

A System for Debugging Taxonomies and their Alignments

Valentina Ivanova and Patrick Lambrix

Department of Computer and Information Science
and Swedish e-Science Research Centre
Linköping University, SE-581 83 Linköping, Sweden

Abstract. Neither developing ontologies nor aligning ontologies are easy tasks, and often the resulting ontologies and alignments are not consistent or complete. Such ontologies and alignments, although often useful, also lead to problems when used in semantically-enabled applications. In this paper we briefly introduce a system that supports domain experts in detecting and repairing wrong and missing is-a relations and mappings.

1 Introduction

Neither developing nor aligning ontologies are easy tasks, and often the resulting ontologies and alignments are not consistent or complete. Such ontologies and alignments, although often useful, also lead to problems when used in semantically-enabled applications. Wrong conclusions may be derived or valid conclusions may be missed.

RepOSE (*Repair of Ontological Structure Environment*) tackles the problem of debugging the is-a structure of a fundamental kind of ontologies, i.e., taxonomies, as well as the debugging of the mappings between taxonomies.

In this demonstration paper we briefly introduce RepOSE¹ (Section 2) and some experiments and projects in which RepOSE was used (Section 3). However, for the theoretical background, algorithms, more detailed descriptions and related work we refer to [3]. For more detailed descriptions for the first and second cases in Section 3, we refer to [3, 2]. In Section 4 we introduce the demonstration at the First International Workshop on Debugging Ontologies and Ontology Mappings.

2 System

The input to RepOSE is an ontology network consisting of taxonomies and alignments between the taxonomies. The debugging process consists of the phases of detecting and validating possible defects, and repairing wrong and missing is-a relations and mappings. At any time during the process, the user can switch between different ontologies and alignments, start earlier phases, or switch between the repairing of wrong and missing is-a relations and mappings. For each of the steps in the debugging process, RepOSE

¹ The version of RepOSE described in this paper is an extension of earlier described versions. Previous versions dealt only with missing and/or wrong is-a relations, but not with mappings.

can recommend possible actions. The process ends when there are no more defects or defect suggestions to deal with.

In the current version of RepOSE we have focused on detecting defects using the knowledge inherent in the network. RepOSE suggests defects in the form of candidate missing is-a relations and mappings. Candidate missing is-a relations in an ontology are is-a relations that can be derived from the network but not from the ontology alone. Candidate missing mappings between two ontologies are mappings that can be derived from the network but not from the ontologies and their alignment alone. These candidate missing is-a relations and mappings are then validated by a domain expert and classified as missing and wrong is-a relations and mappings (Figure 1).

For these defects RepOSE computes repairing actions, i.e., is-a relations or mappings to add to and remove from the ontologies and the alignments such that (i) the missing is-a relations will be derivable from their host ontologies (ii) the missing mappings will be derivable from the host ontologies of the concepts in the mappings, and their alignment, and (iii) the wrong is-a relations and mappings will not be derivable from the ontology network. For wrong is-a relations and mappings RepOSE shows their justification and the domain expert can select is-a relations and mappings to remove (Figure 2)². For missing is-a relations, RepOSE shows two panels, where it is guaranteed that when an is-a relation or mapping is added between an element in the first panel and an element in the second panel, the missing is-a relation or mapping will be repaired (Figure 3). Upon repairing, RepOSE computes all the consequences of the repair.

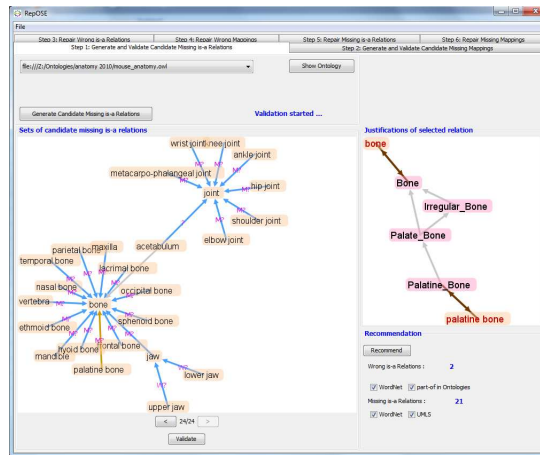


Fig. 1. Generating and validating candidate missing is-a relations and mappings.

