ TBox with $\mathbf{E} = \mathsf{Ext}(\mathcal{T})$, and let $\mathcal{I}'$ be an $\mathbf{E}$-interpretation. Then the trivial expansion $\mathcal{I}$ of $\mathcal{I}'$ to $\mathsf{Sig}(\mathcal{T})$ is a model of $\mathcal{T}$.*

Proof. We need to show that $\mathcal{I} \models \alpha$ for every $\alpha \in \mathcal{T}$. According to Definition 3, every $\alpha \in \mathcal{T}$ has one of the forms: $R^+ \sqsubseteq R$, $\mathsf{Trans}(R^+)$, $C^+ \sqsubseteq C$ or $C \sqsubseteq C^-$, where $R^+ \notin \mathsf{Rol}(\mathbf{E})$, $C^+ \in \mathcal{C}^{+}_{\mathbf{E}}$ and $C^- \in \mathcal{C}^{-}_{\mathbf{E}}$. To prove $\mathcal{I} \models \alpha$ it is then suffices to show that $(R^+)^{\mathcal{I}} = \varnothing$, $(C^+)^{\mathcal{I}} = \varnothing$ and $(C^-)^{\mathcal{I}} = \Delta^{\mathcal{I}}$ for each such axiom. The first property holds since $\mathcal{I}$ is the trivial expansion of an $\mathbf{E}$-interpretation $\mathcal{I}'$. The remaining two properties can be easily shown by induction over the definitions of $\mathcal{C}^{+}_{\mathbf{E}}$ and $\mathcal{C}^{-}_{\mathbf{E}}$ from Definition 3. □

Proof of Lemma 5. Let $\boldsymbol{\mathcal{M}}$ be a set of TBoxes $\mathcal{T}_i$, each of which is local w.r.t. $\mathsf{Ext}(\mathcal{T}_i)$. Property M2 from Definition 2 follows from Definition 3 since every axiom α that is local w.r.t. $\mathbf{E}$ is also local w.r.t. every $\mathbf{E}'$ with $(\mathbf{E}' \cap \mathsf{Sig}(\alpha)) \subseteq \mathbf{E}$.

In order to prove Property M1, let $\mathcal{T}$ be a local $\mathcal{SHIQ}$-TBox. Assume $(\star)$ $\mathcal{T}' \cup \mathcal{T} \models \alpha$ for some TBox $\mathcal{T}'$ and an axiom α with $\mathsf{Sig}(\alpha) \subseteq \mathsf{Sig}(\mathcal{T}')$ and $\mathsf{Sig}(\mathcal{T}') \cap \mathsf{Loc}(\mathcal{T}) = \varnothing$. We have to show that $\mathcal{T}' \models \alpha$.

Assume to the contrary that $\mathcal{T}' \not\models \alpha$. Then, there exists a model $\mathcal{I}'$ of $\mathcal{T}'$ such that $\mathcal{I}' \not\models \alpha$. Let $\mathcal{I}$ be the trivial expansion of $\mathcal{I}'$ to $\mathsf{Sig}(\mathcal{T})$. Then $\mathcal{I} \models \mathcal{T}'$ and $\mathcal{I} \not\models \alpha$ since $\mathsf{Sig}(\alpha) \subseteq \mathsf{Sig}(\mathcal{T}')$. Additionally, by Lemma 6, $\mathcal{I} \models \mathcal{T}$. So $\mathcal{T} \cup \mathcal{T}' \not\models \alpha$, which contradicts our assumption $(\star)$. □

Lemma 5 tells us that, in Example 4, $\mathcal{W} \cup \mathcal{F}$ does not entail new information about food *only*. Even if $\mathcal{F}$ evolves, say by adding the axiom $\mathsf{VealParmesan} \sqsubseteq \exists \mathsf{producedIn}.\underline{\mathsf{Italy}}$ using a third ontology $\mathcal{C}$ of countries, $\mathcal{W}$ will not interfere with $\mathcal{F}$. On the other hand, using the imported concepts from $\mathcal{F}$ allows us to derive some non-trivial properties involving the local and mixed signature of $\mathcal{W}$, such as $\mathsf{Chardonnay} \sqsubseteq \mathsf{RedWine}$ and $\mathsf{Rioja} \sqsubseteq \mathsf{DeliciousProduct}$.

