Diversity of Reason: Equivalence Relations over Description Logic 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 modern ontology languages, such as OWL, there is the possibility for great diversity in the logical content of ontologies. Informally, this can be seen by the constant evolution of reasoners to deal with new sorts of content and the range of optimisations reasoners need in order to be competitive. More formally, the fact that many naturally occurring entailments have multiple justifications (i.e., minimal entailing subsets) indicates that ontologies often overdetermine their consequences, indicating a diversity in supporting reasons. However, the multiplicity of justifications might be due mostly to diverse *material*, not *formal*, grounds for an entailment. That is, the logical form of these multiple reasons could be less diverse than their numbers suggest.

In the present paper, we introduce and explore several equivalence relations over justifications for entailments of OWL ontologies. These equivalence relations range from strict isomorphism to a looser notions which cover similarities between justifications containing different concept expressions or possibly different numbers of axioms. We survey a corpus of ontologies from the bio-medical domain and find that large numbers of justifications can often be reduced to a significantly smaller set of justifications which are isomorphic with respect to one of the given definitions.

1 Introduction

Since its standardisation by the W3C in 2004, the Web Ontology Language OWL has been used to represent domain knowledge from diverse areas, such as medicine and general biology. OWL 2[1] is based on the highly expressive Description Logic $\mathcal{SROIQ}$ [10] as underlying formalism: An OWL 2 ontology corresponds to a set of $\mathcal{SROIQ}$ axioms.

Justifications, minimal entailing subsets of an OWL ontology, provide helpful and easy-to-understand explanation support when repairing unwanted entailments in the ontology debugging process. While previous research has focused on the issue of making individual justifications easier to understand, only little attention has been paid to the occurrence of *multiple justifications* other than

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

the computational problems they often imply. An entailment of an OWL ontology can have a large number of justifications (potentially exponential in the number of axioms in the ontology), with up to several thousands found in large real-life ontologies.

When encountering justifications for a finite set of entailments of an ontology (e.g. the set of entailed atomic subsumptions), we are often faced with a seemingly large and diverse body of reasons why these entailments hold. On closer inspection, however, we frequently find that multiple justifications for different entailments are identical sets of axioms, or that justifications are structurally identical and only diverge in the concept, role, and individual names they use.

In order to draw a clearer picture of the logical diversity of a corpus of justifications, we need to take into account these similarities between justifications. Grouping the set of justifications into subsets according to some similarity measure helps structuring the unsorted pool of justifications. Rather than trying to understand each individual justification, users can focus on understanding the *template* of a particular subset of justifications. This can lead to significantly fewer justifications that the user has to deal with, and therefore to a reduction in user effort.

A well-known syntactical equivalence relation in OWL is *structural equivalence*. The OWL Structural Specification[2] states the condition for two OWL objects (named concepts, roles, or instances, complex expressions, or OWL axioms) to be structurally equivalent. In short, it defines the objects to be equivalent if they contain the same 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.

Justification isomorphism [5] was first introduced in a study of the cognitive complexity of justifications: Two justifications are isomorphic if they are structurally identical,[4] i.e. the axioms contain the same concept expressions and only differ in the concept, role and individual names.

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

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 concept 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

[2] http://www.w3.org/TR/owl2-syntax
[3] http://owlapi.sourceforge.net
[4] Modulo structural equivalence, which takes into account redundant expressions and the commutativity of constructors.

justifications and their entailments can 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. 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.

These examples motivate a looser notion of 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 show how these extended relations are applied to a corpus of justifications from the bio-medical domain.

2 Preliminaries

Justifications in OWL We assume the reader to be familiar with OWL and the underlying Description Logic $\mathcal{SROIQ}$. In what follows, $A, B, \ldots$ denote concept names in an ontology $\mathcal{O}$, r, s role names, and $sig(\alpha)$ denotes the set of concept, role, and individual names in an OWL axiom α.

Justifications [13,11] are a popular type of explanation for entailments of OWL ontologies. A justification 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 an entailment which has only itself as a justification a *self-supporting* entailment, and the justification a *self-justification*.

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 $(\mathcal{J}, \eta)$ where $\mathcal{J}$ is a justification for η.

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 of distinct types [5].

Definition 2 (Justification Isomorphism) *Two justifications* $(\mathcal{J}_1, \eta_1)$, $(\mathcal{J}_2, \eta_2)$ *are* isomorphic $((\mathcal{J}_1, \eta_1) \approx_i (\mathcal{J}_1, \eta_1))$ *if there exists an injective renaming* ϕ *which maps concept, role, and individual names in* $\mathcal{J}_1$ *and* η_1 *to concept, 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. Justification isomorphism preserves the numbers and types of axioms and subexpressions in the justifications:

1. $\mathcal{J}_1 \approx_i \mathcal{J}_2 \rightarrow |\mathcal{J}_1| = |\mathcal{J}_2|$
2. $\mathcal{J}_1 \approx_i \mathcal{J}_2 \rightarrow |sig(\mathcal{J}_1)| = |sig(\mathcal{J}_2)|$
3. The sets of logical constructs used in $\mathcal{J}_1$ and $\mathcal{J}_2$ coincide.

