An Empirical Investigation of Difficulty of Modules of Description Logic Ontologies

Nicolas Matentzoglu, Bijan Parsia, and Uli Sattler

The University of Manchester
Oxford Road, Manchester, M13 9PL, UK
{matentzn,bparsia,sattler}@cs.manchester.ac.uk

Abstract. Very expressive Description Logics in the $\mathcal{SH}$ family have worst case complexity ranging from EXPTIME to double NEXPTIME. In spite of this, they are very popular with modellers and serve as the foundation of the Web Ontology Language (OWL), a W3C standard. Highly optimised reasoners handle a wide range of naturally occurring ontologies with relative ease, albeit with some pathological cases. A recent optimisation trend has been *modular* reasoning, that is, breaking the ontology into hopefully easier subsets with a hopefully smaller overall reasoning time (see MORe and Chainsaw for prominent examples).
However, it has been demonstrated that subsets of an OWL ontology may be harder – even much harder – than the whole ontology. This introduces the risk that modular approaches might have even more severe pathological cases than the normal monolithic ones.
In this paper, we analyse a number of ontologies from the BioPortal repository in order to isolate cases where random subsets are harder than the whole. For such cases, we then examine whether the module nearest to the random subset also exhibits the pathological behaviour.

Keywords: Modular Reasoning, Ontologies, Modules

1 Introduction

Reasoning in popular, very expressive Description Logics is very difficult (e.g., $\mathcal{SROIQ}$ is N2EXPTIME-complete) [9]. Modern reasoners such as FaCT++ [17], Pellet [16] and HermiT [12] generally perform well against naturally occurring inputs such as ontologies downloadable from the Web. However, due to the poor performance in some cases, the quest for optimisations is ongoing.

Modular reasoning techniques have experienced a resurgence in recent years. Intuitively, breaking the input problem into smaller pieces is appealing. From a naive perspective, it seems that keeping the inputs small is a way to deal with high worst case complexity. Furthermore, if there are especially difficult parts of the ontology, perhaps they can be isolated to reduce their impact. However, classifying a subset $\mathcal{S}$ of an ontology $\mathcal{O}$ can be, in fact, much more expensive than classifying $\mathcal{O}$ entirely [6]. There are a number of reasons why classification time could vary in this way. For example, a disjointness axiom may be missing from the random set and thus result in subsumption tests that are computed needlessly (in the worst case almost the entire power set of possible atomic subsumptions of entities in a signature, if the disjointness happens high up in the hierarchy).

This raises a question for modular reasoning: Do modules ever exhibit such a pathological behaviour? The modules currently used in modular reasoners are *depleting*, that is, contain every axiom that could participate in a derivation of an entailment from that module. They are also subsets of $\mathcal{O}$. Our conjecture is that because of this, reasoners should never perform worse on a module than on the whole $\mathcal{O}$. In this paper, we test this conjecture on ontologies from the BioPortal repository. Our results indicate that while generally modules are "easy", there are cases where a module has a classification time worse than that of $\mathcal{O}$. While our current results suggest that some modular reasoning techniques are not vulnerable to this phenomenon, others require further investigation.

2 Preliminaries

Understanding our experiment does not more than a cursory understanding of the syntax, semantics, and proof theories implemented: For the most part, we are treating these as black boxes. Some key points:

- The reasoners under investigations are designed to implement key reasoning services for the Web Ontology Language (OWL).
- The services we test ignore all non-logical aspects of OWL 2, thus OWL 2 ontologies can be regarded as notational variants of $\mathcal{SROIQ}$ theories (with some minor syntactic caveats, e.g., OWL ontologies are *sets* of axioms).
- Given an ontology (a set of axioms) $\mathcal{O}$, the signature of an ontology $\widetilde{O}$ is the set of non-predefined entities (classes, individuals, object properties, data properties) appearing in the axioms in $\mathcal{O}$. We use $CL(\mathcal{O})$ to denote the task of computing the set of entailed atomic subsumptions (i.e., statements of the form $A \sqsubseteq B$ and $A \equiv B$, where A and B are elements of the signature) or *classification* of $\mathcal{O}$. We use $T(X)$ to denote the time taken for a given process to terminate. For example $T(CL(\mathcal{O}))$ denotes the time to classify $\mathcal{O}$ (for a given reasoner).

