Empirical Study of Logic-Based Modules: Cheap Is Cheerful

Chiara Del Vescovo[1], Pavel Klinov[2], Bijan Parsia[1], Ulrike Sattler[1], Thomas Schneider[3], and Dmitry Tsarkov[1]

[1] University of Manchester, UK
{delvescc|bparsia|sattler|tsarkov}@cs.man.ac.uk
[2] University of Ulm, Germany
pavel.klinov@uni-ulm.de
[3] Universität Bremen, Germany
tschneider@informatik.uni-bremen.de

Abstract. For ontology reuse and integration, a number of approaches have been devised that aim at identifying modules, i.e., suitably small sets of "relevant" axioms from ontologies. Here we consider three logically sound notions of modules: MEX modules, only applicable to inexpressive ontologies; modules based on semantic locality, a sound approximation of the first; and modules based on syntactic locality, a sound approximation of the second (and thus the first), widely used since these modules can be extracted from OWL DL ontologies in time polynomial in the size of the ontology.

In this paper we investigate the quality of both approximations over a large corpus of ontologies, using our own implementation of semantic locality, which is the first to our knowledge. In particular, we show with statistical significance that, in most cases, there is no difference between the two module notions based on locality; where they differ, the additional axioms can either be easily ruled out or their number is relatively small. We classify the axioms that explain the rare differences into four kinds of "culprits" and discuss which of those can be avoided by extending the definition of syntactic locality. Finally, we show that differences between MEX and locality-based modules occur for a minority of ontologies from our corpus and largely affect (approximations of) expressive ontologies – this conclusion relies on a much larger and more diverse sample than existing comparisons between MEX and syntactic locality-based modules.

1 Introduction

Some notable examples of ontologies describe large and loosely connected domains, as it is the case for SNOMED CT, the Systematized Nomenclature Of MEDicine, Clinical Terms,[4] which describes the terminology used in medicine including diseases, drugs, etc. Users often are not interested in a whole ontology

[4] http://www.ihtsdo.org/snomed-ct/

$\mathcal{O}$ but rather only in a limited part of it which is relevant to their application. One recently explored technique for addressing this situation is to use *modules*, i.e., suitably small subsets of $\mathcal{O}$ that behave for specific purposes like the original ontology over a given *signature* Σ, i.e., a set of terms (classes and properties).

Using a module rather than a whole ontology aims at improving performance since only information that is relevant to a restricted vocabulary is processed. However, the correctness of the outcome can be guaranteed only if the used modules satisfy certain well-defined properties. For example, reasoning-based tasks require the modules to *provide coverage* for $\mathcal{O}$ over Σ, i.e., preserve *all* the entailments of $\mathcal{O}$ over Σ (they are called *logical modules* [10,4]). Applications of logical modules include reuse of (a part of) well-established ontologies, ontology integration, and computing justifications to debug ontologies [12]. In these scenarios, though, a stronger notion of logical module is required that satisfies also two additional properties [16,20]: *self-containment* and *depletion*. The former means that the module preserves entailments over all terms that occur in the module (not just those used to extract the module). The latter means that $\mathcal{O} \setminus \mathcal{M}$ does not entail any non-tautological axioms over Σ. In this paper we will analyze only depleting and self-contained logical modules.

Interestingly, a *minimal* depleting and self-contained module for a signature Σ is, under some mild conditions, uniquely determined [16]. Extracting such modules is, unfortunately, computationally hard or even undecidable for expressive ontology languages [11,18,19]. In order to identify notions of modules whose extraction is feasible we can follow two alternative strategies. The first one consists of restricting the expressivity of the ontology language, as in the case of the MEX approach [15]: the MEX system allows for the extraction in polynomial time of the minimal self-contained and depleting module from acyclic $\mathcal{ELI}$ terminologies. The second strategy consists of looking for practical sufficient conditions to guarantee the properties of logical modules without imposing minimality on the module $\mathcal{M}$, as it is the case for the family of logical modules known as *locality-based modules* (*LBMs*) [3]; these modules can be extracted from ontologies as expressive as $\mathcal{SROIQ}$, are self-contained and depleting, but can contain axioms that are not relevant to preserve any entailment over the given Σ.

The family of LBMs consists of module notions that are parameterized according to two features: (1) the technique used for identifying which axioms need to be included in the module (*semantic* or *syntactic*); (2) the kind of placeholder(s) used for those terms not included in the signature (*bottom*, *top*, or *nested*). In the next two paragraphs we provide an intuitive discussion of the meaning of these two features.

The extraction of semantic LBMs requires entailment checks against an empty ontology and thus involve reasoning, which makes the computation as hard as reasoning. Moreover, the kind of reasoning service used is rather unusual for DL reasoners.[5] Hence, although algorithms for extracting semantic LBMs are known, until now and to the best of our knowledge they had not been

[5] DL reasoners usually *classify* an ontology: test it for consistency and all concept names for satisfiability/mutual subsumption.

implemented. In contrast, the extraction of syntactic LBMs involves *only* parsing the axioms of the ontology. Algorithms for the extraction of syntactic LBMs are known that run in time polynomial in the size of the ontology (thus much cheaper than reasoning), and are implemented in the OWL API.[6]

The kind of placeholder(s) used for semantic and syntactic LBMs gives a flavour of the different module notions. The bottom variants of LBMs provide a view of $\mathcal{O}$ from Σ "upwards" since they contain all named superclasses of class names in Σ; the top variants instead provide a view of $\mathcal{O}$ from Σ "downwards" since they contain all named subclasses of class names in Σ; finally, the nested variants provide a view of $\mathcal{O}$ "within" Σ since they still provide coverage for Σ as the other variants, but they do not necessarily contain all the sub- or super-classes of the classes in Σ.

This paper empirically studies the seven module notions depicted in Fig. 1 which summarizes their notations and their inclusion relations. Each node represents a module notion; the one for the MEX module is shadowed because this method can be used only for $\mathcal{ELI}$ acyclic ontologies. The MEX notion is in the same column as the nested versions because MEX modules provide a similar view of $\mathcal{O}$ "within" Σ.

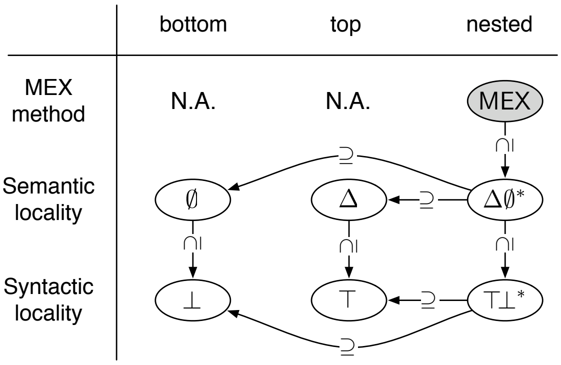

Fig. 1. Inclusion relations between the 7 notions of modules investigated.

As shown in Fig. 1, the MEX module for a signature Σ is a subset of the nested semantic LBM, and for each variant bottom, top, and nested, the semantic LBMs are contained in the corresponding syntactic ones. Hence, syntactic locality can be seen as an approximation of semantic locality which, in turn, is an approximation of MEX modules. This gives rise to the question of how *good* these approximations are: how much larger are the modules extracted by the approximations, and how much faster is the extraction?

This paper provides emprical answers to these questions by comparing different modules systematically extracted from a large corpus of real-life ontologies. Specifically, semantic LBMs are compared with syntatic LBMs and with MEX modules (for acyclic $\mathcal{ELI}$ ontologies). This paper substantially extends the previous experiments reported in [15] where MEX modules were compared

[6] `http://owlapi.sourceforge.net/`

with syntactic bottom modules on a sample of 5000 random signatures and the SNOMED CT ontology. We perform our study on a larger corpus (not restricted to $\mathcal{ELI}$), compare more notions of logical modules, and also provide rigorous statistical significance results.

The main contributions of this paper are summarized as follows:

- We show with statistical significance that, for almost all members of a large corpus of existing ontologies, there is no difference between any syntactic LBM and its semantic counterpart. In the few cases where differences occur, those are extremely modest so that it is questionable whether extracting semantic LBMs is worth the increased computational cost.
- We isolate four *culprits*, i.e., patterns of axioms that completely explain those rare differences. One includes simple tautologies that can be removed in a straightforward preprocessing step.
- Our results show that the extraction of semantic LBMs, which is in principle hard, is feasible in practice: on average, it is between 3 times (for top-modules) and 15 times (for bottom- and nested-modules) slower than the extraction of syntactic LBMs, and both only take milliseconds to seconds for most ontologies below 10K axioms.
- To obtain these results, we use our own implementation of semantic locality which, to the best of our knowledge, is the first ever to be implemented.
- We modify the original corpus to obtain for each ontology an acyclic $\mathcal{EL}$ version suitable for the use with the MEX system. We then compare MEX-modules and the nested-variants of LBMs, and find differences in only ∼27% of the corpus. We explain one reason for the largest differences observed.

2 Preliminaries

We assume the reader to be familiar with Description Logic languages (e.g. $\mathcal{SROIQ}$ [1,14]), and aim here at fixing the notations and at defining the key notions around module extraction, with a focus on locality-based modules [3] and MEX modules [15].

Let $\mathcal{O}$ denote an ontology, $\mathsf{N_C}$ a set of class names, and $\mathsf{N_R}$ a set of property names. A *signature* is a set $\Sigma \subseteq \mathsf{N_C} \cup \mathsf{N_R}$ of *terms*. Given a class, property, or axiom X, we call the set of terms in X the *signature of* X, denoted $\widetilde{X}$. Given a $\mathcal{SROIQ}$ ontology $\mathcal{O}$, a set $\mathcal{M} \subseteq \mathcal{O}$ of axioms from $\mathcal{O}$, and a signature Σ, we say that $\mathcal{O}$ is a *deductive Σ-conservative extension* (*Σ-dCE*) of $\mathcal{M}$ if, for all $\mathcal{SROIQ}$-axioms α with $\widetilde{\alpha} \subseteq \Sigma$, it holds that $\mathcal{O} \models \alpha$ if and only if $\mathcal{M} \models \alpha$. $\mathcal{O}$ is a *model Σ-conservative extension* (*Σ-mCE*) of $\mathcal{M}$ if $\{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{O}\} = \{\mathcal{I}|_\Sigma \mid \mathcal{I} \models \mathcal{M}\}$. Dually, $\mathcal{M}$ is a *dCE-based module* of $\mathcal{O}$ for Σ if $\mathcal{O}$ is a Σ-dCE of $\mathcal{M}$, and it is an *mCE-based module for Σ* if $\mathcal{O}$ is a Σ-mCE of $\mathcal{M}$. All dCE-based modules are also mCE-based modules, whilst the converse is not always true. A module $\mathcal{M} \subseteq \mathcal{O}$ for Σ is called *depleting* if there is no non trivial entailment η over Σ such that $\mathcal{O} \setminus \mathcal{M} \models \eta$; $\mathcal{M}$ is called *self-contained* if $\mathcal{M}$ is a module for $\Sigma = \widetilde{\mathcal{M}}$.

Since $\mathcal{M} \subseteq \mathcal{O}$ the monotonicity of $\mathcal{SROIQ}$ implies that every entailment η over Σ derivable from $\mathcal{M}$ is also derivable from $\mathcal{O}$. Deciding the converse direction is in general computationally hard, or even undecidable for expressive DLs [11,18,19]. Since we do not need to find *all* the subsets of $\mathcal{O}$ that are a module for Σ, we can use easier conditions which guarantee that a set of axioms $\mathcal{M} \subseteq \mathcal{O}$ is a module for Σ.

Let Σ be a signature and $\mathcal{O}$ be an ontology. Let $x \in \{\mathsf{MEX}, \emptyset, \Delta, \bot, \top\}$ be a notion of module. For each such notion, an oracle "x-check" can be defined that determines whether an axiom α may be involved in preserving an entailment η of $\mathcal{O}$ over Σ. Then, the x-module x-mod$(\Sigma, \mathcal{O})$ for Σ in $\mathcal{O}$ can be computed by performing Algorithm 1.

Algorithm 1 Extraction of an x-module for Σ

Input: Ontology $\mathcal{O}$, seed signature Σ, oracle x-check
Output: x-module $\mathcal{M}$ of $\mathcal{O}$ w.r.t. Σ

```
M ← ∅;  O′ ← O
repeat
  changed ← false
  for all α ∈ O′ do
    if the x-check for α against Σ ∪ M̃ is positive then
      M ← M ∪ {α};  O′ ← O′ \ {α};  changed ← true
until changed = false
return M
```

Algorithm 1 is a special case of the one in [3, Figure 4], and its output $\mathcal{M}$ does not depend on the order in which the axioms α are selected [3].

Due to space limitations, we can just briefly sketch the intuition behind the definition of each oracle and the corresponding results of interest for this paper. We refer the interested reader to [3,15] for further details.

The MEX system. In [15], the notion of a MEX-module is defined for acyclic terminologies, i.e., ontologies that satisfy two conditions: (1) they only contain axioms of the form $\mathtt{A} \equiv \mathtt{C}$ or $\mathtt{A} \sqsubseteq \mathtt{C}$ where $\mathtt{A}$ is a class name and $\mathtt{C}$ is a complex class; (2) for each $\mathtt{A}$, there is at most one axiom with $\mathtt{A}$ on the left-hand side; if one such axiom α exists, then $\mathtt{A}$ is said to be *defined*, and to be *directly dependent on* all the terms $\mathtt{X}$ that occur on the right-hand side of α (denoted $\mathtt{A} \succ \mathtt{X}$). The MEX method requires to determine for each defined class $\mathtt{A}$ the set $\mathsf{depend}_{\mathcal{O}}(\mathtt{A})$ of all the terms $\mathtt{X}$ in $\mathcal{O}$ such that the pair $(\mathtt{A}, \mathtt{X})$ belongs to the transitive closure of $\succ$. Intuitively, then, the MEX-check for an axiom α against a signature Σ tests whether either α defines a class $\mathtt{A} \in \Sigma \cup \widetilde{\mathcal{M}}$ and *uses*[7] at least one term $\mathtt{X} \in \mathsf{depend}_{\mathcal{O}}(\mathtt{A}) \cap (\Sigma \cup \widetilde{\mathcal{M}})$ in $\mathcal{O} \setminus \mathcal{M}$, or if every term on which $\mathtt{A}$ depends only via $\equiv$-axioms is *used to define*[7] some term in $\Sigma \cup \widetilde{\mathcal{M}}$. The authors prove that, if $\mathcal{O}$ is an acyclic $\mathcal{ELI}$ ontology, then using the oracle MEX-check in Algorithm 1 generates the minimal depleting self-contained module for a signature Σ in polynomial time.

Semantic locality In [3], the authors define a family of notions of locality with different parameters, the prominent notions being those where the placeholder x belongs to $\{\emptyset, \Delta\}$. These two notions of locality can be intuitively described

[7] The expressions *use* and *used to define* are high-level intuitive descriptions of the two conditions given in [15, Fig. 4], to which we refer the reader since a formal definition goes beyond the scope of this paper.

as follows: a $\mathcal{SROIQ}$ axiom α is $\emptyset$-*local* (resp. Δ-*local*) w.r.t. signature Σ if α' obtained by replacing all terms in $\widetilde{\alpha} \setminus \Sigma$ with $\bot$ (resp. $\top$) is a tautology, in which case the x-check returns negative. This treatment of α independently of the remaining axioms distinguishes the $\emptyset$- and Δ-check (as well as the $\bot$- and $\top$-check introduced in the next paragraph) from the MEX-check; hence the name *local*. The authors of [3] prove that, if all axioms in $\mathcal{O} \setminus \mathcal{M}$ are $\emptyset$-local (or all axioms are Δ-local) w.r.t. $\Sigma \cup \widetilde{\mathcal{M}}$, then $\mathcal{M}$ is an mCE-based (and hence dCE-based) module of $\mathcal{O}$ for Σ. Since deciding $\emptyset$- or Δ-locality requires tautology checks, this problem is as hard as standard reasoning. In some cases, α' is not a $\mathcal{SROIQ}$ axiom, so standard reasoners need to be extended.

