Semantic Expressibility of OWL Ontologies


Data Semantics Lab, Kansas State University, USA
\{aaroneberhart, coganmshimizu, sulognac, mdkamruzzamansarker, hitzler\}@ksu.edu

Abstract. The high expressivity of the Web Ontology Language (OWL) makes it possible to describe complex relationships between classes, properties, and individuals in an ontology. However, this high expressivity can be an obstacle to correct usage and wide adoption. Past attempts to ameliorate this have included the development of specific, presumably human-friendly syntaxes, such as the Manchester syntax or graphical interfaces for OWL axioms, albeit with limited success. If modelers want to develop suitable OWL axioms it is important to study why managing ontology complexity is difficult.

In this paper, we adopt an idea from the Protége plug-in, OWLAx, which provides a simple, clickable interface to automatically input a limited number of axioms by following simple patterns. In particular, each of these axiom patterns contains at most three classes, roles, or data properties. We hypothesize that most of the axioms in existing ontologies can be expressed semantically in terms of simple axiom patterns like these. Our findings, based on an analysis of 518 ontologies from six public ontology repositories, confirm this hypothesis: Over 90% of class axioms in the average ontology are indeed expressible with simple patterns. We provide a detailed analysis of our findings, and are also able to further reduce the number of axiom types that are needed to obtain high expressibility.

1 Introduction

Knowledge graph schema are complex artifacts that can be difficult and expensive to produce and maintain. This is especially true when encoding them in OWL (the Web Ontology Language) as ontologies. The high expressibility of OWL is a boon, in that it makes it possible to describe relationships between classes, properties, and individuals in an ontology. At the same time, however, this high expressibility is often an obstacle to its correct usage that can limit wide adoption. Past attempts to ameliorate this have included the development of specific, presumably human-friendly syntaxes, such as the Manchester syntax \cite{9}, or graphical interfaces for OWL axioms, albeit with modest success \cite{14}. Additionally, certain engineering paradigms and methodologies have been developed, such as eXtreme Design \cite{2} or Modular Ontology Modeling \cite{917}, that try to simplify the modeling process.
In general, these methodologies aim to guide ontology developers through the complex modeling process by either abstracting the complexity away (for example, through the use of Ontology Design Patterns), or by limiting the scope of the model to something immediately applicable and understandable. In this paper we are particularly interested in the latter, especially during the axiomatization process. We believe that it is important to investigate new avenues for improving the approachability of the creation of suitable OWL axioms.

One of the core tenets of the Modular Ontology Modeling methodology is to produce schema diagrams and then systematically axiomatize them, with the input of domain experts. This systematic axiomatization is inspired by the OWLAx plugin for Protégé which provides a simple, clickable interface to automatically input a limited number of axioms that are all created with simple axiom patterns. In particular, each of these simple axiom patterns contains at most three classes or roles. In, it was posited (but left unproven) that the 17 axiom patterns provided by the interface were sufficient for most modeling purposes. In this paper, we test that hypothesis by analyzing 518 ontologies from six public ontology repositories. Concretely, we show the following:

\[ \text{H1. Almost all axioms in OWL are semantically expressible using the set of simple axiom patterns found in Table 1} \]

And indeed, as we will see, it holds for over 90% of class axioms using our relatively straightforward analysis. With a more thorough analysis or with different patterns, the percentage may even be higher.

The rest of this paper is organized as follows. In Section 2 we briefly describe literature related to our analysis. Section 3 presents our research method and then Section 4 presents our evaluation. The results are discussed in Section 5. Finally, in Section 6 we conclude.

2 Related Work

We are aware of only a very limited amount of research that specifically concerns the semantic, not syntactic, composition and expressibility of ontologies. Zhang et al. look at ways to measure the design complexity of ontologies. Their work is focused more on ontology quality evaluation than ontology composition. Some have also attempted to measure the effect that axioms like existential quantifiers have on reasoning time, such as Kang et al. [10], although it is only tangentially related to the work that we are presenting.

