An Ontology Merging Tool to Facilitate Interoperability between Coastal Sensor Networks

Shruthi Bheemireddy, Surya S. Durbha, Roger L. King, Santhosh K. Amanchi, and Nicolas H. Younan
Department of Electrical and Computer Engineering
GeoResources Institute (GRI)
Mississippi State University, Mississippi State, MS 39762-9571, USA
shruthi@gri.mssatte.edu, suryad@gri.msstate.edu, santosh@gri.msstate.edu,
rking@engr.msstate.edu and younan@ece.msstate.edu

Abstract:

Ontologies are considered as key components on the Semantic Web used for solving the problem of information heterogeneity by establishing the semantic correspondences between data entities. In recent years, the growing need for the use of ontology applications by different organizations led to the attention of semantic alignment between ontologies to achieve semantic interoperability among different information systems or services using different individual ontologies in a distributed and heterogeneous environment. Thus, Ontology mapping aims to find semantic correspondences between similar elements of different ontologies become the key point to make the heterogeneous information more accessible. Ontology mapping can be done either manually or using full or semi-automated tools. Manual mapping is error-prone and very laborious. Hence, in recent years different tools and methodologies have been developed for automatic ontology mapping. The available methodological approaches included machine learning techniques, graph based methods, reasoning methods, heuristic and rule based methods etc.

In this paper we present a new instance-based approach combined with string distance metrics for solving ontology mapping problem that employs machine learning techniques to find such mappings. Our approach uses a powerful classification method, Support Vector Machine (SVM), to deal with this ontology mapping problem. Support vector machine (SVM) is a popular technique for classifying any kind of data and is capable of providing superior accuracy compared to other classification methods. The basic idea is to determine the similarity between concepts of two ontologies by finding the similarities between instances contained in each of the

concepts, as in most of the ontologies the semantics of the relations between concepts is defined through the set of their instances. To obtain better matching results, we adopt a hybrid approach of combining the instance-based method with the name-matching methods. String matching tools are used to measure distances between all pairs of concepts contained in both the source and target ontologies. The numerical scores generated are the match features which denote the similarity between the concepts. In many real world ontologies, each concept is associated with set of data instances and the availability of the instance data would be helpful to train a classifier for each category in the source ontology. The methodology consists of extracting suitable numbers of statistical features from the data in the training set for each representation and applying Kernel Principal Components Analysis (KPCA) over the extracted features, followed by a support vector machine classification of each concept contained in the source ontologies. The trained SVM classifier is applied to classify concepts in the target ontology. The focus is on finding mappings between ontologies by analyzing the instances of their concepts as input. A decision on matching concepts is then based on the weights obtained from both SVM classifier and string based techniques. The above methodology is applied to the ontological representations developed for different coastal sensor systems like the National Data Buoy Center (NDBC), and Gulf of Maine Ocean Observing System (GoMoos), whose data are highly heterogeneous in syntax, structure and semantics.