Syntactic locality In order to achieve *tractable* module extraction, the two syntactic notions of x-locality for $x \in \{\bot, \top\}$ have been defined in [3]. Similarly to semantic locality, the x-check for an axiom α against a signature Σ operates on the transformed axiom α' obtained by replacing all terms not in Σ with the placeholder x. However, rather than invoking a reasoner, the x-check of α against Σ makes use of a simple syntactic test [3, Sec. 5.5]. For example, $\bot \sqsubseteq \mathtt{C}$ is clearly a tautology for each class $\mathtt{C}$. If the x-check is negative, α is said to be $\bot$- or $\top$-*local* w.r.t. Σ. The x-check used in syntactic LBMs is sound in identifying non-tautological axioms, but it may fail to spot a tautology, i.e., every $\emptyset$-local (Δ-local, resp.) axiom w.r.t. Σ is also $\bot$-local ($\top$-local, resp.) w.r.t. Σ, but not vice versa. Thus, also $\bot$- and $\top$-modules are mCE- and dCE-based modules for Σ. Applying the syntactic rules requires polynomial time, hence the extraction of this kind of modules is performed in time polynomial in the size of the ontology.

Modules based on syntactic (semantic) locality can be made smaller by iteratively nesting $\top$- and $\bot$-extraction (Δ- and $\emptyset$-extraction), again obtaining mCE- and dCE-based modules [3,20], called $\top\bot^*$- and $\Delta\emptyset^*$-*modules*.

Algorithm 1 guarantees that the module notions considered here are self-contained and depleting: self-containment holds because of the iteration until the signature of $\mathcal{M}$ remains unchanged; depletion holds because the axioms left out of $\mathcal{M}$ are those whose x-check against the enlarged signature is negative.

3 Research questions and experimental design

A natural question arising is whether syntactic and semantic LBMs differ in practice, and, if yes, by how much. A second question is whether semantic module extraction is *noticeably* more costly: the x-check has to be carried out often— once per axiom and signature that the algorithm goes through— and it is hard to predict the feasibility of semantic LBM extraction. Altogether, we want to know whether syntactic LBMs are a *good* approximation of semantic LBMs, and how much they differ in cost. Similarly, for acyclic $\mathcal{ELI}$ ontologies the analogous question arises: how good an approximation of MEX modules are LBMs?

An answer to these questions will allow for a more informed choice of which module extraction technique to select. One can always construct ontologies with huge differences in size and time between syntactic and semantic LBMs and between LBMs and MEX modules. Here, we are interested in these differences in currently available ontologies, and thus we need to design, run, and analyse suitable experiments.

Selection of the corpus. For our experiments, we have built a corpus containing: (1) all the ontologies from the NCBO BioPortal ontology repository,[8] version of November 2012; (2) ontologies from the TONES repository[9] which have already been studied in previous work on modularity [7]: Koala, Mereology, University, People, miniTambis, OWL-S, Tambis, Galen. From this corpus, we have removed ontologies that cannot be downloaded, whose .owl file is corrupted or impossible to parse, or which are inconsistent. Furthermore, we have excluded those large ontologies (exceeding 10K axioms) where the extraction of a semantic LBM repeatedly took more than 2 minutes: for each such ontology, the estimated time needed to perform our experiments could have exceeded 300 hours. However, to include at least one case of a huge ontology, we have kept in the corpus NCI, an $\mathcal{SH}(\mathcal{D})$ ontology with 123,270 axioms.

This selection results in a corpus of 242 ontologies, which even beside NCI greatly vary in expressivity (from $\mathcal{AL}$ to $\mathcal{SROIQ}(\mathcal{D})$) and in size (10–16,066 axioms, 10–16,068 terms) [13]. For a full list of the corpus, please refer to the appendix.

As mentioned above, it is not possible for some ontologies to test Δ-locality (and thus for extracting Δ- and $\Delta\emptyset^*$-modules) using standard DL reasoners. The reason is that, when an axiom α is tested for Δ-locality w.r.t. a signature Σ, it has to be tested whether α is a tautology after all symbols outside Σ have been replaced with $\top$. For some types of axioms, this replacement cannot be expressed in OWL 2 DL: the global restrictions that ensure decidability of the underlying description logic $\mathcal{SROIQ}$ [14] forbid the use of the $\top$-role in number restrictions or on both sides of a role chain inclusion axiom. Hence, for ontologies containing any of these two features (one of the symbols $\mathcal{Q}, \mathcal{N}, \mathcal{F}, \mathcal{R}$ in the expressivity column), we cannot test Δ-locality, or extract Δ- or $\Delta\emptyset^*$-modules, using standard DL reasoners. To cover these cases, we have extended the reasoner FaCT++ to cover the use of the $\top$-role as required by the semantic locality tests.

Since MEX handles only acyclic $\mathcal{ELI}$ ontologies, we created an $\mathcal{ELI}$ version $\mathcal{ELI}(\mathcal{O})$ of each ontology $\mathcal{O}$ in our corpus by filtering unsupported axioms and breaking terminological cycles. A principled way of doing this is beyond the scope of this paper, and we have used the following heuristic. (1) Remove all axioms that are not of the form $\mathtt{A} \sqsubseteq \mathtt{C}$, $\mathtt{A} \equiv \mathtt{C}$ or $\mathtt{r} \sqsubseteq \mathtt{s}$ with $\mathtt{A}$ being a concept name, $\mathtt{r}, \mathtt{s}$ role names and $\mathtt{A}$ an $\mathcal{EL}$ concept description (the latter is determined by the expressivity checker in the OWL API, which cannot distinguish $\mathcal{ELI}$ from $\mathcal{ALCI}$; therefore we needed to resort to $\mathcal{EL}$). The removal includes axioms that are in $\mathcal{EL}$ but are not supported by MEX, such as those containing domain and range restrictions. (2) Break definitional cycles: for each concept name $\mathtt{A}$, iteratively extend their definitions and delete the first axiom that closes a cycle. That is, if $\mathtt{A} \equiv \mathtt{C}_1$ is the definition of $\mathtt{A}$, then successively find definitions $\mathtt{A}_1 \equiv \mathtt{C}_2$,

[8] http://bioportal.bioontology.org

[9] http://owl.cs.manchester.ac.uk/repository/

…, $\mathtt{A}_{n-1} \equiv \mathtt{C}_n$ with $\mathtt{A}_i$ occurring in $\mathtt{C}_i$ until $\mathtt{A}$ occurs in $\mathtt{C}_n$ (a cycle has been detected) or the sequence cannot be extended further. In the former case, delete $\mathtt{A}_{n-1} \equiv \mathtt{C}_n$. It is clear that the overall result depends on the order in which the concept names are traversed. (3) Remove role inclusion axioms $\mathtt{r} \sqsubseteq \mathtt{s}$, which can be consumed by MEX, but the resulting MEX module would be no longer guaranteed to be a minimal mCE-based module. (4) For all roles $\mathtt{r}, \mathtt{s}$ such that $\mathtt{s}$ is the *maximal* role with $\mathtt{r} \sqsubseteq \mathtt{s}$ with respect to (the transitive closure of) the removed role hierarchy, replace all concept descriptions $\exists\mathtt{r}.\mathtt{C}$ with $\exists\mathtt{r}.\mathtt{C} \sqcap \exists\mathtt{s}.\mathtt{C}$. This step "mimics" the behavior of the MEX approach on the original ontology, which first extracts a $\bot$-module from the role hierarchy and then adds to it a minimal mCE-based module for the thus extended signature from the remaining ontology. The additional $\exists\mathtt{s}.\mathtt{C}$ reflects that signature extension.

Comparing modules and locality. In order to compare syntactic and semantic locality, as well as LBMs and MEX modules, we want to understand:

1. whether, for a given seed signature Σ, it is likely that the semantic Σ-module is smaller than the syntactic Σ-module, or the MEX module for Σ is smaller than any of the previous two and, if so by how much,[10]
2. how feasible the extraction of semantic LBMs is.

For this purpose, we compare

- $\emptyset$-semantic locality and $\bot$-syntactic locality,
  Δ-semantic locality and $\top$-syntactic locality,
- $\emptyset$-modules and $\bot$-modules,
  Δ-modules and $\top$-modules,
  $\Delta\emptyset^*$-modules and $\top\bot^*$-modules,
- MEX modules and $\Delta\emptyset^*$-modules.

Due to the recursive nature of Algorithm 1, our investigation is both on a

per-axiom-basis: given axiom α and signature Σ, is it likely that α is $\emptyset$-local (Δ-local, resp.) w.r.t. Σ but not $\bot$-local ($\top$-local, resp.) w.r.t. Σ?

per-module basis: given a signature Σ, is it likely that
- $\bot\text{-mod}(\Sigma, \mathcal{O}) \neq \emptyset\text{-mod}(\Sigma, \mathcal{O})$, or
- $\top\text{-mod}(\Sigma, \mathcal{O}) \neq \Delta\text{-mod}(\Sigma, \mathcal{O})$, or
- $\top\bot^*\text{-mod}(\Sigma, \mathcal{O}) \neq \Delta\emptyset^*\text{-mod}(\Sigma, \mathcal{O})$, or
- $\Delta\emptyset^*\text{-mod}(\Sigma, \mathcal{O}) \neq \mathsf{MEX}\text{-mod}(\Sigma, \mathcal{O})$?

If yes, is it likely that the difference is large?

Clearly we need to pick, for each ontology in our corpus, a suitable set of signatures, and this poses a significant problem. A full investigation is infeasible: if $m = \#\widetilde{\mathcal{O}}$, there are 2^m possible seed signatures, so that testing axioms for locality against *all* the signatures is already impossible for $m \sim 100$. One could assume that comparing modules is easier since many signatures can lead to the same module. In other words, the statistically significant amount of modules

[10] Recall: the MEX module is always a subset of the semantic Σ-module, which is always a subset of the syntactic Σ-module.

w.r.t. the total amount of modules is not larger than that of seed signatures needed w.r.t. the total amount of seed signatures. In previous work [7,9], however, modules have been studied with respect to how numerous they are in real-world ontologies. The experiments carried out suggest that the number of modules in ontologies is, in general, exponential w.r.t. the size of the ontology. Moreover, the extraction of enough *different* modules can be hard, because by looking just at seed signatures there is no chance to avoid the extraction of the same module many times. In particular, for a module $\mathcal{M}$ there can be exponentially many seed signatures w.r.t. $\#\widetilde{\mathcal{M}}$ that generate $\mathcal{M}$ [5].

We will consider seed signatures of two kinds: genuine seed signatures and random seed signatures.

Genuine seed signatures. A module does not necessarily show an internal coherence: e.g., if we had an ontology $\mathcal{O}$ about the domains of geology and philosophy, we could extract the module for the signature $\Sigma = \{\texttt{Epistemology}, \texttt{Mineral}\}$. That module is likely to be the union of the two disjoint modules for $\Sigma_1 = \{\texttt{Epistemology}\}$ and $\Sigma_2 = \{\texttt{Mineral}\}$ [8].

In contrast, *genuine modules* can be said to be coherent: they are those modules that cannot be decomposed into the union of two "$\subseteq$"-uncomparable modules. Interestingly, a module $\mathcal{M}$ is genuine iff there exists an axiom α such that $\mathcal{M} = x\text{-mod}(\widetilde{\alpha}, \mathcal{O})$. As a consequence, there are only linearly many genuine modules in the size of $\mathcal{O}$, and extracting one module per axiom is enough for obtaining all of them. Moreover, all modules of $\mathcal{O}$ are composed from genuine modules [8]. Thus, genuine modules are of special interest, and we can investigate *all* of them, together with the corresponding *genuine signatures.*

Random seed signatures. Since a full investigation of all the signatures is impossible, we compare locality—both on a per-axiom and per-module basis—as well as LBMs and MEX modules on a *random* signature Σ, which we select by setting each named entity E in the ontology to have probability $p = 1/2$ of being included in Σ. This ensures that each Σ will have the same probability to be chosen. This approach has a clear setback: the random variable "size of the seed signature generated" follows a binomial distribution, so a random seed signature is highly likely to be rather large and to contain half the terms of the ontology. However, we do not yet have enough insight into what *typical* seed signatures are for module extraction, so biasing the selection of signatures to, for example, those of a certain size has no *rationale.* In contrast, selecting random seed signatures avoids the introduction of any bias. Moreover, this choice is complementary to the selection of *all* the genuine signatures, which are in general small.

With this in mind, we will analyze the modules obtained by random signatures with $p = 1/2$, and we will see in Section 4 that the module sizes obtained do allow for a reliable statement about the differences observed.

How many seed signatures do we have to sample from a given ontology $\mathcal{O}$ in order to obtain statistically significant statements about modules determined by the real population of *all* signatures from $\mathcal{O}$? We apply the usual statistical model of confidence intervals [21], aiming at a confidence level of 95% that the true proportion of differences between modules – i.e., the proportion of seed signatures that lead to different modules – lies in the confidence interval (±5%) of the observed proportion. Then we can generalize the conclusions for the random sample to the full population because the probability that the proportion of differences among modules for all seed signatures differs by no more than 5% from the proportion observed in the sample (and reported in Section 4) is 95%. In order to reach this confidence level, we need a sample size of at least 385 elements, independently of the size of the full population: for a two-sided test to detect a change in the proportion defective of size δ in either direction, the minimum sample size is

$$N \geqslant \frac{p(1-p)}{\delta^2}\, z^2_{1-\alpha/2}\,,$$

where p is the observed proportion, α the significance level, and $z_{1-\alpha/2}$ the critical value of the underlying distribution [2]. Here, we use the normal distribution as an approximation of the binomial distribution which is usually assumed for proportions in random sampling; hence the significance level of $\alpha = 0.05$ leads to $z_{1-\alpha/2} \approx 1.96$. Furthermore, although we do not know the value p in advance, it is clear that $p(1-p) \leqslant 0.25$ because $0 \leqslant p \leqslant 1$. The confidence interval of ±5% determines the error of $\delta = 0.05$. Therefore, we obtain

$$N \geqslant \frac{0.25}{0.05^2} \cdot 1.96^2 \approx 384.16,$$

that is, a representative sample for these parameters needs at least 385 elements, and this number is independent of the population size. For ontologies with at least 9 elements in the signature, we will therefore draw a sample of size 400. For all other ontologies, we will look at *all* of the $\leqslant 400$ signatures– then, the sample will coincide with the whole population.

Summary. We compare, for every ontology $\mathcal{O}$ in our corpus,

(T1) for random seed signatures Σ from $\mathcal{O}$,
- **(a)** for each axiom α in $\mathcal{O}$, is α
  - $\emptyset$-local w.r.t. Σ but not $\bot$-local w.r.t. Σ?
  - Δ-local w.r.t. Σ but not $\top$-local w.r.t. Σ?
- **(b)** is
  - $\bot$-mod$(\Sigma, \mathcal{O}) \neq \emptyset$-mod$(\Sigma, \mathcal{O})$?
  - $\top$-mod$(\Sigma, \mathcal{O}) \neq \Delta$-mod$(\Sigma, \mathcal{O})$?
  - $\top\bot^*$-mod$(\Sigma, \mathcal{O}) \neq \Delta\emptyset^*$-mod$(\Sigma, \mathcal{O})$?
  - $\Delta\emptyset^*$-mod$(\Sigma, \mathcal{ELI}(\mathcal{O})) \neq \mathsf{MEX}$-mod$(\Sigma, \mathcal{ELI}(\mathcal{O}))$?

