Ontology Merging by Matching a Selection of Ontologies in a Cluster Environment

PhD topic (abstract): Axel Tenschert, tenschert@hlrs.de

High Performance Computing Center Stuttgart, University of Stuttgart

Ontology matching is of high interest for a number of tasks solved in the semantic web, especially when thinking of extending complex information structures. With regard to this work current matching strategies will be considered in order to create a strategy for matching ontologies with the aim to merge them. The purpose of such a merging is to support a user (e.g. scientist, doctor, …) with one priority ontology which is related to his specific requirements. Hereby, the user is enabled to use the knowledge of several ontologies by using only the one priority ontology. Furthermore, when thinking of merging ontologies to one priority ontology, the research field of bioinformatics is of interest because of the need to investigate large scale data within a high number of ontologies. The benefit for such a user is the possibility to receive required information very fast by considering only one data source, the extended priority ontology.

There are several ontology matching strategies and merging tools available. Nevertheless, the complexity of matching ontologies entails the problem of matching them in a scalable way. In order to solve this problem distribution and parallelization techniques are used to speed up the matching by executing it at same time. For this the matching process is distributed to execute as much comparisons between concepts at same time as possible. The concepts from the priority ontology are matched with the concepts from the set of ontologies in parallel. It is obvious that this methodology requires lots of computing resources to execute the large amount of data.

To provide the required computing resources the matching of the concepts and the merging is executed in a cluster environment. The selection of ontologies is copied on the hard drive of the cluster. Afterwards, the matching and merging procedures are executed as several jobs on the nodes of the cluster. When all matching concepts are merged together in the priority ontology, the ontologies on the hard drive of the cluster are deleted automatically expect the extended priority ontology. This priority ontology is prepared to be used by a user, e.g. a doctor or a scientist in the field of medicine/bioinformatics.