Declutter Your Justifications: Determining Similarity Between OWL Explanations

Samantha Bail, Bijan Parsia, Ulrike Sattler

The University of Manchester
Oxford Road, Manchester, M13 9PL
{bails,bparsia,sattler@cs.man.ac.uk}

Abstract. Given the high expressivity of the Web Ontology Language OWL 2, there is a potential for great diversity in the logical content of OWL ontologies. The fact that many naturally occurring entailments of such ontologies have multiple justifications indicates that ontologies often overdetermine their consequences, suggesting a diversity in supporting reasons. On closer inspection, however, we often find that justifications – even for multiple entailments – appear to be structurally similar, suggesting that their multiplicity might be due to diverse *material*, not *formal* grounds for an entailment.
In this paper, we introduce and explore several equivalence relations over justifications for entailments of OWL ontologies which partition a set of justifications into structurally similar subsets. These equivalence relations range from strict isomorphism to looser notions of similarity, covering justifications which contain different class expressions, or even different numbers of axioms. We present the results of a survey of 83 ontologies from the bio-medical domain, showing that OWL ontologies used in practice often contain large numbers of structurally similar justifications.

1 Introduction

Justifications, minimal entailing subsets of an OWL[1] ontology, provide helpful and easy-to-understand explanation support when repairing unwanted entailments in the ontology debugging process. They are currently the prevalent form of explanation in OWL ontology editors such as Protégé 4. While we have some knowledge of how individual justifications can be made easier to understand for human users, e.g. [8,11], we have yet to gain more insights into user interaction with *multiple justifications*. An entailment of an OWL ontology can have a large number of justifications (potentially exponential in the number of axioms in the ontology [3]), with up to several hundreds found in large real-life ontologies [4]. In order to achieve a *minimal repair*, i.e. a modification of the ontology which removes unwanted entailments without losing relevant information, it is often beneficial to consider not only a single, but multiple entailments simultaneously.

[1] We will use the term *OWL* to refer to both OWL and OWL 2 ontologies.

When encountering justifications for a finite set of entailments of an ontology (e.g. unwanted atomic subsumptions, or unsatisfiable classes), we are often faced with a seemingly large and diverse body of reasons why the entailments hold. *Root and derived* justifications [13,14] address this issue by pointing out those justifications which are subsets of others; fixing such a subset (*root*) justifications first will also repair those justifications which are *derived* from it. While proven to be helpful, root and derived justifications are restricted to a very specific kind of relation between justifications. Due to a lack of other suitable interaction strategies, large numbers of multiple justifications may still present themselves to a user as an unordered and often unmanageable list of axiom sets.

On closer inspection, however, we frequently find that sets of justifications are very similar, and often even contain structurally identical axioms, with only class, property, and individual names diverging. Pointing out these similarities and grouping justifications based on their shared structures might greatly assist a user in coping with multiple justifications: Rather than trying to understand each individual material justification, the user can focus on understanding the formal *template* of a particular subset of justifications. Potentially, a user might have to deal with far fewer justifications, thus having a significantly reduced effort when repairing an ontology. This raises two questions: First, how do we determine whether two justifications are structurally similar, and second, how prevalent are such similarities in ontologies used in practice?

A well-known syntactical equivalence relation in OWL is *structural equivalence*. The OWL Structural Specification[2] states the condition for two OWL objects (named classes, properties, or individuals, complex expressions, or OWL axioms) to be structurally equivalent. In short, it defines the objects to be equivalent if they contain the same complex expressions, using identical entity names and constructors, regardless of ordering and repetition (in an unordered association). The OWL API,[3] a Java API which is used to manipulate OWL ontologies, implements this notion of structural equivalence by default.

A looser notion of structural similarity, *justification isomorphism* [6], was first introduced in a study of the cognitive complexity of justifications: Two justifications are isomorphic if there exists a mapping between class, property and individual names of the justifications which makes them structurally equivalent. This equivalence relation covers justifications which contain the same number of axioms, constructors, as well as class, property, and individual names. Justification isomorphism has previously been shown to significantly reduce a corpus of justifications from 64,800 to merely 11,600 justification templates [6].

