

# Debugging the Missing Is-A Structure of Networked Ontologies

Qiang Liu and Patrick Lambrix

Linköpings universitet,  
581 83 Linköping, Sweden

**Abstract.** In parallel with the proliferation of ontologies and their use in semantically-enabled applications, the issue of finding and dealing with defects in ontologies has become increasingly important. Current work mostly targets detecting and repairing semantic defects in ontologies. In our work, we focus on another kind of severe defects, modeling defects, which require domain knowledge to detect and resolve. In particular, we are interested in detecting and repairing the missing structural relations (is-a hierarchy) in the ontologies. Our goal is to develop a system, which allows a domain expert to detect and repair the structure of ontologies in a semi-automatic way.

## 1 Problem Statement

Developing ontologies is not an easy task and often the resulting ontologies are not consistent or complete. Such ontologies, although often useful, also lead to problems when used in semantically-enabled applications. Wrong conclusions may be derived or valid conclusions may be missed. To deal with this problem we may want to repair the ontologies. Until now most work has been performed on finding and repairing the semantic defects such as unsatisfiable concepts and inconsistent ontologies. In this work we tackle the problem of repairing modeling defects and in particular, the repairing of the structure of ontologies.

In addition to its importance for the correct modeling of a domain, the structural information (e.g. is-a or part-of) in ontologies is also important in semantically-enabled applications. For instance, the is-a structure is used in ontology-based search and annotation. It is also important information in ontology engineering research. For instance, most current ontology alignment systems use structure-based strategies to find mappings between the terms in different ontologies (e.g. overview in [11]) and the modeling defects in the structure of the ontologies have an important influence on the quality of the ontology alignment results [1].

As the ontologies grow in size, it is difficult to ensure the correctness and completeness of the structure of the ontologies. Some structural relations may be missing or some existing or derivable relations may be unintended. Detecting and resolving these defects requires, in contrast to semantic defects, the use of domain knowledge. One interesting kind of domain knowledge are the other ontologies and information about connections between these ontologies. For instance, in the case of the Anatomy track in the 2008 and 2009 Ontology Alignment Evaluation Initiative (OAEI) two ontologies,

Adult Mouse Anatomy Dictionary [2] (MA, 2744 concepts) and the NCI Thesaurus - anatomy [15] (NCI-A, 3304 concepts), and a partial reference alignment (PRA, a set of correct mappings between the terms of the ontologies) containing 988 mappings are given. Using one ontology and the mappings as domain knowledge for the other ontology (and vice versa), it was shown in [10] that at least 121 is-a relations in MA and 83 in NCI-A are missing and should be repaired.

In our work, we deal with detecting and repairing the missing is-a structure in ontologies in the context of domain knowledge represented by networked ontologies. Our goal is to develop a system which allows a domain expert to debug and repair the structure of ontologies in a semi-automatic way.

## 2 State of the Art

There is not much work on detecting and repairing modeling defects in networked ontologies. In [3] and [9] similar strategies to detect missing is-a relations are described. Given two pairs of terms between two ontologies which are linked by the same kind of relationship, if the two terms in one ontology are linked by an is-a relation while the corresponding terms in the other are not, it is deemed as a possible missing is-a relation.

Related to the detection of missing relations, there is much work on finding relationships between terms in the text mining area. Much of the work on detecting is-a relations is based on the use of Hearst patterns [5] or extensions thereof (e.g. [4,17]). Most of these approaches have good precision, but low recall. A semi-automatic approach for ontology refinement (including is-a relations) is given in [18]. Another paradigm is based on machine learning and statistical methods, such as k-nearest neighbors approach [12], association rules [13], and clustering techniques [19].

There is more work that addresses semantic defects in ontologies. In [16] minimal sets of axioms are identified which need to be removed to turn an ontology coherent. In [8,7] strategies are described for repairing unsatisfiable concepts detected by reasoners, explanation of errors, ranking erroneous axioms, and generating repair plans. In [6] and [14] the setting is extended to repairing networked ontologies. In this case semantic defects may be introduced by integrating ontologies. Both approaches assume that ontologies are more reliable than the mappings and try to remove some of the mappings to restore consistency. The solutions are often based on the computation of minimal unsatisfiability-preserving sets or minimal conflict sets.

## 3 Approach and Methodology

**Detecting the missing is-a relations.** Given an ontology, a naive way to detect the missing is-a relations would be to take every pair of concepts which have no inferable is-a relation and check whether there should be an is-a relation between them. This requires inspection by domain experts and it usually means a large amount of work. However, as discussed before, other ontologies together with mappings between these ontologies can be used for detecting missing is-a relations. Our research aims to facilitate the detection of missing is-a relations in this way. In particular, we address the following questions:

- (a) How to use other ontologies and PRAs to detect missing is-a relations?
- (b) How to use other ontologies and given (possibly incorrect) mappings to detect missing is-a relations?

