A Criteria for Selecting Background Knowledge for Domain Specific Semantic Matching

Jetendr Shamdasani, Peter Bloodsworth, Tamas Hauer, Andrew Branson, Mohammed Odeh, and Richard McClatchey

CCCS, UWE, Coldharbour Lane, Frenchay, Bristol BS16 1QY, UK
`firstname.lastname@cern.ch`

Abstract. Ontology matching is a problem that is rapidly reaching maturity especially in the context of the emergent semantic web. Semantic matching is the detection of specific correspondences between concepts in two ontologies including but not limited to $\equiv$, $\sqsubseteq$ and $\sqsupseteq$. Many semantic matching approaches today use some form of background knowledge to aid them in discovering correspondences. This paper presents criteria that can be used when selecting a background knowledge source.

1 Introduction

The selection of a correct resource of the semantic matching process is a crucial step since it seeds the final reasoning step in the semantic matching process. Many of these resources are selected manually for example Aleksovski et al used the FMA as a source of background knowledge [1]. The SMatch approach uses WordNet [2] as its background resource. This poster paper deals with the manual selection of a resource for the medical domain. Previously in [3] it was shown that using a resource that is too general may yield in ambiguous alignment results. To aid in background knowledge selection criteria have been developed. The next section will detail the criteria for selection of such a resource.

2 Selection Criteria for Background Knowledge

The selection criteria have been defined to be the following: **coverage**, **semantics**, **granularity** and **meta-information**. Examples from the domain of medicine are used to describe and justify the criteria accordingly.

1. **Coverage** This criterion concerns insuring that sufficient concepts from the medical domain are covered to enable the semantic matching process. For example many medical ontologies today do not cover a single domain or may slightly overlap with another domain. For example, Galen covers the areas of Anatomy with some Pathological information. By using a source of background knowledge which only covers a single subdomain of medicine, some information may be missed. Therefore, the background resource needs to be able to cover as many terms from the medical domain as possible.

2. **Semantics** are required for the derivation of subsumption and equivalence correspondences between ontologies and to create the background theory for the reasoning process. This background theory is present in most semantic matching methods that are available today. For example let us consider two terms search as "Heart" and "Aorta", these two terms have very little in common lexically. Here a source of background knowledge is useful in that it states the relationships between these terms explicitly which is that Heart is more general than Aorta. Therefore relationships such as this need to be explicitly stated in a source of background knowledge.

3. **Granularity** is the availability of terms to a high-level of detail. In comparison to the previous requirement of coverage, granularity is the availability of terms in a background resource which are lower in the semantic hierarchy. Granularity therefore can be thought of as a detailed description of a domain. Therefore to seed the semantic matching process, terms need to be available in a background resource at a high-level of detail. In the medical domain this is especially an issue since many medical terms are related to each other, but they may also have very little lexical information in common. Although the reasoning process may be able to detect a mapping by inference, the relationship from a background resource is given a higher weighting because it is viewed as a more trusted source.

4. **Meta-information** refers to the presence of meta-data within the background knowledge source. For anchoring purposes this can provide a meaningful string to concept mapping. Meta-data needs to be present, either having been derived algorithmically or stated explicitly for this purpose. This is so that information can be extracted from the ontology using it.

3 Conclusion

In this paper the importance of background knowledge in the semantic matching process was stated. It was suggested that background knowledge from a trusted source would be an asset to the semantic matching process, since it seeds the final reasoning process with initial relationships. Criteria were defined for the selection of a background resource for semantic matching. These criteria were justified to be coverage, semantics, granularity and meta-information. These criteria have already been utilised in the creation of the MedMatch algorithm [3].

References

1. Z. Aleksovski et al. Exploiting the structure of background knowledge used in ontology matching. In *Ontology Matching Workshop (ISWC-2006)*, 2006.
2. F. Giunchiglia et al. Semantic Matching: Algorithms and Implementation. *Journal on Data Semantics*, 9:1–38, 2007.
3. J. Shamdasani. *Semantic Matching for the Medical Domain*. PhD thesis, University of the West of England, December 2010.