(T2) for each axiom signature $\widetilde{\alpha}$ from $\mathcal{O}$,
- **(a)** for each axiom β in $\mathcal{O}$, is β
  - $\emptyset$-local w.r.t. $\widetilde{\alpha}$ but not $\bot$-local w.r.t. $\widetilde{\alpha}$?
  - Δ-local w.r.t. $\widetilde{\alpha}$ but not $\top$-local w.r.t. $\widetilde{\alpha}$?
- **(b)** is
  - $\bot$-mod$(\widetilde{\alpha}, \mathcal{O}) \neq \emptyset$-mod$(\widetilde{\alpha}, \mathcal{O})$?
  - $\top$-mod$(\widetilde{\alpha}, \mathcal{O}) \neq \Delta$-mod$(\widetilde{\alpha}, \mathcal{O})$?
  - $\top\bot^*$-mod$(\widetilde{\alpha}, \mathcal{O}) \neq \Delta\emptyset^*$-mod$(\widetilde{\alpha}, \mathcal{O})$?
  - $\Delta\emptyset^*$-mod$(\Sigma, \mathcal{ELI}(\mathcal{O})) \neq \mathsf{MEX}$-mod$(\Sigma, \mathcal{ELI}(\mathcal{O}))$?

Our sample selection includes large as well as small seed signatures: the random seed signatures created to answer T1 will tend to contain around half the terms in the ontology, while the signatures used to answer T2 will range over *all* signatures of *single axioms* and therefore tend to be small.

4 Results of the Experiments

4.1 Semantic versus syntactic locality

No differences in locality. The main result of the experiment is that, for the vast majority of the ontologies in our corpus, no difference between syntactic and semantic locality is observed, for all three variants $\bot$ vs. $\emptyset$, $\top$ vs. Δ, and $\top\bot^*$ vs. $\Delta\emptyset^*$. More precisely, for 209 out of 242 ontologies, we obtain that:

(T1) for random seed signatures, there is no statistically significant difference
  - **(a)** between semantic and syntactic locality of any kind,
  - **(b)** between semantic and syntactic LBMs of any kind;

(T2) given *any* genuine signature $\tilde{\beta}$, there is no difference
  - **(a)** between semantic and syntactic locality of any kind,
  - **(b)** between semantic and syntactic LBMs of any kind.

More specifically, for all randomly generated seed signatures and *all* genuine signatures, the corresponding bottom-modules (and the corresponding top- and nested-modules, respectively) agree, and every axiom is either $\bot$- and $\emptyset$-local, or none of both (and either $\top$- and Δ-local, or none of both).

The 209 ontologies include Galen and People, which are renowned for having unusually large $\bot$-modules [3,9].

In most cases, extracting a semantic and syntactic LBM each took only a few milliseconds, so a performance comparison is not meaningful. For some ontologies, the semantic LBM took considerably longer to extract than the syntactic: up to 5 times for nested-modules in Molecule Role, and up to 34 times in Galen.

Differences in locality. We have observed differences between syntactic and semantic locality for 33 ontologies in our corpus. We call the axioms that cause these differences *culprits* – patterns of axioms which are not $\bot$-local ($\top$-local, respectively) w.r.t. some signature Σ, but which are $\emptyset$-local (Δ-local, respectively) w.r.t. Σ. We have identified four types of patterns, a–d, and we describe them in the following. Sometimes, culprit axioms *pull* additional axioms into the syntactic LBM, due to signature extension during module extraction.

For 6 out of the 33 ontologies, the differences solely consist of axioms of culprit type a: simple tautologies, which are among the inferences that have been pre-computed before the affected ontologies were published in BioPortal. We will first briefly describe this culprit pattern before we focus on the more interesting differences caused by culprits b–d.

We denote class *names* by $\mathtt{A}, \mathtt{B}$, complex classes by $\mathtt{C}, \mathtt{D}$, properties by $\mathtt{r}, \mathtt{s}, \ldots$, nominals by a, non-empty data ranges (e.g., $\mathtt{int}$ or $\mathtt{int}^{0..9}$) by R, possibly with indices. Σ denotes a signature for which a module is extracted or against which an axiom is checked for locality. Terms outside Σ are $\overline{\text{overlined}}$; we further use notation $\mathtt{C}^{\perp}$ and $\mathtt{C}^{\top}$ to denote classes that are bottom- or top-equivalent due to the grammar defining syntactic locality in [3, Fig. 3] and the analogous grammar for semantic locality.

Culprits of type *a* are simple tautologies that accidentally entered the "inferred view" (closure under certain entailments) of an ontology. These axioms do not occur in the original "asserted" versions and could, in principle, be detected in a simple preprocessing step.

Type-a culprits occur in 10 ontologies of the above 33 and are of the kinds $\mathtt{A} \sqsubseteq \mathtt{A}$ or $\mathtt{r} \equiv (\mathtt{r}^{-})^{-}$. Each such tautology is trivially $\emptyset$-local and Δ-local w.r.t. any Σ, but not always $\perp$- or $\top$-local: if Σ contains all terms in that tautology, then both sides of the subsumption (equivalence) are neither $\perp$- nor $\top$-equivalent.

Differences caused not solely by culprits of type *a* have been observed for 27 ontologies. In only 6 of these cases, the differences affect modules; in the remaining 20, they only affect locality of single axioms (tests T1 a and T2 a). We will focus on the former 6, listed in Table 1, and refer to the appendix for details on all 27.

Ontology	Abbreviation	DL expressivity	#axioms	#terms
MiniTambis-repaired	MiniT	$\mathcal{ALCN}$	170	226
Tambis-full	Tambis	$\mathcal{SHIN}(\mathcal{D})$	592	496
Bleeding History Phenotype	BHO	$\mathcal{ALCIF}(\mathcal{D})$	1,925	581
Neuro Behavior Ontology	NBO	$\mathcal{AL}$	1,314	970
Pharmacogenomic Relationsh...	PhaRe	$\mathcal{ALCHIF}(\mathcal{D})$	459	311
Terminological and Ontological...	TOK	$\mathcal{SRIQ}(\mathcal{D})$	466	330

Table 1. Ontologies that exhibit differences in modules

According to Table 1, differences between modules occur for ontologies of medium to large size and medium to high expressivity. Differences in locality alone additionally affect small ontologies such as Koala (42 axioms) and Pilot Ontology (85 axioms), as well as large ontologies such as Galen (4,735 axioms) and Experimental Factor Ontology (7,156 axioms). The number of axioms *causing* these differences (i.e., matching the culprit patterns) in the affected ontologies is small except for Galen, and most of the observed differences are relatively small.

Table 2 gives a representative selection of the differences in *modules* observed, plus the relative sizes of modules extracted for (T1) and (T2). For a complete overview, including differences in locality of single axioms, see the table in the appendix.
Table 2 shows small absolute differences for miniT, BHO, NBO, and TOK. In Tambis, large differences occur only for genuine modules. Finally, in PhaRe, large differences occur only for top-modules, which are hardly used in practice.

For all these ontologies, a single syntactic or semantic module was extracted within only a few milliseconds, making module extraction times roughly equal.

Culprits of type *b* are axioms with an $\exists$-restriction on a set of nominals or a non-empty data range on the right-hand side, such as $\mathtt{A} \sqsubseteq \exists\overline{\mathtt{r}}.\{a_1, \ldots, a_n\}$ or $\mathtt{A} \sqsubseteq \exists\overline{\mathtt{r}}.R$. These axioms are Δ-local w.r.t. a signature that does not contain $\mathtt{r}$

Ontol.	Types affected	#diffs	size of diffs #axs	(rel.)	size of $\Delta\emptyset^*$-modules (%) T1 range	T1 avg.	T2 range	T2 avg.	culprit type	+ freq.
miniT	bot, nested	14–25%	1–7	0–600%[b]	48–79	66	0–8	2	c	3
Tambis	bot, nested	32–57%	2–41[c]	1–62%[c]	75–88	82	0–34	9	c	8
BHO[a]	nested	17%	1–12	0–300%	55–72	65	0–31	4	b	31
NBO[a]	nested	3%	2	0–200%	64–78	71	0–3	0	d	3
PhaRe[a]	top, nested	1–8%	1–326[d]	0–6,520%[d]	50–70	60	0–8	1	d	10
TOK	top, nested	49–100%	1–7	0–9%	48–68	59	9–17	10	d	3

[a]differences only for genuine modules
[b]differences $>5\%$ only for genuine modules
[c]differences >11 axioms ($>2\%$) only for genuine modules
[d]differences >13 axioms ($>1{,}300\%$) only for top-modules

The columns show: ontology name (abbreviations: see Table 1); type of modules affected; relative number of module pairs with differences; number of axioms in the differences (absolute and relative to the $\emptyset$- or Δ- or $\Delta\emptyset^*$-case); type of culprit present and number of axioms of this type involved in differences.

Table 2. Overview of observed differences between modules

because they become tautologies if r is replaced by $\top$. However, they can never be $\top$-local unless $\mathtt{A}$ is replaced by some $\mathtt{C}^\perp$.

Culprit-b axioms affect genuine modules of BHO, and (only) locality of single axioms for 4 more ontologies. We observed a variant $\mathtt{A} \equiv \mathtt{C}^\top \sqcap \exists \overline{\mathtt{r}}.R$.

Culprits of type *c* are axioms α that contain a class description $\mathtt{C}$ such that (a) $\mathtt{C}$ becomes equivalent to $\perp$ (or $\top$) if all terms outside Σ are replaced by $\perp$ (or $\top$); (b) this causes α to be semantically $\perp$-local (or $\top$-local); but (c) the grammars for syntactic locality do not "detect" $\mathtt{C}$ to be a $\mathtt{C}^\perp$ (or $\mathtt{C}^\top$). For example, $\mathtt{C} = \forall \mathtt{r}.\overline{\mathtt{A}} \sqcap \exists \mathtt{r}.\top$ becomes $\perp$-equivalent if $\mathtt{A}$ is replaced by $\perp$; the same holds with cardinality restrictions in place of "$\exists$". Consequently, axioms such as $\mathtt{A}^\perp \equiv \mathtt{B} \sqcap \forall \mathtt{r}.\mathtt{C}^\perp \sqcap \forall \mathtt{s}.\{a\} \sqcap {=}3\,\mathtt{r}.\top$, (taken from Koala) are $\emptyset$-local but not $\perp$-local.

We found this pattern in 8 ontologies. Only in miniT and Tambis, it affects a large proportion of bottom- and nested-modules, with additional axioms "pulled in". Still, the size of the differences is modest, as argued above. Some of the remaining 6 ontologies contain different kinds of complex classes that cause differences in top-locality of single axioms.

Culprits of type *d* are axioms where a class (or property) name from the left-hand side occurs on the right-hand side together with a top-equivalent property (or class), causing differences in top-modules. The simplest such axiom is $\mathtt{A} \sqsubseteq \exists \overline{\mathtt{r}}.\mathtt{A}$, which is Δ-local because replacing r with $\top$ makes it a tautology. The axiom is only $\top$-local if Σ contains neither $\mathtt{r}$ nor $\mathtt{A}$. We have found further, more complex, examples in Adverse Event Reporting Ontology and Galen; see the appendix.

We have observed culprits of type d in 17 ontologies, see the detailed overview in the appendix. Only in 3 cases (NBO, PhaRe, and TOK) are modules affected.

Galen contains 121 culprit-d axioms, but they only affect locality of single axioms. The time differences for Galen are remarkable: checking all axioms for Δ-locality takes up to 70 times longer than checking them for $\top$-locality.

Module sizes. The selection of the signatures for the experiment was designed to allow for the analysis of two, complementary, kinds of modules: 1) genuine modules, which constitute a base of all modules, extracted from generally small axiom signatures; 2) a statistically significant amount of random modules, obtained from random, unbiased signatures which are likely to contain half the terms of the ontology. We argue in what follows that it is neither the case that genuine modules are so small to be almost irrelevant sets to investigate, nor that random modules are so big to leave no space for differences to be observed. We will focus on syntactic modules which contain the other kinds of modules.

During the experiment we have computed and analyzed a high number of genuine modules: more than 380K for the $\bot$-notion, more than 40K for the $\top$-notion, and more than 440K for the $\top\bot^*$-notion of locality. As we mentioned above, these modules tend to be quite small. However, they are not of irrelevant size: $\sim 8\%$ of the genuine $\bot$-modules, $\sim 11\%$ of the genuine $\top$-modules, and $\sim 5\%$ of the genuine $\top\bot^*$-modules contain more than 20% of the axioms of the corresponding ontology. So the low number of differences observed is not due to checking only against very small modules.

With a similar and complementary discussion, we argue that the modules obtained through random, "big" signatures do not necessarily contain almost all of the ontology: e.g., 39% of all random $\top\bot^*$-modules, and 28% of all random $\bot$-modules, contain less that 60% of the axioms of the corresponding ontology.

To sum up, the lack of differences between the modules is not due to too small or to too big sizes of the modules selected.

Discussion. All culprits hardly ever cause significant differences in modules. Only for PhaRe are differences between semantic and syntactic modules not negligible, but we were able to relativize them, see [6].

Table 1 may suggest that culprits occur only in expressive ontologies. However, patterns a, c, d can, in principle, already occur in simple terminologies in $\mathcal{EL}$ and $\mathcal{ALC}$, respectively. Evidently, type-a culprits can easily be filtered out in a preprocessing step. For types c and d, there is no hope for an exhaustive extension to locality because they can (and do) occur in arbitrarily complex shapes and contexts. For this reason, the identification of culprits can only be done "on demand", i.e., by observing the differences in the modules of given ontologies.

Patterns of type b rely on nominals or datatypes – but they are *repairable* by a straightforward extension to the definition of syntactic locality: one can extend the locality definition to distinguish $\bot$- and $\top$-*distinct* classes, by adding appropriate grammars to the definition of syntactic locality, and adding more cases of $\bot$- and $\top$-equivalent classes to the existing grammars. However, from the small numbers of differences observed, we doubt that such an extension of syntactic locality will have any significant effects in practice.

4.2 LBMs vs MEX Results

The results of the experimental comparison of syntactic/semantic LBMs and MEX modules are summarized in Table 3. They show that MEX modules smaller than the corresponding LBMs can be found in ~27% of the preprocessed ontologies, for either random or axiom-based seed signatures. At the same time, unsurprisingly, syntactic and semantic LBMs do not differ at all for these simple $\mathcal{ELI}$ ontologies.

Experiment	#ontol. with diffs.	% tests with diffs.	avg size of diffs #axs	rel.
Random signatures	66	84%	0–26	0–13%
Axiom signatures	61	12%	0–13	0–80%

The results from the third column on are averaged over all ontologies with differences LBM–MEX in at least one module. For example, the last two columns show the average min and max absolute (resp. relative) difference between LBMs and MEX modules.

Table 3. Differences between MEX and LBMs ($\top\bot^*$, $\Delta\emptyset^*$)

In experiments with random seed signatures, it can be seen that for those ontologies where there *are* differences (most notably, Galen), they occur in many tests. Thus, the difference appears to be caused by features of the ontology, not some particular seed signatures. Also, the difference sometimes comes out large in certain tests, also for genuine modules. For example, for the signature of the following axiom in Galen, both $\Delta\emptyset^*$-mod and $\top\bot^*$-mod contain 127 axioms while the MEX-module only contains the axiom itself:[11] $\texttt{RICF} \equiv \texttt{ICF} \sqcap \exists \texttt{ISFO}.\texttt{RSH}$.