While justification isomorphism helps to eliminate the effects of diverging entity names, we can also identify types of justifications which may be considered to be very similar despite their use of different constructors:

[2] http://www.w3.org/TR/owl2-syntax

[3] http://owlapi.sourceforge.net

Example 1

$$\mathcal{J}_1 = \{A \sqsubseteq B \sqcap C, B \sqcap C \sqsubseteq D\} \models A \sqsubseteq D$$
$$\mathcal{J}_2 = \{A \sqsubseteq \exists r.C, \exists r.C \sqsubseteq D\} \models A \sqsubseteq D$$

In this example, the semantics of the complex expressions $B \sqcap C$ in $\mathcal{J}_1$ and $\exists r.C$ in $\mathcal{J}_2$ are not relevant for the respective entailment; their occurrences in the justifications and their entailments could be replaced by freshly generated atomic concept names without affecting the entailment relation. Such a substitution in turn would make the two justifications isomorphic.

Likewise, justifications of different lengths may be considered similar if their general structure of reasoning is identical:

Example 2

$$\mathcal{J}_1 = \{A \sqsubseteq B, B \sqsubseteq C\} \models A \sqsubseteq C$$
$$\mathcal{J}_2 = \{A \sqsubseteq B, B \sqsubseteq C, C \sqsubseteq D\} \models A \sqsubseteq D$$

These two justifications clearly require the same form of reasoning from a user, namely the understanding of simple atomic subsumption. Strict isomorphism only applies to justifications which contain the same number of axioms; it does not cover situations like the above. However, for the purpose of structuring sets of justifications and analysing the logical diversity of a corpus of justifications, capturing those kinds of similarities illustrated in the above examples would be highly desirable.

The idea of finding similarities between concepts in Description Logics has been widely explored in the work on *unification* and *matching*, e.g. [1,2], for the purpose of detecting redundant concept descriptions in knowledge bases. The aim of unification is to find a suitable substitution σ which maps atomic concepts in a concept term C to (possibly non-atomic) concepts in a concept term D such that the two terms are made equivalent.

While unification and matching are very close to our requirements for capturing similarities between justifications, the concepts are not directly applicable. In our case, the inputs are of different shape from the matching problem: The goal is to unify two sets of axioms and the corresponding entailments, rather than matching a given concept pattern containing variables with a concept description.

The above examples motivate a looser notion of justification isomorphism, which allows us to identify justifications as equivalent if they require the same reasoning mechanisms, regardless of size, signature, and logical constructors used. In the present paper, we introduce two new types of equivalence relations based on matching *subexpressions* and *lemmas*, and analyse the effect these extended relations have when applied to a corpus of justifications from the bio-medical domain.

2 Preliminaries

2.1 Justifications in OWL

We assume the reader to be familiar with OWL and the underlying Description Logic $\mathcal{SROIQ}$ [9]. In what follows, $A, B, \ldots$ denote class names in an ontology $\mathcal{O}$, r, s role names, and α denotes an OWL axiom.

The concept of pinpointing minimal entailing subsets of an ontology is the dominant form of explanation for entailments of OWL ontologies [15,13,3]. A justification (also denoted as *MinA*, or MUPS when referring to unsatisfiable classes) is defined as a minimal subset of an ontology $\mathcal{O}$ that causes an entailment η to hold:

Definition 1 (Justification) *$\mathcal{J}$ is a justification for $\mathcal{O} \models \eta$ if $\mathcal{J} \subseteq \mathcal{O}, \mathcal{J} \models \eta$ and, for all $\mathcal{J}' \subset \mathcal{J}$, it holds that $\mathcal{J}' \not\models \eta$.*

For every axiom which is asserted in the ontology, the axiom itself naturally is a justification. We call such a justification a *self-justification*, and an entailment which has only a self-justification and no other justification in $\mathcal{O}$ a *self-supporting* entailment.

It is important to note that a justification is always defined with respect to an entailment η. In the remainder of this paper we will therefore use the term *justification* to describe a justification-entailment pair ($\mathcal{J}$, η) where $\mathcal{J}$ is a minimal entailing axiom set for η.

