Using Semantic Distances for Reasoning with Inconsistent Ontologies[1]

Zhisheng Huang [a] Frank van Harmelen [a]

[a] *Computer Science Department, Vrije Universiteit Amsterdam, The Netherlands*

Abstract

Re-using and combining multiple ontologies on the Web is bound to lead to inconsistencies between the combined vocabularies. Even many of the ontologies that are in use today turn out to be inconsistent once some of their implicit knowledge is made explicit. However, robust and efficient methods to deal with inconsistencies are lacking from current Semantic Web reasoning systems, which are typically based on classical logic. In earlier papers, we have proposed the use of *syntactic relevance functions* as a method for reasoning with inconsistent ontologies. In this paper, we extend that work to the use of semantic distances. We show how Google distances can be used to develop *semantic relevance functions* to reason with inconsistent ontologies. In essence we are using the implicit knowledge hidden in the Web for explicit reasoning purposes. We have implemented this approach as part of the PION reasoning system. We report on experiments with several realistic ontologies. The test results show that a mixed syntactic/semantic approach can significantly improve reasoning performance over the purely syntactic approach. Furthermore, our methods allow to trade-off computational cost for inferential completeness. Our experiment shows that we only have to give up a little quality to obtain a high performance gain.

1 Introduciton

A key ingredient of the Semantic Web vision is avoiding to impose a single ontology. Hence, merging ontologies is a key step. Earlier experiments have shown that merging multiple ontologies can quickly lead to inconsistencies. Other studies have shown how migration and evolution also lead to inconsistencies. This suggests the importance and omnipresence of inconsistencies in ontologies in a truly web-based world.

The classical entailment in logics is *explosive*: any formula is a logical consequence of a contradiction. Therefore, conclusions drawn from an inconsistent knowledge base by classical inference may be completely meaningless. The general task of a system of reasoning with inconsistent ontologies is: given an inconsistent ontology, return *meaningful* answers to queries. In [3] we developed a general framework for reasoning with inconsistent ontologies, in which an answer is "meaningful" if it is supported by a selected consistent sub-ontology of the inconsistent ontology, while its negation is not supported. In that work, we used relevance based selection functions to obtain meaningful answers. The main idea of the framework is: (1) a relevance function is used to select some consistent sub-theory from an inconsistent ontology; (2) then we apply standard reasoning on the selected sub-theory to try and find meaningful answers; (3) if a satisfying answer cannot be found, the relevance degree of the selection function is made less restrictive, thereby extending the consistent sub-theory for further reasoning. In this way the system searches for increasingly large sub-theories of an inconsistent ontology *until the selected sub-theory is large enough to provide an answer, but not yet so large so as to become itself inconsistent*.

2 From syntactic to semantic relevance

In [2], several syntactic relevance based selection functions were developed. However, these approaches suffer several limitations and disadvantages. As we will show with a simple example later in this paper,

[1]The full version of this paper appears in the Proceedings of the International Semantic Web Conference 2008 (ISWC08).

such syntactic relevance functions are very sensitive to the accidental syntactic form of an ontology, which can easily lead to undesired conclusions on one syntactic form. A simple semantics preserving syntactic reformulation would have lead to the appropriate conclusion, but such careful design is unrealistic to require from knowledge engineers.

In this paper, we investigate the approach of *semantic relevance* selection functions as an improvement over the syntactic relevance based approach. We will examine the use of co-occurrence in web-pages, provided by a search engine like Google, as a measure of semantic relevance, assuming that when two concepts appear more frequently in the same web page, they are semantically more relevant. We will show that under this intuitive assumption, information provided by a search engine can be used for semantic relevance based selection functions for reasoning with inconsistent ontologies.

The main contributions of this paper are (1) to define some general formal properties of semantic relevance selection functions, (2) to propose the Google Distance as a particular semantic relevance function, (3) to provide an implementation of semantic relevance functions for reasoning with inconsistent ontologies in the PION system, (4) to run experiments with PION to investigate the quality of the obtained results, and (5) to highlight the cost/performance trade-off that can be obtained using our approach.

3 Google distance as semantic relevance

In [1], the Google Distance is introduced to measure the co-occurrence of two keywords on the Web. Normalised Google Distance (NGD) is introduced to measure the semantic distance between two concepts.

We therefore propose a semantic distance which is measured by the ratio of the summed distance of the difference between two formulas to the maximal distance between two formulas:

$$SD(\phi, \psi) = \frac{sum\{NGD(C_i, C_j)|C_i \in C(\phi)\backslash C(\psi), C_j \in C(\psi)\backslash C(\phi)\}}{(|C(\phi)| * |C(\psi)|)}$$

The intuition behind this definition is to sum the semantic distances between all terms that are not shared between the two formulae, but these must be normalised (divided by the maximum distance possible) to bring the value back to the [0,1] interval.

4 Conclusions

Our research in this paper is the first attempt to introduce the Google Distance for reasoning with inconsistent ontologies. In essence we are using the implicit knowledge hidden in the Web for explicit reasoning purposes.

In our experiment we applied our PION implementation to realistic test data. The experiment used a high-quality ontology that became inconsistent after adding disjointness statements that had the full support of a group of experts. The test showed that the run-time of informed semantic backtracking is much better than that of blind syntactic backtracking, while the quality remains comparable. Furthermore the semantic approach can be parametrised so as to stepwise further improve the run-time with only a very small drop in quality.

References

[1] R. Cilibrasi and P. Vitany. The Google similarity distance. *IEEE/ACM Transactions on Knowledge and Data Engineering*, 19:3:370–383, 2007.

[2] Z. Huang and F. van Harmelen. Reasoning with inconsistent ontologies: Evaluation. Project Report D3.4.2, SEKT, 2006.

[3] Z. Huang, F. van Harmelen, and A. ten Teije. Reasoning with inconsistent ontologies. In *Proceedings of IJCAI'05*, 2005.