

Categorising Logical Differences Between OWL Ontologies

Rafael S. Gonçalves
School of Computer Science
University of Manchester
Manchester, United Kingdom
goncalvj@cs.man.ac.uk

Bijan Parsia
School of Computer Science
University of Manchester
Manchester, United Kingdom
bparsia@cs.man.ac.uk

Ulrike Sattler
School of Computer Science
University of Manchester
Manchester, United Kingdom
sattler@cs.man.ac.uk

ABSTRACT

The analysis of changes between OWL ontologies (in the form of a *diff*) is an important service for ontology engineering. A purely syntactic analysis of changes is insufficient to distinguish between changes that have logical impact and those that do not. The current state of the art in semantic diffing ignores logically ineffectual changes and lacks any further characterisation of even significant changes. We present a diff method based on an exhaustive categorisation of effectual and ineffectual changes between ontologies. In order to verify the applicability of our approach we apply it to 88 versions of the National Cancer Institute (NCI) Thesaurus (NCIt), and demonstrate that all categories are realized throughout the corpus. Based on the outcome of the NCIt study we argue that the devised categorisation of changes is helpful for ontology engineers and their understanding of changes carried out between ontologies.

Categories and Subject Descriptors

I.2 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods

General Terms

Theory, Experimentation

1. INTRODUCTION

The comparison of ontologies is a valuable service whether used as a subroutine in a version control system or to support users in understanding changes. Different diff methods vary in their sensitivity to changes: e.g., a diff method based on character differences will find that two notationally distinct serializations of the same ontology are radically different. If a diff method is too sensitive to irrelevant changes then the user will be faced with determining which reported changes are actually significant. On the other hand, a hard requirement is that the diff method does not miss any change of significance.

The OWL 2 specification [8] defines a high level notion of syntactic equivalence, so-called “structural equivalence” (and thus the associated notion of *structural difference*), which abstracts from

certain neglectable changes such as the order of axioms or concrete syntax. A different syntactic approach is that of an edit-based diff, wherein change records are produced within the ontology editor being used, thereby capturing the history and intent of change, as implemented in Swoop [6]. The diffs mentioned so far, as well as PROMPTDIFF [9], do not perform any further characterisation of reported changes (e.g., whether these have logical impact). This forces the user to analyse each change in the diff, and determine whether it affects the set of entailments of an ontology; thus whether the change is logically *effectual*. When analysing a set of changes it would be useful to not only distinguish between those logically effectual and ineffectual ones, but also to characterise reported changes according to their impact. Certain classes of ineffectual changes are neglectable, e.g., order of conjuncts in an axiom. However other ineffectual changes provide useful insights, such as the introduction of redundant or rewritten axioms. Semantic diffs, such as CEX [7] or ContentCVS [4], regard all ineffectual changes as neglectable, though in general knowing the relative proportions of effectual and meaningful ineffectual changes gives us a better understanding of what has changed between ontologies. For example, if most changes are meaningfully ineffectual it might be a sign of wasted effort, and therefore it would be useful to know if and why this happens. Contrariwise, while those changes might be *logically* ineffectual, other tools might be sensitive to the variant syntactic forms. So on the one hand, syntactic diffs report without distinction both effectual and ineffectual changes, and on the other hand semantic diffs do not present ineffectual changes.

A major problem with the output of change sets is that the user is given a (possibly large) set of axioms (or terms) to analyse, with no indication as to what kind of change each of them represents. A reasonable presentation of changes will cluster changes according to relevant properties. In this paper we propose a diff method, which we refer to as *intentional difference*, incorporating a categorisation of changes. Our categorisation attempts to capture the impact of each change (e.g., whether it is a rewrite of another axiom). Aside from the intuitive appeal of categorised changes, these rectify a problem with existing diffs: the lack of source axioms. Thus users can see what was changed as well as the actual change.

For the purpose of verifying the suitability of our approach, we collected 88 versions of the National Cancer Institute (NCI) Thesaurus (NCIt) available in OWL, and conducted a diachronic study of the corpus. This study aimed at showing the computational feasibility of our approach and confirming that the devised categories are realized throughout the corpus. Additionally we investigate whether the proposed method helps us understand the evolution of the NCIt. Moreover we demonstrate how the devised categorisation can support change analysis by users.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.

Copyright 2011 ACM 978-1-4503-0717-8/11/10 ...\$10.00.

2. PRELIMINARIES

We assume the reader to be reasonably familiar with ontologies and OWL, as well as the underlying description logics (DLs) [3], though detailed knowledge is not required. We do use the notion of entailment [2], which is identical to the standard first order logic entailment (an axiom α entailed by an ontology $\mathcal{O}$ is denoted $\mathcal{O} \models \alpha$). A *justification* $\mathcal{J}$ of a consequence α is a minimal subset of an ontology $\mathcal{O}$ that is sufficient for α to hold [5]. The *signatures* of an axiom α and an ontology $\mathcal{O}$ are denoted $\tilde{\alpha}$ and $\tilde{\mathcal{O}}$, respectively. We also use the notion of a *locality-based module* [1], and denote a $\perp$ -module extracted from an ontology $\mathcal{O}$ for a set of terms (signature) Σ as $\perp\text{-mod}(\Sigma, \mathcal{O})$.

3. ONTOLOGY DIFFERENCE

The problem of computing the difference between pairs of ontologies has been approached both syntactically and semantically. We distinguish two major aspects of ontology diffing: (i) the detection of changes, and (ii) the presentation of changes to the user. We note that most effort has been largely dedicated to (i). It is often the case that the output (ii) of diff operations is the set of axioms or terms in the diff. While this may reflect the desired detection of change, it does not necessarily convey sufficient information to the user w.r.t. the impact of changes.