We analyzed whether the differences observed correlate with the size of the original ontology, its expressivity or the extent of the modification done in the $\mathcal{ELI}$-fication. There is no correlation with size but, as is to be expected, with the other two features, which are closely connected to each other. Table 4 illustrates the observations by dividing the 239 ontologies tested into four groups. The ontologies in Group 1 are in a format MEX can handle, so they have not been modified. The others required more or less heavy modifications (Groups 2–4). Differences between MEX and LBMs as described above occur only for ontologies that required heavy modifications (Group 4).

Group	#axioms removed	#ontologies	ontology size (avg.)
1 unchanged ontologies *no* diff. $\Delta\emptyset^* \setminus$ MEX	0	33 (14%)	19–16,066 (2,176)
2 little-changed ontologies *no* diff. $\Delta\emptyset^* \setminus$ MEX	1–28	36 (15%)	13– 6,587 (466)
3 largely-changed ontologies *no* diff. $\Delta\emptyset^* \setminus$ MEX	31–7,836 (avg. 884)	104 (44%)	51–13,153 (2,373)
4 largely-changed ontologies *with* diff. $\Delta\emptyset^* \setminus$ MEX	30–12,185 (avg. 1,001)	66 (27%)	42–12,344 (1,843)

Table 4. Overview of MEX experiment

As expected, the expressivity among Groups 1 and 2 is generally low: only 21 ontologies in Group 2 use expressivity above $\mathcal{ALE}$ (up to $\mathcal{SHIF}(\mathcal{D})$, which is an outlier). However, the size of some ontologies in Group 1 is already considerable: 22 out of 33 have > 100 axioms; 10 have $> 1,000$ axioms. In contrast, the ontologies in Group 4 have almost always high expressivity, for example 27 out of 66 contain nominals.

[11] The acronyms denote RightIneffectiveCardiacFunction, IneffectiveCardiacFunction, isSpecificFunctionOf, RightSideOfHeart.

Despite the correlation between the impact of the $\mathcal{ELI}$-fication and the differences observed between MEX- and $\Delta\emptyset^*$-modules, we cannot claim that there is a causation between the two events. Indeed, we have investigated the reasons for the differences observed between the two kinds of modules, and we have noticed that in all the cases the culprit is the proliferation of equivalence axioms. For example $\mathtt{A} \equiv \mathtt{B}$ will end up in the $\Delta\emptyset^*$-mod for any seed signature containing either A or B. It is, however, an mCE of $\emptyset$ w.r.t. to either $\{\mathtt{A}\}$ or $\{\mathtt{B}\}$.

The experimental results in view of this insight are summarized as follows:

Random-modules experiment: the 66 ontologies where differences between random MEX- and $\Delta\emptyset^*$-modules were observed, coincide exactly with those where equivalences occur in the $\mathcal{ELI}$-TBox.

Genuine-modules experiment: all 61 ontologies where differences between genuine MEX- and $\Delta\emptyset^*$-modules were observed contain equivalence axioms.

We conjecture that the low expressivity of the $\mathcal{ELI}$-language reduce the possibility of MEX- and $\Delta\emptyset^*$-modules to differ only to the presence of equivalences. In addiction to the empirical evidence for such a claim, we plan to investigate further this aspect in future work.

5 Conclusion and outlook

Summary. We obtain three main observations from our experiments. (1) In general, there is no or little difference between semantic and syntactic locality. Hence, the computationally cheaper syntactic locality is a good approximation of semantic locality. (2) In most cases, there is no or little difference between LBMs and MEX modules. (3) Though in principle hard to compute, semantic LBMs can be extracted rather fast in practice. Still, their extraction often takes considerably longer than that of syntactic LBMs. We cannot make any statement about MEX module extraction times because we use the original MEX implementation, which combines loading and module extraction. Due to results (1) and (2), hardly any benefit can be expected from preferring potentially smaller modules (MEX or semantic LBMs) to cheaper syntactic LBMs. For the ontologies Galen and People, which are "renowned" for having disproportionately large modules, syntactic and semantic LBMs do not differ. Only for Galen are MEX modules considerably smaller than LBMs.

Not only does our study evaluate how good the cheap syntactic locality approximates semantic locality and model conservativity, it also required us to provide the first implementation for extracting modules based on semantic locality. Furthermore, we have been able to fix bugs in the existing implementation of syntactic modularity. A complete report of bugfixes is beyond the scope of the paper; as an example, early runs of the experiment led us to correcting the treatment of reflexivity axioms by the locality checker in the OWL API.

Future work. Two issues are interesting for future work: (1) Sampling seed signatures so that all sizes of signatures are equally likely to be sampled; (2) Comparing LBMs to other types of conservativity-based modules.

As for (1), the current sampling causes small and large signatures to be underrepresented. One might argue that, for big ontologies, the typical module extraction scenario does not require large seed signatures – but it does sometimes require relatively small seed signatures, for example, when a module is extracted to efficiently answer a certain entailment query of typically small size. We therefore plan to conduct a similar experiment using other sampling methods. Concerning (2), one could include, for example, the technique based on reduction to QBF for the OWL 2 QL profile [17] when an off-the-shelf implementation becomes available.

References

1. Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.
2. Carroll Croarkin and Paul Tobias, editors. *NIST/SEMATECH e-Handbook of Statistical Methods*. NIST/SEMATECH, 2012. `http://www.itl.nist.gov/div898/handbook`.
3. Bernardo Cuenca Grau, Ian Horrocks, Yevgeny Kazakov, and Ulrike Sattler. Modular reuse of ontologies: Theory and practice. *J. of Artif. Intell. Research*, 31(1):273–318, 2008.
4. Bernardo Cuenca Grau, Bijan Parsia, Evren Sirin, and Aditya Kalyanpur. Modularity and Web ontologies. In *Proc. of KR-06*. AAAI Press/The MIT Press, 2006.
5. Chiara Del Vescovo, Damian Gessler, Pavel Klinov, Bijan Parsia, Ulrike Sattler, Thomas Schneider, and Andrew Winget. Decomposition and modular structure of BioPortal ontologies. In *Proc. of ISWC-11*, volume 7031 of *LNCS*, pages 130–145, 2011.
6. Chiara Del Vescovo, Pavel Klinov, Bijan Parsia, Ulrike Sattler, Thomas Schneider, and Dmitry Tsarkov. Empirical study of logic-based modules: Cheap is cheerful. Technical report, 2013. `https://sites.google.com/site/cheapischeerful/`.
7. Chiara Del Vescovo, Bijan Parsia, Ulrike Sattler, and Thomas Schneider. The modular structure of an ontology: an empirical study. In *Proc. of DL 2010*, volume 573 of *ceur-ws.org*, 2010.
8. Chiara Del Vescovo, Bijan Parsia, Ulrike Sattler, and Thomas Schneider. The modular structure of an ontology: Atomic decomposition. In *Proc. of IJCAI-11*, pages 2232–2237, 2011.
9. Chiara Del Vescovo, Bijan Parsia, Ulrike Sattler, and Thomas Schneider. The modular structure of an ontology: Atomic decomposition and module count. In *Proc. of WoMO-11*, volume 230 of *FAIA*, pages 25–39, 2011.
10. James Garson. Modularity and relevant logic. *Notre Dame J. of Formal Logic*, 30(2):207–223, 1989.
11. Silvio Ghilardi, Carsten Lutz, and Frank Wolter. Did I damage my ontology? A case for conservative extensions in Description Logics. In *Proc. of KR-06*, pages 187–197. AAAI Press/The MIT Press, 2006.
12. Matthew Horridge, Bijan Parsia, and Ulrike Sattler. Laconic and precise justifications in OWL. In *Proc. of ISWC-08*, volume 5318 of *LNCS*, pages 323–338, 2008.
13. Matthew Horridge, Bijan Parsia, and Ulrike Sattler. The state of bio-medical ontologies. In *Proc. of ISMB-11*, 2011.

14. Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible $\mathcal{SROIQ}$. In *Proc. of KR-06*, pages 57–67, 2006.
15. Boris Konev, Carsten Lutz, Dirk Walther, and Frank Wolter. Semantic modularity and module extraction in description logics. In *Proc. of ECAI-08*, pages 55–59, 2008.
16. Roman Kontchakov, Luca Pulina, Ulrike Sattler, Thomas Schneider, Petra Selmer, Frank Wolter, and Michael Zakharyaschev. Minimal module extraction from DL-Lite ontologies using QBF solvers. In *Proc. of IJCAI-09*, pages 836–841, 2009.
17. Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev. Logic-based ontology comparison and module extraction, with an application to DL-Lite. *Artificial Intelligence*, 174(15):1093–1141, 2010.
18. Carsten Lutz, Dirk Walther, and Frank Wolter. Conservative extensions in expressive Description Logics. In *Proc. of IJCAI-07*, pages 453–458, 2007.
19. Carsten Lutz and Frank Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. of Symbolic Computation*, 45(2):194–228, 2010.
20. Ulrike Sattler, Thomas Schneider, and Michael Zakharyaschev. Which kind of module should I extract? In *Proc. of DL 2009*, volume 477 of `ceur-ws.org`, 2009.
21. Michael Smithson. *Confidence Intervals*. Quantitative Applications in the Social Sciences. Sage Publications, 2003.

A Overview of differences observed for semantic vs. syntactic locality

Altogether, we have observed differences for the following ontologies (see the overview in Table 7 for details).

Abbreviation	Ontology name
AERO	adverse-event-reporting-ontology
BAO	bioassay-ontology
BHO	bleeding-history-phenotype
BOOTSt	gene-regulation-ontology, ID 1106
ChemInf	chemical-information-ontology
EDAM	bioinformatics-operations-types-of-data-data-formats-and-topics
EFO	experimental-factor-ontology
FLU	influenza-ontology
GFOBio	general-formal-ontology-biology
GRO	gene-regulation-ontology, ID 1082
HL7	health-level-seven
KiSAO	kinetic-simulation-algorithm-ontology
IAO	information-artifact-ontology
LiPrO	lipid-ontology
MF	mental-functioning-ontology
MFOEm	emotion-ontology
NBO	neuro-behavior-ontology
NEMO	neural-electromagnetic-ontologies
OBIws	bioinformatics-web-service-ontology
OGI	ontology-for-genetic-interval
PhaRe	phare (Pharmacogenomic Relationships Ontology)
PO	protein-ontology
POL	pilot-ontology
SIO	semanticscience-integrated-ontology
SWO	software-ontology
TEDDY	terminology-for-the-description-of-dynamics
TMO	translational-medicine-ontology
TOK	terminological-and-ontological-knowledge-resources
VSO	vital-sign-ontology
Galen	Galen
Koala	Koala
miniT	miniTambis-repaired
Tambis	Tambis-full

Differences between modules The following table lists all differences observed between syntactic and semantic modules. The columns show: the ontology name (abbreviations are defined above); the type of modules affected; the number of cases with differences (absolute and relative); the number of axioms in the differences (absolute and relative to the $\emptyset$- or Δ- or $\Delta\emptyset^*$-case); the average time ratio semantic : syntactic ("—" indicates that no reliable statement is possible: the time for the syntactic case is only a few, often 0, milliseconds); the type of culprit present and the number of axioms of this type involved in differences. Cases where differences are caused solely by culprits of type a are grayed out.

Table 5: Detailed overview of differences in modules *observed*

Onto-logy	Module types affected	number of diffs #sigs	(rel.)	size of diffs #axs	(rel.)	time ratio average	culprit type (frequency)
AERO	random all	399–400	99–100%	1	0%	—	a 1
	genuine all	17–873	2–100%	1	0%	—	”
BHO	genuine star	329	17%	1–12	0–300%	—	b 31
ChemInf	random all	298–400	75–100%	1	0%	—	a 1
	genuine all	3–1,237	0–100%	1	0–1%	—	”
EDAM	random all	400	100%	1	0%	3.38	a 1
	genuine all	9–535	0–14%	1	0–100%	—	”
EFO	random all	400	100%	14–31	0–1%	1.49	a 31
	genuine all	128–7,156	2–100%	1–31	0–200%	—	”
FLU	random all	400	100%	1	0%	—	a 1
	genuine all	34–1,696	2–100%	1	0–1%	—	”
HL7	random all	400	100%	1–7	0%	1.35	a 7
	genuine all	44–553	0–7%	1–7	0–100%	—	”
IAO	random all	400	100%	22–49	7–27%	—	a 49
	genuine all	201–554	36–100%	1–49	0–200%	—	”
NBO	genuine star	41	3%	2	0–200%	—	d 3
PhaRe	genuine top	5	1%	325–326	0–6,520%	—	d 10
	genuine star	35	8%	1–13	0–1,300%	—	”
SIO	random all	400	100%	1	0%	—	a 1
	genuine all	77–2,044	4–100%	1	0–20%	—	”
SWO	random all	400	100%	20–54	0–2%	1.59	a 54
	genuine all	190–5,507	3–100%	1–54	0–200%	—	”
TEDDY	random all	400	100%	11–14	0%	42.93	a 14
	genuine all	6,956–12,344	56–100%	1–14	0–100%	1.40	”
TOK	random top, star	195	49%	2–7	0–3%	—	d 2
	genuine top, star	459–466	98–100%	1–7	4–9%	—	” 3
miniT	random bot, star	100	25%	1–5	0–5%	—	c 2
	genuine bot, star	23–26	14–15%	1–7	0–300%	—	” 3
Tambis	random bot, star	226	57%	2–11	0–2%	—	c 4

Continued on next page

Table 5 – Continued from previous page

Onto-logy	Module types affected	number of diffs #sigs	(rel.)	size of diffs #axs	(rel.)	time ratio average	culprit type (frequency)
	genuine bot, star	191	32%	4–41	0–62%	—	" 8

Differences in locality only The following table lists all differences observed between syntactic and semantic locality of single axioms. The columns show: the ontology name (abbreviations are defined above); the type of modules affected; the number of cases with differences (absolute and relative); the number of axioms in the differences (absolute and relative to the $\emptyset$- or Δ- or $\Delta\emptyset^*$-case); the average time ratio semantic : syntactic ("—" indicates that no reliable statement is possible: the time for the syntactic case is only a few, often 0, milliseconds); the type of culprit present and the number of axioms of this type involved in differences. Cases where differences are caused solely by culprits of type a are grayed out.

