Ontology Stratification Methods: a Comparative Study*

Ricardo Guimarães[1], Uli Sattler[2], and Renata Wassermann[1]

[1] University of São Paulo, São Paulo-SP, Brazil {ricardof, renata}@ime.usp.br
[2] University of Manchester, Manchester, UK uli.sattler@manchester.ac.uk

Abstract. Comparing the importance of axioms in ontologies is an essential task in a variety of applications such as ontology repair and inconsistency management. It guides the choice of axioms to remain in the ontology or which should prevail when a conflict arises. While evaluating the importance of an axiom is difficult, there are different approaches to stratify ontologies by criteria that act as proxies for importance. However, there is little work about how these methods are related and their adequacy considering different applications. In this work, we evaluate specificity and modularity-based stratification approaches, and work towards understanding how they relate. We compare empirically the result of five stratification methods over a corpus of real-world ontologies. In particular, we investigate correlations between their rankings and their ability to distinguish axioms.

Keywords: Stratification · Ontology modularity · Ontology evolution.

1 Introduction

When managing knowledge bases, users and designers are often confronted with the need to choose among conflicting pieces of information. While the most reliable solution may be to delegate the choice to a domain specialist with information about the particular case at hand, this is not always possible, and thus many approaches propose the use of stratification methods. Stratified or prioritized knowledge bases are such that each axiom has a value that denotes its importance.

In the absence of further information, an axiom's importance is closely related to its generality (or specificity), where none of these notions has a commonly accepted definition for complex axioms. Being able to stratify an ontology by the importance of its axioms, will enable automatic repair algorithms [4] or inconsistent-tolerant [14] reasoning algorithms to choose to preserve more important information and ignore/drop less important one. a pre-determined criterion.

The idea of stratification or prioritization to solve choice problems in knowledge bases is not new. It has appeared in repair and evolution of databases [9],

* The first author has been partially supported by CAPES through grant 001 and partially by FAPESP through grants 2017/04410-0 and 2018/00087-2.

of classical knowledge bases [12] and of ontologies [14, 4]. In Belief Revision, the area of Knowledge Representation that studies change over logical theories, Nebel [12] claims that its impossible to completely define a formal revision operation without resorting to some preference ordering. In the context of default reasoning, Baader and Hollunder [1] discuss and propose approaches to prioritization in terminologies to encode the preference for more specific knowledge when dealing with inconsistencies.

In this paper we consider OWL 2 DL ontologies [7] and regard them as finite sets of axioms in $\mathcal{SROIQ}(\mathcal{D})$. While some approaches to prioritize axioms in these ontologies have been proposed [1, 14, 5, 10], to the best of our knowledge, there is no study that compares stratification techniques. Moreover, there is no work regarding how to measure the quality of a stratification procedure.

The contributions described in this paper are two-fold: firstly, we devise metrics to measure the ability of these methods to distinguish between the axioms of an ontology and suitable statistical methods to measure correlation between these approaches. Secondly, we apply this framework to five different stratification methods and a corpus of existing ontologies, thereby gaining novel insights into their differences, strengths and weaknesses.

Of the five methods selected for this study, two encode specificity and three rely on the modular structure of the ontologies. Besides measuring correlation and axiom dispersion of these methods, we also discuss other aspects such as restrictions to applicability, feasibility and significance of the rankings produced. While we do not validate or compare directly proxies for importance, such as, specificity, generality and module preservation we will often refer to them to understand the stratification methods considered. It turns out that the methods perform very differently according to our metrics, and there are some interesting correlations between some stratification methods.

This paper is accompanied by a technical report with more details about the implementation of the experiments and additional results [11].

2 Preliminaries

Before detailing the methods we evaluate, we present in this section some fundamental concepts and establish conventions used through this work. We also assume the reader to be familiar with description logics and its usual notation (for more details we refer the reader to [2], for example).

We call a set of terms (concept, role and individual names) a *signature*. Moreover, if $\mathcal{O}$ is an ontology, its signature ($\mathrm{sig}(\mathcal{O})$) is the set of terms occurring in $\mathcal{O}$. We extend the notion and definition for axioms and sets of axioms.