As we have seen, our notion of locality from Definition 3 yields a modularity class. This class, however, is not the most general one we can achieve. In particular, there are axioms that are not local, but obviously unproblematic. For example, the axiom $A' \sqsubseteq A' \sqcup C'$ is a tautology, but is disallowed by Definition 3 since it involves external symbols only; another example is the GCI $A_1 \sqcup B' \sqsubseteq A_2 \sqcup B'$ which is implied by the (syntactically) local axiom $A_1 \sqsubseteq A_2$. The limitation of our syntactic notion of locality is its inability to "compare" concepts from the external signature.

A natural question is whether we can generalize Definition 3 to overcome this limitation. Obviously, such generalization cannot be given in terms of syntax only since checking for tautologies in the external signature necessarily involves reasoning. Since our proof of Lemma 5 relies mainly on Lemma 6, we generalize our notion of locality as follows:

Definition 7 (Semantic Locality). Let $\mathbf{E} \subseteq \mathbf{S}$. A $\mathcal{SHIQ}$-axiom α with $\mathsf{Sig}(\alpha) \subseteq \mathbf{S}$ is *semantically local w.r.t.* $\mathbf{E}$ if the trivial expansion $\mathcal{I}$ of every $\mathbf{E}$-interpretation $\mathcal{I}'$ to $\mathbf{S}$ is a model of α. A $\mathcal{SHIQ}$-TBox $\mathcal{T}$ is semantically local if every axiom in $\mathcal{T}$ is semantically local w.r.t. $\mathsf{Ext}(\mathcal{T})$. ◊

Lemma 6 essentially implies that every syntactically local TBox is semantically local. Interestingly, both notions coincide when $\mathbf{E} = \varnothing$ or when α is a non-trivial role inclusion axiom (not of the form $R' \sqsubseteq R'$) or a transitivity axiom. It is easy to check that the conditions for a modularity class in Definition 2 hold for semantic locality as well. The following proposition provides an effective way of checking whether a GCI satisfies Definition 7:

Proposition 8 *Let α be a GCI and $\mathbf{E} \subseteq \mathbf{S}$. Let α' be obtained from α by replacing every subconcept of the form $\exists R.C$, $\geqslant n\, R.C$, and every concept name A in α with $\bot$, where $R \notin \mathsf{Rol}(\mathbf{E})$ and $A \notin \mathbf{E}$. Then α is semantically local w.r.t. $\mathbf{E}$ iff α' is a tautology.*

Proof. The subconcepts of the form $\exists R.C$, $\geqslant n\, R.C$, and A are interpreted by $\varnothing$ in every trivial expansion of every $\mathbf{E}$-interpretation, hence they are indistinguishable from $\bot$ in the context of Definition 7. Replacing all these subconcepts in α with $\bot$ yields α' with $\mathsf{Sig}(\alpha') \subseteq \mathbf{E}$, and thus Definition 7 implies that α is semantically local iff every $\mathcal{SHIQ}$-interpretation satisfies α'. □

As mentioned above, deciding semantic locality involves reasoning; in fact, this problem is PSPACE-complete in the size of the axiom,[7] as opposed to checking syntactic locality, which can be done in polynomial time. We expect the test

[7]This is precisely the complexity of checking subsumption between $\mathcal{SHIQ}$-concepts w.r.t. the empty TBox and without role inclusions and transitivity axioms [Tobies, 2001].

from Proposition 8 to perform well in practice since the size of axioms in a TBox is typically small w.r.t. the size of the TBox, and would like to point out that it can be performed using any existing DL reasoner.

It is worth noting that both notions of locality provide the "black-box" behavior we are aiming at, and both involve only the ontology $\mathcal{T}$ and its external signature. Finally, a natural question arising is whether semantic locality can be further generalized while preserving modularity. The following lemma answers this question negatively.

Lemma 9 [Semantic Locality is Maximal]
If a $\mathcal{SHIQ}$-TBox $\mathcal{T}_1$ is not semantically local, then there exist $\mathcal{SHIQ}$-TBoxes $\mathcal{T}_2$ and $\mathcal{T}'$ such that $\mathcal{T}_2$ is local, $\mathsf{Loc}(\mathcal{T}_2) \subseteq \mathsf{Loc}(\mathcal{T}_1)$, $\mathsf{Loc}(\mathcal{T}_1) \cap \mathsf{Sig}(\mathcal{T}') = \varnothing$, *and* $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}'$ *is not a conservative extension of* $\mathcal{T}'$.

