World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:12, 2017

Clustering-Based Computational Workload Minimization in Ontology Matching

Authors : Mansir Abubakar, Hazlina Hamdan, Norwati Mustapha, Teh Noranis Mohd Aris

Abstract : In order to build a matching pattern for each class correspondences of ontology, it is required to specify a set of attribute correspondences across two corresponding classes by clustering. Clustering reduces the size of potential attribute correspondences considered in the matching activity, which will significantly reduce the computation workload; otherwise, all attributes of a class should be compared with all attributes of the corresponding class. Most existing ontology matching approaches lack scalable attributes discovery methods, such as cluster-based attribute searching. This problem makes ontology matching activity computationally expensive. It is therefore vital in ontology matching to design a scalable element or attribute correspondence discovery method that would reduce the size of potential elements correspondences during mapping thereby reduce the computational workload in a matching process as a whole. The objective of this work is 1) to design a clustering method for discovering similar attributes correspondences and relationships between ontologies, 2) to discover element correspondences by classifying elements of each class based on element's value features using K-medoids clustering technique. Discovering attribute correspondence is highly required for comparing instances when matching two ontologies. During the matching process, any two instances across two different data sets should be compared to their attribute values, so that they can be regarded to be the same or not. Intuitively, any two instances that come from classes across which there is a class correspondence are likely to be identical to each other. Besides, any two instances that hold more similar attribute values are more likely to be matched than the ones with less similar attribute values. Most of the time, similar attribute values exist in the two instances across which there is an attribute correspondence. This work will present how to classify attributes of each class with K-medoids clustering, then, clustered groups to be mapped by their statistical value features. We will also show how to map attributes of a clustered group to attributes of the mapped clustered group, generating a set of potential attribute correspondences that would be applied to generate a matching pattern. The K-medoids clustering phase would largely reduce the number of attribute pairs that are not corresponding for comparing instances as only the coverage probability of attributes pairs that reaches 100% and attributes above the specified threshold can be considered as potential attributes for a matching. Using clustering will reduce the size of potential elements correspondences to be considered during mapping activity, which will in turn reduce the computational workload significantly. Otherwise, all element of the class in source ontology have to be compared with all elements of the corresponding classes in target ontology. K-medoids can ably cluster attributes of each class, so that a proportion of attribute pairs that are not corresponding would not be considered when constructing the matching pattern.

Keywords : attribute correspondence, clustering, computational workload, k-medoids clustering, ontology matching
Conference Title : ICSWBIS 2017 : International Conference on Semantic Web Business and Innovation Systems
Conference Location : Kuala Lumpur, Malaysia
Conference Dates : December 11-12, 2017