Since we consider ontologies here as sets of formulas (bases) in a description logic, we follow the approach adopted by Benferhat et al. [4] (but considering all axioms and not only assertions) and define a *stratified ontology* $\mathcal{O}$ to be a partitioned set of axioms: $\mathcal{O} = \{\mathcal{O}_i\}_{0 \leq i \leq n}$ (with $\mathcal{O}_i \cap \mathcal{O}_j = \emptyset$, if $i \neq j$). We decide by convention that the most important axioms are those at the strata with the lowest index, that is, if $i < j$, then the axioms in $\mathcal{O}_i$ are more important than (or preferred over) the axioms in $\mathcal{O}_j$. Moreover, if an axiom α is at the stratum $\mathcal{O}_i$, we say that it has rank i ($\mathrm{rank}(\alpha) = i$). Through this paper we use the notion of ranks and partitions interchangeably.

The stratification approaches we discuss in Section 3 are induced by a partial order of the axioms. Formally, if S is a set and $\prec$ a partial order on S then, the *stratification of S induced by $\prec$* is such that, for each $a \in S$:

$$\mathrm{rank}(a) = \begin{cases} 0, & \text{if } a \text{ is minimal w.r.t. } \prec \\ 1 + \max_{\{b \in \mathcal{O} \mid b \prec a\}} \mathrm{rank}(b), & \text{otherwise} \end{cases}$$

2.1 Modularity

In this subsection, we briefly discuss the modularity approach underlying the stratification methods that we selected for comparison.

The syntactic locality-based modules (LBMs) [6] constitute a modularity approach which not only has crucial properties for reuse, but also is polynomially extractable. These characteristics, together with its availability through the OWL API[3] explain its current popularity.

Given a signature Σ, the locality-based module of an ontology $\mathcal{O}$ for the signature Σ is a subset of the ontology that has the same entailments as the ontology over Σ and all other terms that compose the module's signature. There are two basic flavours of LBMs: $\top$-locality and $\bot$-locality. They differ in the syntactical rules applied to decide whether an axiom belongs to a module (due to space constrains, we refer the reader to [15] for the actual definition). For $x \in \{\bot, \top\}$, we denote the x-locality-based module of the ontology $\mathcal{O}$ for the signature Σ by x-$\mathrm{LBM}(\mathcal{O}, \Sigma)$.

The atomic decomposition (or simply AD) of an ontology represents the set of all locality-based modules of an ontology. The AD of an ontology consists of a partitioning of the axioms (each partition called *atom*), according to their co-occurrence in LBMs and a dependency relation between these, that is, a partial order indicating that whenever an atom is a subset of a module, so are all atoms that it depends on. As with LBMs, we denote the atomic decomposition of the ontology $\mathcal{O}$ for the locality notion $x \in \{\bot, \top\}$ by x-$\mathrm{AD}(\mathcal{O})$. Each atom in the AD, together with the atoms it depends on, forms a *genuine* module, that is, a locality-based module that is minimal w.r.t. set inclusion [13]. To clarify the definition of the atomic decomposition, consider the illustrative ontology `KoalaS`, a modified version of the ontology `Koala` [8], shown in the first column of Table 1 (the whole table will be discussed in Section 3) and its atomic decompositions in Figures 1a and 1b.

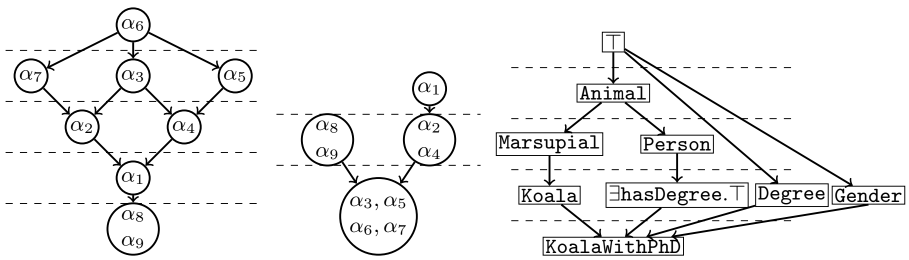
3 Stratification Methods

In this section, we present the stratification methods selected for comparison. We choose these methods because they only use the input ontology as parameter and can be computed without human assistance.