There are many sorts of "logically respectable" modules presented in the literature, most based on deductive conservative extensions. Here we restrict our attention to modules based on *syntactic locality* [7], because they are logically sound, work for all of OWL and often behave well in practice [3], though not always [11]. Among these kinds of modules, we are particularly interested in $\bot$-modules (bottom-modules), because they preserve all subsumptions over their signatures with an element of the signature on the left hand side.

As a consequence, if the union of the signatures of a set of $\bot$-modules equals the signature of $\mathcal{O}$, then the union of the individual classification results equals the classification of the whole ontology.

Hereafter, we will use $\mathcal{M}$ to refer to syntactic locality-based $\bot$-modules. Additionally, we will use $\mathcal{S}$ to denote an arbitrary random subset of $\mathcal{O}$ and $\mathcal{H}$ to denote a random subset whose reasoning time is greater than that of $\mathcal{O}$.

3 Modern Modular Reasoning

We focus on TBox reasoning in locality module based systems for several reasons. First, at least one system (Pellet) has been using them for incremental reasoning in production systems successfully for years [2]. Second, two new systems, MORe [15] and Chainsaw [18], make use of them for non-incremental classification and have seen at least some success. (e.g., each won a category at the 2013 Ontology Reasoner Evaluation [4].)

The overarching rationale for exploiting the modular structure of an ontology for reasoning is the potential in (1) reducing the number of necessary subsumption tests and (2) reducing the hardness of each individual subsumption test, thus improving performance (reasoning time and potentially memory consumption). Additionally, modularisation may make exploiting parallelism easier, though this has not been seriously explored to our knowledge.

Chainsaw and MORe use quite different strategies in exploiting modules. MORe works by building two modules: the largest easy-to-find $\mathcal{EL}$ module and the "complement" module which covers the rest of the signature. The $\mathcal{EL}$ module is fed to a fast $\mathcal{EL}$ specific reasoner (typically ELK[10]) while the complement module is fed to a general purpose OWL 2 reasoner (HermiT or Pellet). If the complement module covers most of the ontology, then there is a significant amount of redundant computation in addition to the overhead and thus MORe will perform comparatively poorly to the monolithic reasoner.

In contrast, Chainsaw employs the "Atomic Decomposition" (AD) [19] of an ontology. The AD is a partitioning of the ontology into so called genuine modules (modules that cannot further be decomposed into the union of two "$\subset$"-incomparable modules). The AD is represented as a directed graph where each node corresponds to an atom and determines the corresponding genuine module. The set of genuine modules covers the signature of $\mathcal{O}$, thus is a suitable candidate for modular classification. While decomposing an ontology can be done in polynomial (quadratic) time, the reality is that the overhead introduced by such a partitioning can be substantial in some cases. Furthermore, the existing implementation can also experience considerable redundancy due to overlapping genuine modules.

3.1 Module Hardness

While the existence for example of Hot Spots [6] suggests that decomposition might have dramatically beneficial effects, it was also empirically observed in [6] (and has been known anecdotally for years) that subsets of an ontology might be dramatically harder than the whole. This fact is not terribly surprising if one recalls the easy-hard-easy pattern for satisfiability of random KCNF propositional formulae where the hard region corresponds to the phase transition from satisfiability to unsatisfiability. As the generated formulae move from low "density" (i.e., a low ratio of number of clauses (L) to number of distinct variables (N)) to high density, the size of the formulae grows (for a fixed N). Thus, we can easily see systems where random subsets of a large problem are much harder than the whole. While there is, as yet, no detailed theory (or even sketch of a theory) for hardness in naturally occurring Description Logic ontologies, we can expect, to a certain extent, that similar mechanisms may be at work.

Obviously, given an ontology $\mathcal{O}$, if $T(CL(\mathcal{M}_i))$ in a decomposition $Dec(\mathcal{O})$ is larger than $T(CL(\mathcal{O}))$, then modular reasoning is counterproductive in that case and thus efforts to tune the decomposition and other overhead is pointless. The goal of this paper is to investigate the general case: Are modules of an ontology ever (or likely) to be harder than $\mathcal{O}$ itself? In contrast with random subsets, there are some *prima facie* reasons for thinking the answer is no if we consider some possible sources of hardness: (1) If the reasoner is using standard enhanced traversal to classify the ontology, a random subset might omit an explicitly asserted subsumption or, perhaps more importantly, disjointness axiom which considerably prunes the search space. Modules must contain all such axioms over their signature in order to capture those entailments. (2) A random subset might break some justification of an atomic subsumption over its signature leaving either only harder ones or leaving enough of a justification that the reasoner has to do a lot of work before it can determine it will not pan out.