3.1 Related Work and Diff Desiderata

Within the detection of changes (i), one would expect a preliminary distinction of axioms in the diff according to their logical effectuality. As such, a purely syntactic change analysis does not suffice to achieve this. Standard semantic diff tools treat all ineffectual changes as neglectable. This goes too far, for example: consider the case where two ontologies differ in a substantial number of axioms, but the axioms in the diff are only equivalences rewritten into subsumptions. Semantic diffs would point out that there are no differences, which may seem counter-intuitive to the user since a shallow inspection of the ontologies would reveal a discrepancy in number of axioms. In this case the ineffectual changes point to possibly unnecessary work. Though if these were intentional, then other developers should be aware of it rather than rewriting the axioms once again. So for this kind of change we would expect a more granular analysis to be helpful.

The second fundamental aspect of any diff is the presentation of changes to the user (ii). Currently ontology diffs return as the output an unstructured, uncharacterised set of changes. As a consequence the task of change analysis is not particularly appealing. Consider the fact that the average diff size across the NCIt corpus is over 6,000 changes; relying on current diff methods for change analysis would be frightening, to say the least. At this point it would be useful to determine further properties of individual changes which might help the user understand changes. A change could relate to, e.g., a newly introduced term, or an adjustment to the class hierarchy. Whatever it may be, the current state-of-the-art in ontology diffing does not carry out such a characterisation of changes.

In terms of computability, one would expect an ontology diff to be efficiently computable for OWL 2 ontologies.² CEX [7], e.g., computes differences efficiently only for a fragment of OWL.³

In Table 1 we summarise such useful features of an ontology diff, and whether existing approaches exhibit such desiderata. These

properties motivate the diff method proposed in this paper, and our categorisation of changes according to their impact.

3.2 Diff Categories

Given the limitations of diff approaches described in Table 1, we address in this paper the problem of categorising changes between ontologies. We start with structural difference as it abstracts from some obviously pointless changes, and is part of the OWL W3C standard. The notion of structural difference is based on OWL's notion of structural equivalence (denoted $\equiv_s$) [8], and rules out differences that an otherwise syntactic equality based diff would detect, such as the concrete syntax of ontologies, the order of axioms or the order of disjuncts and conjuncts in an axiom. Furthermore we would like to distinguish between (a) axioms that do not alter the set of entailments between ontologies (ineffectual), and (b) those axioms that do (effectual). Structural difference does not realize this, and so we build on this notion with an effectuality test. The latter involves checking if changes in the first ontology are entailed by the second (and vice-versa). Finally, we devise a categorisation mechanism of these changes, in order to characterise their purported impact. Let us begin by defining how the change sets are obtained:

DEFINITION 1 (STRUCTURAL DIFFERENCE [4]). *The structural difference between $\mathcal{O}_1$ and $\mathcal{O}_2$ are the following sets:*

- $\text{Adds}(\mathcal{O}_1, \mathcal{O}_2) = \{\beta \in \mathcal{O}_2 \mid \nexists \alpha \in \mathcal{O}_1 \text{ s.t. } \alpha \equiv_s \beta\}$
- $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2) = \{\alpha \in \mathcal{O}_1 \mid \nexists \beta \in \mathcal{O}_2 \text{ s.t. } \alpha \equiv_s \beta\}$

So if there is an axiom β s.t. $\beta \in \text{Adds}(\mathcal{O}_1, \mathcal{O}_2)$, this implies that $\beta \in \mathcal{O}_2 \setminus \mathcal{O}_1$, and analogously for $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$. Intuitively $\text{Adds}(\mathcal{O}_1, \mathcal{O}_2)$ represent additions from $\mathcal{O}_1$ to $\mathcal{O}_2$, and $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$ the removals. Based on these two sets, next, we distinguish between axioms in $\text{Adds}(\mathcal{O}_1, \mathcal{O}_2)$ (or $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$) that affect the set of entailments of $\mathcal{O}_1$ (or $\mathcal{O}_2$). In other words, we distinguish between those axioms in the structural difference which are entailed by $\mathcal{O}_1$ (or $\mathcal{O}_2$), as follows:

DEFINITION 2. *The effectual and ineffectual changes between $\mathcal{O}_1$ and $\mathcal{O}_2$ are the following sets:*

- $\text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2) = \{\beta \in \text{Adds}(\mathcal{O}_1, \mathcal{O}_2) \mid \mathcal{O}_1 \not\models \beta\}$
- $\text{EffRems}(\mathcal{O}_1, \mathcal{O}_2) = \{\alpha \in \text{Rems}(\mathcal{O}_1, \mathcal{O}_2) \mid \mathcal{O}_2 \not\models \alpha\}$
- $\text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2) = \text{Adds}(\mathcal{O}_1, \mathcal{O}_2) \setminus \text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2)$
- $\text{IneffRems}(\mathcal{O}_1, \mathcal{O}_2) = \text{Rems}(\mathcal{O}_1, \mathcal{O}_2) \setminus \text{EffRems}(\mathcal{O}_1, \mathcal{O}_2)$

The resulting sets of ineffectual additions $\text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ and removals $\text{IneffRems}(\mathcal{O}_1, \mathcal{O}_2)$ are composed of those axioms which do not change the set of entailments of $\mathcal{O}_1$ and $\mathcal{O}_2$, but were added or deleted, respectively. The effectual additions and removals, $\text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ and $\text{EffRems}(\mathcal{O}_1, \mathcal{O}_2)$, represent those axioms that have logical impact from $\mathcal{O}_1$ to $\mathcal{O}_2$ (and vice-versa).