There are also, as previously mentioned, tools that attempt to simplify OWL ontology development, such as Manchester Syntax [9], WebVOWL [12], CoModIDE [10], Graffoo [8], and ROWLTab [14]. However, these tools merely simplify the development process and do not measure whether OWL axioms are necessarily complex in everyday usage. It could very well be the case that OWL is unavoidably complicated and these tools are needed to deal with this complexity, although we believe our work demonstrates that this is usually not the case.

\[ \text{1 See } \text{https://protege.stanford.edu/} \]
3 Methodology

Our hypothesis is that most axioms in ontologies either are simple, or else could be expressed mostly with simple axioms. In this section, we will define what we mean by simple axioms, then give an example of a set of patterns for simple axioms, such as those used in the Protégé plugin OWLAx. Following that, we will describe how axiom patterns can be used to analyze how much of an ontology is expressible with simple axioms, and then provide some minimal normalizations that can reduce multiple different ontologies syntactically without modifying their semantics.

3.1 Simple Axioms

The simple axioms we study in this paper are defined below. We consider description logic syntax for OWL DL, that is, we identify it with the description logic $SROIQ(D)$ [7].

Definition 1. A Simple Axiom is any OWL axiom that contains at most three class, role, or data property names, or a data range, and is not a syntactic shortcut for other OWL axioms. Any axiom which is not simple is a Complex Axiom.

Our set of axiom patterns is for class axioms, so we restrict our focus to class axioms in the evaluation, although, in principle, the notion of a simple axiom could apply to role and data property axioms as well. The limitation of three atoms for simple axioms is an intuitive threshold, in terms of size, because it means that nesting in the expression is limited to at most one quantifier, yet the axiom can still participate in complex inferences in combination with other simple axioms. This would not be the case for axioms limited to size two, where one could only express $A \sqsubseteq B$ for classes, or $R \sqsubseteq S$ for roles, which would radically limit the semantic expressibility of the ontology. Axioms with more than three atomic classes or roles may be more expressive, but are often equivalent through normalization to smaller axioms, so they do not make good candidates for simple axioms. Note that negation is not considered complex, since the definition uses the number of classes and roles, and double negations can be eliminated trivially.

3.2 OWLAx

OWLAx [15] is a Protégé plugin that allows users to automatically generate certain simple OWL axioms using a graphical interface. The set of axioms we study in this paper are inspired by the axioms that OWLAx can create, and they are listed in Table 1.

The actual implementation details of the OWLAx plugin are not pertinent to what we want to discuss. Rather, we are interested in what it happens to contain: a set of patterns that only make simple axioms. Since it was designed specifically
to help create ontologies, we speculate that ontologies will be mostly expressible using these patterns. We now show how axiom patterns, like those used in OWL Axiom Patterns, can evaluate the pattern expressibility of axioms in an ontology. If a high percentage of axioms in an ontology are expressible with the patterns from our table, then it must contain a high number of semantically simple axioms, since axioms made from the patterns in Table 1 are simple.

### 3.3 Axiom Pattern Expressibility

To study whether axioms in an ontology are semantically expressible with simple axiom patterns, we first define the term axiom pattern and then show how a set of axiom patterns can be used to study axioms in an ontology, obtaining multiple metrics to evaluate pattern expressibility.

**Definition 2.** An Axiom Pattern is a programmatic template for creating new valid axioms. An axiom pattern may have variable terms that can be instantiated to obtain specific axioms.

For example, the axiom pattern $A \sqsubseteq \exists R.B$ for an existential axiom from Table 1, where $A$, $B$, $R$ are variable terms, can be used to generate the axiom Dog $\sqsubseteq \exists \text{chases}.\text{Squirrel}$ by substitution, where “Dog” and “Squirrel” are classes and “chases” is a role.

