ORIGINAL RESEARCH PAPER | Computer Science | Volume : 6 | Issue : 10 | October 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 74.50

Cross Ontology Based Clustering Framework-A Study

KEYWORDS: Cross ontologies, K-means clustering, concept sharing, ontology alignment.

Dr. S.Vigneshwari

Assistant Professor, Department of Computer Science and Engineering, Sathyabama University, Chennai, India

ABSTRACT *Individual ontologies cannot share huge amound of semantically related data. If the common data can be shared across multiple ontologies then, cross ontology mechanism plays an efficient roll in the fast extraction and retrieval of semantically related linked data. The cross ontologies also share a common data base which is mostly in XML or OWL-RDF format. This paper briefs about the representation of cross ontologies and about the clustering of the relevant concepts using k-means clustering algorithm. It is also proved that the time complexity is very much reduced upon using cross ontologies when compared to the single ontologies*

INTRODUCTION

Ontology is a set which contains a collection of concepts, properties, relations, individuals and events with respect to a particular universe of discourse (Asuncion et al., 2011). Concept in ontology refers to a realistic idea, a native query or a synterm. In the current context, the term 'concept' refers to the synterm, which is the outcome of the WordNet dictionary. Concepts in ontology are sometimes called as classes. The properties are the set of attributes, which refer to the contexts or the relationships among the concepts. The individuals are instances of objects. Ontology is analogous to the Graph theory,, which comprise vertices 'V' and edges 'E'. In an ontological aspect, the vertices are considered as concepts and the edges are considered as relationships.

The concepts are linked to each other based on relations. Various relations utilized in the present ontology model are 'is_a', 'has', 'part_of', 'union_of'.

Single Ontology issues

Ontologies are needed to be designed in a context friendly manner. In single ontology, a common global ontology shares the semantic specification. In case of a varying domain, single ontological approach becomes impossible. There is a need for the integration of heterogeneous resources. Single ontologies does not provide a fine granularity level while dealing with specification, which is an effect of integrating locally designed ontologies under one common global ontology (Stumme and Maedche, 2005). It is also impractical in real time to bring all ontologies under a single global ontology. The resources are subjected to dynamic changes. When a common concept is shared by multiple local ontologies it leads to redundancy. These drawbacks lead to the technique of shared multi ontologies which is preferred to global ontology.

Challenges in Multi Ontologies

Important challenges in real-time multi ontologies are the sparse distribution of data and the huge computational memory space. In order to conserve memory utilization, dynamic linking of the semantic concepts is done across multiple ontologies. The dynamic linking combines the concepts specific to a given query. That is, the object references which refer to the concepts are released when they are not in use. If multiple ontologies are combined and stored in memory, it would lead to huge memory overhead.

Ontologies can be modelled for decentralization and also for meeting the essence of dissolution. Multi-database systems can be developed and organized in collaboration with the semantic web, which helps in modelling local structures and the global data bases. Many difficulties arise in developing an integral structure for ontology merging, especially in the federated type. The process of developing an integral structure for ontology yield child ontology on merging two or more parent ontologies. But, manual ontology merging is more difficult, error prone and it is also expensive and time consuming. Many frameworks and proposals have been developed in the recent decades for pursuing the ontology merging task (Stumme and Maedche, 2005).

Need for cross ontological framework

Social media like Twitter, Facebook and chat-communications including video, text and images are available across the internet, thereby making the world a global village(Saranya et al., 2016). Ontology plays an important role in sharing social knowledge in a semantic way. The currently available tools are not compatible with existing ontology merging tools, inference engines, reasoners, etc. Hence, there is a need for a scalable and compatible framework in the existing real time scenario.

MATERIALS AND METHODS

Multi Ontological Representation

Cross ontological mapping is a crucial phenomenon, in order to achieve semantic interpretation (Yanhui and Chong 2010). The cross ontological mapping could be achieved with the help of efficient cross ontological algorithms (Hariri et al., 2006). While mining the ontologies, the query has to be reformulated and the relevant documents should be indexed. The present ontological framework concentrates on the entire scenario including the secured retrieval of documents. Staab and Studer (2009) and (Gowri et al., 2016)stated that relevant entities and their relationships need to be modelled structurally. It can be done with the help of ontologies. Ontology comprises generalized or précised taxonomy of the concepts. Shared conceptualization is the backbone of any ontology, which can be done both explicitly and implicitly by a knowledge base.

Volume : 6 | Issue : 10 | October 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 74.50

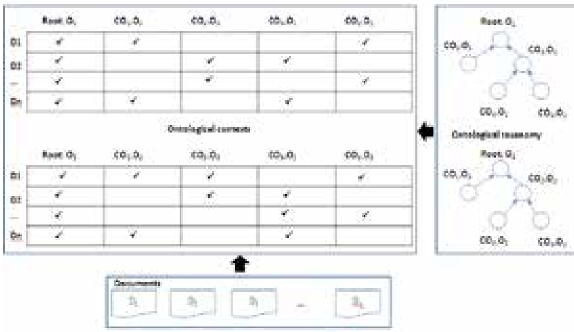

FIGURE 1. REPRESENTATION OF MULTI ONTOLOGIES

Figure 1 shows the representation of terms in documents in the form of ontologies where, D_1, D_2, D_3...D_n represent the documents. Similarly multi document representation is depicted in terms of multi ontologies. O_i represents the i^{th} ontology. Co_j represents the j^{th} concept in the ontology which is the related synterm. The multi ontology representation is crucial for understanding the concept relationships in cross ontology mining.

Use Of Ontology Alignment

Ontological alignment is essential for calculating the semantic similarity among the concepts. The document terms are needed to be mapped as the concepts in ontology and the bondage between the terms is represented by establishing the relations among the ontological concepts. To obtain an integrated relationship among the entities of multiple ontologies, feature selection and assignment of appropriate weights to features are important.