While semantics diffs would consider all ineffectual changes as irrelevant, we here attempt to characterise meaningful ineffectual changes in order to get a sense of how relevant these are. For that purpose, we devise a categorisation of such axioms based on the reasons for the lack of impact, i.e. justifications. Note that this categorisation is not a partitioning of the set of ineffectual changes.

DEFINITION 3 (INEFFECTUAL CHANGE CATEGORISATION). *An axiom $\alpha \in \text{IneffRems}(\mathcal{O}_1, \mathcal{O}_2)$ is:*

- Rewritten, if there is a justification $\mathcal{J}$ for α with $\mathcal{J} \cap \text{Adds}(\mathcal{O}_1, \mathcal{O}_2) \neq \emptyset$, and $\alpha \models \mathcal{J}$. If $\mathcal{J} \subseteq \text{Adds}(\mathcal{O}_1, \mathcal{O}_2)$ then α is a complete rewrite, otherwise a partial rewrite.
- Strengthened, if there is a $\mathcal{J}$ for α with $\mathcal{J} \cap \text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2) \neq \emptyset$.

¹<http://krizik.felk.cvut.cz/km/owldiff/>

²Based on DLs up to *SRQIQ* [3].

³Specifically acyclic *EL*-terminologies [7].

- Redundant, if there is a $\mathcal{J}$ for α with $\mathcal{J} \subseteq (\mathcal{O}_1 \cap \mathcal{O}_2)$. If $\mathcal{J} \subseteq (\mathcal{O}_1 \cap \mathcal{O}_2) \cup \text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ then α is an avoided redundancy.

For axioms $\beta \in \text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$, replace α , $\text{Adds}(\mathcal{O}_1, \mathcal{O}_2)$, $\text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ and $\text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ with β , $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$, $\text{EffRems}(\mathcal{O}_1, \mathcal{O}_2)$ and $\text{IneffRems}(\mathcal{O}_1, \mathcal{O}_2)$ respectively. In $\text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ the label for the criteria of Strengthened axioms changes to Weakened axioms.

The term ‘avoided redundancy’ is used in the sense that, given a removed avoided redundancy $\alpha \in \mathcal{O}_1$, it would have been a redundancy in $\mathcal{O}_2$ if not removed. The ineffectual axiom categorisation returns possibly overlapping sets of axioms, as demonstrated in Section 3.3. Also we note that these categories are exhaustive, in such a way that for any axiom we have that it implies at least one of the defined categories. Consider an axiom α and ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$, with $\alpha \in \mathcal{O}_1$ but $\alpha \notin \mathcal{O}_2$, and $\mathcal{O}_2 \models \alpha$. Then there must be a justification $\mathcal{J} \subseteq \mathcal{O}_2$ for α . If we have that $\mathcal{J} \subseteq (\mathcal{O}_1 \cap \mathcal{O}_2) \cup \text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$, then α is redundant, otherwise if we have that $\mathcal{J} \cap \text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2) \neq \emptyset$, then α is strengthened.

The effectual changes are categorised according to the following definition:

DEFINITION 4 (EFFECTUAL CHANGE CATEGORISATION).
An axiom $\beta \in \text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2)$ is categorised as follows:

- If $\tilde{\beta} \subseteq \tilde{\mathcal{O}}_1$, β is an element of:
 - $\text{Strgth}(\mathcal{O}_1, \mathcal{O}_2)$, if there exists an axiom $\alpha \in \text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$ s.t. $\beta \models \alpha$.
 - $\text{PAdd}(\mathcal{O}_1, \mathcal{O}_2)$, otherwise, i.e., if there is no $\alpha \in \mathcal{O}_1$ s.t. $\beta \models \alpha$.
- If $\tilde{\beta} \not\subseteq \tilde{\mathcal{O}}_1$, let $T := \tilde{\beta} \setminus \tilde{\mathcal{O}}_1$. Then β is an element of:
 - $\text{StrgthNT}(\mathcal{O}_1, \mathcal{O}_2)$, if there exists an axiom $\alpha \in \text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$ s.t. $\beta \models \alpha$.
 - $\text{NewDesc}(\mathcal{O}_1, \mathcal{O}_2)$, if we have that $\perp\text{-mod}(T, \{\beta\}) \neq \emptyset$.
 - $\text{PAddNT}(\mathcal{O}_1, \mathcal{O}_2)$, otherwise, i.e., if we have that $\beta \notin \text{NewDesc}(\mathcal{O}_1, \mathcal{O}_2)$ and there is no $\alpha \in \mathcal{O}_1$ s.t. $\beta \models \alpha$.

$\text{Strgth}(\mathcal{O}_1, \mathcal{O}_2)$ represents strengthenings from $\mathcal{O}_1$ to $\mathcal{O}_2$, $\text{PAdd}(\mathcal{O}_1, \mathcal{O}_2)$ pure additions, and $\text{NewDesc}(\mathcal{O}_1, \mathcal{O}_2)$ new descriptions. Categories ending in ‘NT’ represent axioms that use new terms. For axioms $\alpha \in \text{EffRems}(\mathcal{O}_1, \mathcal{O}_2)$ replace β , $\mathcal{O}_1$, and $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2)$ with α , $\mathcal{O}_2$, and $\text{Adds}(\mathcal{O}_1, \mathcal{O}_2)$, accordingly. We denote effectual removals as $\text{Weakng}(\mathcal{O}_1, \mathcal{O}_2)$ for weakenings and $\text{RetDesc}(\mathcal{O}_1, \mathcal{O}_2)$ for retired descriptions, while categories involving axioms with retired terms are followed by *RT* (e.g., $\text{PRemRT}(\mathcal{O}_1, \mathcal{O}_2)$ represents pure removals with retired terms).