**Definition 3.** The Axiom Pattern Expressibility of an axiom $\alpha$ w.r.t. a set of axiom patterns $P$ is the set of patterns $p \in P$ that can generate $\alpha$ with the fewest substitutions, written $ae_P(\alpha)$.

Note that the requirement for pattern expressibility to use the fewest substitutions means that, for our set of patterns at least, axioms will only be expressible with at most one pattern. A functional axiom $\top \sqsubseteq \leq 1S.\top$, for instance, technically can be created by all of the functional patterns, because $\top$ is a class. However its axiom pattern expressibility is the functional pattern because there are fewer variable terms to substitute for the that pattern than the scoped functional, qualified functional, and scoped qualified functional patterns. A range
axiom appears to fit the scoped range pattern for the same reason, but the range pattern has fewer variable terms, so that is its expressibility, and so on.

**Definition 4.** The **Ontology Pattern Expressibility** for an ontology $O$ and set of axiom patterns $\mathcal{P}$ is the sum of the sizes of the axiom pattern expressibility sets for axioms in $O$ divided by the number of axioms it contains $|O|$. 

$$\text{oe}_{\mathcal{P}}(O) = \frac{1}{|O|} \sum_{\alpha \in O} |\text{ae}_{\mathcal{P}}(\alpha)|$$

The cardinality $|\text{ae}_{\mathcal{P}}(\alpha)|$ is necessary because we are now evaluating a sum of different axioms that may have different expressibilities. However, recall $|\text{ae}_{\mathcal{P}}(a)|$ is either 1 or 0 for any axiom using our patterns, so the sum of the expressibility sizes of all the axioms in an ontology is always less than or equal to the number of axioms it contains (there are no duplicates). In a future project it may be possible to expand on this definition to accommodate a more advanced set of patterns, but for this study the current notion suffices.

And now, naturally we can define averages for these measures

**Definition 5.** For a set of $n$ ontologies $O$ and set of axiom patterns $\mathcal{P}$, the **Average Ontology Pattern Expressibility**, $\overline{\text{oe}}_{\mathcal{P}}(O)$, is given by

$$\overline{\text{oe}}_{\mathcal{P}}(O) = \frac{1}{n} \sum_{O \in O} \text{oe}_{\mathcal{P}}(O)$$

and the **Average Axiom Pattern Expressibility**, $\overline{\text{ae}}_{\mathcal{P}}(O)$, is given by

$$\overline{\text{ae}}_{\mathcal{P}}(O) = \frac{\sum_{O \in O} \sum_{\alpha \in O} |\text{ae}_{\mathcal{P}}(\alpha)|}{\sum_{O \in O} |O|},$$

where $|O|$ denotes the number of axioms in $O$.

It is important to note that average ontology pattern expressibility evaluates a set of ontologies, and average axiom pattern expressibility evaluates all the axioms in a set of ontologies. Because the axiom measurement has fewer groupings it is in certain respects more precise, however we cannot produce a standard deviation since it simply has a single value.

It is also important to emphasize that an axiom can only ever match at most one of our axiom patterns with the fewest substitutions. In every case where there appear to be multiple solutions, there is always a pattern with more instantiated terms than the others. This means that the sum operations do not contain any duplicates.

### 3.4 Normalization

We have discussed a method to evaluate the axiom pattern expressibility of an ontology. However, there remains an issue that ontologies often vary radically in
the way they are syntactically expressed, even if semantically they mean similar things. Ontologies are written for completely different purposes and at differing levels of complexity; some ontologies are developed for complex reasoning applications, while others are used for more straightforward data integration. Even within a single ontology, different authors may express equivalent statements in non-equivalent ways based on personal preference or style. In order to evaluate the semantic expressibility of a large number of ontologies uniformly, we therefore need at least a minimal syntactic normalization strategy taken from community standards that allows us to compare disparate sources without biasing the evaluation in favor of any particular style. For this, we use multiple strategies derived from common OWL practices.