In general, since a module contains *everything* from the ontology relevant to entailments over its signature, it seems reasonable to think that reasoning should be no harder. No shortcuts can be missing and we have removed a lot of potential distractions. However, these considerations are merely speculative. In this paper, we attempt to establish whether there are hard subsets that are modules and to start to sketch their prevalence and other properties. If we can establish the phenomenon, then we have laid the groundwork for causal investigations. We would like to note that we are not engaged in either MORe or Chainsaw algorithm optimisation at this point. It could well be that, if there are hard modules, that they are not the sort of modules that either technique uses. Similarly, even if there are no hard modules, that does not mean that modules generally will be "easy enough" to overcome the overhead challenges.

4 Materials and Experimental Design

The conjecture under test is that given an ontology, $\mathcal{O}$, reasoner R, then there is no module $\mathcal{M}$ from that ontology such that the $T(CL(\mathcal{O})) < T(CL(\mathcal{M}))$ for R. It's trivially the case that there is a $\mathcal{M}$ such that $T(CL(\mathcal{O})) = T(CL(\mathcal{M}))$ since $\mathcal{O}$ is a module for itself.

The number of modules in an ontology can be very large (in principle, exponential in the number of axioms or the signature; this has been verified empirically [14]), thus pure uniform sampling is unlikely to detect counterexamples to our conjecture if they are rare. Given that empirically hard arbitrary subsets are not particularly common either within ontologies (i.e., most subsets are not harder than the whole) or across ontologies (i.e., most ontologies do not have hard subsets), we need a more directed approach. Our basic approach is to look for hard arbitrary subsets using the techniques described in [6]. We then extract the module for the signature of that hard subset, that is, the nearest superset of the hard subset that is a module. If the speculations on sources of hardness are right, at least in spirit, we should see a big drop in classification time going from the hard subset to the module. Of course, this procedure cannot rule out that other modules might be hard (it is, intentionally, a biased sample), but at least the modules we extract are in a known region of hard subsets.

4.1 Corpus Selection

We conducted our study on a corpus of 363 BioPortal ontologies (August 2013), serialised into OWL/XML by the OWL API (3.4.8) [8]. BioPortal [13] is a open repository for accessing and sharing biomedical ontologies, developed in languages such as OWL, RDF or OBO. It is the most prominent and active collection of naturally occurring ontologies on the web, and serves as the go-to point for many empirical investigations, such as studies on modularity or reasoner profiling [5]. BioPortal contains a range of interesting sizes and expressivities (see Figure 1).

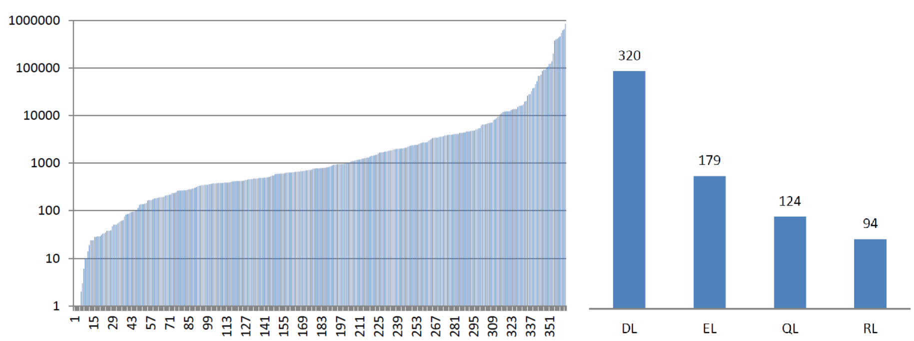


Fig. 1. Distribution of ontology sizes in Bioportal (logical axiom counts), and OWL 2 profiles.

4.2 Corpus Curation and Filtering

As we are primarily interested in classification, we considered only the TBoxes of the ontologies (we stripped $\mathcal{O}$ of all ABox axioms). In order to handle large numbers of subsets, we indexed the axioms using annotation properties and physically manifested the subsets (or modules) only as lists of axiom identifiers.