TABLE -1
RELATION TABLE

Input Concept	Relation	Matched Concept
Active_DB	is_a	DB_Modelling
Active_DB	is_a	Database
Cloud_DB	is_a	DB_Modelling
Cloud_DB	is_a	Database
DBA	part_of	DBMS
DBA	part_of	DB_Languages
DBA	part_of	DB_Systems
DBMS	has	DB_Design
DBMS	has	DB_Developer
DBMS	has	Entity_Relational
DBMS	has	Hierarchical
DBMS	has	Network
DBMS	has	Relational

The data for constructing cross ontology is obtained from recent search links based on the topic 'database'. By referring 250 documents the cross ontology is constructed. The relationships are manually applied to the portion of of the cross ontology and are tested. Table 1 describes the relation table where the input concept is matched onto the matched concept using various relations. From Table 1 it is inferred that for the concept 'DBMS', there of 6 'Has' relationships and 1 'Part_of' relationship. Is-a and Part-of relations are null. According to the priority and the count of the relation, the mining algorithm extracts the information regarding DBMS as follows:

Clustering the semantically linked concepts with K-Means clustering

Let 'K' be the number of clusters.

Input: Dataset to be clustered

Output: A set of 'k' clusters.

Step 1: Initial partition of 'K' clusters is selected.

Step 2: Repeat the following:

Step 2.1: Each pattern is assigned a closest cluster center and a new partition is generated.

Step 2.2: Centroid of the cluster is calculated using the cluster centers.

Step 2.3: Repeat Step 2.2 until stabilization of cluster membership is attained.

Step 3: Clusters are updated after performing merge and split operations over the smaller clusters.

Step 4: Cluster points are adjusted based on the square-error criterion function SE (C_i,k).

End

$$SE(C_i, k) = \sum_{i=1}^{k} \sum_{f \in C_i} \|f - mean_i\|^2 \quad (1)$$

In Equation (1), SE is the square error of the features in the dataset. '' refers to the feature in the given data space and '' is the mean of cluster ''.

Similarly, the designed clustering algorithm groups up all the associations with concepts according to the procedural mining method. Time based evaluation is calculated against the various thresholds assigned manually. The sample concepts considered are DBMS, DBA and Data Model. From the graph in Fig 2, it is inferred that, when the threshold is gradually increased, the computation time is reduced. When the threshold reaches the maximum value, the computation time gets saturated for all the concepts, due to the preciseness of human interpretations on the relationships. The experimental evaluation on synthetic dataset has shown that the highest time registered by cross ontological mining method is 10469ms and the optimum time is noted as 6513ms. Memory usage is mostly correlated with the computation time. As the threshold value gradually increases and reaches the maximum value, the memory occupied by the sample concepts DBMS, DBA and DB Model from the cross ontology gets reduced. The memory occupied is measured in terms of bytes. The computation time and the memory utilization shows a balancing behaviour when the threshold reaches the maximum value. The threshold values are manually fixed in these experiments.

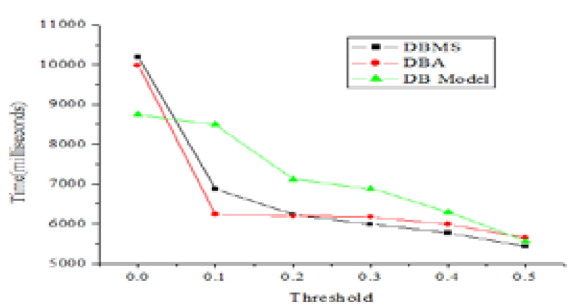

Fig 2 Computation time

CONCLUSION

Cross ontology based information extraction plays a major role in extracting the linked data across the semantic web. The main features of this framework is to improve the time complexity and space complexity of the information retrieval scenario. An efficient information clustering algorithm is designed as an enhancement of this framework. The computation time is very much reduced and stabilized when the manualy threshold reaches 0.5.

ACKNOWLEDGEMENT
I would like to thank Sahtyabama University for providing me opportunity to utilize the DST sponsored cloud computing lab for conducting experiments related to my research in an efficient manner.

REFERENCES
1. Asuncion Gomez-Perez, Mariano Fernandez-Lopez, Oscar Corcho (2011), "Ontological Engineering", Springer International Edition, India.
2. Stumme Gerd, Maedche Alexander (2005), "Ontology Merging for Federated Ontologies on the Semantic Web", Institute for Applied Computer Science and Formal Description Methods (AIFB), University of Karlsruhe, Vol. 3 (12), pp. 1-9.
3. Yanhui Lv, Chong Xie (2010), "A Framework for Ontology Integration and Evaluation", Proceedings of Intelligent Networks and Intelligent Systems (ICINIS), pp. 521 – 524.
4. Hariri Babak Bagheri, Hassan Abolhassani, Ali Khodaei (2006), "A new Structural Similarity Measure for Ontology Alignment", Proceedings of the 2006 International Conference on Semantic Web & Web Services, Vol. 1, pp. 36-42.
5. Staab Steffen, Studer Rudi (2009), "Handbook on Ontologies", International Handbooks on Information Systems, Springer, 2nd edition.
6. Gowri, S., Vigneshwari, S., Sathiyavathi, R., Lakshmi, T.R.K.(2016)," A framework for group decision support system using cloud database for broadcasting earthquake occurrences", Advances in Intelligent Systems and Computing-978-981-10-0767-5, Vol. 38, pp. 611-615.
7. Saranya, R.,Gowri, S.,Monisha, S.,Vigneshwari, S.(2016)," An ontological approach for originating data services with hazy semantics ",Indian Journal of Science and Technology,Vol.9(23), pp. 1-6.