Our normalization begins by filtering out all axioms except class axioms, property axioms, and HasKey axioms. This is necessary because there are many OWL axioms for which our pattern study will not apply, but are semantically equivalent. Included in this are assertion (ABox) axioms, since the notion of an axiom pattern has little relevance for a fact, but also axiom types such as annotation axioms, declaration axioms, and datatype definitions, that carry no or few formal semantics. HasKey axioms are counted because they are logical axioms and not facts or annotations, but their semantics is different from class and property axioms so they are simply added to the total number of axioms and not normalized or used. The remaining axioms are then transformed according to the following procedures.

The first transformation that we perform is an equivalence transformation based on the syntactic shortcuts defined in the OWL Structural Specification [13]. Whenever an axiom is found that has one of the forms in Column 1 of Table 2, we perform the designated substitution. These substitutions are equivalent rewritings so they do not alter the semantics of the ontology. It is also possible that other simple transformations of class axioms according to the equivalences defined in the structural specification could reduce false negative matches. Thus for EquivalentClasses, DisjointClasses, and DisjointUnion we convert them to sets of SubClass axioms in the standard way.

The second transformation that we perform is obtaining negation normal form (NNF) of all class axioms in an ontology. By using the NNF we can transform all of the class axioms in an ontology into simple syntactic forms that are stripped of unnecessary information that might be due to coincidence rather than semantic equivalence.

The last transformation we apply is splitting SubClass axioms with conjunctions in the consequent, or disjunctions in the antecedent, into separate axioms. This is a standard procedure in many normalizations, and we simply replace the axiom with a set of axioms formed from the conjuncts or disjuncts whenever an axiom of this type is found. There is a special case that occurs only when the consequent is an ExactCardinality expression whose value is equal to 1. In this case, we do not use a MinCardinality 1 substitution but instead add an existential, since that is equivalent and more compact.

Property axioms include both role axioms and data property axioms.
Ontology Axiom | Substituted Axiom
--- | ---
ReflexiveObjectProperty(R) | true ⊑ ∃R.Self
IrreflexiveObjectProperty(R) | ∃R.Self ⊑ false
FunctionalObjectProperty(R) | true ⊑ ⩽1R.true
FunctionalDataProperty(S) | true ⊑ ⩽1S.true
InverseFunctionalObjectProperty(R) | true ⊑ ⩽1R−true
ObjectPropertyRange(R C) | true ⊑ ∀R.C
DataPropertyRange(S D) | true ⊑ ∀S.D
ObjectPropertyDomain(R C) | true ⊑ ∀R.true ⊑ C
DataPropertyDomain(S C) | true ⊑ ∀S.true ⊑ C

where R is a Role, S is a Data Property, C is a Class, and D is a Data Range

Table 2: Axiom Transformations

All normalizations are performed sequentially and axioms are output into a separate collection for evaluation. This compartmentalizes the data and ensures that no duplicates are created, even when a single axiom is transformed into a set of axioms.

4 Evaluation

We analyze a set of 518 ontologies from various sources, normalizing them, and testing them for semantic axiom pattern expressibility according to the principles described in the previous section. Ontologies were selected from diverse sources with unique design requirements: benchmark ontologies, Ontology Design Patterns (ODPs), ontologies extracted from Linked Open Vocabulary (LOV) [20], as well as medical domain ontologies. It would technically be possible to integrate other types of linked data in this analysis, though semantically it may not be straightforward to interpret the results if the data was mixed, so we use only OWL ontologies. Average statistics about the original ontologies gathered before normalization can be found in Table 3.