Proof. Let $\mathcal{T}_1$ be not semantically local, and define $\mathcal{T}_2$ and $\mathcal{T}'$ as follows: $\mathcal{T}_2$ consists of the axioms of the form $A \sqsubseteq \bot$ and $\exists R.\top \sqsubseteq \bot$ for every $A, R \in \mathsf{Loc}(\mathcal{T}_1)$; $\mathcal{T}'$ consists of axioms of the form $\bot \sqsubseteq A'$ and $\bot \sqsubseteq \exists R'.\top$ for every $A', R' \in \mathsf{Ext}(\mathcal{T}_1)$. Note that (i) $\mathcal{T}_2$ is local, (ii) for every model $\mathcal{I}$ of $\mathcal{T}_2$, we have $A^{\mathcal{I}} = \varnothing$ and $R^{\mathcal{I}} = \varnothing$ for every $A, R \in \mathsf{Loc}(\mathcal{T}_1)$ and (iii) every $\mathsf{Ext}(\mathcal{T}_1)$-interpretation is a model of $\mathcal{T}'$, and (iv) $\mathcal{T}'$ uses all symbols from $\mathsf{Ext}(\mathcal{T}_1)$.

In order to show that $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}'$ is not a conservative extension of $\mathcal{T}'$, we construct an axiom α' over the signature of $\mathcal{T}'$ such that $\mathcal{T}' \not\models \alpha'$ but $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}' \models \alpha'$. Since $\mathcal{T}_1$ is not semantically local, there exists an axiom $\alpha \in \mathcal{T}_1$ which is not semantically local w.r.t. $\mathsf{Ext}(\mathcal{T}_1)$, and which we use to define α'. If α is a role axiom of the form $\alpha = (R' \sqsubseteq R)$, we set $\alpha' = (\top \sqsubseteq \forall R'.\bot)$; if $\alpha = (R' \sqsubseteq S')$ with $R' \neq S'$ or $\alpha = \mathsf{Trans}(R')$, we set $\alpha' = \alpha$; and if α is a GCI, we define α' from α as in Proposition 8. As a result, α' uses $\mathsf{Ext}(\mathcal{T}_1) = \mathsf{Sig}(\mathcal{T}')$ only, and $\mathcal{T}' \not\models \alpha'$ since α' is not a tautology (for the last case this follows from Proposition 8). Since $\mathcal{T}_1$ contains α and because of the property (ii) above for $\mathcal{T}_2$, we have $\mathcal{T}_1 \cup \mathcal{T}_2 \models \alpha'$, and so $\mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}' \models \alpha'$. □

Lemma 9 shows that semantic locality cannot be generalized without violating the properties in Definition 2. Indeed, condition M2 implies that the union $\mathcal{T}$ of two local TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ is a local TBox, and condition M2 implies that $\mathcal{T} \cup \mathcal{T}'$ is conservative over every $\mathcal{T}'$ with $\mathsf{Sig}(\mathcal{T}') \cap \mathsf{Loc}(\mathcal{T}) = \varnothing$. The results in Lemma 5 and Lemma 9 are summarized in the following theorem:

Theorem 10 *A set of semantically local $\mathcal{SHIQ}$ TBoxes is a maximal class of modular TBoxes.*

5 Modularity of $\mathcal{SHOIQ}$ ontologies

When trying to extend the results in the previous section to the more expressive logic $\mathcal{SHOIQ}$, we soon encounter difficulties. Nominals are interpreted as singleton sets and, thus, a straightforward extension of Definition 7 fails since nominals cannot be interpreted by the empty set.

A notion of modularity, however, can still be achieved if *all* nominals in a TBox $\mathcal{T}$ are treated as external concepts; the intuitive reason for this is that the interpretation of nominals is already very constrained, and hence we have little control over it. Under this assumption, Definitions 3 and 7 can be reused for $\mathcal{SHOIQ}$. Such notions of locality still allow for non-trivial uses of nominals in $\mathcal{T}$. For example, the following axiom is semantically local w.r.t. $\mathbf{E} = \{elvis\}$, even if $elvis$ is used as a nominal:

$$\mathsf{ElvisLover} \equiv \mathsf{MusicFan} \sqcap \exists \mathsf{likes}.elvis$$

Indeed, the trivial expansion of every $\mathbf{E}$-Interpretation to $\mathbf{S} = \{\mathsf{ElvisLover}, \mathsf{MusicFan}, \mathsf{likes}\}$ is a model of this axiom.

Definition 11 (Locality for $\mathcal{SHOIQ}$). A $\mathcal{SHOIQ}$-TBox $\mathcal{T}$ with $\mathsf{Sig}(\mathcal{T}) = \mathbf{R} \uplus \mathbf{C} \uplus \mathbf{I}$ is syntactically (semantically) local w.r.t. $\mathbf{E}$ if $\mathcal{T}$ is syntactically (semantically) local w.r.t. $\mathbf{E} \cup \mathbf{I}$ as in Definition 3 (Definition 7). ◊

Lemma 12 [Semantic Locality Implies Modularity] *The set of semantically local $\mathcal{SHOIQ}$ TBoxes is a modularity class for* $\mathbf{E}$.

