

SEMANTIC ENRICHMENT OF ONTOLOGY MAPPINGS

Schema and ontology matching play an important part in the field of data integration and semantic web. Given two heterogeneous data sources, meta data matching usually constitutes the first step in the data integration workflow, which refers to the analysis and comparison of two input resources like schemas or ontologies. The result is a list of correspondences between the two schemas or ontologies, which is often called mapping or alignment. Many tools and research approaches have been proposed to automatically determine those correspondences. However, most match tools do not provide any information about the relation type that holds between matching concepts, for the simple but important reason that most common match strategies are too simple and heuristic to allow any sophisticated relation type determination.

Knowing the specific type holding between two concepts, e.g., whether they are in an equality, subsumption (is-a) or part-of relation, is very important for advanced data integration tasks, such as ontology merging or ontology evolution. It is also very important for mappings in the biological or biomedical domain, where is-a and part-of relations may exceed the number of equality correspondences by far. Such more expressive mappings allow much better integration results and have scarcely been in the focus of research so far.

In the doctoral thesis “Semantic Enrichment of Ontology Mappings”, the determination of the correspondence types in a given mapping is the focus of interest, which is referred to as semantic mapping enrichment. In contrast to previous approaches, linguistic laws and linguistic insights are highly exploited. The approach comprises various strategies that make use of these linguistic laws and are able to calculate the semantic type between two matching concepts. The observations and insights gained from this research go far beyond the field of mapping enrichment and can be also applied to schema and ontology matching in general.

Since generic strategies have certain limits and may not be able to determine the relation type between more complex concepts, like a *laptop* and a *personal computer*, background knowledge plays an important role in this research as well. The thesis shows how background knowledge can be effectively used, how it is possible to draw conclusions even if a concept is not contained in it, how the relation types in complex paths can be resolved and how time complexity can be reduced by a so-called bidirectional search. The developed techniques go far beyond the background knowledge exploitation of previous approaches, and are now part of the semantic repository SemRep, a flexible and extendable system that combines different lexicographic resources. Additionally, an approach to automatically extract semantic concept relations by parsing Wikipedia articles is introduced.

The mapping enrichment tool STROMA could achieve very favorable results. In a comparative evaluation, it was able to achieve considerably better results compared to related applications. This corroborates the overall usefulness and strengths of the implemented strategies, which were developed with particular emphasis on the principles and laws of linguistics.