<table>
<thead>
<tr>
<th>LOV</th>
<th>Hydrography</th>
<th>Anatomy</th>
<th>Conference</th>
<th>ODP</th>
<th>Ontobee</th>
<th>Misc</th>
</tr>
</thead>
<tbody>
<tr>
<td>classes</td>
<td>20075</td>
<td>341</td>
<td>6048</td>
<td>498</td>
<td>577</td>
<td>509925</td>
</tr>
<tr>
<td>roles</td>
<td>10106</td>
<td>121</td>
<td>5</td>
<td>226</td>
<td>600</td>
<td>10453</td>
</tr>
<tr>
<td>data properties</td>
<td>8340</td>
<td>77</td>
<td>0</td>
<td>85</td>
<td>71</td>
<td>480</td>
</tr>
<tr>
<td>axioms</td>
<td>274991</td>
<td>8873</td>
<td>41407</td>
<td>3037</td>
<td>7378</td>
<td>5071892</td>
</tr>
<tr>
<td>logical axioms</td>
<td>82450</td>
<td>5463</td>
<td>16383</td>
<td>2153</td>
<td>3223</td>
<td>963880</td>
</tr>
<tr>
<td>ontologies</td>
<td>250</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>80</td>
<td>171</td>
</tr>
</tbody>
</table>

Table 3: Ontology Statistics

In this section, we report the result for all ontologies we tested, then go into details about each source, reporting a separate evaluation for each. Next we
break down the results by profile and report the numbers for those as well. In all cases, the expressibility numbers are reported for All Axioms, Class Axioms, and Simple Axioms. Our axiom patterns can only express simple class axioms, thus the values for Class Axioms and All Axioms are a reflection of the number of complex class axioms and the number of property and HasKey axioms in ontologies that, by definition, will not match our patterns. In Figures and Tables, the term “miss” is used to indicate a simple axiom that was observed but was inexpressible using our patterns. The last value we report, which is a byproduct of calculations that produce expressibility numbers, is the percent subclass and percent existential, as well as their combination. By this we mean, what percent of all of the axioms in an ontology are expressible with the subclass pattern, the existential pattern, or both. It will turn out in nearly every case that a surprisingly high proportion of most ontologies are expressible with just these two axiom patterns. OWL files and source code for the evaluation, except the gene ontology which can’t be uploaded due to size restrictions, can be found on the project GitHub page \(^3\) and the raw data can be inspected in this spreadsheet \(^4\).

### 4.1 Overall Expressibility

The average axiom expressibility and the average ontology expressibility for our simple axiom patterns over all ontologies is included in Table 4 as well as the standard deviation for the ontology expressibility.

<table>
<thead>
<tr>
<th>Axiom</th>
<th>Ontology</th>
<th>StdDev</th>
<th>Ont.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Axioms</td>
<td>82.2%</td>
<td>82.9%</td>
<td>0.206</td>
</tr>
<tr>
<td>Class Axioms</td>
<td>83.8%</td>
<td>92.9%</td>
<td>0.165</td>
</tr>
<tr>
<td>Simple Class Axioms</td>
<td>99.8%</td>
<td>96.5%</td>
<td>0.153</td>
</tr>
</tbody>
</table>

Table 4: Overall Average Expressibility

Figure 1 shows the overall percent of axioms that are expressible as each pattern for the entire collection of ontologies. Simple subclass is 54.8%, and existential is 23.9%, totaling 78.8%. This is almost the same as the axiom expressibility value for all axioms. A more detailed view of axiom type distributions can be found in the next section in Figure 2.

### 4.2 Source Expressibility

As they are all from very different domains, each source was analyzed independently from the whole. Our evaluation includes 250 OWL files that were automatically pulled from LOV using a script that can be found on the project GitHub. We obtained benchmark ontologies that are used for ontology alignment

\(^3\) https://github.com/aaronEberhart/owlax
\(^4\) https://tinyurl.com/eswc2021
evaluation. There are 4 ontologies from Hydrography, 2 from Anatomy, and 7 from the Conference domains, and each appear in their own column in Tables 5 and 6. We also obtained and evaluated 80 ODPs\cite{18}, as well as a collection of 171 OWL files that are mainly from the medical domain from the ontobee\cite{22} website. Additionally, we gathered 4 ontologies that did not fall neatly into any of these categories but nonetheless seemed better to include in the overall result than omit. These ontologies are General Formal Ontology \cite{5}, Gene Ontology \cite{4}, GeoLink base ontology \cite{11}, and The Enslaved Ontology \cite{19}, and their average is labeled Misc in the tables.