[3] `http://owlcs.github.io/owlapi/`

Table 1: The `KoalaS` ontology and five distinct rankings of its axioms

Axiom	AxS	CNS	BAH	BAD	TAH
α_1 : $\mathtt{Animal} \sqsubseteq \exists\mathtt{hasGender.Gender}$	4	1	1	3	2
α_2 : $\mathtt{Person} \sqsubseteq \mathtt{Animal}$	3	1	2	2	1
α_3 : $\mathtt{Person} \sqcap \mathtt{Marsupial} \sqsubseteq \bot$	1	1	3	1	0
α_4 : $\mathtt{Marsupial} \sqsubseteq \mathtt{Animal}$	3	2	2	2	1
α_5 : $\mathtt{Koala} \sqsubseteq \mathtt{Marsupial}$	2	1	3	1	0
α_6 : $\mathtt{KoalaWithPhD} \equiv \mathtt{Koala} \sqcap \exists\mathtt{hasDegree}.\{PhD\}$	1	0	4	0	0
α_7 : $\exists\mathtt{hasDegree}.\top \sqsubseteq \mathtt{Person}$	2	1	3	1	0
α_8 : $\mathtt{Degree}(MS)$	0	1	0	4	1
α_9 : $\mathtt{Degree}(PhD)$	0	1	0	4	1



(a) ⊥-AD of `KoalaS` (b) ⊤-AD of `KoalaS` (c) Inferred concept hierarchy

Fig. 1: Atomic decompositions and inferred concept hierarchy of `KoalaS`

3.1 Axiom Specificity

Axiom specificity (AxS) was described by Qi and Pan [14], based on the work of Benferhat, Baida and Cuppens [3] that used this notion of importance in access-control applications. A similar notion was applied in default logic terminologies [1]. This criterion considers the most specific axioms as the most important ones. According to the definition in [14] $\alpha_1 = \mathtt{C}_1 \sqsubseteq \mathtt{D}_1$ is more specific than $\alpha_2 = \mathtt{C}_2 \sqsubseteq \mathtt{D}_2$ (written $\alpha_1 \prec_{\text{AxS}} \alpha_2$) in $\mathcal{O}$ if $\mathcal{O} \models \mathtt{C}_1 \sqsubseteq \mathtt{C}_2$, but $\mathcal{O} \not\models \mathtt{C}_2 \sqsubseteq \mathtt{C}_1$. Therefore, AxS is the stratification of axioms induced by $\prec_{\text{AxS}}$.

We extend the definition of of specificity ($\prec_{\text{AxS}}$) in [14] to other types of axioms, such as role (chain) inclusions for example, using an analogous methodology for comparison (e.g. check inclusion between the left-hand sides of two axioms). ABox axioms (concept and role assertions) are minimal axioms, as suggested by Cóbe [5]. That is, given an ontology $\mathcal{O}$, if α is a class or role assertion, then α is minimal w.r.t. $\prec_{\text{AxS}}$. Excluding relations with ABox axioms, axioms from different types are incomparable.

Given the differences between $\mathcal{SROIQ}(\mathcal{D})$ axioms and their OWL 2 DL equivalents, in our implementation we handled each axiom type in a straightforward way, more details are included in the technical report [11].

This method is syntax dependent, i.e., two ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$ can have equivalent axioms (for every $\alpha_1 \in \mathcal{O}_1$ there exists $\alpha_2 \in \mathcal{O}_2$, such that $\alpha_1 \equiv \alpha_2$, and vice versa), and yet have distinct stratifications. It is possible to change the rank of an axiom by rewriting it in another way. Consider the axiom $\mathtt{A} \sqcap \exists\mathtt{r}.\mathtt{B} \sqsubseteq \mathtt{X}$, a simple *absorption*, a common optimization in reasoners, would rewrite it equivalently as $\mathtt{A} \sqsubseteq \forall\mathtt{r}.\neg\mathtt{B} \sqcup \mathtt{X}$ that is (usually) less specific than its original version. Moreover, this criterion ignores the right-hand side of axioms, hence in the ontology $\mathcal{O} = \{\mathtt{A} \sqsubseteq \mathtt{C}_1, \mathtt{B} \sqsubseteq \mathtt{C}_2, \mathtt{C}_1 \sqsubseteq \mathtt{C}_2\}$, the axioms $\mathtt{A} \sqsubseteq \mathtt{C}_1$ and $\mathtt{B} \sqsubseteq \mathtt{C}_2$ would be placed in the same stratum, even though we could say that the first one is more specific than the because since $\mathtt{A}$ and $\mathtt{C}_1$ are subconcepts of $\mathtt{C}_2$ [5].

The dependency on classification to compute AxS also interferes with their application on inconsistent ontologies, as the hierarchies collapse. But still, it can be adapted to such cases [14].