Observe that the categories encompass two dimensions of changes; those changes involving new (or retired) terms, and changes involving shared terms between both ontologies (i.e. terms in $\tilde{\mathcal{O}}_1 \cap \tilde{\mathcal{O}}_2$). This duality of categories depending on the signature of axioms is illustrated in Table 2. We can see that there cannot be a strengthening which involves retired terms, or analogously a weakening involving new terms. Similarly new or retired descriptions are based on new or retired terms, accordingly.

Table 2: Effectual change categories.

Axiom terms	Shared terms	New terms	Retired terms
Strengthening	✓	✓	N/A
Weakening	✓	N/A	✓
Pure change	✓	✓	✓
New Description	N/A	✓	N/A
Retired Description	N/A	N/A	✓

3.3 Diff Walkthrough

In order to demonstrate the usefulness of the devised categories, let us compare ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$, defined in Table 3.

Table 3: Example ontologies.

$\mathcal{O}_1$	$\mathcal{O}_2$
$\alpha_1 : A \sqsubseteq C$	$\beta_1 : A \sqsubseteq B \sqcup C$
$\alpha_2 : B \sqsubseteq C$	$\beta_2 : A \sqsubseteq B$
$\alpha_3 : E \equiv D$	$\beta_3 : B \sqsubseteq C$
$\alpha_4 : D \sqsubseteq F$	$\beta_4 : E \sqsubseteq D$
$\alpha_5 : F \sqsubseteq G$	$\beta_5 : D \sqsubseteq E$
$\alpha_6 : G \sqsubseteq H \sqcap \exists s.H$	$\beta_6 : E \sqsubseteq B \sqcup \exists r.C$
$\alpha_7 : F \sqsubseteq I$	$\beta_7 : D \sqsubseteq E \sqcup G$
$\alpha_8 : F \sqsubseteq G \sqcap I \sqcap J$	$\beta_8 : G \sqsubseteq \exists s.H \sqcap H$
	$\beta_9 : F \sqsubseteq G \sqcap I$
	$\beta_{10} : K \sqsubseteq \exists r.F$
	$\beta_{11} : D \sqsubseteq F \sqcap \exists s.A$
	$\beta_{12} : D \sqsubseteq F \sqcap \exists p.T$
	$\beta_{13} : B \sqsubseteq K$

From $\mathcal{O}_1$ and $\mathcal{O}_2$ we have the following structural differences:

- ◊ $\text{Adds}(\mathcal{O}_1, \mathcal{O}_2) = \{\beta_1, \beta_2, \beta_4, \beta_5, \beta_6, \beta_7, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}\}$
- ◊ $\text{Rems}(\mathcal{O}_1, \mathcal{O}_2) = \{\alpha_1, \alpha_3, \alpha_4, \alpha_5, \alpha_7, \alpha_8\}$

Note that α_6 is not syntactically equal to β_8 ($\alpha_6 \neq \beta_8$), however they are structurally equivalent ($\alpha_6 \equiv_s \beta_8$). Therefore these axioms are not reported as changes. Given the sets of structural additions and removals from $\mathcal{O}_1$ to $\mathcal{O}_2$ we have that:

- ◊ $\text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2) = \{\beta_2, \beta_6, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}\}$
- ◊ $\text{EffRems}(\mathcal{O}_1, \mathcal{O}_2) = \{\alpha_8\}$
- ◊ $\text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2) = \{\beta_1, \beta_4, \beta_5, \beta_7, \beta_9\}$
- ◊ $\text{IneffRems}(\mathcal{O}_1, \mathcal{O}_2) = \{\alpha_1, \alpha_3, \alpha_4, \alpha_5, \alpha_7\}$

There are several ineffectual changes in the change set, while effectual changes are mostly additions (and a single removal). The

Table 1: Desiderata of ontology diffing approaches.

Properties	Structural Diff	ContentCVS	CEX	OWLDiff ^l	PROMPTDIFF	Swoop Diff
Difference Detection						
Syntactic analysis	✓	✓	✗	✓	✓	✓
Semantic analysis	✗	✓	✓	✓	✗	✗
Effectively computable	✓	✓	✗	✓	✓	✓
OWL 2 adequacy	✓	✓	✗	✓	✗	✓
Difference Output						
Axiom-based	✓	✓	✓	✓	✗	✓
Term-based	✗	✗	✓	✗	✓	✓
Change characterisation	✗	✗	✗	✗	✗	✗

Table 4: Categorisation of axioms in $\text{diff}(\mathcal{O}_1, \mathcal{O}_2)$.