Table 6: Detailed overview of differences in locality of single axioms *observed*

Onto-logy	Module types affected	number of diffs #sigs	(rel.)	size of diffs #axs	(rel.)	time ratio average	culprit type (frequency)
AERO	random all	201–285	50–71%	1–4	0–1%	—	d 3, a 1
	genuine all	4–11	0–1%	1	0–1%	—	"
BAO	random top	179	45%	1–6	0–1%	—	b 7
	genuine top	1774	99%	3–7	0–11%	—	"
BHO	random top	370	93%	1–48	0–4%	—	b 57
	genuine top	1925	100%	6–57	1–7%	—	"
BOOTSt	random top	96	24%	1	0%	—	d 1
	genuine top	2	<1%	1	0–1%	—	"
ChemInf	random all	213–326	53–82%	1–2	0%	—	b 1, a 1
	genuine all	3–1,209	0–98%	1–2	0–2%	—	"
EDAM	random all	206	52%	1	0%	—	a 1
	genuine all	8	<1%	1	0–50%	—	"
EFO	random all	400	100%	8–23	0–1%	1.42	a 31
	genuine all	128	2%	1–2	0–25%	—	"
FLU	random all	188	47%	1	0%	—	a 1
	genuine all	5	<1%	1	0–1%	—	"
GFOBio	random top	60	15%	1	0–1%	—	c 1
	genuine top	8	2%	1	0–2%	—	"
GRO	random top	104	26%	1	0%	—	d 1

Continued on next page

Table 6 – Continued from previous page

Onto-logy	Module types affected	number of diffs #sigs	(rel.)	size of diffs #axs	(rel.)	time ratio average	culprit type (frequency)
	genuine top	2	<1%	1	0–1%	—	”
HL7	random all	399	99%	1–7	0%	1.30	*a* 7
	genuine all	44	<1%	1	0–100%	—	”
IAO	random all	400	100%	14–35	5–17%	—	*a* 49
	genuine all	201	36%	1–2	0–8%	—	”
KiSAO	random top	109	27%	1–2	0%	—	*d* 2
	genuine top	200	28%	1–2	0–1%	—	”
LiPrO	random bot	1	<1%	1	0%	—	*c* 1
MF	random top	246	62%	1–7	0–3%	—	*d* 12
	genuine top	69	13%	1–4	0–5%	—	”
MFOEm	random top	256	64%	1–6	0–2%	—	*d* 12
	genuine top	70	10%	1–4	0–5%	—	”
NBO	random top	131	33%	1–3	0%	—	*d* 3
	genuine top	20	2%	1	0–50%	—	”
NEMO	random top	105	26%	1–13	0–1%	—	*b* 32, *d* 1
	genuine top	2,734	96%	20–92	0–30%	—	*b* 91, *d* 1
OBIws	random top	14	4%	1	0%	—	*d* 1
	genuine top	3	<1%	1	0–10%	—	”
OGI	random top	91	23%	1	0%	—	*d* 1
	genuine top	20	4%	1	0–1%	—	”
PhaRe	random top	204	51%	1–10	0–5%	—	*d* 12
	genuine top	52	11%	1–2	0–22%	—	”
PO	random top	145	36%	1–2	0%	—	*d* 2
	genuine top	9	1%	1–2	0–1%	—	”
POL	random top	88	22%	1	0%	—	*d* 1
	genuine top	8	9%	1	0–3%	—	”
SIO	random all	207–324	52–81%	1–5	0%	—	*b* 3, *d* 1, *a* 1
	genuine all	9-2,025	0–99%	1–4	0–25%	—	”
SWO	random all	400	100%	15–38	1–2%	1.69	*a* 54
	genuine all	190	3%	1–2	0–100%	—	”
TEDDY	random bot	400	100%	2–12	0%	—	*c* 1, *d* 6, *a* 14
	random top	400	100%	3–16	0%	28.48	”

Continued on next page

Table 6 – Continued from previous page

Onto-logy	Module types affected	number of diffs #sigs	(rel.)	size of diffs #axs	(rel.)	time ratio average	culprit type (frequency)
	genuine bot	167	1%	1–3	0–29%	—	”
	genuine top	587	5%	1–6	0%	29.96	”
TMO	random top	196	49%	1	0%	—	*b* 1
	genuine top	491	98%	1	0–2%	—	”
TOK	random top	227	57%	1–4	0–2%	—	*d* 4
	genuine top	69	15%	1–4	0–3%	—	”
VSO	random top	258	65%	1–7	0–2%	—	*d* 12
	genuine top	88	12%	1–4	0–5%	—	”
Galen	random top	393	98%	1–99	0–3%	—	*d* 121
	genuine top	717	15%	1–33	0–16%	—	”
Koala	random all	10–200	3–50%	1	0–8%	—	*b* 1, *c* 1
	genuine top	38	90%	1	0–12%	—	*b* 1
miniT	random bot	66	17%	1–2	0–3%	—	*c* 3
	genuine bot	11	6%	1–2	0–50%	—	” 2
Tambis	random bot	46	12%	1–2	0–1%	—	*c* 9
	genuine bot	12	2%	1–2	21–40%	—	” 8

B Examples of culprit axioms

We have $\overline{\text{overlined}}$ terms that must *not* be in the signature Σ in order for the axiom to be semantically but not syntactically local. For abbreviations of the ontology names, consult the table above.

Culprit type b.

– BAO contains the axioms

$$\overline{\mathsf{BAO\_0000346}} \equiv \overline{\mathsf{BAO\_0000200}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>49} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{<51}$$
$$\overline{\mathsf{BAO\_0000347}} \equiv \overline{\mathsf{BAO\_0000201}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>49} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{<51}$$
$$\overline{\mathsf{BAO\_0000348}} \equiv \overline{\mathsf{BAO\_0000201}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>79} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{<81}$$
$$\overline{\mathsf{BAO\_0000349}} \equiv \overline{\mathsf{BAO\_0000202}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>49} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{<51}$$
$$\overline{\mathsf{BAO\_0000588}} \equiv \overline{\mathsf{BAO\_0000006}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>49} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{<51}$$
$$\overline{\mathsf{BAO\_0001106}} \equiv \overline{\mathsf{BAO\_0000096}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>49} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{<51}$$
$$\overline{\mathsf{BAO\_0001108}} \equiv \overline{\mathsf{BAO\_0000096}} \sqcap \exists \overline{\mathsf{BAO\_0000195}}.\mathtt{float}^{>98},$$

where the term IDs have the following meaning.

Term ID	Type	Meaning
BAO_0000006	concept name	percent cytotoxicity
BAO_0000096	concept name	percent growth inhibition
BAO_0000195	role name	has percent response value
BAO_0000200	concept name	percent activation
BAO_0000201	concept name	percent inhibition
BAO_0000202	concept name	percent viability
BAO_0000346	concept name	50 percent activation
BAO_0000347	concept name	50 percent inhibition
BAO_0000348	concept name	80 percent inhibition
BAO_0000349	concept name	50 percent viability
BAO_0000588	concept name	50 percent cytotoxicity
BAO_0001106	concept name	50 percent growth inhibition
BAO_0001108	concept name	100 percent growth inhibition

– BHO contains axioms such as

$$\begin{aligned}\overline{\texttt{person}} \equiv\ & \exists \overline{\texttt{has\_demographic\_characteristic}}.\overline{\texttt{ethnicity}} \\ & \sqcap \exists \overline{\texttt{has\_demographic\_characteristic}}.\overline{\texttt{race}} \\ & \sqcap \exists \overline{\texttt{has\_sex}}.\overline{\texttt{sex}} \\ & \sqcap \exists \overline{\texttt{has\_age}}.\texttt{int}^{\geqslant 0}.\end{aligned}$$

– ChemInf contains the axiom

$$\begin{aligned}\texttt{CHEMINF\_000501} &\sqsubseteq \exists \overline{\texttt{CHEMINF\_000012}}.(\texttt{decimal} \sqcup \texttt{float} \sqcup \texttt{integer} \sqcup \texttt{long}) \\ \texttt{numeric\_data\_item} &\sqsubseteq \exists \overline{\texttt{has\_value}}.(\texttt{decimal} \sqcup \texttt{float} \sqcup \texttt{integer} \sqcup \texttt{long}) \quad \text{(without IDs).}\end{aligned}$$

– Koala contains the axiom

$$\texttt{GraduateStudent} \sqsubseteq \exists \overline{\texttt{hasDegree}}.\{\texttt{BA}, \texttt{BS}\}.$$

– TMO contains the axiom

$$\begin{aligned}\texttt{TMO\_0179} &\sqsubseteq \exists \overline{\texttt{TMO\_0185}}.\texttt{int}^{0..9} \\ \texttt{age\_in\_years} &\sqsubseteq \exists \overline{\texttt{has\_value}}.\texttt{int}^{0..9} \quad \text{(without IDs).}\end{aligned}$$

– SIO contains 3 instances of the b pattern:

$$\begin{aligned}\texttt{SIO\_000765} &\sqsubseteq \exists \overline{\texttt{SIO\_000300}}.\texttt{double}^{0...1} \\ \texttt{SIO\_001081} &\sqsubseteq \exists \overline{\texttt{SIO\_000008}}.(\overline{\texttt{SIO\_001074}} \sqcap \exists \overline{\texttt{SIO\_000300}}.\texttt{double}^{>0}) \\ \texttt{SIO\_001082} &\sqsubseteq \exists \overline{\texttt{SIO\_000008}}.(\overline{\texttt{SIO\_001074}} \sqcap \exists \overline{\texttt{SIO\_000300}}.\texttt{double}^{<0})\end{aligned}$$

The term IDs have the following meaning.

Term ID	Type	Meaning
SIO_000008	role name	has attribute
SIO_000300	role name	has value
SIO_000765	concept name	p-value
SIO_001074	concept name	t-statistic
SIO_001081	concept name	t-statistic based increased differential gene expression
SIO_001082	concept name	t-statistic based decreased differential gene expression

– NEMO contains 91 complex instances of the *b* pattern:

- The following axiom and 31 more of the same structure cause differences for locality of single axioms w.r.t. random *and* genuine seed signatures.

$$
\begin{aligned}
\overline{\texttt{NEMO\_6184000}} \equiv {} & \overline{\texttt{NEMO\_0877000}} \\
& \sqcap \exists \overline{\texttt{has\_proper\_part}}.(\overline{\texttt{NEMO\_6334000}} \sqcap \exists \overline{\texttt{OBI\_0000298}}.\overline{\texttt{NEMO\_9658000}} \\
& \qquad \sqcap \exists \overline{\texttt{NEMO\_7943000}}.\texttt{decimal}^{>0})
\end{aligned}
$$

- The following axiom and 42 more of the same structure cause differences for locality of single axioms w.r.t. genuine seed signatures *only*.

$$
\begin{aligned}
\overline{\texttt{NEMO\_3553000}} \equiv {} & \overline{\texttt{NEMO\_0000093}} \\
& \sqcap \exists \overline{\texttt{has\_proper\_part}}.(\overline{\texttt{NEMO\_6902000}} \\
& \quad \sqcap \exists \overline{\texttt{IAO\_0000221}}.(\overline{\texttt{PATO\_0000049}} \\
& \qquad \sqcap \exists \overline{\texttt{inheres\_in}}.(\overline{\texttt{NEMO\_8225000}} \sqcap \exists \overline{\texttt{unfolds\_in}}.\overline{\texttt{NEMO\_0000005}})) \\
& \quad \sqcap \exists \overline{\texttt{NEMO\_7943000}}.\texttt{decimal}^{\leqslant -.4}) \\
& \sqcap \exists \overline{\texttt{proper\_part\_of}}.(\overline{\texttt{NEMO\_0000495}} \\
& \quad \sqcap \exists \overline{\texttt{IAO\_0000136}}.(\overline{\texttt{NEMO\_8225000}} \\
& \qquad \sqcap \exists \overline{\texttt{NEMO\_0367000}}.(\overline{\texttt{NEMO\_4762000}} \\
& \qquad\quad \sqcap \exists \overline{\texttt{NEMO\_1568000}}.(\overline{\texttt{Object}} \sqcap \exists \overline{\texttt{OBI\_0000298}}.\overline{\texttt{NEMO\_0000452}} \\
& \qquad\qquad \sqcap \exists \overline{\texttt{OBI\_0000316}}.\overline{\texttt{NEMO\_0000468}})))) \\
& \sqcap \exists \overline{\texttt{has\_proper\_part}}.(\overline{\texttt{NEMO\_0745000}} \sqcap \exists \overline{\texttt{NEMO\_7943000}}.\texttt{decimal}^{150\ldots 275})
\end{aligned}
$$

Such an axiom causes differences only for top-locality and only if *all* overlined terms are not in the seed signature – then the axiom is Δ-local but not $\top$-local. Since randomly selected seed signatures are likely to be significantly larger than axiom signatures, the chance that they include one of the terms in the axiom dramatically increases, which might explain why these axioms do not cause differences for random seed signatures.

- The following axiom and 15 similar ones behave analogously.

$$
\begin{array}{l}
\overline{\mathtt{NEMO\_6279000}} \equiv \\
\quad \overline{\mathtt{NEMO\_0877000}} \\
\qquad \sqcap \exists \overline{\mathtt{OBI\_0000312}}.(\overline{\mathtt{NEMO\_0660000}} \\
\qquad\quad \sqcap \exists \overline{\mathtt{OBI\_0000293}}.(\overline{\mathtt{NEMO\_0000495}} \\
\qquad\qquad \sqcap \exists \overline{\mathtt{IAO\_0000136}}.(\overline{\mathtt{NEMO\_8225000}} \sqcap \exists \overline{\mathtt{NEMO\_0367000}}.(\overline{\mathtt{NEMO\_4762000}} \\
\qquad\qquad\quad \sqcap \exists \overline{\mathtt{NEMO\_1568000}}.(\overline{\mathtt{Object}} \sqcap \exists \overline{\mathtt{OBI\_0000298}}.\overline{\mathtt{NEMO\_0000444}} \sqcap \exists \overline{\mathtt{OBI\_0000316}}.\overline{\mathtt{NEMO\_0000468}}) \\
\qquad\qquad\quad \sqcap \exists \overline{\mathtt{proper\_part\_of}}.(\overline{\mathtt{NEMO\_6332000}} \sqcap \exists \overline{\mathtt{NEMO\_6442000}}.(\overline{\mathtt{NEMO\_6892000}} \\
\qquad\qquad\qquad \sqcap \exists \overline{\mathtt{OBI\_0000294}}.(\overline{\mathtt{NEMO\_0000382}} \sqcap \exists \overline{\mathtt{OBI\_0000316}}.\overline{\mathtt{NEMO\_7752000}})))))) \\
\qquad\quad \sqcap \exists \overline{\mathtt{OBI\_0000293}}.(\overline{\mathtt{NEMO\_0000495}} \\
\qquad\qquad \sqcap \exists \overline{\mathtt{IAO\_0000136}}.(\overline{\mathtt{NEMO\_8225000}} \sqcap \exists \overline{\mathtt{NEMO\_0367000}}.(\overline{\mathtt{NEMO\_4762000}} \\
\qquad\qquad\quad \sqcap \exists \overline{\mathtt{NEMO\_1568000}}.(\overline{\mathtt{Object}} \sqcap \exists \overline{\mathtt{OBI\_0000298}}.\overline{\mathtt{NEMO\_4353000}} \sqcap \exists \overline{\mathtt{OBI\_0000316}}.\overline{\mathtt{NEMO\_0000468}}) \\
\qquad\qquad\quad \sqcap \exists \overline{\mathtt{proper\_part\_of}}.(\overline{\mathtt{NEMO\_6332000}} \sqcap \exists \overline{\mathtt{NEMO\_6442000}}.(\overline{\mathtt{NEMO\_6892000}} \\
\qquad\qquad\qquad \sqcap \exists \overline{\mathtt{OBI\_0000294}}.(\overline{\mathtt{NEMO\_0000382}} \sqcap \exists \overline{\mathtt{OBI\_0000316}}.\overline{\mathtt{NEMO\_2813000}}))))))) \\
\qquad \sqcap \exists \overline{\mathtt{has\_proper\_part}}.(\overline{\mathtt{NEMO\_0745000}} \sqcap \exists \overline{\mathtt{NEMO\_7943000}}.(\mathtt{decimal}^{200\ldots350})) \\
\qquad \sqcap \exists \overline{\mathtt{has\_proper\_part}}.(\overline{\mathtt{NEMO\_6902000}} \\
\qquad\quad \sqcap \exists \overline{\mathtt{IAO\_0000221}}.(\overline{\mathtt{PATO\_0000049}} \\
\qquad\qquad \sqcap \exists \overline{\mathtt{inheres\_in}}.(\overline{\mathtt{NEMO\_8225000}} \sqcap \exists \overline{\mathtt{unfolds\_in}}.(\overline{\mathtt{NEMO\_0000007}} \sqcup \mathtt{NEMO\_0000010}))) \\
\qquad\quad \sqcap \exists \overline{\mathtt{OBI\_0000298}}.\overline{\mathtt{NEMO\_9658000}} \sqcap \exists \overline{\mathtt{proper\_part\_of}}.\overline{\mathtt{NEMO\_0000495}} \sqcap \exists \overline{\mathtt{NEMO\_7943000}}.\mathtt{decimal}^{<0})
\end{array}
$$

All IDs in the above NEMO axioms have the following meaning.

