Direct computation of diagnoses for ontology alignment*

Kostyantyn Shchekotykhin, Philipp Fleiss, Patrick Rodler, and Gerhard Friedrich

Alpen-Adria Universität, Klagenfurt, 9020 Austria
firstname.lastname@aau.at

Abstract. Modern ontology debugging methods allow efficient identification and localization of faulty axioms in an ontology. However, in many use cases such as ontology alignment the ontologies might include many conflict sets, i.e. sets of axioms preserving the faults, thus making ontology diagnosis infeasible. In this paper we present a debugging approach based on a direct computation of diagnoses that omits calculation of conflict sets. The evaluation results show that the approach is practicable and is able to identify a fault in adequate time.

1 Algorithm details and evaluation

Most of the modern debugging approaches apply the model-based diagnosis [3] and compute diagnoses using conflict sets CS, i.e. irreducible sets of axioms ax_i in an ontology $\mathcal{O}$ that preserve a fault. A user should modify at least all axioms of a diagnosis in order to be able to formulate the intended (target) ontology $\mathcal{O}_t$. The computation of the conflict sets can be done within a polynomial number of calls to the reasoner, e.g. by QUICKXPLAIN algorithm [2]. To identify a diagnosis of cardinality $|\mathcal{D}| = m$ the hitting set algorithm suggested in [3] requires computation of m conflict sets. In the use cases when an ontology is generated by an ontology matching system the number of conflict sets m can be large, thus making the ontology debugging practically infeasible.

In this paper we present two algorithms INV-HS-TREE and INV-QUICKXPLAIN, which inverse the standard model-based approach and compute diagnoses directly, rather than by means of conflict sets. INV-QUICKXPLAIN partitions the initial set of axioms a given faulty ontology into two equal subsets. The algorithm continues to partition the sets until it identifies that the set $\mathcal{D}'$ such that $\mathcal{O} \setminus \mathcal{D}'$ fulfills all requirements and its partitions are not. In further iterations the algorithm minimizes the $\mathcal{D}'$ by splitting it into sub-problems of the form $\mathcal{D} = \mathcal{D}' \cup \mathcal{O}_\Delta$, where $\mathcal{O}_\Delta$ contains only one axiom. In the case when $\mathcal{D}$ is a diagnosis and $\mathcal{D}'$ is not, the algorithm decides that $\mathcal{O}_\Delta$ is a subset of the sought diagnosis. Just as the original algorithm, INV-QUICKXPLAIN always terminates and returns either a diagnosis $\mathcal{D}$ or "no diagnosis". In order to enumerate all possible diagnoses we modified the HS-TREE algorithm [3] to accept diagnoses as node labels instead of conflict sets.

In the diagnosis discrimination settings [4] the ontology debugger acquires new knowledge by asking the user whether some axiom should be entailed by the target ontology $\mathcal{O}_t$ or not. Given the answer the algorithm can invalidate some of the diagnoses that are used as labels of tree nodes. Given such a node, INV-HS-TREE removes its label and places it to the list of open nodes. Moreover, the algorithm removes all the nodes of a subtree originating from this node. After all nodes with invalid labels are cleaned-up,

* This research is funded by Austrian Science Fund (Project V-Know, contract 19996).

the algorithm attempts to reconstruct the tree by reusing the remaining valid diagnoses. In the direct approach limiting the number of diagnoses used to compute a query to some reasonable number. e.g. $n = 10$ results in a small size of the search tree, thus, using less memory in comparison to the standard approach.

We evaluated the direct ontology debugging technique using aligned ontologies generated in the framework of OAEI 2011 [1]. These ontologies represent a real-world scenario in which a user generated ontology alignments by means of some (semi-)automatic tools. The Conference test suite we included 146 classifiable ontologies and computed 1, 9 and 30 diagnoses with both HS-TREE and INV-HS-TREE. For 133 ontologies both approaches were able to compute the required amount of diagnoses. In the experiment where only 1 diagnosis was requested, the direct approach outperforms the HS-TREE as it was expected. In the next two experiments the time difference between the approaches decreases. However, the direct approach was able to avoid a rapid increase of computation time for very hard cases. In the 13 cases HS-TREE was unable to find all requested diagnoses in each experiment. Within 2 hours the algorithm calculated only 1 diagnosis for `csa-conference-ekaw` and for `ldoa-conference-confof` it was able to find 1 and 9 diagnoses, whereas INV-HS-TREE required 9 sec. for 1, 40 sec. for 9 and 107 sec. for 30 diagnoses on average.

Moreover, in the first experiment we evaluated the efficiency of the interactive direct debugging approach applied to the 13 "hard" ontologies. We selected the target diagnosis randomly among all diagnoses that included only invalid alignments suggested by a system. The latter can be computed using the set of correct alignments provided by the organizers of OAEI 2011. In the experiment the used the Entropy scoring function [4] with prior fault probabilities of axioms corresponding to aliments set to $1 - v$, where v is the confidence value of the matcher. All axioms of the aligned ontologies were assumed to be correct and were assigned small probabilities. The debugging was then applied to the set of all alignments returned by a matcher. The experiment shows that the system was able to identify the target diagnosis efficiently requiring less than 4 sec. in 75% of all cases to compute a query. The system's performance decreased only in the cases when a reasoner required much time to verify the consistency of an ontology.

In the second scenario we applied the direct method to unsatisfiable and classifiable within 2 hours ontologies, generated for the Anatomy problem. The source ontologies $\mathcal{O}_1$ and $\mathcal{O}_2$ include 11545 and 4838 axioms correspondingly, whereas the size of the alignments varies between 1147 and 1461 axioms. The target diagnosis selection process was performed in the same way as in the first experiment. The results of the experiment show that the target diagnosis can be computed within 40 second in an average case. Moreover, INV-HS-TREE slightly outperformed HS-TREE.

References

1. Euzenat, J., Ferrara, A., van Hage, W.R., Hollink, L., Meilicke, C., Nikolov, A., Ritze, D., Scharffe, F., Shvaiko, P., Stuckenschmidt, H., Sváb-Zamazal, O., dos Santos, C.T.: Final results of the Ontology Alignment Evaluation Initiative 2011. In: Proceedings of the 6th International Workshop on Ontology Matching. pp. 1–29. CEUR-WS.org (2011)
2. Junker, U.: QUICKXPLAIN: Preferred Explanations and Relaxations for Over-Constrained Problems. In: Proc. 19th National Conference on Artificial Intelligence. pp. 167–172 (2004)
3. Reiter, R.: A Theory of Diagnosis from First Principles. Artificial Intelligence 32(1), 57–95 (1987)
4. Shchekotykhin, K., Friedrich, G., Fleiss, P., Rodler, P.: Interactive ontology debugging : two query strategies for efficient fault localization. Journal of Web Semantics 12, 88–103 (2012)