3.2 Concept Name Specificity

We also define an alternative to AxS. First, we establish a partial order of concepts: $\mathtt{C}$ is more specific than $\mathtt{D}$ ($\mathtt{C} \prec_{\sqsubseteq} \mathtt{D}$) in $\mathcal{O}$ if $\mathcal{O} \models \mathtt{C} \sqsubseteq \mathtt{D}$, but $\mathcal{O} \not\models \mathtt{D} \sqsubseteq \mathtt{C}$. Let $\mathrm{rank}(\mathtt{C})$ be the rank of the concept name $\mathtt{C}$ for all concept names in $\mathcal{O}$ in the stratification obtained from $\prec_{\sqsubseteq}$, then we say that α_1 is more *concept-specific* than α_2 (written $\alpha_1 \prec_{\mathrm{CNS}} \alpha_2$) in $\mathcal{O}$ if $\min_{\mathtt{C} \in \mathrm{sig}(\alpha_1)} \mathrm{rank}(\mathtt{C}) < \min_{\mathtt{C} \in \mathrm{sig}(\alpha_2)} \mathrm{rank}(\mathtt{C})$. The *concept name specificity* (CNS) is the stratification of axioms induced by $\prec_{\mathrm{CNS}}$.

Even though this approach does not display the same syntax sensitivity as axiom specificity for absorption, it fails to capture differences between any pair of axioms whose most specific concepts are at the same level of specificity. Moreover, it also has the same issues when applied to inconsistent ontologies as AxS. This dependency on the hierarchy of concept names also implies that the axioms without concepts (such as role inclusions) cannot be ranked by this method.

One notable way in which it differs from AxS is the handling of the assertions: instead of considering all of them equally important, it also considers concepts involved. For instance if we have $\mathcal{O} = \{\mathtt{A} \sqsubseteq \mathtt{B}, \mathtt{A}(a), \mathtt{B}(b)\}$, the axioms $\mathtt{A}(a)$ and $\mathtt{B}(b)$ would attain the same rank by axiom specificity, however CNS would tell us that $\mathtt{A}(a)$ is more specific than $\mathtt{B}(b)$.

3.3 *x*-AD Height

The last stratification method we consider, proposed by Guimarães and Wassermann [10], is based on the atomic decomposition (discussed in Subsection 2.1). We use the partial order determined by the dependency relation in the x-AD of $\mathcal{O}$: α_1 is *x-preferred* to α_2 (written $\alpha_1 \prec_{x\text{-AD}} \alpha_2$) in $\mathcal{O}$ if $\mathfrak{a}_1 \prec \mathfrak{a}_2$ in x-AD($\mathcal{O}$), where $\mathfrak{a}_i$ is the atom that contains the axiom α_i (since each axiom belong to exactly one atom, this relation is well-defined). The *x-AD height* is the stratification of axioms induced by $\prec_{x\text{-AD}}$. To shorten the notation, we will refer to $\bot$-AD height as BAH and $\top$-AD height as TAH.

In [10], the authors focus on $\top\bot^*$-locality (other type of locality) to define height, since it produces smaller modules than $\top$ or $\bot$. However, we restrict ourselves to $\top$ and $\bot$-locality, since we are more interested in smaller atoms and good dependency relations than in small modules. Compared to $\top\bot^*$-locality,

$\top$ and $\bot$-locality preserve more information about class hierarchies that seem to be beneficial for specificity/generality and importance analysis [8] (e.g. the $\top\bot^*$-AD of a pure taxonomy has only unconnected atoms). In particular, for $\bot$-locality we have that the module preserves all superclass relationships for the concepts in its signature. In this way, axioms from lower atoms in the $\bot$-AD tend to deal with more general concepts, while higher deal with more specific ones, and the opposite happens with $\top$-AD. Hence, we expect that BAH displays a strong negative correlation with specificity-based metrics, while TAH displays a positive one. To get a clearer picture, we include also the $\bot$-*AD depth* (BAD) in our investigation; it is defined analogously to $\bot$-AD height, but replacing $\prec_{\bot\text{-AD}}$ with $\succ_{\bot\text{-AD}}$ in the definition of rank.

The height is a way to estimate which axioms participate of more or fewer modules. An axiom with low height is more likely to be part of more genuine modules and thus also of more general modules, via the dependency relations [10]. Proposition 1 considers the case of removal of an axiom and the question whether modules can be preserved and shows that this metric acts as proxy for importance, if one intends to preserve as many modules as possible.