2.2 Justification Isomorphism

Isomorphism between justifications was first introduced as a method to reduce the number of similar justifications when sampling from a large corpus to justifications [6].

Definition 2 (Justification Isomorphism) *Two justifications $(\mathcal{J}_1, \eta_1)$, $(\mathcal{J}_2, \eta_2)$ are* isomorphic *($(\mathcal{J}_1, \eta_1) \approx_i (\mathcal{J}_2, \eta_2)$) if there exists an injective renaming ϕ which maps class, role, and individual names in $\mathcal{J}_1$ and η_1 to class, role, and individual names in $\mathcal{J}_2$ and η_2, respectively, such that $\phi(\mathcal{J}_1) = \mathcal{J}_2$ and $\phi(\eta_1) = \eta_2$.*

Example 3 (Isomorphic Justifications)

$$\begin{aligned}
\mathcal{J}_1 &= \{A \sqsubseteq B \sqcap \exists r.C, B \sqcap \exists r.C \sqsubseteq D\} \models A \sqsubseteq D \\
\mathcal{J}_2 &= \{E \sqsubseteq B \sqcap \exists s.F, B \sqcap \exists s.F \sqsubseteq D\} \models E \sqsubseteq D \\
\phi &= \{A \mapsto E, C \mapsto F, r \mapsto s\}
\end{aligned}$$

The relation $\approx_i$ is symmetric, reflexive and transitive, from which it follows that $\approx_i$ is an equivalence relation and thus partitions a set of justifications.

In the remainder of this paper, we may refer to the isomorphism defined above as *strict* isomorphism in order to distinguish it from the other equivalence relations.

3 Subexpression-Isomorphism

From the above definition of isomorphism it follows that only justifications which have the same number and types of axioms and subexpressions can be isomorphic. It is easy to see, however, that justifications can have a similar structure despite their use of different concept expressions, as demonstrated in Example 1. This motivates a notion of isomorphism which allows not only the mapping of concept names, but also that of complex subexpressions.

We introduce a *justification template* Θ, which functions as the *unifying justification* for the isomorphic justifications:

Definition 3 (Subexpression-Isomorphism) *Two justifications* $(\mathcal{J}_1, \eta_1)$, $(\mathcal{J}_2, \eta_2)$ *are* s-isomorphic $((\mathcal{J}_1, \eta_1) \approx_s (\mathcal{J}_2, \eta_2))$ *if there exists a justification* (Θ, η), *called a* template, *and two injective substitutions* ϕ_1, ϕ_2, *such that*

1. $\Theta \models \eta$
2. $\phi_1(\Theta) = \mathcal{J}_1$ *and* $\phi_2(\Theta) = \mathcal{J}_2$
3. $\phi_1(\eta) = \eta_1$ *and* $\phi_2(\eta) = \eta_2$.

The mappings ϕ_1 *and* ϕ_2 *map class, role, and individual names in the template* (Θ, η) *to subexpressions of* $(\mathcal{J}_1, \eta_1)$ *and* $(\mathcal{J}_2, \eta_2)$, *respectively.*

Lemma 1
1. *The relation* $\approx_s$ *is reflexive, transitive and symmetric; it is therefore an equivalence relation and thus partitions a set of justifications.*
2. *S-isomorphism is a more general case of strict isomorphism:* $\mathcal{J}_1 \approx_i \mathcal{J}_2$ *implies* $\mathcal{J}_1 \approx_s \mathcal{J}_2$.

For a complete proof of Lemma 1 we refer the reader to the supporting materials page[4] for this paper.

4 Lemma-Isomorphism

While s-isomorphism covers a number of justifications that can be regarded as equivalent due to them requiring the same type of reasoning to reach the entailment, it only applies to justifications which have the same number of axioms. This does not take into account cases where the justifications differ only marginally in some subset, but where the general reasoning may be regarded as similar nonetheless. We therefore introduce the notion of *lemma-isomorphism*, which extends subexpression-isomorphism with the substitution of subsets of justifications through intermediate entailments, so-called *lemmas* [7]. The general motivation behind lemma-isomorphism is demonstrated by the following example:

Example 4

$$\mathcal{J}_1 = \{A \sqsubseteq \exists r.B, B \sqsubseteq C, \exists r.C \sqsubseteq D\} \models A \sqsubseteq D$$
$$\mathcal{J}_2 = \{A \sqsubseteq \exists r.B, B \sqsubseteq C, C \sqsubseteq D, D \sqsubseteq E, \exists r.E \sqsubseteq F\} \models A \sqsubseteq F$$

[4] http://owl.cs.manchester.ac.uk/research/publications/supporting-material/

It is straightforward to see that both $\mathcal{J}_1$ and $\mathcal{J}_2$ require the same type of reasoning from a human user. As the justifications only differ in the length of the atomic subsumption chains they contain, we can certainly consider them to be *similar* with respect to *some* similarity measure. However, the two justifications are not considered isomorphic with respect to the definitions for strict isomorphism or subexpression-isomorphism. We therefore introduce a new type of isomorphism which takes into account the fact that subsets of justifications can be replaced with intermediate entailments which follow from them.

4.1 Lemmas in OWL

Lemmas of OWL justifications have previously found use in the extension of justifications to *justification-oriented proofs* [7]. The following definitions introduce simplified variants of the definitions [7] of justification lemmas and lemmatisations. Please note that for the purpose of illustrating the effect of lemma-isomorphism, we will simplify the lemmatisations to a more specific type of lemmas in the next section.

Definition 4 (Lemma) *Let $\mathcal{J}$ be a justification for an entailment η. A* lemma *of $(\mathcal{J}, \eta)$ is an axiom λ for which there exists a subset $S \subseteq \mathcal{J}$ such that $S \models \lambda$. A* summarising *lemma of $(\mathcal{J}, \eta)$ is a lemma λ for which there exists an $S \subseteq \mathcal{J}$ such that $\mathcal{J} \setminus S \cup \{\lambda\} \models \eta$ for $S \models \lambda$.*

Definition 5 (Lemmatisation) *Let $(\mathcal{J}, \eta)$ be a justification, let $S_1 \dots S_k$ be subsets of $\mathcal{J}$, and let $\lambda_1 \dots \lambda_k$ be axioms satisfying $S_i \models \lambda_i$ for $i \in \{1, \dots, k\}$. Then the set $\mathcal{J}^{\Lambda} := (\mathcal{J} \setminus \bigcup S_i) \cup \bigcup \{\lambda_i\}$ for $i \in \{1, \dots, k\}$ is called a* lemmatisation *of $\mathcal{J}$ if $\mathcal{J}^{\Lambda} \models \eta$. A* summarising lemmatisation *comprises only summarising lemmas.*

4.2 Lemma-Isomorphism

Given the definitions for lemmatisations, we can now define lemma-isomorphism as an extension to subexpression-isomorphism:

Definition 6 (Lemma-isomorphism) *Two justifications $(\mathcal{J}_1, \eta_1)$, $(\mathcal{J}_2, \eta_2)$ are* ℓ-isomorphic *($(\mathcal{J}_1, \eta_1) \approx_\ell (\mathcal{J}_2, \eta_2)$) if there exist lemmatisations $\mathcal{J}_1^{\Lambda_1}, \mathcal{J}_2^{\Lambda_2}$ which are s-isomorphic: $\mathcal{J}_1^{\Lambda_1} \approx_s \mathcal{J}_2^{\Lambda_2}$.*

Lemma-isomorphism using arbitrary lemmas as defined above carries some undesirable properties: First, unlike the previously defined relations, it describes a relation which is *not transitive*. This isssue can be adressed by allowing only *summarising* lemmatisations. Second, the lemmatisation might differ strongly from the original justifications; in the most extreme case, the lemmatisation of a justification can be the entailment itself. We therefore have to introduce some *constraints* on the admissible lemmatisations in order to preserve the nature of the original justifications. In order to restrict the lemmatisations to justifications which do not differ too much from the original justification, we focus on substituting *obvious steps* with their lemmas.

4.3 Lemmatisations and Obvious Steps

The notion of *obvious proof steps* [10,5] describes how proof steps which are intuitively *obvious* can be replaced with their conclusion, thereby shortening the proof without omitting important information. We loosely base the lemma restriction on this obviousness and choose one such example of an obvious and frequently occurring constellation of axioms in OWL justifications, namely atomic subsumption chains.