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 subexpressions.

The idea of finding similarities between concepts in Description Logics has been widely explored in the work on *unification* and *matching*, e.g. [2,1,3], 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 the basic idea behind extended subexpression-isomorphism is based on unification and matching, the concepts are not directly applicable to the given problem of matching justifications. We therefore introduce a *unifying justification* $\mathcal{J}$, which functions as the *template* 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* $(\mathcal{J}, \eta)$ *and two injective substitutions* ϕ_1, ϕ_2*, such that* $\phi_1(\mathcal{J}) = \mathcal{J}_1$*,* $\phi_2(\mathcal{J}) = \mathcal{J}_2$*,* $\phi_1(\eta) = \eta_1$*, and* $\phi_2(\eta) = \eta_2$*. The mappings* ϕ_1 *and* ϕ_2 *map concept, role, and individual names in* $(\mathcal{J}, \eta)$ *to subexpressions of* $(\mathcal{J}_1, \eta_1)$ *and* $(\mathcal{J}_2, \eta_2)$*, respectively.*

It can be shown that the relation $\approx_s$ is reflexive, transitive and symmetric; it is therefore an equivalence relation and thus partitions a set of justifications.

The substitutions ϕ_1, ϕ_2 are injective, but not surjective, as the set of subexpressions in the justifications $\mathcal{J}_1$ and $\mathcal{J}_2$ can be of higher arity than the set of concept names in the unifying justification $\mathcal{J}$ (unless the justifications themselves contain no complex expressions).

S-isomorphism can easily be shown to be a more general case of strict isomorphism between two justifications $\mathcal{J}_1$ and $\mathcal{J}_2$: The unifying justification $\mathcal{J}$ is set to $\mathcal{J}_1$, ϕ_1 is the identity relation *id* which maps $\mathcal{J}_1$ to itself, and ϕ_2 corresponds to the mapping ϕ, which maps concept, role, and individual names in $\mathcal{J}_2$ to the respective names in $\mathcal{J}_1$. It follows that $\mathcal{J}_1 \approx_i \mathcal{J}_2 \rightarrow \mathcal{J}_1 \approx_s \mathcal{J}_2$.

In order to be s-isomorphic, the justifications may differ in the number of subexpressions. They must, however, have the same number of axioms: $\mathcal{J}_1 \approx_s \mathcal{J}_2 \rightarrow |\mathcal{J}_1| = |\mathcal{J}_2|$

While we focus on entailed atomic subsumptions in the present paper, we point out that s-isomorphism between justifications is not restricted to a specific axiom type as entailment, as the substitutions ϕ_1, ϕ_2 preserve all entailment relations regardless of the axiom type and constructor usage.

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

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

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 that lead to the entailment, 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.

Lemmas of OWL justifications have previously found use in the extension of justifications to *justification-oriented proofs* [9]. 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 $\mathcal{J}^{\Lambda} := (\mathcal{J} \setminus \bigcup S_i) \cup \bigcup \lambda_i$ is called a* lemmatisation *of $\mathcal{J}$ if $\mathcal{J}^{\Lambda} \models \eta$.*

If clear from the context, a lemmatisation $\mathcal{J}^{\Lambda}$ may also be called a *lemmatised justification*. 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 *($(\mathcal{J}_1, \eta) \approx_\ell (\mathcal{J}_2, \eta)$) 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. Further, the lemmatised justifications 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 guarantee the transitivity of the isomorphism relation, as well as preserve the nature of the original justifications. In what follows we will focus on lemmatisations through *obvious steps* in justifications.

4.1 Lemmatisations and Obvious Steps

The notion of *obvious proof steps* [4,12] 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 select an example of obvious proof steps which commonly occur in DL justifications, namely chains of atomic subsumptions.