² The screenshots in the figures are for tabs related to is-a relations. Similar tabs exist related to mappings.

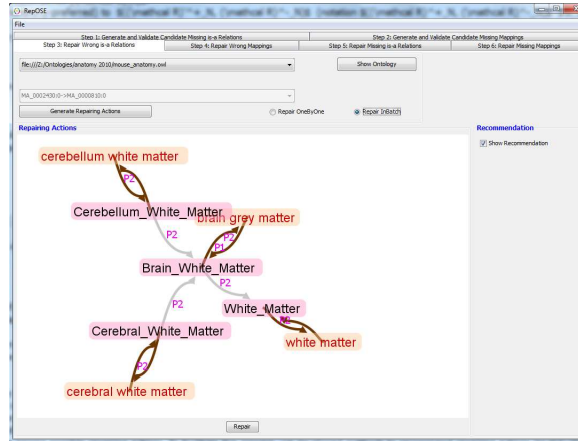


Fig. 2. Repairing wrong is-a relations.

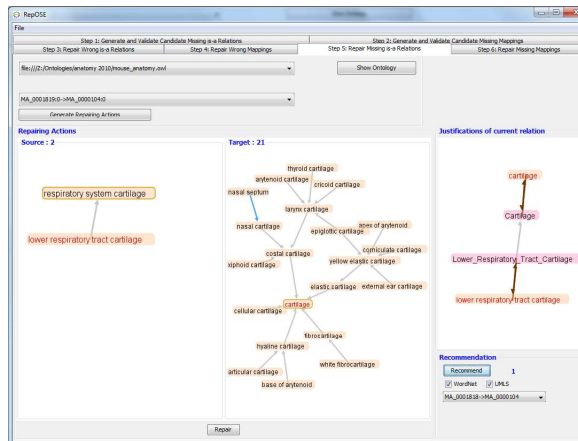


Fig. 3. Repairing missing is-a relations.

3 Uses of RepOSE

The current version of RepOSE has been used in a number of cases. The first case is work that was performed for the Swedish National Food Agency. The input data contained a large ontology, a small ontology and an alignment. The existing structure was of good quality. Most defects related to missing is-a relations and wrong mappings. The second case uses the ontologies and alignment of a track in the Ontology Alignment Evaluation Initiative. The input data contained two larger ontologies and an alignment. Defects in the structures of the ontologies and the alignment were repaired. In this case using our approach also new knowledge was added to the network. For the third case the input data contained five smaller ontologies and four alignments. In this case also missing mappings could be found and new alignments were generated.

ToxOntology - MeSH. RepOSE has been used to debug ToxOntology and an alignment to MeSH [4]. ToxOntology is a toxicology ontology and was created within an informatics system development at the Swedish National Food Agency as part of an initiative to facilitate the identification of adequate substance-associated health effects. ToxOntology is an OWL2 ontology, encompassing 263 concepts and 266 asserted is-a relations. Further, an alignment with MeSH was desired to obtain an indirect index to the scientific literature. MeSH is a thesaurus of the National Library of Medicine. As MeSH contains many concepts not related to the domain of toxicology, a part of MeSH was used. This part contained 9,878 concepts and 15,786 asserted is-a relations.

In the initial detection phase RepOSE generated 12 non-redundant candidate missing is-a relations for ToxOntology (34 in total) of which 9 were validated by the domain experts as missing and 3 as wrong. For MeSH, RepOSE generated 17 non-redundant candidate missing is-a relations (among which 2 relations represented one equivalence relation - 32 candidate missing is-a relations in total) of which 5 were validated as missing and the rest as wrong. For the 3 wrong is-a relations for ToxOntology and the 12 wrong is-a relations for MeSH, the justifications contained at least one mapping that the domain expert validated to be wrong or related and the wrong is-a relations were repaired by removing these mappings. The 9 missing is-a relations in ToxOntology and the 5 missing is-a relations in MeSH were repaired by adding the missing is-a relations themselves. In all but three cases this was what RepOSE recommended based on external knowledge from WordNet and UMLS. After this repairing, one new candidate missing is-a relation was detected in MeSH, which was validated as a wrong is-a relation and resulted in the removal of one more mapping.