In atomic subsumption chains of the type $A_0 \sqsubseteq A_1, A_1 \sqsubseteq A_2 \ldots A_{n-1} \sqsubseteq A_n$ only the relation between the subconcept A_0 in the first axiom and the superconcept A_n in the last axiom are relevant for understanding the subsumption chain; i.e. the step from the subconcept to the final superconcept is *obvious*. We can say that it is only important to understand *that* there is a connection between the subconcept and the final superconcept, but we do not need to know *what* this connection is. Therefore, it seems reasonable to substitute the chain with its conclusion in the form of a single axiom $A_0 \sqsubseteq A_n$. Please note that it is possible for such a substitution to generate a non-summarising lemma; therefore, we will only allow *summarising lemmatisations based on atomic subsumption chains*.

Atomic subsumption chains represent only one of many examples of such lemmatisations which preserve both transitivity and the original style of the justification. For the purpose of introducing lemma-isomorphism as an equivalence relation in this paper, we focus on this particular type of lemmatisations, as it captures a frequently occurring pattern in OWL justifications.

5 Diversity of Reason in the NCBO BioPortal Ontologies

5.1 Test Corpus

We performed a survey of equivalence relations in OWL- and OBO-ontologies from the NCBO BioPortal.[5] The purpose of this study was to determine the prevalence of the different types of isomorphism across an independently motivated (as opposed to hand selected) corpus of OWL ontologies used in practice.

At the time of downloading (January 2012), the BioPortal listed 278 OWL- and OBO-ontologies, of which 241 could be downloaded, merged with their imports, and serialised as OWL/XML. 15 of those ontologies could not be processed in the given time frame of 30 minutes using the selected reasoner, and another 25 did not contain any relevant entailments (direct subsumptions between named classes). For the remaining 201 ontologies, we computed justifications for all entailments with a maximum of 500 justifications per entailment. Self-supporting entailments and self-justifications were excluded from the survey, which led to the discarding of further ontologies.

The final corpus of justifications consisted of 6,744 justifications from 83 ontologies, covering a very broad spectrum of sizes and complexity. Half of the ontologies had less than 1000 named concepts and axioms, with the other half

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

reaching a maximum of 13,959 concepts and 70,015 axioms. Likewise, the expressivity of the ontologies ranged from $\mathcal{AL}$ to several highly expressive samples in $\mathcal{SROIQ}$. A detailed listing of all surveyed ontologies alongside the study results is available online.[6]

5.2 Isomorphism on the Entailment Level

We first analysed how the equivalence relations affected the set of justifications for a *single* entailment. For this purpose we focused exclusively on those 39 ontologies in the corpus which produced entailments with multiple justifications. 5,647 justifications were computed for the 3,264 entailments of those ontologies (including those entailments which had only 1 justification).

Strict Isomorphism On average, an entailment in the reduced corpus has 1.7 justifications, with a maximum of 122 justifications for an entailment from the *Orphanet Ontology of Rare Diseases*. Strict isomorphism shows a significant reduction by 23.6% to an average of 1.3 templates per entailment. Overall, however, only few ontologies are visibly affected by this reduction: In 11 ontologies, an average of 3 justifications for an individual entailment is covered by a single template, in 13 ontologies a template covers an average of 2 justifications, and in the remaining 15 ontologies strict isomorphism does not affect the numbers of justifications per entailment.

Of those 11 ontologies which do show some significant reduction, entailments of the Orphanet and *Cognitive Atlas* ontology reveal the most striking regularities: The 122 justifications from the Orphanet ontology were covered by only 2 templates, with 61 justifications each:

$$\Theta_1 = \{A \sqsubseteq \exists r.B, Domain(r,\ C)\} \models A \sqsubseteq C$$
$$\Theta_2 = \{A \sqsubseteq \exists s.B, s \sqsubseteq r, Domain(r,\ C)\} \models A \sqsubseteq C$$

This pattern is repeated by a large number of entailments across the Orphanet ontology; as we will see in the next section, almost all entailments in this ontology have justifications which are covered by these two templates.