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 $A_0 \sqsubseteq A_n$. Restricting lemmatisations to atomic subsumption chains leads to the definition for a transitivity-preserving type of l-isomorphism:

Definition 7 (Transitivity-preserving l-isomorphism) *Two justifications* $(\mathcal{J}_1, \eta_1)$, $(\mathcal{J}_2, \eta_2)$ *are* ℓt-isomorphic $((\mathcal{J}_1, \eta) \approx_{\ell t} (\mathcal{J}_2, \eta))$ *if there exist summarising lemmatisations* $\mathcal{J}_1^{\Lambda_1}, \mathcal{J}_2^{\Lambda_2}$ *which are s-isomorphic, and every* $S_i \subseteq \mathcal{J}_i$ *where* $S_i \models \lambda_i \in \Lambda_1 (\lambda_i \in \Lambda_2$*, respectively) is of type* $\{A_i \sqsubseteq A_{i+1} \mid i \in \{0, \ldots, n\}\}$ *where* n *is the number of axioms in the respective chain of subsumption axioms, and* A_i *a concept name.*

Atomic subsumption chains represent only one of many examples of such lemmatisations which preserve transitivity. For the purpose of introducing lemma-isomorphism as an equivalence relation, it suffices to 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. The BioPortal provides ontologies from various groups from the biomedical domain, including the full set of daily updated OBO Foundry[5] ontologies, which are built based on common design principles. OBO ontologies use a flat-file format, which can be translated into OWL 2 and were therefore included in the test corpus.

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.

[5] http://obofoundry.org

The final corpus of justifications consisted of 6,744 justifications from 83 ontologies, covering a broad spectrum of sizes and complexity. Half of the ontologies had less than 1000 named concepts and axioms, with the other half reaching a maximum of 13,959 concepts and 70,015 axioms. Likewise, the expressivity of the corpus ranged from AL to several highly expressive samples in $\mathcal{SROIQ}$, which corresponds to the OWL 2 language. A detailed listing of the surveyed ontologies, as well the complete data from the study is available online.[6]

5.2 Implementation

The algorithm for detecting structural similarities between justifications (and therefore, OWL axioms) uses a purely syntactic search for matching subexpressions of OWL axioms. The axioms are first parsed into a tree form, which allows for pairwise comparison of the nodes which represent connectives as well as concept and role names. The implementation uses the OWL API (v3.2.4.) and the HermiT (v1.3.6.) reasoner. Our experiments were performed on a 3GHz Intel Core2 Duo machine with 4GB of RAM allocated to the Java Virtual Machine. The 7051 justifications in the test corpus could be analysed for strict isomorphism in approximately 3 minutes.

5.3 Results of the Survey

Isomorphic Justifications Strict isomorphism applied to all justifications of the individual ontologies drastically reduces the number of regular 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 of Rare Diseases*, where the a single template covers 901 (of 1139) justifications for *distinct* entailments. These justifications are all of the type $\{gen\_x \sqsubseteq (\exists geneOf.pat\_y), Domain(geneOf,\ gen\_id\_1)\} \models gen\_x \sqsubseteq gen\_id\_1$ where x and y denote identifiers used in the ontology.

Likewise, in the *Cognitive Atlas* ontology, all 401 justifications for a *single* entailment are reduced to only one template. Intuitively, it seems that we can observe a much stronger reduction on a large set of justifications than those ontologies that produce only 1 or 2 justifications. And indeed, there exists a strong correlation between the number of justifications of an ontology and the amount of reduction (Spearman's $\rho = 0.78$). However, even some of the larger sets of justifications are only reduced by a small proportion. The 95 justifications of the *Gene Regulation Ontology* are represented by 61 distinct templates, which indicates a fairly diverse corpus.

When applied across all justifications from the corpus, strict isomorphism reduces the corpus from 6,744 justifications to only 614 templates, a reduction

[6] http://owl.cs.manchester.ac.uk/research/publications/supporting-material/just-iso-dl2012

to only 9.1% of the original corpus. On average, 11 justifications share the same template, with the most frequent template occurring 1,603 times across 18 different ontologies; coincidentally, this template is of the same form as the Orphanet Ontology described above. The most prevalent template in the corpus, based on its occurrence in 37 distinct ontologies, is an atomic subsumption chain with 2 axioms.

Subexpression-Isomorphism The reduction from strict isomorphism to s-isomorphism is clearly 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 (11.0%) and 32 to 14 templates (43.8%), respectively. Recall that two justifications are s-isomorphic, if their different complex subexpressions can be mapped to an atomic variable name. The significant reduction therefore suggests that in these two ontologies complex expressions are frequently used in the same way as atomic concepts.

Closer inspection reveals, however, that a large number of justifications in the Lipid Ontology consist of one axiom of the form $A \equiv B \sqcap x$, with the entailment being $A \sqsubseteq B$. Here, x represents a number of complex expressions of varying nesting depth. While s-isomorphism captures these types of justifications as equivalent, the actual reason for their similarity lies in their identical *cores*, with the remainder x being a *superfluous* part (with respect to the given entailment).

