ISSN No. 0976-5697

Volume 6, No. 2, March-April 2015

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

To Resolve Heterogeneity on Ontology Merging in Semantic Web - Survey

S. Yuvalatha
Assistant Professor
Sree Sakthi Engineering College
Coimbatore, Tamil Nadu, India

Dr. P. Nandhakumar
Senior Software Engineer
Easy Design Systems (P) Ltd.
Coimbatore, Tamil Nadu, India

Abstract: Ontology is a formal specification of knowledge by a set of concepts with in a domain and their semantic relationship. Different Ontologies are devolved by developer for a same domain differently. Furthermore, ontology tools using different language. Ontology merging in semantic web is used to resolve the heterogeneity problem among the source ontology. The language level and ontology level mismatches are the heterogeneity problem in the ontology merging. Lexical matching strategies resolve the syntax based on Jaro distance. The word net is a lexical database with synset which solve the problem of semantic mismatches. A knowledge base is constructed by using Semantic Web Rule Language (SWRL) for ontology level mismatches. A merging framework identifies the similarities and dissimilarities of source ontologies are merged to resolve the heterogeneity problem.

Keywords: Ontology, Heterogeneity problem, word net, Lexical, Semantic, SWRL

I. INTRODUCTION

The semantic web describes the meaning of information that can be understood by people and computer machine. The semantic web uses RDF to describe web resources with background in logic and artificial intelligences. Ontology merging concept is a key issues in semantic web infrastructure. The utility of semantic web depends on three issues such as Existences of data; user can retrieve the data, quality.

Ontology is a formal explicit representation of concepts in domain properties of each concept describes the characteristics and attributes of the concepts known as slots. Conceptualization is a description of concepts and relationship exists. Different ontologies are created by people in different format for a specific domain which causes heterogeneity problem.

Ontology is the platform for sharing the knowledge of domain that helps the machine to make intelligent decision. The semantic heterogeneity is caused by different meaning that are resolved by ontology management includes mapping, alignment and merging. Ontology matching is a solution for heterogeneity problem that find the semantic relation between the entities of ontologies. Ontology mapping aims to find the relationship between pair of concepts of different domain ontologies. Ontology alignment is made if the sources become consistent with each other but are kept separate. Ontology merging is the process of merging the source ontologies in to the global ontology. The similar concepts are merged and dissimilar concepts are added directly in to global ontology.

```
<owl: class rdf: id=”Book”>
</owl: class>
<owl: Data type Property rdf:id="Publication">
<rdfs: domain rdf: resource="#Book"/>
</owl: Data type Property>
<owl: object Property rdf:ID="Publi _name"> <rdfs: domain rdf: resource="#Book"/>
</owl: Data type Property>
```

Figure 1 Ontology structure of book

Fig 1 Ontology model is represented by using owl .The advantage of owl is to share and reuse the knowledge among the domain. Owl is an XML – based language which include sub-languages such as owl Lite, owl Dl, owl FULL. Ontology consists of two main properties Data type properties and Object properties. The relationship link between classes is an object property. The data type properties are a data in similar XML schema data type.

The OWL abstract syntax presents a sequence of annotations, axioms and facts. OWL ontology is interpreted as asset of axioms that provide semantics by allowing system to infer additional information based on the data explicitly provided. Axioms specify the characteristic of classes and properties.

SWRL is a standard OWL language[1] to detect the similar rules and then cluster based on their similarity. SWRL is based on the combination of the ontology sublanguage. It is represented in the form of antecedent and consequent. SQWRL is a SWRL based language using a SWRL library that builds a query language on the SWRL.SQWRL queries operate with SWRL conjunction can be used to retrieve knowledge by SWRL rule.

This paper is organized as follows. Some related research works are briefly reviewed in Section 2. Conclusion is drawn in section 3.

II. RELATED WORK

This section deals with the issues of ontology merging. The merging is the bottleneck in the research of semantic field. Recently, some interesting techniques and methodologies are focus on the interoperability among the domain specific data sources.