We classified the entire set of 363 ontology TBoxes using Pellet 2.3.1 and HermiT 1.3.8, and then filtered them, separately for each reasoner. We excluded ontologies that (1) not OWL DL, (2) for a given reasoner classified in less than 5 seconds (relatively easy), (3) for a given reasoner did not classify within 20 minutes (feasible), (4) that merely contained a class hierarchy (only atomic SubClassOf axioms and EquivalentClasses axioms) (expressive).

43 ontologies were filtered out as being OWL Full using the profile checker in the OWL API tools ([1]). In order to classify an interesting amount of subsets, and given our informed assumption that there are subsets that are harder than the whole, an infeasible amount of resources would be required if ontologies were included that were significantly harder than 20 minutes.

4.3 Finding Hard Subsets

In order to find hard subsets of ontologies, we employed the technique described in [6]. We randomly sampled 3 *paths* from each of the ontologies in order to identify ontologies with *hard regions*. A path was sampled by *slicing* the ontology $\mathcal{O}$ (the set of logical axioms) into size regions relative to the size of $\mathcal{O}$, in our case eighths. The first slice corresponds to 1/8 of the ontology, the second slice to 2/8 and so on, the full TBox to 8/8. A random eighth of the set of (logical) axioms was picked from the ontology and exported into a new ontology. Then a second eighth of the remaining axioms was drawn (randomly) and added to the first eighth to get the 2/8 slice of the ontology, and so on. A *path* thus consists of eight cumulatively grown subsets.

We then classified all the subsets generated this way to identify hard subsets (subsets for which the reasoning time exceeds the reasoning time of the whole TBox). Our assumption was that if we find such a hard subset, it is an indicator for a *hard region*. For example, if we find a hard subset that correspond to 6/8 of the ontology, we assume that there might be more hard subsets to be found in that region. Once we identified hard subsets, we sampled 50 random subsets from their corresponding regions and obtained their classification time.

4.4 Classifying the Modules

For each hard subset $\mathcal{H}$ identified in the previous step, we extracted a $\bot$-module $\mathcal{M}$ with the syntactic locality-based module extractor embedded in FaCT++, using the signature of the hard subset as seed. We then obtained the classification time for $\mathcal{M}$. In principle $\mathcal{M}$ should be a superset or equal to $\mathcal{H}$. However, if $\mathcal{H}$ contains axioms that are tautologies, they generally do not make it into the module. The most prominent cases are axioms of the form $A \sqsubseteq (\geq 0 R.C)$, or simply $A \sqsubseteq \top$.

4.5 Experiment Machine Specification

A Dell Optiplex 790, with Ubuntu Release 12.10 (quantal) 64-bit, Memory: 15.6 GiB (1333MHz, DDR3), Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz x 4 was used for the actual benchmarking (obtaining the classification times). Every single classification was done in a separate isolated virtual machine (Java 7, -Xms2G, -Xmx12G). Network and other unnecessary system services were disabled to prevent background processes from affecting reasoning performance.

5 Results

5.1 Hard Subset Finding

The filtering process described in Section 4.2 resulted in 21 ontologies for HermiT and 15 ontologies for Pellet from the BioPortal corpus. Out of these, we found hard subsets using the technique described in Section 4.3 for three ontologies with Pellet and three ontologies for HermiT. For Pellet we had to check a total of 315 subsets to find 4 hard subsets (all in region 7), for HermiT a total of 441 to find 3 hard subsets (two in region 7, one in region 6). It should be noted that there were a minor amount of failed classifications (6 for Pellet, 1 for HermiT), which were either due to a stack overflow or a timeout (fixed at 20 min, wall clock time). They were excluded from the analysis. There are two things to observe at this point: (1) hard subsets are relatively rare and (2) they are found in the higher regions that correspond to 6/8 or 7/8 of the ontology. As for the first observation, it should be noted that a lot of subsets under consideration corresponded to less than half of the size of the ontology (252 for HermiT, 180 for Pellet). Although it was not clear whether hard subsets might be hidden in these regions, it was also not surprising that they did not turn out to be. In particular, a random sample is likely to select unrelated axioms, thus small subsets are unlikely to be tightly constrained.