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

Lemma-Isomorphism As with s-isomorphism, the effects of l-isomorphism were not as significant as the first reduction through strict isomorphism. The justifications were 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. This shows that, while not very frequent, there are indeed ontologies in which justifications with subsumption chains of differing lengths occur.

In the *Plant Ontology*, l-isomorphism reduces the 74 justifications to 12 templates, with one template capturing 32 justifications of varying sizes. These justifications contain atomic subsumption chains ranging from 2 to 6 atomic subsumption axioms, which can all be reduced to a single axiom (namely the entailment of the subsumption chain) in the lemmatised version of the justification.

Across the corpus, l-isomorphism reduces the 6,744 justifications to a mere 384 templates, which is an overall reduction of 94.3%. 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.

5.4 Counting Distinct Reasons

For the purpose of detecting how many distinct *types* of justifications there are for a given set of entailments, we have seen that it is crucial to focus not on the *material* form of a justification, but rather on the justification templates in an ontology. By only considering the abstract template of a set of justifications, we can represent the reasoning that underlies not only one, but a whole class of justifications in the ontology.

Surprisingly, while the newly introduced equivalence relations, s-isomorphism and l-isomorphism, could be shown to capture some of the structural similarities in the surveyed corpus, a large number of justifications were indeed strictly isomorphic. We can argue, however, that even a small reduction can be beneficial when presenting multiple justifications to OWL ontology developers for the purpose of explaining entailments, as it prevents repetitive actions and gives a higher-level view of the set of justifications. This is made obvious in the example of the Plant Ontology, where a large number of justifications that differed only in the length of their atomic subsumption chains, could be captured by a single template.

Another aspect to take into account when dealing with OWL justifications is the *superfluousness* of expressions, as we have seen in the case of the Lipid Ontology. Two justifications may be nearly identical and only differ in expressions that are not necessary for the entailment to hold; these superfluous parts would prevent them from being isomorphic with respect to any of the above definitions, but it is clear that their *form* is the same. This situation describes one of several types of justification *masking* [8]. In order to prevent distortion caused by masking effects, we may want to focus on a type of justifications which is minimal with respect to its subexpressions.

A laconic justification [6] is a justification which does not contain any superfluous parts, with every subexpression being as *weak* as possible. A comparison of the equivalence relations between the regular justifications for the above set of ontologies and their laconic versions as part of future work will allow us to gain further insight into the effect of superfluousness on the diversity of justifications.

6 Conclusions and Future Work

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. Since the present paper demonstrates the effects of the different equivalence relations on only a subset of the justifications from the BioPortal corpus, we are planning a survey of the entire set of justifications, taking into account the differences between laconic and non-laconic justifications. Finally, we aim to obtain more detailed knowledge about the application of ontology design patterns in the surveyed ontologies, which will allow us to investigate the relations between these design patterns and the form and frequency of justification templates.

References

1. Baader, F., Küsters, R., Borgida, A., McGuinness, D.: Matching in description logics. Journal of Logic and Computation 9(3), 411–447 (1999)
2. Baader, F., Narendran, P.: Unification of concept terms in description logics. In: Proc. of ECAI-98. pp. 331–335 (1998)
3. Brandt, S.: Implementing matching in $\mathcal{ALE}$—first results. In: Proc. of DL-03 (2003)
4. Davis, M.: Obvious logical inferences. In: Proc. of IJCAI-81. pp. 530–531 (1981)
5. Horridge, M., Bail, S., Parsia, B., Sattler, U.: The cognitive complexity of OWL justifications. In: Proc. of ISWC-11. pp. 241–256 (2011)
6. Horridge, M., Parsia, B., Sattler, U.: Laconic and precise justifications in OWL. In: Proc. of ISWC-08. pp. 323–338 (2008)
7. Horridge, M., Parsia, B., Sattler, U.: Lemmas for justifications in OWL. In: Proc. of DL 2009 (2009)
8. Horridge, M., Parsia, B., Sattler, U.: Justification masking in OWL. In: Proc. of DL 2010 (2010)
9. Horridge, M., Parsia, B., Sattler, U.: Justification oriented proofs in OWL. In: Proc. of ISWC-10. pp. 354–369 (2010)
10. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible $\mathcal{SROIQ}$. In: Proc. of KR-06. pp. 57–67 (2006)
11. Parsia, B., Sirin, E., Kalyanpur, A.: Debugging OWL ontologies. In: Proc. of WWW-05. pp. 633–640 (2005)
12. P.N., Johnson-Laird: Mental models in cognitive science. Cognitive Science 4(1), 71–115 (1980)
13. Schlobach, S., Cornet, R.: Non-standard reasoning services for the debugging of description logic terminologies. In: Proc. of IJCAI-03. pp. 355–362 (2003)