Proposition 1. *Let $\mathcal{O}$ be an ontology, $x \in \{\bot, \top\}$ a locality notion, $\alpha \in \mathcal{O}$ an axiom and $\Sigma \subseteq \mathrm{sig}(\mathcal{O})$ a signature such that $\alpha \notin x\text{-LBM}(\mathcal{O}, \Sigma)$. Then, $x\text{-LBM}(\mathcal{O} \setminus \{\alpha\}, \Sigma) = x\text{-LBM}(\mathcal{O}, \Sigma)$.*

This proposition is an immediate consequence of the fact that x-locality w.r.t. Σ (and any of its extensions) during the computation of x-LBMs, is decided on a "per axiom" basis. As a consequence of Proposition 1, if $\alpha \prec_{x\text{-AD}} \beta$ then the removal of α affects more x-LBMs than the removal of β. Therefore, axioms with lower heights can be said to be more influential, thus more important.

As the locality-based modules and atomic decomposition are syntax-dependent, so are the x-AD heights and depths of the axioms. Therefore, all approaches we consider are syntax-dependent: AxS by ignoring the right-hand sides, CNS by considering only at the most specific concept in the class signature and the ones based on the atomic decomposition due to the definition of syntactic locality.

To conclude this section, we compare the selected stratification methods illustratively with Example 1, which uses the ontology `KoalaS` from Table 1 and that is inspired by the ontology repair scenario.

Example 1. Consider the ontology `KoalaS` shown in Table 1 and suppose that we would like to restore coherence (the concept `KoalaWithPhD` is unsatisfiable) by removing a minimal set of axioms. Therefore, the candidates for removal are: α_3, α_5, α_6 and α_7. Removing any of them is enough to make the ontology coherent. TAH is unable to distinguish between them. Instead, if we follow axiom specificity, we would either remove α_5 or α_7, using CNS or BAD we would remove any of them, except α_6, as `KoalaWithPhD` becomes the most specific concept due to its unsatisfiability. In contrast, since BAH is more likely to favour more general axioms (see Figure 1), it would tell us to remove (or fix) α_6.

In Table 1 and Figure 1 we also observe that BAH and AxS are mostly inverses of each other as we expected, if we ignore the concept assertions α_8 and

α_9. Also, BAD is very similar to AxS and CNS. In Section 5, we will see whether these correlations hold in general.

4 Methodology

Our first methodological step is to formulate the research questions we aim to answer. The first question is due to the fact that all stratification methods produce partial orders and thus, do not guarantee that all the axioms have different ranks. It consists in determining how well these methods separate the axioms, i.e., how likely a stratification approach is to help us pick a suitable axiom for a certain task. The second, is to verify if these methods are correlated.

Regarding dispersion, for each method, we are interested in measuring the number of strata in the whole corpus, their relative size and how many pairs of equivalent axioms exist after stratification. The number of strata and their distribution by relative size (the percentage of axioms from the original ontology inside a stratum) show how often small and large strata occur for each method. The *indistinguishable axioms rate* (IAR) of a stratified ontology is the proportion of pairs of axioms in the ontology that have axioms with the same rank, i.e., it measures how far the ranking is from a strict total order (the 0 IAR situation).

In the case of the ontology `KoalaS` from Table 1, we have for $\bot$-AD height: 5 strata, with the smaller strata with approximately 11.11% of the axioms and the largest with around 33.33% of the axioms, and an IAR of roughly 0.1389.

For the second question, we test correlation between methods in two ways. First, we look at each ontology and measure the correlation between the rankings in each. Second, we normalize the ranks of the axioms in each ontology to the $[0, 1]$ interval, join the results for the whole corpus and only then measure correlation. We call them *local correlation* and *corpus correlation*, respectively.

We choose Spearman's rank-order coefficient to measure correlation. This is a metric of monotonic association between two rankings that does not make assumptions about the distribution of the input data (i.e. is non-parametric). Its value ranges from -1 to 1 and expresses whether the association is direct or inverse. Values close 0 indicate weak or absent correlation. We will also refer to the p-values associated to the correlation measurements. The p-value indicates the probability of obtaining the same coefficient given that there is no monotonic association between the rankings, the null hypothesis for the Spearman's test. Smaller p-values mean that we have more grounds to reject the null hypothesis, thus the correlation is more statistically significant.

4.1 Experiment Design

We want to answer our two research questions by looking at the stratification over a corpus of realistic ontologies. Moreover, as we rely on reasoning to compute AxS and CNS, and on current module extraction algorithms, we only handle legal OWL 2 DL ontologies.