<table>
<thead>
<tr>
<th>LOV</th>
<th>Hydrography</th>
<th>Anatomy</th>
<th>Conference</th>
<th>ODP</th>
<th>Ontobee</th>
<th>Misc</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Axioms</td>
<td>78.7%</td>
<td>77.1%</td>
<td>99.9%</td>
<td>89.3%</td>
<td>85.7%</td>
<td>88.7%</td>
</tr>
<tr>
<td>Class Axioms</td>
<td>92.6%</td>
<td>84.5%</td>
<td>100%</td>
<td>96.2%</td>
<td>97.3%</td>
<td>89.9%</td>
</tr>
<tr>
<td>Simple Class Axioms</td>
<td>99.2%</td>
<td>96.5%</td>
<td>100%</td>
<td>99.2%</td>
<td>99.1%</td>
<td>99.8%</td>
</tr>
</tbody>
</table>

Table 5: Average Axiom Expressibility By Source

The Gene Ontology tends to dominate the other sources in Misc due to its extremely large size. It also contains a much higher percentage of complex class

\cite{http://ontologypedesignpatterns.org} \cite{http://ontobee.org}
axioms than any other ontology we tested, which likely accounts for the difference in Misc between simple class axioms and class axioms. In LOV there are also a considerable number of property axioms. This can explain why the expressibility is so much higher for class axioms than all axioms.

In Table 7, we see the range of percent subclass and existential among sources. Anatomy, Ontobee, and Misc all contain medical domain ontologies, which may account for the increase in percent existential if they contain more ontologies in the EL profile. LOV and Hydrography, on the other hand, are expressible with very little subclass at all, and both contain many property axioms. Except for the Anatomy Benchmarks, which is actually only two ontologies so a disproportionately small sample size, it does not appear to be the case that any sources are entirely existential and subclass. Neither are any sources completely lacking the two axiom patterns. When we break the results down by profile in the next section, things will look quite a bit different.

Figure 2 shows the actual counts of each axiom pattern used to calculate expressibility in logarithmic scale. In this chart we can see how sources like Ontobee and Misc do contain some of the less common patterns. They are just so large that smaller sources, like ODPs and benchmarks, tend to have higher percentages. The previously mentioned high percentage of complex class axioms for Misc can be seen in the third column. Also, the two ontology sources with the highest percent expressibility of subclass and existential are Anatomy and Ontobee, both medical type ontology sources. If we move farther down the chart to the less common axiom patterns, the larger ontology sources are less prevalent and now the benchmarks and ODPs start to dominate. The last two axiom patterns were never detected by our program. For the inverse qualified functional axiom pattern it is conceivable that authors rarely had occasion to write axioms like this. Disjoint classes may seem surprising, however we investigated the evaluation and found that, even though our pattern, \( A \cap B \sqsubseteq \bot \), is expressible in profiles that do not contain negation, authors are likely using Protégé or the OWLAPI to state disjoint classes axioms, which will normalize to a subclass axiom containing negation. This can cause disjoint classes to
match the subclass pattern, so it is not a false negative, but it is technically a misclassification due to our differing patterns. The program is working correctly and the reported results are correct. There is just no pattern in our set that can discriminate between a simple subclass axiom and a simple disjoint classes axiom with negation. This irregularity actually illustrates rather well how an expanded study of patterns may yield interesting insights into the semantics of pattern design.