Removals	Axioms	Pairing
Rewritten (I)	α_3	$\{\beta_4, \beta_5\}$
Strengthened (I)	α_1	$\{\beta_2, \beta_3\}$
	α_4	$\{\beta_{11}\}, \{\beta_{12}\}$
Redundant (I)	α_1	$\{\beta_2, \beta_3\}, \{\beta_1, \beta_3\}$
	α_3	$\{\beta_4, \beta_5\}$
	α_5, α_7	$\{\beta_9\}$
WeakeningRT	α_8	of $\beta_9 (\alpha_8 \models \beta_9 \wedge J \notin \tilde{\mathcal{O}}_2)$
Additions	Axioms	Pairing
Rewritten (I)	β_9	$\{\alpha_5, \alpha_7\}$
Weakened (I)	β_9	$\{\alpha_8\}$
Redundant (I)	β_1	$\{\alpha_1\}$
	β_4, β_5	$\{\alpha_3\}$
	β_7	$\{\alpha_3\}, \{\alpha_4, \alpha_5\}$
	β_9	$\{\alpha_5, \alpha_7\}$
Strengthening	β_{11}	of $\alpha_4 (\beta_{11} \models \alpha_4)$
StrengtheningNT	β_{12}	of $\alpha_4 (\beta_{12} \models \alpha_4 \wedge p \notin \tilde{\mathcal{O}}_1)$
NewDescription	β_{10}	$K \notin \tilde{\mathcal{O}}_1 \wedge \perp\text{-mod}(\{K\}, \{\beta_{10}\}) \neq \emptyset$
PureAddition	β_2	$\tilde{\beta}_2 \subseteq \tilde{\mathcal{O}}_1 \wedge \beta_2 \notin \text{NewDesc}(\mathcal{O}_1, \mathcal{O}_2)$ $\wedge \nexists \alpha \in \mathcal{O}_1 \text{ s.t. } \beta_2 \models \alpha$
PureAdditionNT	β_6	$K \notin \tilde{\mathcal{O}}_1 \wedge \beta_6 \notin \text{NewDesc}(\mathcal{O}_1, \mathcal{O}_2)$ $\wedge \nexists \alpha \in \mathcal{O}_1 \text{ s.t. } \beta_6 \models \alpha$
	β_{13}	$r \notin \tilde{\mathcal{O}}_1 \wedge \beta_{13} \notin \text{NewDesc}(\mathcal{O}_1, \mathcal{O}_2)$ $\wedge \nexists \alpha \in \mathcal{O}_1 \text{ s.t. } \beta_{13} \models \alpha$

changes are categorised as shown in Table 4, where ineffectual categories are labeled with (I).

Let us consider two ineffectual additions; β_9 is a rewrite of $\{\alpha_7, \alpha_5\}$, as well as an avoided redundancy (i.e., had it been added to $\mathcal{O}_1$ it would be redundant). The axiom is also weakened, due to α_8 . This may seem like an unintentional change, since now we face a loss of information regarding J , which is no longer mentioned in $\mathcal{O}_2$. Such a change may be worth revising. The axiom β_1 is redundant, since we have from $\mathcal{O}_1$ that $A \sqsubseteq C$, which is also entailed from $\mathcal{O}_2$. Therefore the user can dispose of this axiom.

Bear in mind that the existence of a rewritten axiom from $\mathcal{O}_1$ to $\mathcal{O}_2$ does not imply that the same holds in the opposite direction.

This is applicable to all categories. Also we can have that an axiom is in more than one category. Consider axiom α_1 ; a justification $\mathcal{J}_1$ for α_1 is $\mathcal{J}_1 = \{\beta_2, \beta_3\}$, which indicates a strengthening (since we have that $\beta_2 \in \text{EffAdds}(\mathcal{O}_1, \mathcal{O}_2)$), as well as a redundancy ($\beta_3 \in \mathcal{O}_1 \cap \mathcal{O}_2$). Another justification $\mathcal{J}_2 = \{\beta_1, \beta_3\}$ indicates a redundancy; $\beta_1 \in \text{IneffAdds}(\mathcal{O}_1, \mathcal{O}_2)$.

In terms of effectual changes there is only one removal, and six additions. The effectual removal (α_8) represents a weakening of β_9 with retired terms (J is not mentioned in $\mathcal{O}_2$). In the analysis of the ineffectual changes it was already noted that axiom α_8 should be revised. The pure additions appear to be adjustments to the class hierarchy, some associated with new terms in $\mathcal{O}_2$. Both axioms β_{11} and β_{12} are strengthenings of α_4 , which suggests that they could be merged, especially since there is intra-axiom redundancy. Finally there is a new term K in $\mathcal{O}_2$ being described via axiom β_{10} .

Generally speaking, with such a categorisation it becomes conceivably more intuitive to navigate and understand sets of changes. In addition to this, we gathered from the analysis of ineffectual changes useful information about the changes between $\mathcal{O}_1$ and $\mathcal{O}_2$, e.g., that axiom α_4 is strengthened in two distinct, yet partially superfluous axioms (β_{11} and β_{12}). Similarly we discover that axiom β_9 is weakened, from α_8 , which should be reconsidered as we now have that $\mathcal{O}_2 \not\models F \sqsubseteq J$ (J becoming a retired term).

4. CASE STUDY

In order to evaluate our approach to ontology diffing, we carried out a study of the NCIt corpus. We extracted all pairwise, consecutive diffs between NCIt versions, and show that all categories are instantiated throughout the corpus.

The NCIt archive⁴ contains 88 versions of the ontology in OWL format, freely downloadable from the web. Two of these versions were unparseable (releases 05.03F and 05.04d) with the OWL API⁵ and consequently Protégé⁶. The experiment machine is an Intel Xeon Quad-Core 3.20GHz, with 12Gb DDR3 RAM dedicated to the Java Virtual Machine (JVM v1.5). The system runs Mac OS X 10.6.7, and all tests were run using the OWL API (v3.1). All gathered test data is available from <http://owl.cs.manchester.ac.uk/ncit>, a part of it is published on Google Public Data Explorer,⁷ and can be visualised at <http://bit.ly/leZ6fM>.