Term ID	Type	Meaning
IAO_0000136	role name	is_about
IAO_0000221	role name	is_quality_measurement_of
NEMO_0000005	concept name	frontal_scalp_surface_region
NEMO_0000007	concept name	occipital_scalp_surface_region
NEMO_0000010	concept name	posterotemporal_scalp_surface_region
NEMO_0000093	concept name	scalp_recorded_ERP_component
NEMO_0000382	concept name	experiment_condition
NEMO_0000444	concept name	stimulus_valence
NEMO_0000452	concept name	auditory
NEMO_0000468	concept name	stimulus_role
NEMO_0000495	concept name	averaged_EEG_data_set
NEMO_0367000	role name	prime_stimulus
NEMO_0660000	concept name	vocalization
NEMO_0745000	concept name	right_mastoid_reference
NEMO_0877000	concept name	scalp_recorded_ERP_diffwave_component
NEMO_1568000	role name	has_object
NEMO_2813000	concept name	condition_of_interest
NEMO_3553000	concept name	auditory_frontocentral_P200
NEMO_4353000	concept name	affectively_arousing

(continued on next page)

Term ID	Type	Meaning
NEMO_4762000	concept name	onset_stimulus_presentation
NEMO_6184000	concept name	unnamed_positive_RATEMP_effect
NEMO_6279000	concept name	EPN_effect
NEMO_6332000	concept name	experimental_condition_execution
NEMO_6334000	concept name	mean_intensity_RATEMP
NEMO_6442000	role name	is_realization_of
NEMO_6892000	concept name	plan
NEMO_6902000	concept name	intensity_measurement_datum
NEMO_7752000	concept name	condition_for_comparison
NEMO_7943000	role name	has_numeric_value
NEMO_8225000	concept name	scalp_recorded_ERP
NEMO_9658000	concept name	statistical_significance

Term ID	Type	Meaning
OBI_0000293	role name	has_specified_input
OBI_0000294	role name	is_concretization_of
OBI_0000298	role name	has_quality
OBI_0000312	role name	is_specified_output_of
OBI_0000316	role name	has_role
PATO_0000049	concept name	intensity

Culprit type c.

- GFOBio contains the axiom

$$\begin{aligned}\overline{\texttt{Individual}} \equiv (&\overline{\texttt{Abstract}} \sqcup \texttt{Concrete} \sqcup \texttt{Space\_time})\\ &\sqcap (\texttt{Dependent} \sqcup \overline{\texttt{Independent}})\\ &\sqcap (\texttt{Role} \sqcup (\overline{\texttt{Individual}} \sqcap \neg\texttt{Role})).\end{aligned}$$

  Then (1) the first two conjuncts are equivalent to $\top$ if at least one of their respective disjuncts is replaced by $\top$; (2) the last conjunct is equivalent to $\bot$ if Individual is replaced by $\bot$, but this is not "recognized" by the definition of syntactic locality. Hence, the axiom is Δ-local, but not $\top$-local.
- Koala contains the axiom

$$\overline{\texttt{MaleStudentWith3Daughters}} \equiv \texttt{Student} \sqcap \forall\,\texttt{hasChildren}.\overline{\texttt{Female}} \sqcap \forall\,\texttt{hasGender}.\{\texttt{male}\} \sqcap {=}3\,\texttt{hasChildren}.\top,$$

  where the concept description $\forall\,\texttt{hasChildren}.\overline{\texttt{Female}} \sqcap {=}3\,\texttt{hasChildren}.\top$ is equivalent to $\bot$ if Female is replaced by $\bot$, but this is not "recognized" by the definition of syntactic locality. Hence, the axiom is $\emptyset$-local, but not $\bot$-local.
- miniT contains the following axioms, which show the same pattern as the above axiom from Koala.

$$\begin{aligned}\overline{\texttt{small\_nuclear\_rna}} \equiv\ & \texttt{macromolecular\_compound} \sqcap \exists\,\texttt{polymer\_of}.\texttt{ribo\_nucleotide}\\ &\sqcap \exists\,\texttt{transcribed\_from}.\texttt{gene} \sqcap \forall\,\texttt{part\_of}.\overline{\texttt{spliceosome}}\\ &\sqcap \forall\,\texttt{polymer\_of}.\texttt{ribo\_nucleotide} \sqcap \forall\,\texttt{transcribed\_from}.\texttt{gene} \sqcap {\geqslant}1\,\texttt{part\_of}.\top\\ \overline{\texttt{EnzymeReaction}} \equiv\ & \texttt{Reaction} \sqcap \forall\,\texttt{relatedTo}.\overline{\texttt{Enzyme}} \sqcap {\geqslant}1\,\texttt{relatedTo}.\top\\ \overline{\texttt{Peptide}} \equiv\ & \forall\,\texttt{part\_of}.\overline{\texttt{Protein}} \sqcap {\geqslant}1\,\texttt{part\_of}.\top\end{aligned}$$

– Tambis contains 9 similar axioms.
– LiPrO contains the following axiom.

$$\begin{aligned}\overline{\mathtt{LC\_Alpha\_prime\_mycolic\_acids}} \equiv\ & \mathtt{LC\_Alpha\_1\_mycolic\_acid} \\ & \sqcap \exists\, \mathtt{hasPart.proximal\_Alkenyl} \\ & \sqcap \forall\, \mathtt{hasPart.}(\overline{\mathtt{Alcohol}} \sqcup \overline{\mathtt{Alkenyl\_Group}} \sqcup \overline{\mathtt{Alpha\_Hydroxy\_Acid\_Group}} \\ & \qquad \sqcup \overline{\mathtt{Carboxylic\_Acid}} \sqcup \overline{\mathtt{Meromycolic\_Chain}})\end{aligned}$$

If all overlined terms are replaced by $\bot$, then both sides of the axiom are equivalent to $\bot$; hence, the axiom is $\emptyset$-local. This is not "recognized" by the definition of syntactic $\top$-locality.
– TEDDY contains the culprit-c axiom

$$\begin{aligned}\mathsf{TEDDY\_0000069} \sqsubseteq\ & (\neg\exists\, \mathsf{TR\_0017}.\overline{\mathsf{TEDDY\_0000086}} \sqcap \exists\, \overline{\mathsf{TR\_0018}}.\overline{\mathsf{TEDDY\_0000095}}) \\ & \sqcup (\exists\, \mathsf{TR\_0017}.\overline{\mathsf{TEDDY\_0000129}} \sqcap \exists\, \overline{\mathsf{TR\_0018}}.\overline{\mathsf{TEDDY\_0000095}}),\end{aligned}$$

whose right-hand side becomes equivalent to $\top$ if all overlined terms are replaced by $\top$ – then the axiom is Δ-local but not $\top$-local. The term IDs have the following meaning.

Term ID	Type	Meaning
TR_0017	role name	creates
TR_0018	role name	destroys
TEDDY_0000069	concept name	local bifurcation
TEDDY_0000086	concept name	fixed point
TEDDY_0000095	concept name	Liapunov stable fixed point
TEDDY_0000129	concept name	unstable fixed point

Culprit type d.

– The following axioms cause differences only in locality in NEMO, OGI, PO, GRO, POL.

$$\begin{aligned}\mathtt{NEMO\_8529000} &\sqsubseteq \exists\overline{\mathtt{proper\_part\_of}}.\mathtt{NEMO\_8529000} \\ \mathtt{Biological\_Interval} &\sqsubseteq \exists\overline{\mathtt{hasIntervalRelation}}.\mathtt{Biological\_Interval} \\ \mathtt{DisulphideBond} &\sqsubseteq \exists\overline{\mathtt{DisulphideBondBindRef1}}.\mathtt{DisulphideBond} \\ \mathtt{DisulphideBond} &\sqsubseteq \exists\overline{\mathtt{DisulphideBondBindRef2}}.\mathtt{DisulphideBond} \\ \mathtt{PocketDomain} &\sqsubseteq \exists\overline{\mathtt{partOf}}.\mathtt{PocketDomain} \\ \mathtt{Experiments} &\sqsubseteq \exists\overline{\mathtt{fromExperiment}}.\mathtt{Experiments},\end{aligned}$$

where the term IDs NEMO_8529000 means "frontotemporal scalp surface".
– The following four axioms occur in TOK.

$$\begin{aligned}\mathtt{TOK\_Entity} &\sqsubseteq \exists\overline{\mathtt{aligned\_to}}.\mathtt{TOK\_Entity} \\ \mathtt{TOK\_Resource} &\sqsubseteq \exists\overline{\mathtt{aligned\_to}}.\mathtt{TOK\_Resource} \\ \mathtt{TOK\_Resource} &\sqsubseteq \exists\overline{\mathtt{related\_to}}.\mathtt{TOK\_Resource} \\ \mathtt{TOK\_Resource} &\sqsubseteq \exists\overline{\mathtt{collection\_of}}.\mathtt{TOK\_Resource}\end{aligned}$$

– The 12 axioms of culprit type d in PhaRe are

$$\mathtt{A} \sqsubseteq \exists\overline{\mathtt{modified}}.\mathtt{A},$$

where A ranges over the terms Drug, DrugDose, DrugDoseVariation, Gene, GeneOrGeneProductActivity, GeneProduct, GeneProductActivityChange, GenomicRegion, GenomicVariation, Mrna, Protein, Variant.

– The ontologies MF, MFOEm and VSO contain the same 12 slightly more complex instances of the culprit-d pattern.

$$\exists\overline{\mathsf{BFO\_0000053}}.\mathsf{BFO\_0000016} \equiv \exists\overline{\mathsf{BFO\_0000112}}.\mathsf{BFO\_0000016}$$
$$\exists\overline{\mathsf{BFO\_0000053}}.\mathsf{BFO\_0000019} \equiv \exists\overline{\mathsf{BFO\_0000086}}.\mathsf{BFO\_0000019}$$
$$\exists\overline{\mathsf{BFO\_0000053}}.\mathsf{BFO\_0000023} \equiv \exists\overline{\mathsf{BFO\_0000087}}.\mathsf{BFO\_0000023}$$
$$\exists\overline{\mathsf{BFO\_0000053}}.\mathsf{BFO\_0000034} \equiv \exists\overline{\mathsf{BFO\_0000085}}.\mathsf{BFO\_0000034}$$

$$\exists\overline{\mathsf{BFO\_0000158}}.\mathsf{BFO\_0000016} \equiv \exists\overline{\mathsf{BFO\_0000162}}.\mathsf{BFO\_0000016}$$
$$\exists\overline{\mathsf{BFO\_0000158}}.\mathsf{BFO\_0000019} \equiv \exists\overline{\mathsf{BFO\_0000159}}.\mathsf{BFO\_0000019}$$
$$\exists\overline{\mathsf{BFO\_0000158}}.\mathsf{BFO\_0000023} \equiv \exists\overline{\mathsf{BFO\_0000161}}.\mathsf{BFO\_0000023}$$
$$\exists\overline{\mathsf{BFO\_0000158}}.\mathsf{BFO\_0000034} \equiv \exists\overline{\mathsf{BFO\_0000160}}.\mathsf{BFO\_0000034}$$

$$\overline{\mathsf{BFO\_0000016}} \sqcap \exists\overline{\mathsf{BFO\_0000052}}.\mathsf{BFO\_0000004} \equiv \exists\overline{\mathsf{BFO\_0000107}}.\mathsf{BFO\_0000004}$$
$$\overline{\mathsf{BFO\_0000019}} \sqcap \exists\overline{\mathsf{BFO\_0000052}}.\mathsf{BFO\_0000004} \equiv \exists\overline{\mathsf{BFO\_0000080}}.\mathsf{BFO\_0000004}$$
$$\overline{\mathsf{BFO\_0000023}} \sqcap \exists\overline{\mathsf{BFO\_0000052}}.\mathsf{BFO\_0000004} \equiv \exists\overline{\mathsf{BFO\_0000081}}.\mathsf{BFO\_0000004}$$
$$\overline{\mathsf{BFO\_0000034}} \sqcap \exists\overline{\mathsf{BFO\_0000052}}.\mathsf{BFO\_0000004} \equiv \exists\overline{\mathsf{BFO\_0000079}}.\mathsf{BFO\_0000004}$$

The term IDs have the following meaning.

Term ID	Type	Meaning
BFO_0000004	concept name	independent continuant
BFO_0000016	concept name	disposition
BFO_0000019	concept name	quality
BFO_0000023	concept name	role
BFO_0000034	concept name	function
BFO_0000052	role name	inheres in at all times
BFO_0000053	role name	bearer of at some time
BFO_0000079	role name	function of at all times
BFO_0000080	role name	quality of at all times
BFO_0000081	role name	role of at all times
BFO_0000085	role name	has function at some time
BFO_0000086	role name	has quality at some time
BFO_0000087	role name	has role at some time
BFO_0000107	role name	disposition of at all times
BFO_0000112	role name	has disposition at some time
BFO_0000158	role name	bearer of at all times
BFO_0000159	role name	has quality at all times
BFO_0000160	role name	has function at all times
BFO_0000161	role name	has role at all times
BFO_0000162	role name	has disposition at all times

– KiSAO has two culprit-d axioms:

$$\mathtt{KISAO\_0000000} \sqcap \exists \mathtt{KISAO\_0000245}.\mathtt{KISAO\_0000311} \sqsubseteq \exists \mathtt{KISAO\_0000245}.\overline{\mathtt{KISAO\_0000102}}$$
$$\mathtt{KISAO\_0000064} \sqcap \neg\mathtt{KISAO\_0000367} \sqcap \exists \mathtt{KISAO\_0000245}.\mathtt{KISAO\_0000366} \sqsubseteq \exists \mathtt{KISAO\_0000245}.\overline{\mathtt{KISAO\_0000240}}$$

The term IDs have the following meaning.

Term ID	Type	Meaning
KISAO_0000000	concept name	modeling and simulation algorithm
KISAO_0000064	concept name	Runge-Kutta based method
KISAO_0000102	concept name	spatial description
KISAO_0000240	concept name	implicit method type
KISAO_0000245	role name	has characteristic
KISAO_0000311	concept name	type of domain geometry handling
KISAO_0000366	concept name	symplecticness
KISAO_0000367	concept name	partitioned Runge-Kutta method

Note that it suffices for these axioms to be Δ-local if KISAO_0000102 or KISAO_0000240, respectively, are not in the seed signature.

– NBO contains three equivalences of type culprit-d.

$$\mathtt{NBO\_0000130} \equiv \mathtt{NBO\_0000130} \sqcap \exists \overline{\mathtt{has\_participant}}.\overline{\mathtt{PATO\_0000145}}$$
$$\mathtt{NBO\_0001536} \equiv \mathtt{NBO\_0001536} \sqcap \exists \overline{\mathtt{in\_response\_to}}.\overline{\mathtt{GO\_0009314}}$$
$$\mathtt{NBO\_0000338} \equiv \mathtt{NBO\_0000338} \sqcap \exists \overline{\mathtt{by\_means}}.\overline{\mathtt{GO\_0050881}}$$

Each of these is implied by one of the following axioms in NBO:

$$\mathtt{NBO\_0000130} \sqsubseteq \exists \mathtt{has\_participant}.\mathtt{PATO\_0000145}$$
$$\mathtt{NBO\_0001536} \sqsubseteq \exists \mathtt{in\_response\_to}.\mathtt{GO\_0009314}$$
$$\mathtt{NBO\_0000338} \sqsubseteq \exists \mathtt{by\_means}.\mathtt{GO\_0050881}$$

The term IDs have the following meaning.