S-Isomorphism Subexpression-isomorphism affects the justifications of only 12 of the 3,264 entailments. Most of these stem from the *Bleeding History Phenotype* ontology, where the template Θ_1 also covers justifications of the type $\{A \sqsubseteq \exists r.(B \sqcup D), Domain(r,\ C)\}$, i.e. they contain a disjunction instead of the atomic class name B as the filler of the existential restriction.

L-Isomorphism Similarly, lemma-isomorphism only affects 39 entailments, with the most notable effects in the *Human Developmental Anatomy* ontology, where justifications comprising of atomic subsumption chains of lengths 2 and 3, respectively, are covered by a single template.

[6] See footnote 4.

5.3 Isomorphism Across Multiple Entailments

Strict Isomorphism When applied to the justifications for *all* entailments of the individual ontologies, strict isomorphism drastically reduces the number of justifications from an average of 81.3 ($\sigma = 185.5$) justifications per ontology to 10.5 ($\sigma = 18.0$) templates for equivalent justifications. The mean number of justifications per template is 7.7 ($\sigma = 41.7$), which means that in each ontology nearly 8 justifications have an identical structure. This effect is highly visible in the Orphanet ontology, where the above template Θ_1 covers 901 (of 1139) justifications for *distinct* entailments.

S-Isomorphism The reduction from strict isomorphism to s-isomorphism is less drastic than the difference between the main pool and the non-isomorphic pool. The justifications of the 83 ontologies are reduced from an average of 81.3 justifications to 8.8 templates ($\sigma = 13.1$), which is a reduction by 1.7 templates compared to strict isomorphism. An average of 9.2 justifications ($\sigma = 46.6$) in an ontology share the same template. Surprisingly, the majority of ontologies (67) does not show any difference between strict isomorphism and s-isomorphism. Only 2 ontologies, the *Lipid Ontology* and Bleeding History Phenotype, are significantly affected by s-isomorphism, with a reduction from 118 to 13 templates (an 89% reduction from strict isomorphism) and 32 to 14 templates (46.2% reduction from strict isomorphism), respectively.

L-Isomorphism As with s-isomorphism, the effects of ℓ-isomorphism are not as significant as the first reduction through strict isomorphism. The justifications are further reduced to an average of 7.4 templates per ontology ($\sigma = 11.4$), with 11 justifications per template ($\sigma = 51.5$). Still, 35 of the 83 ontologies show at least a minor difference between s-isomorphism and ℓ-isomorphism, which indicates that they contain at least 1 atomic subsumption chain. L-isomorphism reduces the 106 justifications generated for the *Cereal Plant Gross Anatomy* ontology to only 14 templates, compared to 29 templates for s-isomorphism.

5.4 Similarities Across Multiple Ontologies

Strict Isomorphism When applied across all justifications from the corpus, strict isomorphism reduces the corpus from 6,744 justifications to only 614 templates, a reduction to only 9.1% of the original set of justifications. On average, 11 justifications share the same template, with the most frequent template occurring 1,603 times across 18 different ontologies (that is, in about a fifth of all ontologies); this template is of the same form as the Orphanet Ontology described above.

S-Isomorphism Subexpression-isomorphism reduces the corpus from 6,744 to 456 templates (6.8% of the corpus), which is a further reduction by 25.7% compared to the 614 templates for strict isomorphism. The most frequent templates in terms of number of justifications and prevalence across all ontologies are the same as for strict isomorphism, with numbers differing only slightly.

L-Isomorphism Finally, lemma-isomorphism reduces the 6,744 justifications to a mere 384 templates, which is an overall reduction of 94.3%, and a further reduction by 15.8% compared to subexpression-isomorphism. The effect of lemma-isomorphism is visible when we look at the most prevalent justification, an atomic subsumption chain of size 2, which occurs in 44 (compared to previously 37) ontologies. This chain represents all 701 atomic subsumption chains of differing sizes that can be found in the corpus.

