Web Explanations for Semantic Heterogeneity Discovery*

Pavel Shvaiko[1] **Fausto Giunchiglia**[1] **Paulo Pinheiro da Silva**[2] **Deborah L. McGuinness**[2]
[1]University of Trento, Povo, Trento, Italy [2]Stanford University, Stanford, USA
{pavel|fausto}@dit.unitn.it {pp|dlm}@ksl.stanford.edu

Abstract

Managing semantic heterogeneity is a complex task. One solution involves matching like terms to each other. We view *Match* as an operator that takes two graph-like structures (e.g., concept hierarchies or ontologies) and returns a mapping between the nodes of the graphs that correspond semantically to each other. State of the art matching systems (e.g., Cupid) perform well for many real world applications. However, matching systems may produce mappings that may not be intuitively obvious to human users. In order for users to trust the mappings (and thus use them), they need information about them. Users need access to the sources that were used to determine semantic correspondences between terms and potentially they need to understand how deductions are performed. In this paper we describe how matching systems can explain their answers using the Inference Web (IW) infrastructure. There, *S-Match*, a *semantic matching* system, produces proofs for mappings it has discovered. Using the IW browser, users may visualize different explanations including provenance information, the derivation path of the mappings, etc. Thus users can make informed decisions about them.

1 Introduction

The progress of information and communication technologies, and in particular of the Web, has made a huge amount of disparate information available. The number of different information resources is growing significantly, and therefore the problem of managing semantic heterogeneity is increasing. Many solutions to this problem include identifying terms in one information source that "match" terms in another information source. We view *Match* as one of the key operators for enabling the Semantic Web since it takes two graph-like structures (e.g., database schemas, concept hierarchies or ontologies) and produces a mapping between the nodes of the graphs that correspond semantically to each other. Matching, however, requires explanations because mappings between terms are not always intuitive to human users. Indeed, if Semantic Web users are going to trust the fact that two terms may have the same meaning, then they need to understand the reasons leading a matching system to produce such a result.

*Long version of this paper is available at http://www.dit.unitn.it/research/publications/techRep?id=529 as [Shvaiko *et al.*, 2004].

2 Semantic Matching

In this paper, we discuss semantic matching as introduced in [Giunchiglia and Shvaiko, 2003], and implemented within the *S-Match* system [Giunchiglia *et al.*, 2004]. Semantic matching uses the following relations between terms: *equivalence* (=); *more general* ($\sqsupseteq$); *less general* ($\sqsubseteq$); *mismatch* ($\perp$); *overlapping* ($\sqcap$). The semantic relations are calculated by mapping meaning which is codified in the element descriptions and the graphs in two steps: (i) by obtaining a representation of the node meaning and (ii) by determining the meaning of the node position in the graph. In order to obtain some information about the node labels, our initial implementation accesses WordNet. Semantic matching translates the matching problem into a validity check of the appropriate propositional formula. The algorithm then checks for sentence validity by proving that its negation is unsatisfiable. Our implementation uses the JSAT, propositional satisfiability (SAT), reasoner.

3 Explaining Matching using Inference Web

Inference Web [McGuinness and Pinheiro da Silva, 2003] enables applications to generate portable and distributed explanations for answers. In order to explain semantic matching and thereby increase the trust level of its users, we need to provide information about background theories (initially Wordnet), the JSAT manipulations of sentences, and the semantic matching translations of graphs into propositional sentences. This paper addresses the first two issues.

IW proofs and explanations are represented in PML [Pinheiro da Silva *et al.*, 2004] and are composed of PML *node sets*. Node sets are subclasses of the W3C's OWL class [McGuinness and van Harmelen, 2004] and they are the building blocks of OWL documents describing proofs and explanations for application answers published on the Web. Exploiting PML properties, meaningful fragments of the *S-Match* proofs can be loaded on demand. Users can browse an entire proof or they can limit their view and refer only to specific, relevant parts of proofs since each node set has its own URI that can be used as an entry point for proofs and proof fragments. Also IW has a registry of meta-information about proof elements, such as sources (e.g., publications, ontologies), inference engines and their rules. In the *S-Match* case, it contains meta information about WordNet and JSAT.

Users may need different types of explanations. For example, if negotiating agents trust each other's information sources, explanations should focus on the *S-Match* manipulations. If on the other hand, the sources may be suspect, explanations should focus on meta information about sources. If a user wants an explanation of the inference engine(s) embedded in a matching system, a more complex explanation may be required. Our current version of *S-Match* uses JSAT, and in particular the Davis-Putnam-Longemann-Loveland (DPLL) procedure [Davis and Putnam, 1960]. The basic DPLL procedure recursively implements three rules: *unit resolution*, *pure literal* and *split*. Explanations of the unit resolution and pure literal rules are straightforward since they directly use results of each step of the DPLL procedure. Explanations of the split rule are tuned to the specificity of the SAT tasks.