The proof is analogous to the one of Lemma 5. Unfortunately, an important use of nominals in DLs, namely ABox assertions, is non-local according to our definition. For example, the assertion $elvis \sqsubseteq$ Singer (typically written as $elvis$:Singer) is not local since $elvis$ is treated as an external element. In fact, it is not possible to extend the definition of locality to capture assertions and retain modularity:

Proposition 13 [Assertions Cannot be Local]
For every assertion $\alpha = (i \sqsubseteq A)$ *there exists a syntactically local TBox* $\mathcal{T}$ *such that* $\mathcal{T} \cup \{\alpha\}$ *is inconsistent.*

Proof. Take $\mathcal{T} = \{A \sqsubseteq \bot\}$. □

Proposition 13 implies that no TBox $\mathcal{T}_1$ containing an assertion α can be declared as local without braking either property M1 or M2 of modularity from Definition 2. Indeed, by taking $\mathcal{T}$ as in the proof of Proposition 13, we obtain an inconsistent merge $\mathcal{T}_1 \cup \mathcal{T}$ which should be local according to M2 if $\mathcal{T}_1$ is local. However, no inconsistent TBox can be local since it implies all axioms and hence violates condition M1.

Even if the merge of a TBox and a set of assertions is consistent, new subsumptions over the external signature may still be entailed. For example, consider the TBox $\mathcal{T}$ consisting of the axiom:

$$\mathsf{Frog} \sqsubseteq \exists \mathsf{hasColor}.\underline{green} \sqcap \forall \mathsf{hasColor}.\underline{\mathsf{Dark}}$$

which is local w.r.t. $\mathbf{E} = \{green, \mathsf{Dark}\}$. If we add the assertion $kermit \sqsubseteq$ Frog to $\mathcal{T}$, then we obtain that green is a dark color ($green \sqsubseteq$ Dark), as a new logical consequence.

To sum up, we have shown that locality can be extended to $\mathcal{SHOIQ}$, but not in the presence of assertions. An open question is whether semantic locality for $\mathcal{SHOIQ}$ is maximal in the sense of Lemma 9.

6 Field Study

In order to test the adequacy of our conditions in practice, we have implemented a (syntactic) locality checker and run it over ontologies from a library of 300 ontologies of various sizes and complexity some of which import each other [Gardiner *et al.*, 2006].[8] Since OWL does not allow to declare

[8]The library is available at `http://www.cs.man.ac.uk/~horrocks/testing/`

symbols as local or external, we have used the following "informed guess work": for an ontology $\mathcal{T}$, we define the set $\mathsf{Loc}(\mathcal{T})$ as the set of symbols in $\mathcal{T}$ that do not occur in the signature of the ontologies imported (directly or indirectly) using the owl:imports construct by $\mathcal{T}$, and defining $\mathsf{Ext}(\mathcal{T})$ to be the complement of $\mathsf{Loc}(\mathcal{T})$ in $\mathsf{Sig}(\mathcal{T})$. The owl:imports construct allows to include, in an ontology $\mathcal{T}$, the axioms of another ontology $\mathcal{T}'$ published on the Web by reference. The usage of owl:imports $\mathcal{T}'$ in $\mathcal{T}$ produces the (logical) *union* $\mathcal{T} \cup \mathcal{T}'$.

It turned out that 96 of the 300 ontologies used the owl:imports construct, and that all but 11 of these 96 ontologies are syntactically local (and hence also semantically local). From the 11 non-local ontologies, 7 are written in the OWL-Full species of OWL to which our framework does not yet apply. The remaining 4 non-localities are due to the presence of so-called *mapping axioms* of the form $A \equiv B'$, where $A \in \mathsf{Loc}(\mathcal{T})$ and $B' \in \mathsf{Ext}(\mathcal{T})$, which are not even semantically local. We were able to fix these non-localities as follows: we replace every occurrence of A in $\mathcal{T}$ with B' and then remove this axiom from $\mathcal{T}$. After this transformation, all 4 non-local ontologies turned out to be local.

7 Discussion and Related Work

In the last few years, a rapidly growing body of work has been developed under the names of *Ontology Mapping and Alignment*, *Ontology Merging* and *Ontology Integration*; see [Kalfoglou and M.Schorlemmer, 2003; Noy, 2004] for surveys. This field is rather diverse, has originated from different communities, and is concerned with two different problems: (i) how to (semi-automatically) detect correspondences between terms in the signatures of the ontologies to be integrated (e.g. Instructor corresponds to Professor), and, (ii) how to assess and predict the (logical) consequences of the merging. Typically, when integrating ontologies, one first solves (i) and then (ii).