5.5 Summary

The results of our survey indicate that the effects of the three equivalence relations vary strongly between the ontologies in the corpus. In contrast to strict isomorphism, subexpression- and lemma-isomorphism have almost no effect on the justifications for *individual* entailments. For multiple entailments, however, some ontologies show a clear reaction to s- or ℓ-isomorphism. Across the corpus, the logical diversity could be shown to be significantly smaller than the number of justifications would suggest, as lemma-isomorphism reduced the over 6000 justifications to only around 600 distinct templates.

6 An Application Scenario

The methods proposed in this paper were motivated by an example from the well-known *Pizza* tutorial ontology.[7] An example entailment for this ontology is $\mathsf{Fiorentina} \sqsubseteq \mathsf{InterestingPizza}$, which has over 200 justifications. An ontology engineer wanting to understand why this entailment holds, for example because it is considered incorrect, would have to go through a list of several hundred justifications, inspecting each one and deciding which axiom to modify or remove in order to 'break' the entailment.

Closer inspection, however, reveals significant similarities between the justifications for this entailment: All justifications are of the form

$$\{\mathsf{S}_1, \mathsf{S}_2, \mathsf{InterestingPizza} \equiv \mathsf{Pizza} \sqcap (\geqslant 3\ \mathsf{hasTopping}.\top)\}$$

where S_1 is one of several axiom sets entailing that $\mathsf{Fiorentina} \sqsubseteq \mathsf{Pizza}$, and S_2 a set of axioms entailing that $\mathsf{Fiorentina} \sqsubseteq\ \geqslant 3\ \mathsf{hasTopping}.\top$, which originates from the fact that the $\mathsf{Fiorentina}$ pizza is defined to have six disjoint toppings. While the large number of justifications may seem daunting at first, once the structural similarities have been spotted, understanding the different reasons why the entailment holds requires significantly less effort—both mentally, and in terms of a 'click-count'.

Integrating the proposed equivalence relations into a user interface could support users in spotting these patterns. In the case of the *Pizza* ontology, we can apply techniques that make use of both strict isomorphism and lemma-isomorphism: Strict isomorphism may be used to group those justifications which

[7] http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/

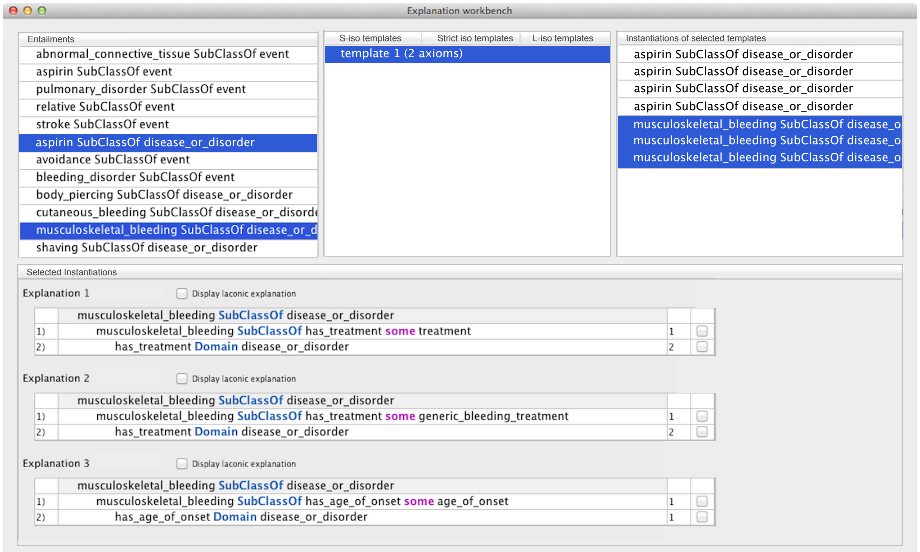

Fig. 1: Screenshot of a justification template exploration tool

have identical subexpressions modulo the names of the different toppings. L-isomorphism (extended by additional types of 'obvious' proof steps) could further group those axiom sets which lead to the same lemmas. A basic interaction mechanism for navigation such lemmatisations has been suggested in the work on *justification-based proofs* [7]. Going beyond the example of the *Pizza* ontology, s-isomorphic justifications could easily be highlighted by covering up irrelevant expressions similar to the strike-out techniques for superfluous expressions implemented in the Swoop ontology editor [12]. This would prevent users from getting distracted by complex expressions, thus allowing them to focus on understanding the relevant axioms and expressions in a set of justifications.