5.2 Performance Heterogeneity

Gonçalves et al [6] observed that ontologies which have non-linear path profiles are highly likely to be performance heterogenous, and thus have Hot Spots. Performance heterogeneity also seems to correlate with having hard subsets. We observed the performance profiles exhibited by the paths (i.e. the classification times of the cumulative grown 1/8, 2/8, ..., 8/8 of $\mathcal{O}$). In principle we could count a local peak in a path that grows non-monotonically as a candidate for a region of hard subsets, but for experiment resource related reasons we restricted our attention to regions were a path had a peak that was harder than the whole ontology – in such cases we knew that there was at least one hard subset.

For HermiT, 3 out of the 63 paths (3 for each of the 21 ontologies in the set) contain hard subsets and 5 have some non-monotonic characteristics (i.e. there was a drop in reasoning time from one subset to its successor). The profiles of the ontologies that contained hard subsets can be seen in Figure 2. 16 paths have a Pearson product-moment correlation coefficient higher than 0.85 (between the relative size of the subset and the reasoning time), i.e. $T(CL(\mathcal{S}))$ and $|S|$ have a strong linear dependence, while a total of 7 paths have a coefficient between

0.5 and -0.5, indicating a relatively weak linear relationship. For Pellet, 4 out of the 45 paths have hard subsets, 10 of the respective performance graphs have non-monotonic characteristics, 11 paths have a high Pearsons coefficient and 9 a low one. Figure 2 shows the profiles of the three ontologies that contain hard subsets.

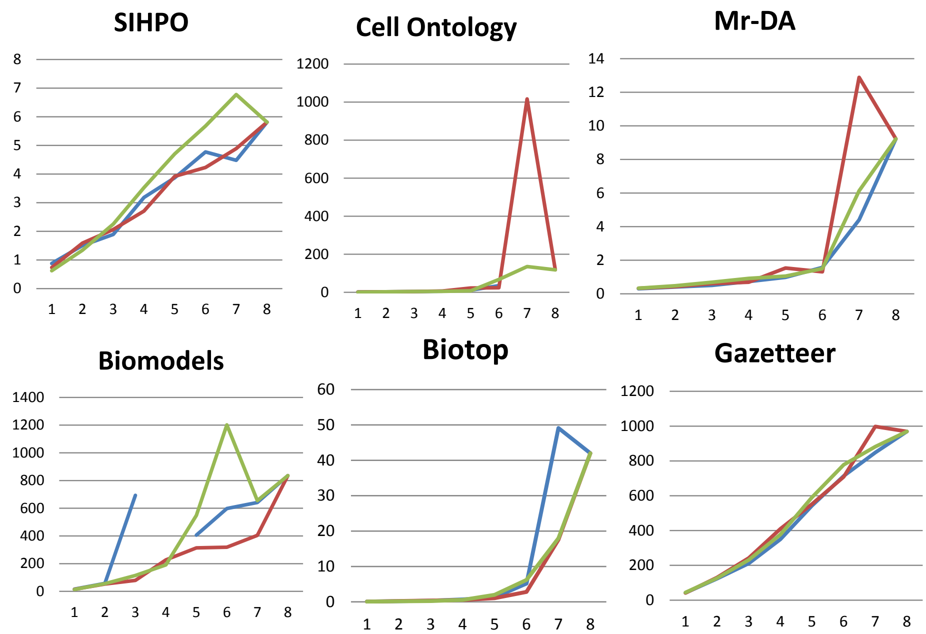


Fig. 2. Performance profiles for SIHPO, CO and MDA with Pellet (top) and Biomodels, Biotop and Gazetteer with HermiT (bottom). Y axis is $T(CL(\mathcal{O}))$ in seconds

5.3 Performance Profile of Hard Subsets and their Corresponding Modules

We selected a sample of 50 subsets from the identified hard regions for each ontology/reasoner pair for a total of 300 subsets. We then tested each subset for hardness. For HermiT, a total of 25 hard subsets were identified, for Pellet 42. We subsequently classified these subsets and their corresponding modules. To mitigate unanticipated experimental effects, we classified the subset / module pairs three times, as well as the ontology. Details can be found in Figure 5.3. The timeout was set to 30 minutes.