4 Experimental Study

The main goal of the experiments being conducted is to obtain a vision of how the *S-Match* explanations potentially scale to requirements of the Semantic Web.

In order to conduct tests in a real environment, we used the IW web service of KSL at Stanford University to generate proofs in PML, while the (modified, i.e., proof producing) JSAT was run at the University of Trento. All the tests were performed without any optimizations: for each single task submitted to JSAT, the IW web service was invoked, no compression methods were used while transferring files, etc. Our modified JSAT produces proof information on a single mapping element requiring, on the average, less then 1 millisecond. Thus, results of the experimental study look promising and demonstrate their potential to scale to requirements of the Semantic Web, providing adjustable answers in real time.

5 Discussion

While there are a number of other efforts in semi-automated schema/ontology matching [Rahm and Bernstein, 2001], we are not aware that any provide explanations. As the use of matching systems for managing semantic heterogeneity grows, it becomes very important to explain mappings in order to make the Semantic Web more transparent and trustable. The key distinctions of *S-Match* proofs are:

- They are produced by a modified version of JSAT in *S-Match* that implements the Barrett and Berezin approach [Barrett and Berezin, 2003] for generating proofs;
- They are formatted in PML and consequently they are designed for use in a distributed Web environment;
- Their sentence propositions are mapped into meaningful terms rather than numbers in sentences using the DIMACS format;
- They are supported by the Inference Web tools for explanation and interactive proof presentation;
- They potentially scale to requirements of the Semantic Web.

By extending *S-Match* to use the IW infrastructure, we have demonstrated our approach for explaining matching systems that use background ontological information and reasoning engines. In particular, we presented DPLL-based IW explanations of the SAT engine used in the *S-Match* system. The DPLL procedure discussed in the paper constitutes a basic (without heuristics and optimizations) propositional satisfiability search procedure for state of the art SAT engines. Thus, one could consider using another optimized SAT reasoner that may be chosen for particular matching problems and using the approach discussed for generating explanations. The explanations can be presented in different styles allowing users to understand the mappings and consequently to make informed decisions about them. The paper also demonstrates that *S-Match* users can leverage the Inference Web tools, for example, for sharing, combining, browsing proofs, and supporting proof meta-information including knowledge provenance information, in particular about WordNet and JSAT. Future work includes using more expressive background ontologies and other SAT engines as well as other non-SAT DPLL-based inference engines, e.g., DLP, FaCT [Horrocks and Patel-Schneider, 1998].

References

[Barrett and Berezin, 2003] C. Barrett and S. Berezin. A proof-producing boolean search engine. In *Proceedings of workshop on PDPAR*, 2003.

[Davis and Putnam, 1960] M. Davis and H. Putnam. A computing procedure for quantification theory. In *Journal of the ACM*, number 7, pages 201–215, 1960.

[Giunchiglia and Shvaiko, 2003] F. Giunchiglia and P. Shvaiko. Semantic matching. In *KER Journal*, number 18(3), pages 265–280, 2003.

[Giunchiglia *et al.*, 2004] F. Giunchiglia, P. Shvaiko, and M. Yatskevich. S-match: an algorithm and an implementation of semantic matching. In *Proceedings of ESWS*, pages 61–75, 2004.

[Horrocks and Patel-Schneider, 1998] I. Horrocks and P. F. Patel-Schneider. Fact and dlp. In *Automated Reasoning with Analytic Tableaux and Related Methods: Tableaux'98*, pages 27–30, 1998.

[McGuinness and Pinheiro da Silva, 2003] D. L. McGuinness and P. Pinheiro da Silva. Infrastructure for web explanations. In *Proceedings of ISWC*, pages 113–129, 2003.

[McGuinness and van Harmelen, 2004] D.L. McGuinness and F. van Harmelen. Owl web ontology language overview. Technical report, W3C, February 10 2004. http://www.w3.org/TR/owl-features/.

[Pinheiro da Silva *et al.*, 2004] P. Pinheiro da Silva, D. L. McGuinness, and R. Fikes. A proof markup language for semantic web services. Technical report, KSL-04-01, Stanford University, 2004.

[Rahm and Bernstein, 2001] E. Rahm and P. Bernstein. A survey of approaches to automatic schema matching. In *VLDB Journal*, number 10(4), pages 334–350, 2001.