OAEI 2010 Anatomy. RepOSE was also used during an experiment on a network consisting of the two ontologies and the alignment from the Anatomy track in the Ontology Alignment Evaluation Initiative (OAEI, [6]) 2010. The Adult Mouse Anatomy Dictionary (AMA, [1]) contains 2744 concepts and 1807 asserted is-a relations, while the NCI Thesaurus anatomy (NCI-A, [5]) contains 3304 concepts and 3761 asserted is-a relations. The alignment contains 986 equivalence and 1 subsumption mapping between AMA and NCI-A. These ontologies as well as the alignment were developed by domain experts. The experiment was performed by a domain expert.

The system detected 200 candidate missing is-a relations in AMA of which 123 were non-redundant. Of these non-redundant candidate missing is-a relations 102 were validated to be missing is-a relations and 21 were validated to be wrong is-a relations. For NCI-A 127 candidate missing is-a relations of which 80 were non-redundant, were detected. Of these non-redundant candidate missing is-a relations 61 were validated to be missing is-a relations and 19 were validated to be wrong is-a relations. To repair these defects 85 is-a relations were added to AMA and 57 to NCI-A, 13 is-a relations were removed from AMA and 12 from NCI-A, and 12 mappings were removed from the alignment. In 22 cases in AMA and 8 cases in NCI-A a missing is-a relation was repaired using a more informative repairing action (not derivable from the network), thereby adding new knowledge to the network.

The recommendations seemed useful. Regarding candidate missing is-a relations, 81 and 27 recommendations that the relation should be validated as a missing is-a relation, were accepted for AMA, respectively NCI-A, while 8 and 2 were rejected. When the system recommended that a candidate missing is-a relation should be validated as a wrong is-a relation, the recommendation was accepted in 7 out of 20 cases for AMA and 6 out of 8 cases for NCI-A. The recommendations regarding repairing missing is-a relations were accepted in 69 out of 85 cases for AMA and 43 out of 57 cases for NCI-A.

OAEI 2010 Bibliography. Another experiment is based on the bibliography case in the OAEI 2010. This case consists of 5 smaller ontologies in the bibliography domain (called 101, 301, 302, 303 and 304 in the OAEI set), with between 13 and 56 concepts each. The ontologies are connected in a star shape. (Ontology 101 has 22, 23, 18 and 30 mappings to 301, 302, 303 and 304, respectively. There are no mappings among the other pairs of ontologies.) Initially, RepOSE found 6 candidate missing is-a relations in ontology 101, 5 in ontology 304 and 1 in each of the other ontologies. During the repairing 2 additional candidate missing is-a relations were found for ontology 101, 1 for 302 and 3 for 304. Of all these 14 were validated as missing and 5 as wrong. Further, for the pairs of ontologies for which no alignment existed, candidate missing mappings were generated of which 187 were validated as missing and 15 as wrong. These defects were repaired by adding 19 is-a relations in ontologies and 181 mappings, and removing 7 is-a relations in ontologies and 10 mappings. The missing is-a relations and mappings were in 18 cases repaired by more informative (with respect to the network) repairing actions. We note that using RepOSE alignments were generated for each pair of ontologies for which no alignment existed previously.

4 Demonstration

In the demonstration we guide the visitors through a debugging session using (parts of) the ontologies from OAEI 2010 Anatomy, as well as a debugging session for OAEI 2010

Bibliography. Further, we explain the algorithms for the detection and validation of defects, as well as the generation, recommendation and execution of repairing actions.

Acknowledgements. We thank the Swedish Research Council (Vetenskapsrådet) and the Swedish e-Science Research Centre (SeRC) for financial support.

References

1. AMA. Adult mouse anatomical dictionary. http://www.informatics.jax.org/searches/AMA_form.shtml.
2. V Ivanova, J Laurila Bergman, U Hammerling, and P Lambrix. Debugging taxonomies and their alignments: the ToxOntology - MeSH use case. *1st International Workshop on Debugging Ontologies and Ontology Mappings*, 2012.
3. P Lambrix and V Ivanova. A unified approach for debugging is-a structure and mappings in networked taxonomies. *Report*, 2012.
4. MeSH. Medical subject headings. <http://www.nlm.nih.gov/mesh/>.
5. NCI-A. National cancer institute - anatomy. <http://www.cancer.gov/cancerinfo/terminologyresources/>.
6. OAEI. Ontology alignment evaluation initiative. <http://oei.ontologymatching.org/>.