Each grouped column on the x-axis in the figures represents the classification times of a pair (the subset and its corresponding module). The y-axis represents the classification time in seconds. The red line $CT(O)$ (CT(X) as an equivalent shorter form of $T(CL(X))$) represents the median classification time of the entire ontology measured in the three runs mentioned earlier. If one of the other runs for classification of $\mathcal{O}$ deviated from the median by more than 5%, we included the value in the plot (CT(O)+ standing for the highest classification time obtained for the ontology, and CT(O)- the lowest).

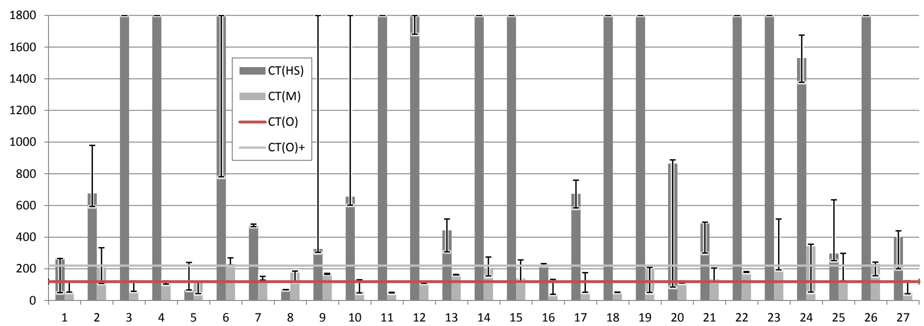
6 Discussion

We call a case (a subset/module/ontology triplet) *strictly confirmatory* of our hypothesis if the maximum classification time measured for a module is lower than the minimum classification time measured for the entire TBox. We call a case *tendentiously confirmatory* if the median classification time measured for a module is lower than the median classification time measured for the entire TBox.

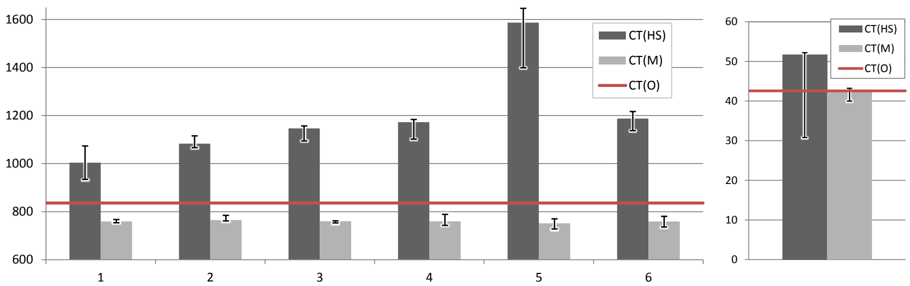
The first, and most striking, observation is that only three of the six ontology/reasoner pairs are strictly confirmatory of our hypothesis as a whole (contain only strictly confirmatory cases, Biomodels, MDA, SIHPO) and only one other ontology is tendentiously confirmatory as a whole (Biotop). Biotop/HermiT shows modules that are easier (at least for the median value) than their generating subset, but one measurement was taken were the subset was classified faster. Cell ontology/Pellet was strictly confirmatory in 21 out of 27 cases, and tendentiously confirmatory in another 4 (25/27 in total).

Gazetteer/HermiT clearly contradict the conjecture. We must thus, barring experiment error, regard it as false. We found this surprising.