4.3 Profile Expressibility

During the analysis we also tested each ontology to see if it was in the OWL profiles EL, QL, RL, or DL, and report the expressibility information for each profile separately. In Tables 8 and 9 we reproduce the overall results in the column Full, since all ontologies will be in OWL Full. There were 15 ontologies that could be loaded into the evaluation, but the OWLAPI could not test their profile; these ontologies were not included in the profile results. Interestingly, for the EL, QL, and RL profiles we see around a ten percent expressibility boost over the overall result. All three also have perfect expressibility for simple class axioms, and nearly perfect expressibility for all class axioms. The expressibility numbers for OWL DL are also slightly higher than the overall numbers, though significantly less so than for the other profiles.

Unlike the different sources, where the percent subclass and existential numbers were mostly near the average, we get a much more skewed result when we break the ontologies down by profile in Table 10. EL and DL ontologies seem to be expressible with a similar percent of subclass axioms as the overall result,
though EL has many more existential expressions. QL ontologies, on the other hand, are eighty percent expressible with the simple subclass pattern. And the RL profile ontologies are almost entirely expressible with simple subclass. It is no surprise, then, that EL, QL, and RL ontologies have such high expressibility numbers.

In Table 11, we mark which of our axiom patterns are expressible in each profile with an X symbol, using the OWL 2 Profiles [21] document as a reference. Comparing the different profiles with the expressibility numbers, we can see how close their sums are, even though the expressible axioms are rather different. For instance, RL expressibility is almost entirely subclass, and the existential pattern is inexpressible in that profile so the 0 value from Table 10 makes sense. EL seems to be evenly divided between subclass and existential, which again aligns with the types of statements permitted in the language. The DL profile allows all the types of expressions and it understandably has a similar result to the overall average.

For the EL profile we can also observe a unique result, because our axiom patterns have almost complete overlap with the 4 normal form class axioms defined for $\mathcal{EL}^{++}$ in [1], as shown in Table 12. The only exception is conjunction, which can only match our disjoint classes axiom pattern when the consequent is equal to $\bot$. If we were to define a conjunction axiom pattern, it might be possible to completely express this profile with simple axiom patterns. This could also be done for class axioms in the QL profile, where our axiom patterns could express all simple axioms, and could express any set of QL axioms that was normalized to remove nested quantifiers. Simple class axioms for the RL profile could be

<table>
<thead>
<tr>
<th>EL</th>
<th>QL</th>
<th>RL</th>
<th>DL</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subclass</td>
<td>52.5%</td>
<td>78.2%</td>
<td>95.1%</td>
<td>60.9%</td>
</tr>
<tr>
<td>Existential</td>
<td>46.1%</td>
<td>21.1%</td>
<td>0%</td>
<td>29.2%</td>
</tr>
<tr>
<td>Subclass + Existential</td>
<td>98.6%</td>
<td>99.3%</td>
<td>95.1%</td>
<td>90.2%</td>
</tr>
</tbody>
</table>

Table 10: Percent Subclass and Existential By Profile
expressed in much the same way as EL, missing only conjunction axioms that do not have ⊥ in the consequent. With the addition of a conjunction pattern and by normalizing nested quantifiers we could also obtain complete class axiom pattern expressibility for RL.