Although (i) has been the focus of intensive research in the last few years [Kalfoglou and M.Schorlemmer, 2003], and tools for ontology mapping are available, to the best of our knowledge, the problem of predicting and controlling the consequences of ontology integration has been addressed only very recently in [Ghilardi *et al.*, 2006] and [Grau *et al.*, 2006]. In [Ghilardi *et al.*, 2006], the authors point out the importance of the notion of a conservative extension for ontology evolution and merging, and provide decidability and complexity results for the problem of deciding conservative extensions in the basic DL $\mathcal{ALC}$. In [Grau *et al.*, 2006], the authors identify two basic ontology integration scenarios. For each of them, the authors established a set of semantic properties (including being conservative extensions) to be satisfied by the integrated ontology, and presented a set of syntactic constraints on the component ontologies to ensure the preservation of the desired semantic properties.

The results in this paper generalize those from [Grau *et al.*, 2006], since the integration scenarios presented there are particular cases of Definition 3. Also, in contrast to both [Ghilardi *et al.*, 2006], and [Grau *et al.*, 2006], our notion of modularity implies a "black box" behavior with respect to the external signature: instead of considering a pair of ontologies $\mathcal{T}, \mathcal{T}'$, our approach takes an ontology $\mathcal{T}$ with specified sets of local and external symbols, and provides guarantees for the merge of $\mathcal{T}$ with *any* ontology $\mathcal{T}'$ which does not use local symbols. In contrast, the approach described in [Ghilardi *et al.*, 2006] considers the problem of whether, for two given $\mathcal{ALC}$ TBoxes $\mathcal{T}, \mathcal{T}'$, their merge $\mathcal{T} \cup \mathcal{T}'$ is conservative over $\mathcal{T}'$. It turns out that this problem is decidable in 2EXPTIME in the size of $\mathcal{T} \cup \mathcal{T}'$, and thus it is significantly harder than standard reasoning tasks (such as deciding ontology consistency). A solution to the latter problem can be used to decide whether $\mathcal{T}$ and $\mathcal{T}'$ can be safely merged—which can be the case without any of them being local. If $\mathcal{T}$ or $\mathcal{T}'$ are changed, however, then this test would need to be repeated—which is not the case in the approach presented here (see the above discussion of its black box behavior). As a consequence, these two approaches can be used in different scenarios: ours can be used to provide guidelines for ontology engineers who want to design modular ontologies that show black box behavior, whereas the one described in [Ghilardi *et al.*, 2006] can be used to check safe integrability for a given, fixed set of TBoxes.

Summing up, we have proposed a logic-based framework for modularity, which we have instantiated in a plausible and practically applicable way for $\mathcal{SHIQ}$, and in a preliminary way for $\mathcal{SHOIQ}$. We believe that our results will be useful as the foundations of tools that support both the collaborative development of complex ontologies and the integration of independently developed ontologies on the Semantic Web.

References

[Gardiner *et al.*, 2006] T. Gardiner, I. Horrocks, and D. Tsarkov. Automated benchmarking of description logic reasoners. In *Proc. of DL*, 2006.

[Ghilardi *et al.*, 2006] S. Ghilardi, C. Lutz, and F. Wolter. Did I damage my ontology? a case for conservative extensions in description logics. In *Proc. of KR-06*, 2006.

[Grau *et al.*, 2006] B. Cuenca Grau, I. Horrocks, O. Kutz, and U. Sattler. Will my ontologies fit together? In *Proc. of DL*, 2006.

[Horrocks and Sattler, 2005] I. Horrocks and U. Sattler. A tableaux decision procedure for $\mathcal{SHOIQ}$. In *Proc. of IJCAI*, 2005.

[Kalfoglou and M.Schorlemmer, 2003] Y. Kalfoglou and M.Schorlemmer. Ontology mapping: The state of the art. *The Knowledge Engineering Review*, 18:1–31, 2003.

[Noy, 2004] N. Noy. Semantic integration: A survey on ontology-based approaches. *SIGMOD Record*, 2004.

[Patel-Schneider *et al.*, 2004] P.F. Patel-Schneider, P. Hayes, and I. Horrocks. OWL web ontology language semantics and abstract syntax. W3C Recommendation, 2004.

[Tobies, 2001] S. Tobies. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation*. PhD thesis, RWTH Aachen, 2001.