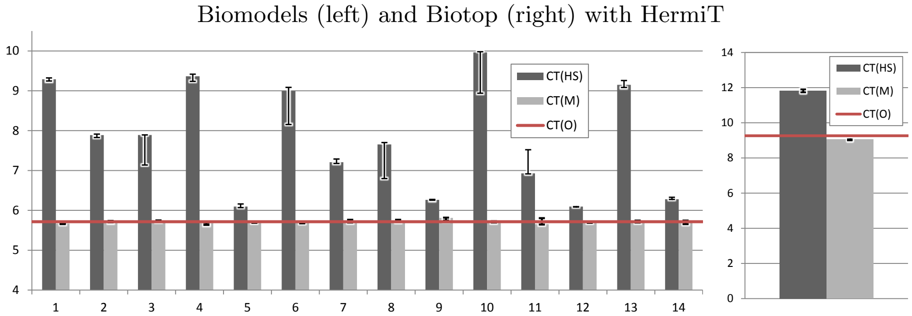
In all cases, both the hard subset and the module were significant fractions of the ontology, as expected. This makes the results more relevant for reasoners like MORe which tries to generate large modules. Hard large modules suggest that MORe may have surprisingly pathological cases. In conversation with the MORe developers, the threat to performance that they saw was that the non-$\mathcal{EL}$ reasoner might essentially perform as it would on the whole ontology (or perhaps a little better, but still close) but MORe would also have the cost of running the $\mathcal{EL}$ reasoner. That is, they were concerned with *duplicated* work. Our results show that the complement module might take much longer than the whole ontology. Determining how often this sort of pathological behaviour may occur is part of our future work. One possibility to explain the hardness of subsets over their generating module is the presence or absence of search space pruning axioms. It is at least possible that the existence of an axiom or a group of axioms in the module reduces the amount of necessary subsumption tests, either by pruning the traversal space (for example through a disjointness axiom) or by asserting a subsumption that makes a lot of tests redundant (for example by making a concept unsatisfiable) or reduces the amount of non-deterministic choices. We tried to gather some very initial evidence for these kinds of explanations by counting the number of SAT calls for a given reasoner/ontology pair. For Biomodels/HermiT we can confirm that the subset requires between 11057 and 25271 more SAT calls then its module for a complete classification



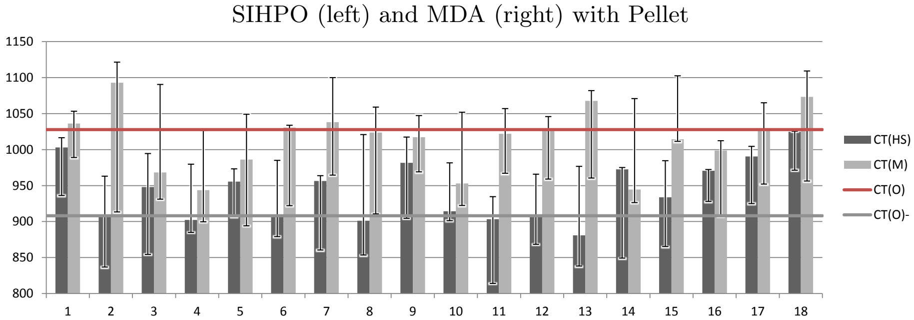
Cell Ontology (Pellet). Bars reaching the top of the chart timed out at 30 min.



Biomodels (left) and Biotop (right) with HermiT



SIHPO (left) and MDA (right) with Pellet



Gazetteer (HermiT)

Fig. 3. Subset and module comparisons for respective ontologies. Y axis is $T(CL(\mathcal{O}))$ in seconds.

(Median 15066.5). In the Biotop/HermiT case, we have 110, in the MDA/Pellet case 9 and in the SIHPO/Pellet case between 7606 and 42296 (Median 28053.5) more SAT calls. All of the former were at least tendentiously confirming of our hypothesis. In the Gazetteer case, where in all but one cases, $T(CL(M))$ was higher than $T(CL(\mathcal{S}))$ (strongly contradicting the hypothesis), we can observe exactly the reverse: we have at between 541 and 624 (Median 578) fewer SAT calls during the subset classification, which at least strengthens the evidence that the amount of SAT calls correlates somehow with the overall classification time. The CO/Pellet combination appears to draw a more mixed picture. For a random third (9) of the subset/module pairs, all of which are strictly confirmatory of our hypothesis, the differences in SAT calls tend from 6065 more to 6009 less calls, with a Median of 2740 less (ie, does not support this explanation).

The primary threat to internal validity is that we occasionally observed non-deterministic reasoning times for the same input. While we believe our current results are stable, further work needs to be done to ensure that the way we feed inputs to the reasoner ensures a like behaviour. It could be that we accidentally have modules triggering a "hard state" of the reasoner while the ontology is triggering an easy state due to irrelevant features such as order of axioms in the file (which can have an effect on absorption). If the reasoner can be kicked into a state where the reasoning times are reversed, then it is possible that a refined version of our conjecture is true.