For the first question it is assumed that the given mappings are correct. The case where the structure of the ontologies is assumed to be correct and the mappings are 1-1 equivalence mappings, is relatively straightforward (e.g. [9]). However, this approach needs to be extended for other kinds of mappings. When we cannot assume that the structure of the ontologies is correct, we will only be able to compute suggestions for missing is-a relations and additional validation by a domain expert will be needed.

The second question adds the additional difficulty that the mappings are not necessarily correct (e.g. they are generated by an ontology alignment system and not yet validated). In this case we will need to also deal with the repairing of the mappings (semantic defects) such as in [6] and [14], and analyze the interaction between the modeling defects and the semantic defects.

**Repairing the missing is-a relations.** Once missing is-a relations are found, we need to repair the structure. This can be done by adding a set of is-a relations (called a *structural repair* in [10]) to the ontology such that when these are added, all missing is-a relations can be derived from the extended ontology. Clearly, the missing is-a relations themselves constitute a structural repair, but this is not always the most interesting solution for the domain expert. For instance, in a real case based on the Anatomy track from the OAEI 2008, we know that an is-a relation between *wrist joint* and *joint* is missing in MA and thus  $\{wrist\ joint\ is-a\ joint\}$  is a structural repair. However, knowing that there is an is-a relation between *wrist joint* and *limb joint*, a domain expert will most likely prefer to use the structural repair  $\{limb\ joint\ is-a\ joint\}$  instead. This is correct from a modeling perspective as well as more informative and would lead to the fact that the missing is-a relation between *wrist joint* and *joint* can be derived. In this particular case, using the second structural repair would actually also lead to the repairing of 6 other missing is-a relations in MA (e.g. between *elbow joint* and *joint*).

Our research needs to address the following questions.

- (a) How can we generate structural repairs?
- (b) How can we recognize structural repairs that are interesting for a domain expert?
- (c) How can we recommend structural repairs?
- (d) How can we execute repairs?

For the first question we need to find sets of is-a relations that would allow us to derive the missing is-a relations. In our approach we use a description logic knowledge base to check this. Further, not all structural repairs are interesting from a modeling point of view and therefore we need to define ways to recognize relevant structural repairs. We do this by defining preference relations between structural repairs and develop algorithms that take these preference relations into account.

A domain expert will repair the ontologies in a semi-automatic way and essentially chooses between the generated structural repairs. As there may be many of these we need to define ways to recommend structural repairs. We do this by involving other domain knowledge (e.g. WordNet or UMLS). Further, as there may be many missing is-a relations, it may not be practical to deal with all at the same time, and methods for deciding and executing repairs in an iterative way need to be developed.

**Integrated approach.** Up to now we have used ontologies and mappings as background knowledge for detecting missing is-a relations in another ontology. However, in the case

of networked ontologies we can actually detect and repair missing is-a relations for all ontologies at the same time. The repairing of missing is-a relations in one ontology may lead to the detection and repairing of missing is-a relations in other ontologies. Therefore, our definitions should be generalized to deal with this case and algorithms need to be extended or new algorithms developed.

**Evaluation.** Our approaches will be implemented in a system for detecting and repairing missing is-a relations. We will evaluate our approaches with existing networked ontologies (as we did in [10]). We will evaluate properties of the algorithms, quality of the generated and recommended repairs, as well as usability issues of the system.

## 4 Results

In our work until now we have studied the case where the ontologies are defined using named concepts and subsumption axioms. Most ontologies contain this case and many of the most well-known and used ontologies, e.g. in the life sciences, are covered by this setting. In [9] we discussed the use of a PRA in the setting of ontology alignment. One of the approaches included detecting missing is-a relations by using the structure of the ontologies and the PRA. Missing is-a relations were found by looking at pairs of equivalence mappings. If there is an is-a relation between the terms in the mappings belonging to one ontology, but there is no is-a relation between the corresponding terms in the other ontology, then we concluded that an is-a relation is missing in the second ontology. The detected missing is-a relations were then added to the ontologies. This is the simplest kind of structural repair.

In [10] we focused on the repairing of the missing structure. First, we defined the notion of structural repair. Then, we defined three preference relations that relate to heuristics that may be used by a domain expert when deciding which structural repair is more interesting. The first preference prefers structural repairs without redundant or non-contributing is-a relations. The second one prefers structural repairs with more informative ways of repairing. The third one prefers structural repairs which do not change existing is-a relations in the original ontology into equivalence relations. Further, we developed algorithms for generating structural repairs that take the preference relations into account. We also developed an algorithm for recommending structural repairs using other domain knowledge as well as algorithms for executing repairs. Based on these algorithms we implemented a prototype system RepOSE (*Repair of Ontological Structure Environment*), that allows a user to repair the structure of an ontology in a semi-automatic way. We evaluated our approach using MA and NCI-A.