⁴http://evs.nci.nih.gov/ftp1/NCI_Thesaurus

⁵<http://owlapi.sourceforge.net/>

⁶<http://protege.stanford.edu/>

⁷<http://www.google.com/publicdata/home>

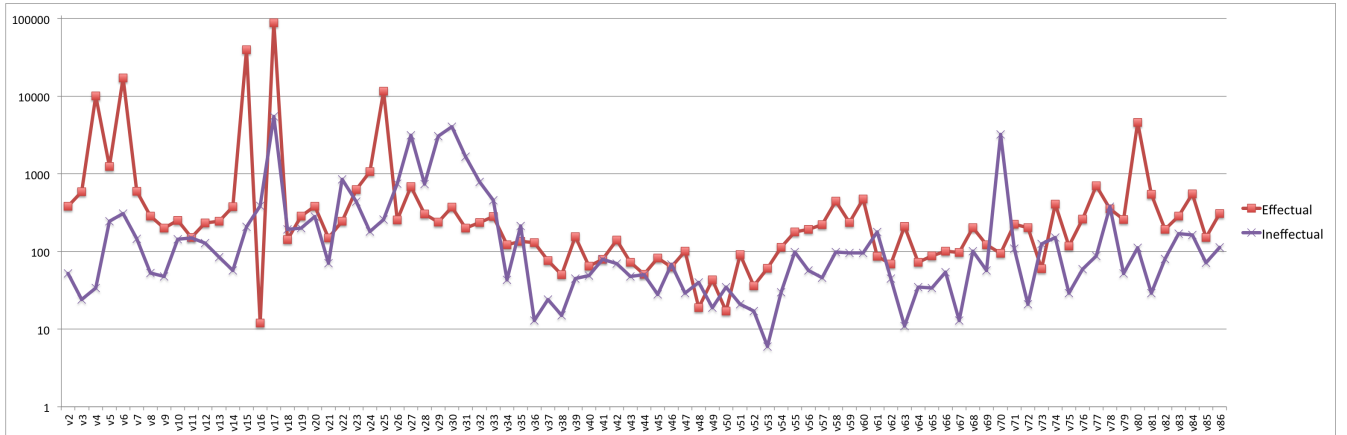


Figure 1: Breakdown of effectual vs ineffectual removals (logarithmic scale, number of axioms).

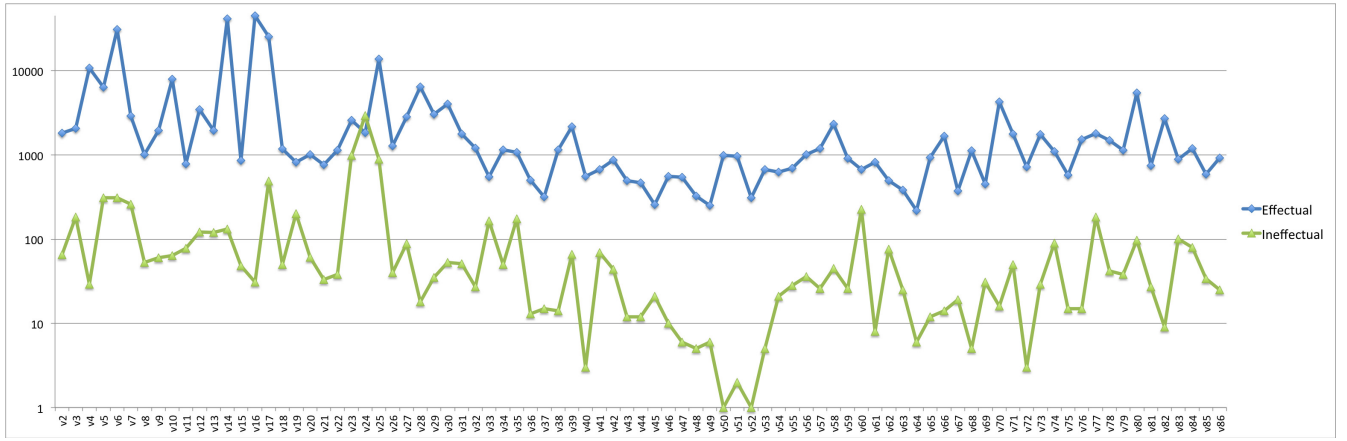


Figure 2: Breakdown of effectual vs ineffectual additions (logarithmic scale, number of axioms).

Table 5: Sum of the results of $\text{diff}(\mathcal{O}_i, \mathcal{O}_{i+1})$, for $1 \leq i \leq 86$.

Change type		Additions			Removals		
Effectual	Axiom terms	Strengthening	New Description	Pure Addition	Weakening	Retired Description	Pure Removal
	Shared terms	4,649	N/A	36,499	1,122	N/A	25,022
	New terms	4,948	170,095	69,333	N/A	N/A	N/A
	Retired terms	N/A	N/A	N/A	508	52,229	23,634
Ineffectual		Rewrite	Weakened	Redundant	Rewrite	Strengthened	Redundant
		115	8,576	4,186	114	30,370	8,012

4.1 Effectual vs Ineffectual Changes

The logical difference throughout the NCIt time-line consists of an average of 87% effectual changes, while the remaining are ineffectual (see Figures 1 and 2 and Table 5). There are, however, cases where the number of effectual and ineffectual changes is more balanced, such as $\mathcal{O}_{24}$ with 48% effectual changes, or $\mathcal{O}_{29}$ and $\mathcal{O}_{30}$ with 52% effectual changes out of all changes. Note that in $\mathcal{O}_{14}$ and $\mathcal{O}_{16}$ the number of effectual changes is (unusually) well above the number of ineffectual changes. In these two NCIt versions there is an annotation property and a data property both named “code”, which in the version of the OWL API used results in the parsing of only data property assertions.⁸

Throughout the NCIt study we have excluded a class of ineffectual changes: changes to annotations. Since there are ontologies whose major focus is the annotations, as the NCIt, these represent a topic of study in themselves. E.g., in the NCIt the average proportion of axiomatic changes throughout the corpus is 15%, while the remaining are annotation changes. But for the purposes of this paper we focus on axiom changes only.