Among the possible corpus options, the set of ontologies hosted by Bioportal[4] stands out. This repository hosts ontologies from the biomedical domain, many of them used in practical applications. The corpus of ontologies used in this work is a subset of 272 ontologies out of the 438 from the Bioportal snapshot at 30/03/2017[5]. For brevity, we omit here other details about the corpus and experiments which are available with additional plots, tables and discussion regarding influence of expressive power in the technical report [11].

5 Results

5.1 Axiom Dispersion

We start our analysis of dispersion by looking at the cumulative histogram in Figure 2. For each method there is a line that can be read as the proportion of strata it produced that have size up to a certain fraction of the original ontology. Each bin corresponds to increments of 10% of relative size. In addition, since each method produces a different number of strata, the total number of strata in the corpus is shown as the N near the method label in the legend.

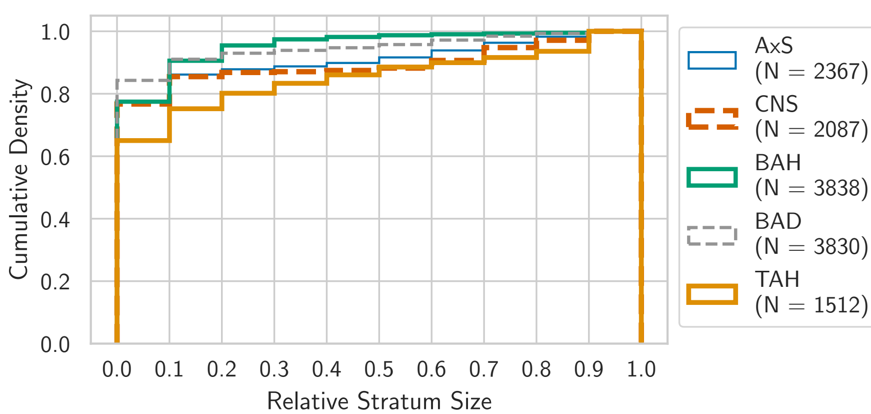


Fig. 2: Cumulative density histograms of stratum relative sizes over the corpus.

For all methods, at least 60% of strata produced has size up to 10% of their original ontologies. This is a good indicator that they produce useful rankings. We notice that BAH and BAD are clear winners regarding the total number

[4] https://bioportal.bioontology.org/

[5] http://doi.org/10.5281/zenodo.439510

of strata generated[6], while TAH is notably the worst method in this aspect. Furthermore, BAH and BAD have the largest proportions of strata below the 20% mark and are the ones which gain the least as the relative size increases: the proportion of their strata with size 80% or more of the original ontology is negligible. In contrast, TAH has the largest proportions of bigger strata, almost 10% of the strata produced by this method has relative size of 90% or more.

The specificity-based approaches not only generated similar number of strata, but also resemble each other in the distribution of strata by relative size, as we can observe by the closeness between their histograms' curves and their trends.

In the technical report [11], we repeated our studies regarding axiom dispersion and correlations diving the ontologies according to expressive power. The major difference from what we see in Figure 2 is that BAH and BAD perform extremely well in "rich" ontologies, i.e., on the ontologies that cannot be expressed in RDFS (RDFS only allows concept inclusions, domain and range restrictions on roles, and assertions). In these "rich" ontologies, BAH and BAD have largest proportion of strata with relative size of 20% or less, and practically no strata with relative size of 80% or more. The reasons for this behaviour, as well as deeper analysis of expressive power in other metrics are part of our future work.

As discussed in Section 4, the IAR is one of the most important characteristics that we evaluate in this work, as it gives us a clear indication of how likely it is that each method improves our ability to choose between two axioms. Figure 3 displays one violin plot for each method, their external shape (the "violin") is the estimated probability density function (using kernel density estimation), cut at the range of the observations. The width of the violin indicates the probability of an ontology having the given IAR value, considering the sample distribution. The internal elements are boxplots marking from the centre to the extremities median, interquartile range and 95% confidence interval.

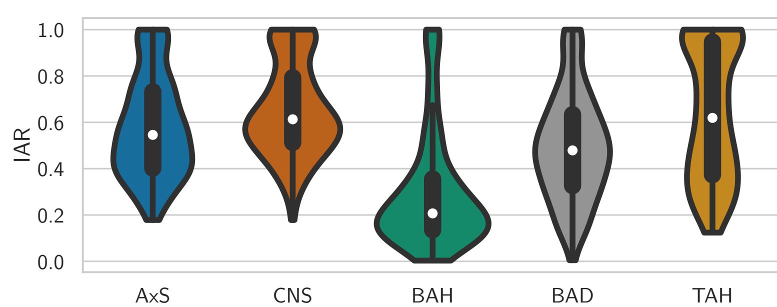


Fig. 3: Distribution of the indistinguishable axioms rates.