## 5 Conclusions and Future Work

Motivated by the impact of defects in ontologies on their use as domain models and in semantically-enabled applications, we work on debugging the missing is-a structure in ontologies. We have already proposed approaches, developed algorithms and implemented a prototype that allow a domain expert to repair the is-a structure of ontologies in a semi-automatic way for the basic cases discussed in section 3.

In future work we will address the issues discussed in section 3. This will include dealing with ontologies represented in more expressive representation languages as well

as finding ways to optimize the generation of results. For instance, as the generation of structural repairs may take a lot of time, we may want to investigate ways to partition the set of missing is-a relations into parts that can be processed independently. Another interesting track is to investigate the integration of this work with ontology alignment.

## References

1. Preliminary results of the OAEI 2009 (2009), <http://oaei.ontologymatching.org/2009/results>
2. AMA. Adult mouse anatomical dictionary, [http://www.informatics.jax.org/searches/AMA\\_form.shtml](http://www.informatics.jax.org/searches/AMA_form.shtml)
3. Bada, M., Hunter, L.: Identification of OBO nonalignments and its implications for OBO enrichment. *Bioinformatics*, 1448–1455 (2008)
4. Cimiano, P., Staab, S.: Learning by googling. *ACM SIGKDD Explorations Newsletter* 6(2), 24–33 (2004)
5. Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: 14th Int. Conference on Computational Linguistics, July 1992, pp. 539–545 (1992)
6. Ji, Q., Haase, P., Qi, G., Hitzler, P., Stadtmüller, S.: RaDON - repair and diagnosis in ontology networks. In: Demo at the 6th European Semantic Web Conference, pp. 863–867 (2009)
7. Kalyanpur, A., Parsia, B., Sirin, E., Cuenca-Gray, B.: Repairing unsatisfiable concepts in OWL ontologies. In: Sure, Y., Domingue, J. (eds.) *ESWC 2006*. LNCS, vol. 4011, pp. 170–184. Springer, Heidelberg (2006)
8. Kalyanpur, A., Parsia, B., Sirin, E., Hendler, J.: Debugging unsatisfiable classes in OWL ontologies. *Journal of Web Semantics* 3(4), 268–293 (2006)
9. Lambrix, P., Liu, Q.: Using partial reference alignments to align ontologies. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) *ESWC 2009*. LNCS, vol. 5554, pp. 188–202. Springer, Heidelberg (2009)
10. Lambrix, P., Liu, Q., Tan, H.: Repairing the missing is-a structure of ontologies. In: 4th Asian Semantic Web Conference, pp. 371–386 (2009)
11. Lambrix, P., Strömbäck, L., Tan, H.: Information integration in bioinformatics with ontologies and standards. In: *Semantic Techniques for the Web: The REWERSE perspective*, pp. 343–376 (2009)
12. Maedche, A., Pekar, V., Staab, S.: Ontology learning part one - on discovering taxonomic relations from the web. In: *Web Intelligence*. Springer, Heidelberg (2002)
13. Maedche, A., Staab, S.: Discovering conceptual relations from text. In: 14th European Conference on Artificial Intelligence, pp. 321–325 (2000)
14. Meilicke, C., Stuckenschmidt, H., Tamilin, A.: Repairing ontology mappings. In: 20th National Conference on Artificial Intelligence, pp. 1408–1413 (2007)
15. NCI-A. National cancer institute - anatomy, <http://www.cancer.gov/cancerinfo/terminologyresources/>
16. Schlobach, S.: Debugging and semantic clarification by pinpointing. In: Gómez-Pérez, A., Euzenat, J. (eds.) *ESWC 2005*. LNCS, vol. 3532, pp. 226–240. Springer, Heidelberg (2005)
17. van Hage, W.R., Katrenko, S., Schreiber, G.: A method to combine linguistic ontology-mapping techniques. In: 4th Int. Semantic Web Conference, pp. 732–744 (2005)
18. Völker, J., Hitzler, P., Cimiano, P.: Acquisition of OWL DL axioms from lexical resources. In: Franconi, E., Kifer, M., May, W. (eds.) *ESWC 2007*. LNCS, vol. 4519, pp. 670–685. Springer, Heidelberg (2007)
19. Zavitsanos, E., Paliouras, G., Vouros, G.A., Petridis, S.: Discovering subsumption hierarchies of ontology concepts from text corpora. In: *IEEE / WIC / ACM Int. Conference on Web Intelligence*, pp. 402–408 (2007)