A. Ontology merging concepts using word net:

Merging the ontologies based on vertical and horizontal approach. These approaches are possible to create a single ontology and also ontology alignment is established by link between them and reuses the information. The merging is done semi automatically. The two approaches have different concepts to merge the ontologies.

The horizontal approaches [2] mapping the similar concepts among ontologies in the same level. The concepts are described as a sense in a word net.

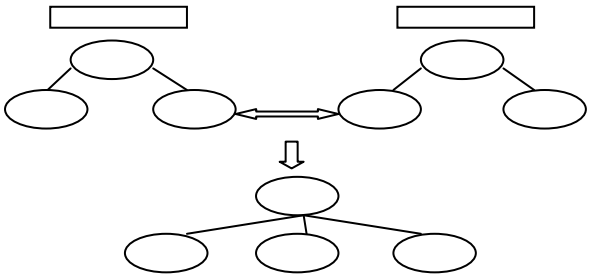

Figure 1 Horizontal Approach

The vertical approach is based on the similarity measure between concepts in the different level of ontologies. The similarity measure is not possible in the horizontal.

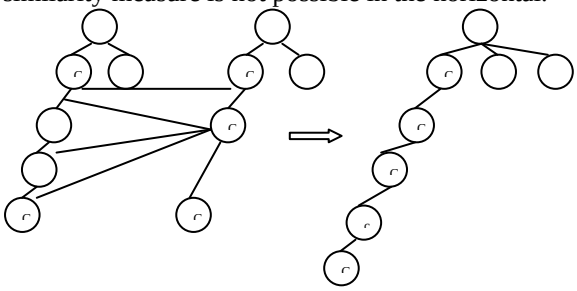

Figure 2 Vertical Approach

The approaches source the semi automatically merging of domain specific ontology. It merges the similar concepts and dissimilarities are added directly in to the global ontology. In this case fill the resulting ontology with concepts from both ontologies similarity measure is calculated in order to define the hierarchy between these concepts in the resulting tree. The merit of word net is the approaches are purely semantic, It resolves the semi automatic merging The approaches do not provides the accurate solution of automatic merging the ontologies of domain specific.

B. Automated ontology merging using hybrid strategy:

Automated merging of ontology using hybrid strategy consists of lexical, semantic, checking similarity and heuristic function. Ontology merging is a process of creating global ontology by defining concepts, common vocabulary. The hybrid strategy [3] begins the merging process from top in one owl file and bottom from another file. Lexical and semantic matching is used to compare similarities of classes among ontologies.

a. Lexical Matching:

Lexical matching performs stemming and the string compare. The stemming is performed for the class names to obtain the root values. The root values of the class are compared using the string compare method. If match is found return the Boolean values.

b. Semantic Matching:

The string comparison is performed with the class name of ontologies using the meaning of the word. Word net is a lexical database of English. It contains set of synonyms called synset. Word net is a textual description of the concepts. Every synset is compared with the other class name using lexical analysis. It is used to find the similarities word.

c. Similarity checking:

Properties are stored in array. It compares the property of class with the other. Lexical and semantic matching is used to compare the properties of class among ontologies and then compute the similarities of the class properties.

d. Heuristic function:

It checks the lexical analysis and semantic match is found then it written the class file along with the properties.

The similarity checking is performed. The number of properties is compared to find the super class.

These strategies provide a fully automated merging framework for improving the semantic interoperability in heterogeneous system. The algorithm using the hybrid strategies increase the hope that is possible to merge the ontologies automatically.

The four different similarity measures perform the merging semi automatic with less human intervention. The user job to merge source ontologies only by giving input as owl file and global ontology is produced as output.