[6] The small different occurs because in atomic decompositions where the distance between the nodes at the top and the bottom is not unique (it depends on the pair of nodes), depth and height produce different number of strata.

From Figure 3, we can infer that the ⊥-AD height is the best approach regarding IAR, with more than a quarter of the 272 ontologies having IAR of less than 0.2. Additionally, the IAR of BAH in at least 75% of the corpus is below 0.5 and the probability density decreases rapidly at higher IAR values. While ⊥-AD depth also has 0 IAR in some cases, its distribution is centred on the median, which is close to 0.5 IAR, and it has also attained 1 IAR in the corpus. Excluding BAH, all methods reach 0.4 IAR in at least 50% of the ontologies.

As expected, AxS and CNS have similar IAR distributions. Also, we observe a small dominance of AxS over CNS, as the median and quartiles are marked at lower IARs. Still, given the proximity of both distributions and the higher variability of AxS when compared to CNS, we are not able to decide which one is better than the other from IAR alone. Neither specificity-based approaches nor ⊤-AD height were able to produce any ranking with IAR smaller than 0.1, this is shown by the bottom limits of their (cut) violin plots.

5.2 Correlation

We start our correlation analysis with the local version (as defined in Section 4). Figure 4 shows violin plots (similar to ones presented in Figure 3). Each violin plot depicts the distribution of the Spearman's correlation coefficient between two metrics for each ontology in the corpus. To avoid counting results with low significance, we consider only instances of correlation for which the p-value was below 0.05, i.e., we only count situations whose probability of obtaining the same Spearman's coefficient under the null hypothesis was less than 5%. The value of N in the legend indicates, for each method, how many rankings out of the 272 for passed this criterion, for each pair of methods.

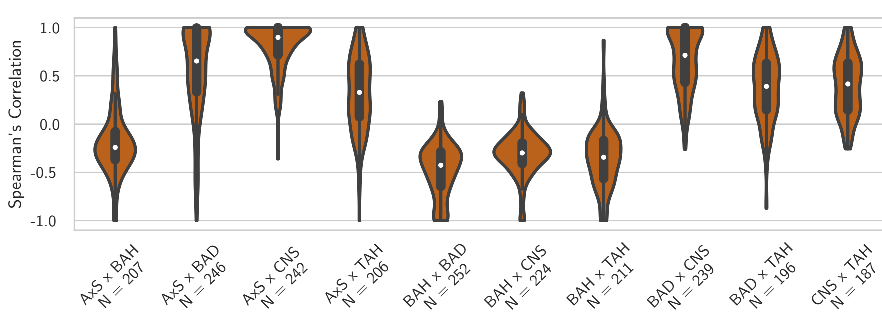


Fig. 4: Distribution of correlation between the methods (for p-value < 0.05).

As anticipated, CNS often has a very strong positive correlation with AxS and BAD, while AxS and BAD usually display a weaker, but considerable positive correlation. Generally, TAH correlates positively with CNS, AxS and BAD, but weakly. The correlations involving BAH are typically negative and weak, even though we would expect it to display stronger negative associations with BAD and TAH. We are confident with these results since in all cases most ontologies displayed a significant correlation (p-value < 0.05).

Table 2: Corpus correlation values between stratification methods.

Methods	AxS × BAH	AxS × BAD	AxS × CNS	AxS × TAH	BAH × BAD	BAH × CNS	BAH × TAH	BAD × CNS	BAD × TAH	CNS × TAH
Spearman's	-0.1847	0.3108	0.6110	-0.1080	-0.3372	-0.1646	-0.0120	0.4383	0.0633	0.0532

Now we direct our attention to the corpus correlation (Table 2). We remind that we normalized each axiom rank to the interval $[0, 1]$, took the union of all rankings and only then calculated the associations. For brevity, we omit the p-values because they were sufficiently small (the largest at the order of 10^{-16}).

We can observe in Table 2 some similar relationships to those seen in Figure 4. For example, the strong positive correlation between AxS and CNS and the average negative one between BAH and BAD. Also, there is a noticeable positive association between BAD and both AxS and CNS. The other interactions have values of correlation too close to zero to indicate any consistent association.