The primary threat to external validity is the relatively small sample sets and ontologies examined as well as the inherent biasing of our sampling strategy. We will continue to cover more hard subsets and candidate ontology reasoner pairs to get a fuller picture even though our current work suffices to refute the conjecture. Given that we start with random, thus not coherent, subsets, it may be the case that we can find 'interestingly hard" modules by looking at genuine modules or other guided seed signature strategies. While not telling for the conjecture, they might help provide insight into sources of difficulty.

With respect to field significance, it seems that it is not unreasonable for modular reasoner developers to neglect this threat. The comparatively low number of hard subsets in general and of hard modules suggests that other issues still dominate.

In addition to broadening our sampling as wide as possible, the key future task is to investigate the hard subsets and their modules to see if we can isolate the sources of pathological or benign behaviour. One strategy is to search for minimal supersets of hard subsets or modules that are easy. In this way, we can provide minimal test cases for pathological hardness to reasoner developers.

References

1. owlapitools - a set of independent add-ons for owl api. `https://github.com/owlcs/owlapitools`. Accessed: 2014-06-09.
2. B. Cuenca Grau, C. Halaschek-Wiener, and Y. Kazakov. History matters: Incremental ontology reasoning using modules. In *Lecture Notes in Computer Science*

(including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 4825 LNCS, pages 183–196, 2007.
3. C. Del Vescovo, P. Klinov, B. Parsia, U. Sattler, T. Schneider, and D. Tsarkov. Empirical study of logic-based modules: Cheap is cheerful. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8218 LNCS, pages 84–100, 2013.
4. R. S. Gonçalves, S. Bail, E. Jiménez-Ruiz, N. Matentzoglu, B. Parsia, B. Glimm, and Y. Kazakov. Owl reasoner evaluation (ore) workshop 2013 results: Short report. In *ORE*, pages 1–18, 2013.
5. R. S. Gonçalves, N. Matentzoglu, B. Parsia, and U. Sattler. The empirical robustness of description logic classification. In *CEUR Workshop Proceedings*, volume 1014, pages 197–208, 2013.
6. R. S. Gonçalves, B. Parsia, and U. Sattler. Performance heterogeneity and approximate reasoning in description logic ontologies. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7649 LNCS, pages 82–98, 2012.
7. B. Grau, I. Horrocks, Y. Kazakov, and U. Sattler. Modular Reuse of Ontologies: Theory and Practice. *J. Artif. Intell. Res.(JAIR)*, 31:273–318, 2008.
8. M. Horridge and S. Bechhofer. The OWL API: A Java API for OWL ontologies. *Semantic Web*, 2:11–21, 2011.
9. Y. Kazakov. RIQ and SROIQ are harder than SHOIQ. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, pages 274–284, 2008.
10. Y. Kazakov, M. Krötzsch, and F. Simancik. The incredible elk - from polynomial procedures to efficient reasoning with ;; ontologies. *J. Autom. Reasoning*, 53(1):1–61, 2014.
11. B. Konev, C. Lutz, D. Walther, and F. Wolter. Model-theoretic inseparability and modularity of description logic ontologies. *Artif. Intell.*, 203:66–103, 2013.
12. B. Motik, R. Shearer, and I. Horrocks. Hypertableau reasoning for description logics. *J. Artif. Intell. Res. (JAIR)*, 36:165–228, 2009.
13. N. F. Noy, N. H. Shah, P. L. Whetzel, B. Dai, M. Dorf, N. Griffith, C. Jonquet, D. L. Rubin, M. A. Storey, C. G. Chute, and M. A. Musen. BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37, 2009.
14. B. Parsia and T. Schneider. The modular structure of an ontology: An empirical study. In *KR*, 2010.
15. A. A. Romero, B. C. Grau, and I. Horrocks. MORe: Modular Combination of OWL Reasoners for Ontology Classification. In *International Semantic Web Conference (1)*, pages 1–16, 2012.
16. E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y. Katz. Pellet: A practical OWL-DL reasoner. *Web Semantics*, 5:51–53, 2007.
17. D. Tsarkov and I. Horrocks. Fact++ description logic reasoner: System description. In *IJCAR*, pages 292–297, 2006.
18. D. Tsarkov and I. Palmisano. Divide et Impera: Metareasoning for Large Ontologies. In *OWLED*, 2012.
19. C. D. Vescovo, B. Parsia, U. Sattler, and T. Schneider. The Modular Structure of an Ontology: Atomic Decomposition. In *IJCAI*, pages 2232–2237, 2011.