Harnessing the power of folksonomies for formal ontology matching on-the-fly

Theodosia Togia, Fiona McNeill and Alan Bundy

School of Informatics, University of Edinburgh, EH8 9LE, Scotland

Abstract. This paper is a short introduction to our work on building and using folksonomies to facilitate communication between Semantic Web agents with disparate ontological representations. We briefly present the *Semantic Matcher*, a system that measures the semantic proximity between terms in interacting agents' ontologies at run-time, fully automatically and minimally: that is, only for semantic mismatches that impede communication. The system is designed to allow agents to "understand" the meanings of terms to be matched by comparing their folksonomy-based "mental representations".

1 Introduction

The Semantic Matcher is an extension of the *Ontology Repair System* (ORS) [2], a plug-in for a service-requesting agent (*requester*) in the Semantic Web. Terms unknown to the requester which are encountered during interaction with a service-providing agent (*provider*) are mapped to terms in the former's ontology. We assume that the requester, and therefore ORS, has no access to the provider's ontology beyond what is revealed during interation: we are thus concerned with matching not two full ontologies but only individual terms from the provider's ontology to the most relevant terms of the requester's ontology.

We believe that the most serious obstacle for meaning sharing between agents is the lack of symbol grounding in ontologies: ontology terms are unable to refer to the objective world without human interpretation. We argue [3] that this can be dealt with if we allow agents to interprete the meanings of their terms by building a mental representation (sense) [1] for each one of these terms. In our work, senses are simulated by broad folksonomies [4] which annotate physical or abstract resources as opposed to digital resources (e.g. the set of cats in the world vs. a web-page about cats). Folksonomies are created and related to the requester's (formal) ontology. We show that combining ontologies and folksonomies in this way can allow fast and effective matching to be done on-the-fly and provides a way of grounding terms in ontologies to real-world entities.

2 Using Folksonomies for Ontology Matching

The architecture of our matcher is inspired by that of search engines. Broad folksonomies (comparable to "virtual documents" [5] or bags of words) are built for every candidate term (i.e. name of a relation, class or individual) in the requester's ontology with our *sense creation algorithm* (which extracts information from databases such as WordNet and SUMO and manipulates it with techniques such as stemming and stopping[1]). During agent interaction, when ORS diagnoses a semantic mismatch, a sense must be created for the unknown term, which will act as a query to the search engine. This step must be performed on-the-fly without interrupting normal interaction more than necessary. Our search engine then takes the sense representing the unknown term and a list of senses representing the requester's candidate terms as input and returns as output a ranking of the candidate terms.

3 Implementation and Evaluation

The system briefly discussed here has been fully implemented. Evaluation was performed using different versions of the SUMO ontology and its sub-ontologies from the Sigmakee repository[2]. When terms are changed between SUMO versions (e.g. "Corn" becomes "Maize"), we have an objective way of measuring the performance of the matcher because we can safely regard terms and their renamings as synonyms and compare these pairings with our system's prediction. Initial results are encouraging, with 57% of correct matches chosen as the best by the system, 19% as the second-best.

4 Further Work and Conclusions

This paper briefly introduced our work on integrating folksonomies with formal ontologies to perform matching on-the-fly, whenever the need becomes apparent. We believe these ideas could be a major step forward in the problem of ontology matching in an agent communication environment, and in providing symbol grounding for ontology terms. Furthermore, they can provide a framework for the design of matchers which exploit the vast amount of tag data available on the web. Full details of the theory on which this work is based, together with full descriptions of the implementation and evaluation, can be found in [3]. We are currently extending this work and evaluating it more fully.

References

1. G. Frege (1892). On sense and reference. In P. Ludlow, editor, *Readings in the Philosophy of Language*. MIT, 1997.
2. F. McNeill and Alan Bundy. Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. *IJSWIS (International Journal on Semantic Web and Information Systems) special issue on Ontology Matching*, 3:1–35, 2007.
3. T. Togia. Automated ontology evolution: Semantic matching. Master's thesis, University of Edinburgh, 2010. Available online at http://dream.inf.ed.ac.uk/projects/dor/ (unpublished).
4. T. Vander Wal. Explaining and showing broad and norrow folksonomies, 2005. available online at http://www.vanderwal.net/random/entrysel.php?blog=1635.
5. W. Hu Y. Qu and G. Cheng. Constructing virtual documents for ontology matching. In *Proceedings of the 15th International WWW Conference*, pages 23–31, 2006.

[1] Supplementing this data with folksonomies discovered on the web (e.g. tags that often co-occur with the tag "cat") is also possible though not currently implemented.

[2] http://sigmakee.cvs.sourceforge.net/viewvc/sigmakee/KBs/