Term ID	Type	Meaning
GO_0009314	concept name	response to radiation
GO_0050881	concept name	musculoskeletal movement
NBO_0000130	concept name	liquid consumption
NBO_0001536	concept name	behavioral response to radiation
NBO_0000338	concept name	kinesthetic behavior
PATO_0000145	concept name	liquid substance

– TEDDY contains 6 culprit-d axioms of different complexity.

$$\mathtt{TEDDY\_0000092} \sqsubseteq \exists \overline{\mathtt{TR\_0012}}.\mathtt{TEDDY\_0000092}$$
$$\mathtt{TEDDY\_0000106} \sqsubseteq \exists \overline{\mathtt{TR\_0012}}.\mathtt{TEDDY\_0000106}$$

$$\mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\mathtt{TEDDY\_0000032} \sqsubseteq \mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\overline{\mathtt{TEDDY\_0000130}}$$
$$\mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\mathtt{TEDDY\_0000067} \sqsubseteq \mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\overline{\mathtt{TEDDY\_0000066}}$$
$$\mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\mathtt{TEDDY\_0000066} \sqsubseteq \mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\overline{\mathtt{TEDDY\_0000067}}$$

$$\mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\mathtt{TEDDY\_0000006} \sqcap \exists \mathtt{TR\_0004}.\mathtt{TEDDY\_0000067} \sqsubseteq \mathtt{TEDDY\_0000083} \sqcap \exists \mathtt{TR\_0004}.\overline{\mathtt{TEDDY\_0000066}}$$

The term IDs have the following meaning.

Term ID	Type	Meaning
TR_0004	role name	hasCharacteristic
TR_0012	role name	adjacentTo
TEDDY_0000006	concept name	oscillation
TEDDY_0000032	concept name	sigmoid shape
TEDDY_0000066	concept name	periodic oscillation
TEDDY_0000067	concept name	period
TEDDY_0000083	concept name	temporal behaviour
TEDDY_0000092	concept name	non-isolated fixed point
TEDDY_0000106	concept name	non-isolated cycle
TEDDY_0000130	concept name	inflexion point

– SIO contains a single culprit-d axiom with a *disjunction*:

$$\mathtt{SIO\_010506} \sqsubseteq \exists \overline{\mathtt{SIO\_000273}}.(\mathtt{SIO\_010506} \sqcup \mathtt{SIO\_011125})$$
$$\mathtt{molecular\_complex} \sqsubseteq \exists \overline{\mathtt{has\_direct\_part}}.(\mathtt{molecular\_complex} \sqcup \mathtt{molecule}) \quad \text{(without IDs)}$$

– Each of the two single culprit-d axioms in AERO and OBIws contain a *universal* quantifier:

$$\mathtt{AERO\_0000044} \sqsubseteq \exists \overline{\mathtt{has\_part}}.(\mathtt{AERO\_0000044} \sqcap \forall \mathtt{precedes}.\overline{\mathtt{AERO\_0000027}})$$
$$\mathtt{OBIws\_0000045} \sqsubseteq \forall \mathtt{OBI\_0000299}.(\overline{\mathtt{OBIws\_0000029}} \sqcap \exists \overline{\mathtt{OBIws\_0000130}}.(\overline{\mathtt{IAO\_0000098}} \sqcap \exists \overline{\mathtt{OBI\_0000295}}.\mathtt{OBIws\_0000045})),$$

where the term IDs have the following meaning.

Term ID	Type	Meaning
AERO_0000027	concept name	clonic motor manifestation
AERO_0000044	concept name	tonic motor manifestation
IAO_0000098	concept name	data format specification
OBI_0000295	role name	is_specified_input_of
OBI_0000299	role name	has_specified_output
OBIws_0000029	concept name	multiple sequence alignment data set
OBIws_0000045	concept name	retrieve alignment result execution
OBIws_0000130	role name	is_encoded_in

– More complex shapes of this pattern have been observed in Galen:

$$A \sqsubseteq \exists \overline{r}.(\overline{B} \sqcap \exists \overline{s}.A)$$

$$A \sqcap \exists r.B \sqsubseteq \exists \overline{s}.B$$

$$A \sqcap \exists r.(B \sqcap \exists s.C) \sqsubseteq \exists \overline{t}.C$$

$$A \sqcap X \sqcap \exists r.B \sqsubseteq \exists r.\overline{C}$$

$$A \sqcap \exists r.(B \sqcap \exists s.(C \sqcap \exists t.D \sqcap X)) \sqcap Y \sqsubseteq \exists \overline{u}.(\overline{E} \sqcap \exists t.\overline{F})$$

$$A \sqcap X \sqcap \exists r.(\overline{B} \sqcap \exists \overline{s}.(\overline{C} \sqcap \exists t.D \sqcap \exists u.(\overline{E} \sqcap \exists v.F))) \sqsubseteq \exists \overline{s}.(\overline{C} \sqcap \exists t.\overline{G} \sqcap \exists u.(\overline{E} \sqcap \exists v.\overline{B}))$$

Six of the 121 axioms of culprit type d in Galen are

$$\begin{aligned}\mathtt{Abdomen} \sqsubseteq \exists \overline{\mathtt{hasSurfaceDivision}}.(&\overline{\mathtt{AnatomicalSurfaceQuadrant}}\\ &\sqcap \exists.\, \overline{\mathtt{isSurfaceDivisionOf}}.\mathtt{Abdomen})\end{aligned}$$

$$\begin{aligned}&\mathtt{Behaviour} \sqcap \exists \mathtt{hasIntrinsicAbnormalityStatus}.\mathtt{nonNormal}\\ &\quad\sqsubseteq \exists \overline{\mathtt{hasAbnormalityStatus}}.\mathtt{nonNormal}\end{aligned}$$

$$\begin{aligned}\mathtt{AcquiredLesion} \sqcap \exists \mathtt{hasLocation}.(&\mathtt{BodyStructure}\\ &\sqcap \exists \mathtt{hasSpecificationLevel}.\mathtt{atLeastWellSpecified})\\ &\sqsubseteq \exists \overline{\mathtt{hasSpecificationLevel}}.\mathtt{atLeastWellSpecified}\end{aligned}$$

$$\begin{aligned}\mathtt{NAMEDInternalBodyPart} &\sqcap \exists \mathtt{hasLateralPosition}.(\dots)\\ &\sqcap \exists \mathtt{isPairedOrUnpaired}.\mathtt{exactlyPaired}\\ &\sqcap \exists \mathtt{hasSpecificationLevel}.\mathtt{atLeastWellSpecified}\\ &\sqsubseteq \exists \mathtt{hasSpecificationLevel}.\overline{\mathtt{uniquelySpecified}}\end{aligned}$$

$$\mathtt{D} \sqcap \exists \mathtt{e}.(\mathtt{A} \sqcap \exists \mathtt{s}.(\mathtt{T} \sqcap \exists \mathtt{t}.\mathtt{R} \sqcap \exists \mathtt{w}.(\mathtt{P} \sqcap \exists \mathtt{p}.\mathtt{L}))) \sqcap \exists \mathtt{g}^{-}.\mathtt{L} \sqsubseteq \exists \overline{\mathtt{f}}.(\overline{\mathtt{E}} \sqcap \exists \mathtt{t}.\overline{\mathtt{I}})$$

$$\mathtt{B} \sqcap \exists \mathtt{g}.(\mathtt{S} \sqcap \exists \mathtt{a}.\mathtt{L}) \sqcap \exists \mathtt{e}^{-}(\overline{\mathtt{A}} \sqcap \exists \overline{\mathtt{s}}.(\overline{\mathtt{T}} \sqcap \exists \mathtt{t}.\mathtt{R} \sqcap \exists \mathtt{w}.(\overline{\mathtt{P}} \sqcap \exists \mathtt{p}.\mathtt{L}))) \sqsubseteq \exists \overline{\mathtt{s}}.(\overline{\mathtt{T}} \sqcap \exists \mathtt{t}.\overline{\mathtt{V}} \sqcap \exists \mathtt{w}.(\overline{\mathtt{P}} \sqcap \exists \mathtt{p}.\overline{\mathtt{A}}))$$

where, in the last two axioms, the abbreviations have the following meaning.

Concept names

Abb	Meaning
A	Antimicrobial
B	Bacterium
D	Degradation
E	Effectiveness
I	ineffective
L	BetaLactamase
P	presence
R	resistant
S	Secretion
T	Sensitivity
V	sensitive

Role names

Abb	Meaning
a	actsSpecificallyOn
e	hasExposureTo
$\mathtt{e}^{-}$	isExposedTo
f	hasEffectiveness
g	hasFunction
$\mathtt{g}^{-}$	isFunctionOf
p	isPresenceAbsenceOf
s	hasSensitivity
t	hasState
w	isWithReferenceTo

C Ontology corpus

The following table lists all ontologies in our corpus, with the BioPortal ID given in the column "BP ID" where applicable. The entries in the "DL expressivity" column have been computed using a standard method implemented in the OWL API, which does not perform a check for inclusion in any of the OWL profiles. Ontologies containing one of the features $\mathcal{Q}, \mathcal{N}, \mathcal{F}, \mathcal{R}$ – and thus requiring extensions to reasoners for checking semantic Δ-locality – have been marked "!".

Table 7: Corpus of ontologies included in experiments

Ontology	BP ID		DL express.	#axs	#terms
aba-adult-mouse-brain	1290		$\mathcal{ALCI}$	3441	916
adverse-event-reporting-ontology	1580	!	$\mathcal{SHOIQ(D)}$	873	492
african-traditional-medicine	1099		$\mathcal{ALE}$	208	211
amino-acid	1054	!	$\mathcal{ALCF(D)}$	477	54
amphibian-gross-anatomy	1090		$\mathcal{ALE}$	2673	1612
amphibian-taxonomy	1370		$\mathcal{ALE}$	12163	6137
anatomical-entity-ontology	1568		$\mathcal{ALE}$	368	253
animal-natural-history-and-life-history	1530	!	$\mathcal{ALCOF(D)}$	638	475
ascomycete-phenotype-ontology	1222		$\mathcal{AL}$	294	298
basic-formal-ontology	1332		$\mathcal{ALC}$	95	40
basic-vertebrate-anatomy	1056	!	$\mathcal{SHIF}$	388	179
bilateria-anatomy	1114		$\mathcal{ALEH}+$	138	120
bio-health-ontological-knowledge-base-cystic-fibrosis	3155	!	$\mathcal{ALCHIF}$	660	440
bioassay-ontology	1533	!	$\mathcal{SROIQ(D)}$	1797	1484
bioinformatics-operations-types-of-data-data-formats-and-topics	1498		$\mathcal{ALEH}$	3814	2239
bioinformatics-web-service-ontology	3119	!	$\mathcal{SROIQ(D)}$	430	238
biological-imaging-methods	1023		$\mathcal{S}$	548	519
biomedical-resource-ontology	1104	!	$\mathcal{SHIF(D)}$	634	598
biopax	1522	!	$\mathcal{SHIN(D)}$	391	171
bioportal-metadata	1148	!	$\mathcal{ALUHIN(D)}$	822	348
biotop	1134	!	$\mathcal{SRI}$	922	472
birnlex	1089		$\mathcal{AL}$	3572	3581
bleeding-history-phenotype	1116	!	$\mathcal{ALCIF(D)}$	1925	585
body-system	1487		$\mathcal{AL}$	28	30
book	3059	!	$\mathcal{ALCHOIN(D)}$	529	248
breast-cancer-grading-ontology	1304	!	$\mathcal{SHOIN(D)}$	690	383
breast-tissue-cell-lines	1438		$\mathcal{ALCH(D)}$	2734	414
brenda-tissue-enzyme-source	1005		$\mathcal{ALE}$	6451	5423
c-elegans-development	1049		$\mathcal{AL}$	71	73
c-elegans-gross-anatomy	1048		$\mathcal{ALE}$	12341	6778
c-elegans-phenotype	1067		$\mathcal{AL}+$	2366	2073
cao	1582	!	$\mathcal{SHIQ(D)}$	442	244
carelex	3008		$\mathcal{ALH(D)}$	327	333

Continued on next page

Table 7 – Continued from previous page

Ontology	BP ID	DL express.	#axs	#terms
cell-behavior-ontology	1158	$\mathcal{ALUO}$	13	14
cell-culture-ontology	3108 !	$\mathcal{SHOIF}$	9467	6097
cell-line-ontology	1541	$\mathcal{ALCH}(\mathcal{D})$	3996	841
cell-type	1006	$\mathcal{SH}$	2975	1938
cereal-plant-development	1047	$\mathcal{ALE}$	235	236
cerebrotendinous-xanthomatosis	3025 !	$\mathcal{ALCOIN}(\mathcal{D})$	1969	612
chemical-information-ontology	1444 !	$\mathcal{SROIN}(\mathcal{D})$	1237	685
clinical-measurement-ontology	1583	$\mathcal{ALE}+$	1382	1204
cognitive-atlas	1633	$\mathcal{ALC}$	4100	1707
common-anatomy-reference-ontology	1063	$\mathcal{ALE}+$	54	53
common-terminology-criteria-for-adverse-events	1415	$\mathcal{AL}(\mathcal{D})$	6940	3891
comparative-data-analysis-ontology	1128 !	$\mathcal{SROIQ}(\mathcal{D})$	462	253
computational-neuroscience-ontology	3003 !	$\mathcal{SHOIF}$	1121	264
dendritic-cell	1144	$\mathcal{ALC}$	313	158
dengue-fever-ontology	3174 !	$\mathcal{SROIF}$	5534	4969
dictyostelium-discoideum-anatomy	1008	$\mathcal{ALE}+$	379	141
dikb-evidence-ontology	1672 !	$\mathcal{ALCHOIN}(\mathcal{D})$	660	248
drosophila-development	1016	$\mathcal{ALEH}+$	410	132
eagle-i-research-resource-ontology	3016 !	$\mathcal{SHOIF}(\mathcal{D})$	4378	3643
electrocardiography-ontology	1146 !	$\mathcal{ALCIF}(\mathcal{D})$	1274	1144
emotion-ontology	1666 !	$\mathcal{SROIQ}$	728	429
environment-ontology	1069	$\mathcal{S}$	1752	1548
enzyme-mechanism-ontology	1626 !	$\mathcal{ALCRQ}(\mathcal{D})$	931	345
epilepsy	1639	$\mathcal{ALH}(\mathcal{D})$	145	149
event-inoh-pathway-ontology-	1011	$\mathcal{ALEH}+$	7131	3686
evidence-codes	1012	$\mathcal{ALE}$	363	285
experimental-conditions-ontology	1585	$\mathcal{ALE}+$	269	246
experimental-factor-ontology	1136 !	$\mathcal{ALHIF}+$	7156	5889
exposure-ontology	1575 !	$\mathcal{ALER}+$	101	91
family-health-history-ontology	1126 !	$\mathcal{ALCHIF}(\mathcal{D})$	1103	684
fda-medical-devices-2010-	1576	$\mathcal{AL}$	4907	4927
fly-taxonomy	1064	$\mathcal{AL}$	6587	6585
flybase-controlled-vocabulary	1017	$\mathcal{ALE}+$	793	686
fungal-gross-anatomy	1019	$\mathcal{ALEI}+$	106	84
gene-regulation-ontology	1082 !	$\mathcal{ALCHIQ}(\mathcal{D})$	962	544
gene-regulation-ontology	1106 !	$\mathcal{ALCHIQ}(\mathcal{D})$	962	544
general-formal-ontology-biology	1440 !	$\mathcal{SHIN}$	466	241
general-formal-ontology	1439 !	$\mathcal{SHIQ}$	212	87
genomic-clinical-decision-support-genomic-cds	3179 !	$\mathcal{ALCQ}$	4322	2268
health-level-seven	1343	$\mathcal{AL}$	8072	7502
health_indicators	1581	$\mathcal{AL}$	548	539
hom-datasource_oshpd	1648	$\mathcal{AL}$	351	353
hom-datasource_oshpdsc	1667	$\mathcal{AL}$	351	353