All results in this section reflect the behaviour of the methods in a particular corpus. It may be the case that the overall results differ if we consider ontologies from a different domain, or include more ontologies from the same corpus.

6 Conclusion and Future Work

In this work we discussed desirable traits for stratification procedures and evaluated five distinct methods according to these criteria. We focused on axiom dispersion and correlation, devised suitable metrics to evaluate the methods and discussed the results obtained. The methodology we followed gives insight into a crucial component for automatic debug and repair of ontologies.

We observed that AxS and CNS displayed similar behaviour and significant positive association following our expectations. However, the same cannot be said about BAH and TAH. Besides the low correlation, their behaviour is very different regarding axiom dispersion: $\bot$-AD height being one of the most successful according to our metrics, and $\top$-AD height the worst in number of strata and relative size distribution. Interestingly, $\bot$-AD depth had similar behaviour regarding dispersion as $\bot$-AD height, but displayed a significant correlation with the specificity-based methods.

As future work we propose the extension of this study covering other stratification approaches, including other metrics related to the atomic decomposition and other definitions of specificity. Moreover, other characteristics of the stratification methods such as ease of interpretation and computational performance also need to be studied in detail. Such extended experiments will also include a larger and more diversified corpus and analyse whether the expressive power (i.e. the OWL 2 DL fragment considered) affects performance and correlation.

Furthermore, the ability to distinguish axioms, feasibility and ideal of importance can vary according to the application, thus evaluating the effects of a range of stratification approaches in a particular setting, such as ontology repair, is another possible direction. In this sense, some applications may also provide a way to compare stratification methods which depend on other parameters besides the ontology itself (such as syntactic connectivity [5, 10]).

References

1. Baader, F., Hollunder, B.: Priorities on defaults with prerequisites, and their application in treating specificity in terminological default logic. J. of Automated Reasoning **15**(1), 41–68 (1995)
2. Baader, F., Horrocks, I., Lutz, C., Sattler, U.: An Introduction to Description Logic. Cambridge University Press (2017)
3. Benferhat, S., Baida, R.E., Cuppens, F.: A stratification-based approach for handling conflicts in access control. In: 8^{th} ACM Symposium on Access Control Models and Technologies. pp. 189–195 (2003)
4. Benferhat, S., Bouraoui, Z., Papini, O., Würbel, E.: Prioritized assertional-based removed sets revision of DL-Lite belief bases. Annals of Mathematics and Artificial Intelligence **79**(1-3), 45–75 (2017)
5. Cóbe, R.: Integração entre múltiplas ontologias: reúso e gerência de conflitos. Ph.D. thesis, Universidade de São Paulo (2014)
6. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular Reuse of Ontologies: Theory and Practice. J. Artificial Intelligence Research **31**, 273–318 (2008)
7. Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P.F., Sattler, U.: OWL 2: The next step for OWL. J. of Web Semantics **6**(4), 309–322 (2008)
8. Del Vescovo, C.: The Modular Structure of an Ontology: Atomic Decomposition and its Applications. Ph.D. thesis, University of Manchester (2013)
9. Fagin, R., Ullman, J.D., Vardi, M.Y.: On the Semantics of Updates in Databases. In: ACM Symposium on Principles of Database Systems. pp. 352–365 (1983)
10. Guimarães, R., Wassermann, R.: Local Change in Ontologies with Atomic Decomposition. In: 3^{rd} Joint Ontology Workshops (JOWO). CEUR, vol. 2050 (2017)
11. Guimarães, R., Sattler, U., Wassermann, R.: Ontology stratification: a comparative study. Tech. rep., University of São Paulo, University of Manchester (2018), `https://www.ime.usp.br/~ricardof/womocoe2018_tr.pdf`
12. Nebel, B.: Syntax-Based Approaches to Belief Revision. In: Belief Revision. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press (1992)
13. Parsia, B., Schneider, T.: The Modular Structure of an Ontology: An Empirical Study. In: 12^{th} International Conference on Principles of Knowledge Representation and Reasoning (KR). AAAI Press (2010)
14. Qi, G., Pan, J.Z.: A stratification-based approach for inconsistency handling in description logics. In: International Workshop on Ontology Dynamics-IWOD (2007)
15. Sattler, U., Schneider, T., Zakharyaschev, M.: Which Kind of Module Should I Extract? In: Proceedings of the 22^{nd} International Workshop on Description Logics (DL 2009), Oxford, UK, July 27-30, 2009. CEUR Workshop Proceedings, vol. 477. CEUR-WS.org (2009)