### Table 12: $\mathcal{EL}^+^+$ Expressibility

<table>
<thead>
<tr>
<th>Axiom Pattern</th>
<th>$\mathcal{EL}^+^+$ Normal Class Axiom</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\exists R.A \sqsubseteq B$</td>
<td>$\exists R.A \sqsubseteq B$</td>
</tr>
<tr>
<td>$\top \sqsubseteq \forall R.B$</td>
<td>$\top \sqsubseteq \forall R.B$</td>
</tr>
<tr>
<td>$A \sqsubseteq \forall R.B$</td>
<td>$A \sqsubseteq \forall R.B$</td>
</tr>
<tr>
<td>$\exists R.A \sqsubseteq B$</td>
<td>$\exists R.A \sqsubseteq B$</td>
</tr>
<tr>
<td>$\top \sqsubseteq \exists R.\top$</td>
<td>$\top \sqsubseteq \exists R.\top$</td>
</tr>
<tr>
<td>$A \sqsubseteq \exists R.\top$</td>
<td>$A \sqsubseteq \exists R.\top$</td>
</tr>
<tr>
<td>$\exists R.A \sqsubseteq B$</td>
<td>$\exists R.A \sqsubseteq B$</td>
</tr>
<tr>
<td>$A \sqsubseteq \exists R.\top$</td>
<td>$A \sqsubseteq \exists R.\top$</td>
</tr>
<tr>
<td>$\top \sqsubseteq \exists R.\top$</td>
<td>$\top \sqsubseteq \exists R.\top$</td>
</tr>
<tr>
<td>$A \sqsubseteq \exists R.\top$</td>
<td>$A \sqsubseteq \exists R.\top$</td>
</tr>
</tbody>
</table>

5 Discussion

Our motivation for this study is that we believe *simple axioms*, in general, are easier for non-logicians (for example, domain experts) to understand and utilize for modeling. Alongside improved comprehension, they come with a number of added benefits: attempting to measure the non-local effects of ontological commitments may be easier, they can be easily and automatically created by tools that allow users to specify statements in a graphical interface without a deep
technical understanding of the inner-workings of OWL, and simple axioms often do not require normalization before being input to a reasoner. To support this, we determine the current usage characteristics of axioms in existing ontologies that are semantically expressible as simple axioms, as well as how well this relates to the different OWL profiles.

To be specific, 313 of the ontologies we analyze are exclusively in the OWL Full profile, yet collectively they have an ontology expressibility above 90% for our simple axiom patterns on class axioms. There are exceptions, of course. Some ontologies have many more complex axioms than usual, such as the Gene Ontology, or consist primarily property axioms, so they are expressibility outliers. However, in general, if even these more complex ontologies are for the most part expressible as simple axioms, then efforts to improve the modeling process should start with the simple axioms.

5.1 Future Work

In the future there are many potential next steps that could build on this study. One interesting approach would be to test different sets of simple axiom patterns and see how the expressibility numbers compare between them. OWLAX was a good basis to create an initial set of simple axiom patterns but there are some obvious common ones that it lacks, for instance conjunction, disjunction, negation, as well as multiple variations on cardinality and property axioms. For the current study we only use simple axioms because there is no clearly defined way to categorize complex axioms, which can be arbitrarily large. However, it may also be interesting to analyze the complex axioms to see if there are any new patterns that can be included.

We also admit that our definition of expressibility is quite simple, intentionally kept this way for clarity. However it may be possible with some more comprehensive statistical tools that a better understanding of semantic axiom expressibility in ontologies is possible. In a future study we may look into different evaluations besides expressibility, perhaps it will be informative to compare.

Our method did normalize many axioms, however it is likely that complex axioms existed in the ontologies we studied that could have been normalized but weren’t because our method only obtained NNF and then split up appropriate conjunction and disjunction axioms. By introducing new terms in the normalization to syntactically split some expressions we might be able to even further increase the expressibility detection capability. Though, as previously mentioned, this would require the addition of new terms, so it would be equivalent but would also contain more entities, so the comparison would be less obviously appropriate.

6 Conclusion

In this paper we demonstrate that most class axioms in OWL ontologies are expressible with a small set of simple axiom patterns. This has implications for
how we can approach ontology management and development. If ontologies are mostly semantically simple then focusing on supporting and explaining these types of axioms can lead to easier adoption and maintenance. Complex axioms will of course always be a part of OWL, but we can improve our ontologies most easily by first making sure that the simple axioms are well understood and used correctly.

Acknowledgement The authors acknowledge partial support from the financial assistance award 70NANB19H094 from U.S. Department of Commerce, National Institute of Standards and Technology and partial support from the National Science Foundation under Grant No. 1936677.

References