Continued on next page

Table 7 – Continued from previous page

Ontology	BP ID	DL express.	#axs	#terms
hom-dxprocs_mdcdrg	1642	$\mathcal{AL}$	774	776
hom-dxvcodes2_oshpd	1654	$\mathcal{AL}$	16064	16067
hom-harvard	1631	$\mathcal{AL}$	189	180
hom-icd9_dxandvcodes_oshpd	1647	$\mathcal{AL}$	16066	16068
hom-icd9_procs_oshpd	1643	$\mathcal{AL}$	4642	4644
hom-icd9cm-ecodes	1641	$\mathcal{AL}$	1490	1492
hom-icd9pcs	1625	$\mathcal{AL}$	4643	4645
hom-mdcdrg	3046	$\mathcal{AL}$	790	792
hom-oshpd-sc	1668	$\mathcal{AL}$	266	268
hom-oshpd_usecase	1652	$\mathcal{AL}$	393	395
hom-procs2_oshpd	1653	$\mathcal{AL}$	4642	4644
hom-ucare	1629	$\mathcal{AL}$	64	64
hom_elixhauserscores	1578	$\mathcal{AL}$	29	31
hom_mdcs-drgs	1596	$\mathcal{AL}$	774	776
homerun-ontology	1627	$\mathcal{AL}$	1194	1088
host-pathogen-interactions-ontology	1569	$\mathcal{SHI}$	403	309
human-developmental-anatomy-abstract-version	1021	$\mathcal{ALE}$	2336	2317
human-developmental-anatomy-timed-version	1022	$\mathcal{ALE}$	8339	8342
human-disease-ontology	1009	$\mathcal{ALE}$	6743	6289
human-phenotype-ontology	1125	$\mathcal{AL}$	13153	9904
hymenoptera-anatomy-ontology	1362 !	$\mathcal{SR}$	3944	1970
icps-network	1509	$\mathcal{AL}$	24	26
imgt-ontology	1491 !	$\mathcal{SHIN}(\mathcal{D})$	2114	321
immune-disorder-ontology	3127	$\mathcal{AL}$	1676	1416
infectious-disease-ontology	1092 !	$\mathcal{SROIF}$	1233	570
influenza-ontology	1417 !	$\mathcal{SROIN}(\mathcal{D})$	1696	856
information-artifact-ontology	1393 !	$\mathcal{ALRIF}+(\mathcal{D})$	554	250
interaction-network-ontology	1515	$\mathcal{ALC}$	1034	979
interaction-ontology	1614	$\mathcal{AL}$	39	42
international-classification-for-nursing-practice	1401 !	$\mathcal{SHIF}$	11891	3324
kinetic-simulation-algorithm-ontology	1410 !	$\mathcal{ALCRIQ}(\mathcal{D})$	710	245
linkingkin2pep	1406 !	$\mathcal{SHIF}(\mathcal{D})$	30	19
lipid-ontology	1183 !	$\mathcal{ALCHIN}$	2375	763
loggerhead-nesting	1024	$\mathcal{ALE}$	347	312
maize-gross-anatomy	1050	$\mathcal{ALE}$	217	182
malaria-ontology	1311 !	$\mathcal{ALER}+$	3212	2463
mammalian-phenotype	1025	$\mathcal{AL}+$	10934	8885
mass-spectrometry	1105	$\mathcal{ALE}+$	2518	2017
measurement-method-ontology	1584	$\mathcal{AL}$	332	323
medaka-fish-anatomy-and-development	1027	$\mathcal{ALE}$	4402	4361
mego	1257	$\mathcal{ALE}+$	421	356
mental-functioning-ontology	3002 !	$\mathcal{SROIQ}$	514	238
microrna-ontology	3139	$\mathcal{ALEI}$	638	567

Continued on next page

Table 7 – Continued from previous page

Ontology	BP ID	DL express.	#axs	#terms
minimal-anatomical-terminology	1152	$\mathcal{ALE}$	504	461
mixs-controlled-vocabularies	3000	$\mathcal{AL}$	518	520
molecule-role-inoh-protein-name-family-name-ontology-	1029	$\mathcal{ALE}+$	9629	9222
mosquito-gross-anatomy	1030	$\mathcal{ALE}+$	2733	1868
mosquito-insecticide-resistance	1077	$\mathcal{ALE}+$	4413	4326
mouse-adult-gross-anatomy	1000	$\mathcal{ALE}+$	3776	2985
mouse-experimental-design-ontology	3180	$\mathcal{ALH}$	86	86
mouse-pathology	1031	$\mathcal{ALE}+$	808	708
multiple-alignment	1026	$\mathcal{ALE}+$	168	171
neglected-tropical-disease-ontology-ntdo-	3153 !	$\mathcal{SRIQ}$	1237	625
neomark-oral-cancer-ontology	1686 !	$\mathcal{SHIQ}$	399	352
neural-electromagnetic-ontologies	1321 !	$\mathcal{SHIQ}(\mathcal{D})$	2843	1972
neural-immune-gene-ontology	1539	$\mathcal{SH}$	8835	4842
neuro-behavior-ontology	1621	$\mathcal{ALE}$	1314	971
nif-cell	1402 !	$\mathcal{SROIF}(\mathcal{D})$	3570	2939
nif-subcellular	3126 !	$\mathcal{SROIF}(\mathcal{D})$	4061	2968
nmr-instrument-specific-component-of-metabolomics-investigations	1033	$\mathcal{SH}(\mathcal{D})$	599	494
obo-relationship-types	1042 !	$\mathcal{ALR}+$	33	21
oboe	1524 !	$\mathcal{SRIQ}(\mathcal{D})$	265	87
ontologia-proj-alternativa	3035 !	$\mathcal{ALUIN}+(\mathcal{D})$	270	117
ontology-for-disease-genetic-investigation	1086 !	$\mathcal{SHIN}(\mathcal{D})$	1867	861
ontology-for-drug-discovery-investigations	1540 !	$\mathcal{SHOIN}(\mathcal{D})$	996	746
ontology-for-general-medical-science	1414	$\mathcal{ALCO}$	216	153
ontology-for-genetic-interval	1100 !	$\mathcal{SHIN}(\mathcal{D})$	509	289
ontology-for-microrna-target-prediction	1505 !	$\mathcal{ALCHIQ}(\mathcal{D})$	2364	592
ontology-for-parasite-lifecycle	1190 !	$\mathcal{SHOIF}$	885	391
ontology-of-clinical-research-ocre-	1076 !	$\mathcal{ALCHIF}(\mathcal{D})$	51	30
ontology-of-data-mining	1638 !	$\mathcal{SHOIQ}(\mathcal{D})$	2353	1052
ontology-of-experimental-variables-and-values	3006	$\mathcal{ALCO}(\mathcal{D})$	189	122
ontology-of-general-purpose-datatypes	1588	$\mathcal{SHOI}$	773	340
ontology-of-geographical-region	1087	$\mathcal{AL}$	38	40
ontology-of-glucose-metabolism-disorder	1085	$\mathcal{AL}$	132	133
ontology-of-homology-and-related-concepts-in-biology	1328	$\mathcal{ALC}$	83	66
ontology-of-language-disorder-in-autism	1398	$\mathcal{AL}$	35	37
ontology-of-medically-related-social-entities	1565 !	$\mathcal{ALCHOIQ}$	218	145
ontology-of-physics-for-biology	1141 !	$\mathcal{ALCHIQ}(\mathcal{D})$	954	717
pathogen-transmission	1094	$\mathcal{AL}$	24	26
pathway-ontology	1035	$\mathcal{ALE}$	1432	1146
pediatric-terminology	1640	$\mathcal{AL}$	893	887
phare	1550 !	$\mathcal{ALCHIF}(\mathcal{D})$	459	309
phenotype-fragment-ontology	3049	$\mathcal{ALUHI}+$	28	28
phenotypic-quality	1107	$\mathcal{SH}$	1916	1441
phenx-terms	3078	$\mathcal{AL}$	339	369

Continued on next page

Table 7 – Continued from previous page

Ontology	BP ID		DL express.	#axs	#terms
phylogenetic-ontology	1616		$\mathcal{ALCH(D)}$	194	140
physical-medicine-and-rehabilitation	3015		$\mathcal{ALU}$	163	152
physicalfields	1369		$\mathcal{ALI}$	136	79
physico-chemical-methods-and-properties	1014		$\mathcal{ALE}$	1684	1166
physico-chemical-process	1043		$\mathcal{ALE}$	734	555
pilot-ontology	1399	!	$\mathcal{ALCIF(D)}$	85	41
pko_re	1409	!	$\mathcal{ALCF}$	771	771
plant-anatomy	1108		$\mathcal{ALE}+$	2128	1208
plant-environmental-conditions	1036		$\mathcal{AL}$	499	484
plant-ontology	1587		$\mathcal{S}$	2545	1487
plant-structure-development-stage	1038		$\mathcal{ALE}+$	281	277
plant-trait-ontology	1037		$\mathcal{ALE}$	1429	1271
platynereis-stage-ontology	1490		$\mathcal{ALE}$	31	19
pma-2010	1497		$\mathcal{AL(D)}$	10	10
protein-modification	1041		$\mathcal{ALE}+$	1986	1319
protein-ontology	1052	!	$\mathcal{ALCF(D)}$	691	232
protein-protein-interaction	1040		$\mathcal{ALE}+$	1007	909
proteomics-data-and-process-provenance	1039	!	$\mathcal{SHOIN(D)}$	732	492
proteomics-pipeline-infrastructure-for-cptac	1192	!	$\mathcal{ALCF(D)}$	1118	451
prov-o	3131	!	$\mathcal{ALCRIN(D)}$	185	89
pseudogene	1135		$\mathcal{AL}$	19	24
quantitative-imaging-biomarker-ontology	1671	!	$\mathcal{ALUIF(D)}$	1699	1383
rapid-phenotype-ontology	3114	!	$\mathcal{ALF(D)}$	2047	1713
rat-strain-ontology	1150		$\mathcal{ALE}$	4739	3345
reproductive-trait-and-phenotype-ontology	1552		$\mathcal{AL}$	91	96
rna-ontology	1500	!	$\mathcal{SRIQ}$	666	347
sample-processing-and-separation-techniques	1044		$\mathcal{AL}$	193	195
sanou	3090		$\mathcal{ALC}$	400	452
sanou	3091		$\mathcal{ALC}$	400	452
semanticscience-integrated-ontology	1532	!	$\mathcal{SRIQ(D)}$	2044	1449
sequence-types-and-features	1109		$\mathcal{SHI}$	2627	1977
skin-physiology-ontology	1122	!	$\mathcal{ALERIF}+$	678	374
sleep-domain-ontology	1651		$\mathcal{ALCO}$	204	141
smoking-behavior-risk-ontology	1249		$\mathcal{ALEI}+$	185	134
snp-ontology	1058	!	$\mathcal{SHOIN(D)}$	11199	2379
software-ontology	1413		$\mathcal{ALHI}+(\mathcal{D})$	5507	3845
solanaceae-phenotype-ontology	3029		$\mathcal{ALE}$	422	391
soyontology	3028		$\mathcal{AL}$	1816	1818
spatial-ontology	1078		$\mathcal{ALEHI}+$	236	161
spider-ontology	1091		$\mathcal{ALE}+$	778	557
student-health-record	1665		$\mathcal{ALH(D)}$	418	385
subcellular-anatomy-ontology-sao-	1068	!	$\mathcal{SHIN(D)}$	2935	913
symptom-ontology	1224		$\mathcal{AL}$	839	839

Continued on next page

Table 7 – Continued from previous page

Ontology	BP ID		DL express.	#axs	#terms
syndromic-surveillance-ontology	1394	!	$\mathcal{ALIF}(\mathcal{D})$	1684	369
sysmo-jerm	1488		$\mathcal{SHI}(\mathcal{D})$	482	328
systems-biology	1046		$\mathcal{AL}$	587	538
systems-chemical-biology-chemogenomics	1615	!	$\mathcal{SHIN}(\mathcal{D})$	489	218
taxonomic-rank-vocabulary	1419		$\mathcal{AL}$	58	60
teleost-anatomy-ontology	1110	!	$\mathcal{ALERI}+$	5188	3248
terminological-and-ontological-knowledge-resources	1418	!	$\mathcal{SRIQ}(\mathcal{D})$	466	329
terminology-for-the-description-of-dynamics	1407	!	$\mathcal{SRIQ}(\mathcal{D})$	12344	184
thesaurus-alternativa	3037		$\mathcal{AL}(\mathcal{D})$	138	142
thesaurus	3034		$\mathcal{AL}(\mathcal{D})$	138	142
thomcan-upper-level-cancer-ontology	3178		$\mathcal{AL}$	51	53
tick-gross-anatomy	1065		$\mathcal{ALE}+$	948	631
time-event-ontology	3042	!	$\mathcal{SROIQ}(\mathcal{D})$	1042	760
tissue-microarray-ontology	1494		$\mathcal{ALI}(\mathcal{D})$	60	36
translational-medicine-ontology	1461	!	$\mathcal{SRIN}(\mathcal{D})$	502	361
uni-ece	3048	!	$\mathcal{SOIF}(\mathcal{D})$	3133	712
units-of-measurement	1112		$\mathcal{ALE}$	371	363
units-ontology	1650		$\mathcal{AL}$	63	66
variation-ontology	3159		$\mathcal{ALEI}+$	390	384
vertebrate-homologous-organ-groups	1574		$\mathcal{ALE}+$	1688	1188
vertebrate-skeletal-anatomy-ontology	1555	!	$\mathcal{ALER}+$	455	289
vertebrate-trait-ontology	1659		$\mathcal{AL}+$	3691	3155
vital-sign-ontology	3124	!	$\mathcal{SROIQ}$	743	346
web-service-interaction-ontology	1632	!	$\mathcal{ALER}+$	29	30
wheat-trait	1545		$\mathcal{AL}$	175	177
xeml-environment-ontology	3176		$\mathcal{ALE}$	237	148
xenopus-anatomy-and-development	1095		$\mathcal{ALE}+$	4819	1263
yeast-phenotypes	1115		$\mathcal{AL}$	266	270
zebrafish-anatomy-and-development	1051		$\mathcal{ALE}+$	10600	2774
Galen		!	$\mathcal{ALEHIF}+$	4,735	3,161
Koala		!	$\mathcal{ALCON}(\mathcal{D})$	42	36
Mereology		!	$\mathcal{SHIQ}$	44	30
MiniTambis-repaired		!	$\mathcal{ALCN}$	170	227
OWL-S Profile		!	$\mathcal{ALCHOIN}(\mathcal{D})$	276	154
People		!	$\mathcal{ALCHOIN}$	108	99
Tambis-full		!	$\mathcal{SHIN}(\mathcal{D})$	592	497
University		!	$\mathcal{SOIN}(\mathcal{D})$	52	45