4.2 Ineffectual Changes

Throughout the corpus there is a high number of ineffectual removals, reaching values of 93% in $\mathcal{O}_{29}$ or 97% in $\mathcal{O}_{16}$, and with an average of 35% of all removals (see Figure 1). Out of these ineffectual removals 92% turned out to be strengthened axioms, which indicates a continuous refining of information throughout the corpus. E.g. $\mathcal{O}_{27}$ has 3,104 strengthened axioms out of 3,843 removals. There is a high number of removed redundancies as well, constituting 49% of ineffectual removals. This tells us that there is some pruning of redundant information going on in the corpus. Taking into account that only 23% of these removed redundancies are avoided redundancies, we see that there is some redundancy removal throughout. On average 5% of additions are ineffectual (see Figure 2), yet there are some high values such as 61% in $\mathcal{O}_{24}$.

Among these, 73% are added redundancies, and 82% are weakened axioms. The latter occur typically due to adjustments in the class hierarchy. Despite being a high percentage, the number of ineffectual additions is generally low (see Table 5).

We also identified a number of rewrites in the corpus. Particularly in $\mathcal{O}_{33}$, there are 227 rewritten axioms. Upon inspecting the rewritten axioms, we noticed that these changes are not only syntactic but also trivial and easily detected. While ideally the underlying structural diff would not include these, at least with our categorisation and alignment with source axioms, it is easy to spot and recognize the triviality. In this version we also see a number of redundancies above average, with 202 removed redundancies and 157 added. Clearly certain ineffectual changes are in fact “refactorings”, albeit in the case of strengthened and weakened axioms the refactoring would have to be of a set of axioms rather than a single axiom. Thus a strengthened axiom does not necessarily mean strengthening of the ontology since the change might either introduce a redundancy or redistribute information from other axioms. Consider an ontology $\mathcal{O}_1 = \{\alpha_1 : A \sqsubseteq B, \alpha_2 : A \sqsubseteq C\}$, and a change of α_1 into $A \sqsubseteq B \sqcap C$. The axiom α_1 was strengthened, but the resulting ontology $\mathcal{O}_2 = \{\alpha_1 : A \sqsubseteq B \sqcap C, \alpha_2 : A \sqsubseteq C\}$ was not. However, if we change $\alpha_2 \in \mathcal{O}_2$ into $A \sqsubseteq C \sqcap D$, then we can say both the axiom α_2 and the ontology $\mathcal{O}_2$ are strengthened.

We noted a somewhat high number of added redundancies in the corpus, with more incidence up until $\mathcal{O}_8$ (average of 166 added redundancies per comparison). However there are some outliers in the rest of the corpus, such as $\mathcal{O}_{35}$ with 174 added redundant axioms. The highest value found is in $\mathcal{O}_{17}$, where 482 redundant axioms were added. Upon investigating these axioms, we found that such added redundancies are, in many cases, entailments from previous versions. These entailments are those derived from the transitivity of the subclass relationship, but are not atomic subsumptions e.g. $\mathcal{O}_1 = \{\alpha_1 : A \sqsubseteq \exists r.B, \alpha_2 : C \sqsubseteq A\}$, $\mathcal{O}_2 = \{\alpha_1, \alpha_2, \alpha_3 : C \sqsubseteq \exists r.B\}$. From the example we see that α_3 is redundant; $C \sqsubseteq A$ suffices for $C \sqsubseteq \exists r.B$ to hold, yet entailments of this form constitute the majority of added redundancies.

⁸In turn, this lead to a mass-generation of individuals.

In cases where the number of ineffectual changes is quite high (e.g., $\mathcal{O}_{24}$ where 52% of changes are ineffectual, $\mathcal{O}_{27}$, $\mathcal{O}_{29}$ and $\mathcal{O}_{30}$ with 48% each) note that semantic diffs would significantly understate the amount of activity performed. While structural diff captures this, it does not analyse the logical impact of such changes.

4.3 Effectual Changes

We see in Figures 1 and 2 that the number of effectual changes is typically higher than the number of ineffectual changes. There are more additions than removals throughout the corpus, with an average of 60% additions versus 40% removals, which is hardly surprising for a constantly evolving ontology. Among the categories of effectual changes we discovered that the majority of these are new and retired descriptions. In terms of effectual additions the average of new descriptions is 60%, while retired descriptions average 51% of effectual removals. The highest values of new descriptions are in $\mathcal{O}_{14}$ (40,417), $\mathcal{O}_{16}$ (43,522) and $\mathcal{O}_6$ (18,912). This high number of new descriptions is unsurprising, as the terminology keeps increasing. We see in $\mathcal{O}_{15}$ and $\mathcal{O}_6$ high values of retired descriptions, with 39,584 in $\mathcal{O}_{15}$ and 11,835 in $\mathcal{O}_6$. This does not prove that such concepts were deleted, it could instead suggest a high number of concept renamings. Bear in mind that the number of changes in $\mathcal{O}_{14}$, $\mathcal{O}_{16}$, and adjacent versions is distorted by the parsing anomaly discussed in Section 4.1.