The second task for which our notions of isomorphism may be useful is the exploration and understanding of an ontology, without focusing on a specific entailment. In this case, the user could be offered a browser-type interface as the one shown in Figure 1 (displaying entailments from the Bleeding History Phenotype ontology). The browser-style exploration tool consists of three top panels, which show a list of entailments, a list of the formal templates which cover the justifications for the selected entailment(s), and a list of material instantiations of the selected template (identified by the entailment they stand for). The bottom panel displays the selected instantiations or templates. A user seeking to understand the structure of an ontology could gain a high-level view of the ontology by selecting a set of entailments which then displays the set of their justification templates and respective instantiations of those templates. Entailments which share a template (or a number of templates) then highlight regularities in the axiom structure of the ontology, as well as the prevalence of such regularities.

7 Conclusions

In this paper, we introduced new types of equivalence relations between OWL justifications, subexpression-isomorpism and lemma-isomorphism. We demonstrated how a seemingly diverse corpus of justifications from the NCBO BioPortal could be reduced by over 90% to a much smaller set of non-isomorphic justifications. We have found that, surprisingly, most justifications are in fact strictly isomorphic, with only a few ontologies being affected by the other equivalence relations.

Future work will involve exploring further notions of obvious proof steps in order to extend lemma-isomorphism beyond atomic subsumption chains. We will also consider the issue of overlapping chains, i.e. subsumption chains which lead to non-summarising lemmas. Finally, we aim to fully implement the proposed tool which orders and groups justifications based on their isomorphism relations, and conduct user studies that investigate the usefulness of the tool for various tasks in the ontology development process.

References

1. Baader, F., Küsters, R., Borgida, A., McGuinness, D.L.: Matching in description logics. J. of Logic and Computation 9(3), 411–447 (1999)
2. Baader, F., Morawska, B.: Unification in the description logic EL. In: Proc. of RTA-09. pp. 350–364 (2009)
3. Baader, F., Peñaloza, R., Suntisrivaraporn, B.: Pinpointing in the description logic $\mathcal{EL}^+$. In: Proc. of KI-07. pp. 52–67 (2007)
4. Bail, S., Parsia, B., Sattler, U.: The justificatory structure of OWL ontologies. In: Proc. of OWLED-10 (2010)
5. Davis, M.: Obvious logical inferences. In: Proc. of IJCAI-81. pp. 530–531 (1981)
6. Horridge, M., Bail, S., Parsia, B., Sattler, U.: The cognitive complexity of OWL justifications. In: Proc. of ISWC-11 (2011)
7. Horridge, M., Parsia, B.: From justifications to proofs for entailments in OWL. In: Proc. of OWLED-09 (2009)
8. Horridge, M., Parsia, B., Sattler, U.: Laconic and precise justifications in OWL. In: Proc. of ISWC-08. pp. 323–338 (2008)
9. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible $\mathcal{SROIQ}$. In: Proc. of KR-06 (2006)
10. Johnson-Laird, P.N.: Mental models in cognitive science. Cognitive Science 4(1), 71–115 (1980)
11. Kalyanpur, A., Parsia, B., Cuenca Grau, B.: Beyond asserted axioms: Fine-grain justifications for OWL-DL entailments. In: Proc. of DL-06 (2006)
12. Kalyanpur, A., Parsia, B., Sirin, E., Cuenca Grau, B., Hendler, J.: Swoop: A web ontology editing browser. J. of Web Semantics 4(2), 144–153 (2006)
13. Kalyanpur, A., Parsia, B., Sirin, E., Hendler, J.: Debugging unsatisfiable classes in OWL ontologies. J. of Web Semantics 3(4), 268–293 (2005)
14. Meyer, T., Moodley, K., Varzinczak, I.: First steps in the computation of root justifications. In: Proc. of ARCOE-10 (2010)
15. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Proc. of IJCAI-03. pp. 355–362 (2003)