Strengthenings average around 4% of all additions, indicating refinements of concepts with additional constraints. There are cases where the number of strengthenings is high, such as $\mathcal{O}_{27}$ which has 142 with new terms and 1065 without. Also in $\mathcal{O}_{29}$ there are 102 strengthenings with new terms, and 891 without. It is natural that upon introducing new terms, others need to be re-described, thus explaining the strengthenings with new terms.

Pure additions account for 37% of all additions, divided between 24% of additions with new terms and 13% without. The highest number of pure additions with new terms is in $\mathcal{O}_{25}$, with 11,746 axioms. This version also contains the most pure removals with retired terms, amounting to 10,645 axioms. With respect to pure additions with shared terms $\mathcal{O}_{80}$ contains the most, amounting to 4,489 axioms. Coincidentally this same version also has the highest number of pure removals with shared terms, with a total of 4,583 such axioms. Typically pure changes with shared terms suggest adjustments to the class hierarchy, while pure additions involving new terms point to the insertion of said terms and subsequent re-adjustment of the hierarchy accordingly.

The average of pure removals throughout the corpus is 47%, split between 23% with retired terms and 24% without. There are not as many weakenings in the corpus as there are strengthenings. This tells us that typically there is not much reduction of information from version to version. The average of weakenings (with or without retired terms) throughout is below 2% of all removals, with the highest value being 69% (739) in $\mathcal{O}_{24}$. The next version, $\mathcal{O}_{25}$, contains the highest number of weakenings with retired terms (508).

Table 5 demonstrates that all effectual categories are instantiated in the NCIt corpus. For instance the diff between versions $\mathcal{O}_{32}$ and $\mathcal{O}_{33}$ contains axioms in nearly all categories, with the exception of weakenings with retired terms.

5. DISCUSSION AND OUTLOOK

The diachronic study of the NCIt revealed that all the categories of changes occur throughout the corpus, providing indications as to the impact of changes. We showed that such categorisation of change sets is useful for change analysis, with particular benefits for division of labour by ontology engineers. By means of this categorisation we can group changes according to their impact, al-

lowing users to shift their attention to specific types of changes, rather than going through a change set while inspecting both ontologies. With our correspondence of changes between ontologies we can show the changed axioms and what they are a change of. Consequently, by analysing changes in this way there is no need for constantly having to inspect ontologies manually. As such, we can support users in understanding the impact of their changes (or lack thereof), and refine these before publishing newer versions.

We found that ineffectual changes account for a significant amount of changes throughout the NCIt. Despite the fact that semantic diffs ignore these changes in their output, we show that they provide helpful modelling insights, and thus are worth examining. For instance we discovered a high number of redundant axioms in the NCIt, some of which could be disposed of. Also we found a number of structurally distinct ineffectual changes that are clearly neglectable: the rewrites in the NCIt. These indicate the need for improvement of the underlying diff. In general the inspection of ineffectual changes is helpful to prevent, e.g., re-doing work or introducing redundancy. Relying on semantic diffs one would be missing out on these meaningful ineffectual changes, which in turn could help users recognize the impact of certain types of change.

The next step in our study is to evaluate the diff tool with users. In particular we expect to confirm that the categorisation helps users in understanding changes between ontologies. In terms of further analysis of the NCIt corpus we intend to inspect the history of axioms, as in checking the progress of each axiom throughout the corpus since it was introduced (e.g., changes in constructors or strengthenings that the axiom went through). Another future survey involves checking for patterns of change throughout the corpus, e.g., if all strengthenings exhibit a common form.

6. REFERENCES

- [1] B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler. Modular reuse of ontologies: Theory and practice. *J. of Artificial Intelligence Research*, 31:273–318, 2008.
- [2] B. Cuenca Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler. OWL 2: The next step for OWL. *J. of Web Semantics*, 2008.
- [3] I. Horrocks, O. Kutz, and U. Sattler. The even more irresistible *SR*OTQ. In *Proc. of KR-06*, pages 57–67, 2006.
- [4] E. Jiménez-Ruiz, B. Cuenca Grau, I. Horrocks, and R. Berlanga Llavori. Supporting concurrent ontology development: Framework, algorithms and tool. *Data and Knowledge Engineering*, 70(1):146–164, 2011.
- [5] A. Kalyanpur, B. Parsia, M. Horridge, and E. Sirin. Finding all justifications of OWL DL entailments. In *Proc. of ISWC/ASWC*, 2007.
- [6] A. Kalyanpur, B. Parsia, E. Sirin, B. Cuenca Grau, and J. Hendler. Swoop: A Web ontology editing browser. *J. of Web Semantics*, 4(2):144–153, 2006.
- [7] B. Konev, D. Walther, and F. Wolter. The logical difference problem for description logic terminologies. In *IJCAR-08*, volume 5195 of *LNCS*, 2008.
- [8] B. Motik, P. Patel-Schneider, and B. Parsia. *OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax*. W3C Cand. Rec., 11 June 2009. <http://www.w3.org/TR/2009/CR-owl2-syntax-20090611/>.
- [9] N. F. Noy and M. A. Musen. PROMPTDIFF: A fixed-point algorithm for comparing ontology versions. In *Proc. of AAAI-02*, pages 744–750, 2002.