C. Syntactico –Semantic Algorithm:

Ontologies are an efficient model for information representation and storage. Syntactico [4] semantic algorithm is used for automatic merging. It combines the syntactic and semantic measure for identifying the similar concepts. Synonyms and homonym problem can be resolved. Ontologies provide a bridge for sharing the information and semantic interoperability in environment. Ontology merging is a process of building a single ontology from a set of sources ontologies. Ontologies tools are used for ontology merging such as prompt, Chimaeras, HCONE.

a. Syntactic Similarity:

It is based on the computation of a distance between two strings to obtain the similarity measures. Similarity measures are computed by hamming distance, Jaro distance and Jaro Winkler distance. The similarity values between 0 and 1. The threshold value of 0.5 is set to compute the similarity of the string.

b. Semantic similarity:

Semantic similarity is done by word net. Word net is a lexical dictionary. It uses two different means to define the meaning of the world, the synset and lexical relation. A word is defined by a set synonyms and a definition. Constructing vectors which describes the synset for the concept. Semantic based approach using a novel approach to capture the relationships between tagged data for learning object stored from the repository. The syntactic similarity issues are overcome by semantic matching using synset of word net.

$$\text{Simsem}(c1, c2) = \frac{2*(\text{synset}(c1) \cap \text{synset}(c2))}{\text{Synset}(c1) + \text{synset}(c2)}$$

If SIMsem(c1,c2) is greater than the threshold the similarity is obtained. Once the similarities of lexical and semantic matching are computed then compute the Syntactico-semantic similarities among the results.

$$SimSynset(c1,c2)= \frac{Simsyn(c1,c2) +2*Simset(c1,c2)}{3}$$

Similar concepts are merged in to the single concept and the dissimilar concepts are directly copied in to the resulting ontology. It increases the accracy search result of ontology merging. Reduce the space and time complexity of ontology matching. It cannot find the all possible alignments entities between ontologies because compare only the super classes.

D. Resolving Terminological Heterogeneity in Ontologies:

Heterogeneity problem rises due to creating ontology in different formats and languages to represent the data and Meta data. The same data format differ in structure s and semantic of the terminology. The semantic among the ontologies is obtained by using articulation rules. The rule gives the relationship between the concepts of ontologies.

Ontology is represented as a graph with set of logical rules O= (G, R) .G represents a directed label graph and R is represented set of rule. The matcher is to find the terms and assign the similarity scores for the terms matched. The word sense similarity is constructed based on the word relator[5].

a. Thesaurus based word relator:

The heterogeneity is resolved based on thesaurus based word relator. The similarity between the attributes is computed based on external resources such as Nexus and word net. Word net gives the synset for word. If two words are equals then it return a score 1.0. If words are not equal then the similarity measure is calculated based on the definition for that word.

b. Corpus based word relator:

Linguistics matcher is developed for similarities using a corpus based word relator. Word similarities score are based on the context of words appear in the document. Instance based heuristics are used to find the match schema in database. Similarities are determined based on the data type and also extract attributes vectors.

Corpus based method provide better results than the thesaurus based method. The method increases the recall to 70%.The drawback is similarity is calculated based on the threshold values.

E. Automatic Ontology Merging by Hierarchical Clustering and Inference Mechanism:

Ontology merging algorithm resolves the mismatches between the ontologies without human intervention. The hierarchical clustering and inference mechanisms [6] are combined to develop the global ontology. The clustering methodology based on the semantic matching to define the entities of the global ontology.

a. Hierarchical clustering:

Hierarchical clustering algorithm based on the OWL-DL semantic language. Clustering algorithm is used to find the equivalent entities such as concepts, properties, instance of different ontologies. Ontology is a triplet(C, R, I) where C is an OWL classes, R is the OWL classes relationship. I is an instance of owl classes. Semantic between two concepts have high probability because of same properties, relationship. The semantic similarity between two attributes is determined based on the word net thesaurus. The word net returns the synonym set between two attributes. When the similarity between the attributes is computed need to set threshold to eliminate the dissimilarities among the ontologies

b. Inference:

It establishes the semantic relationship between the entities of different classes of different categories. Inference mechanisms detect the similar property of classes among the ontologies. The relation is determined as owl: Same As property, Owl: equivalent, owl: sub class of property.

Clustering different entities of local ontology and inference establish the similar property link of entities. The methodologies allow the reuse of ontologies for future ontology merging.

F. A framework for ontology matching techniques in semantic web

Ontology matching is the process of finding semantically related entities of different ontology. Multi matching techniques reduce the search space and time requirement by removing the entities like classes, properties. Ontology provides a knowledge that represents the particular domain which contains classes and properties and also relationship between the concepts. People develop different ontologies for specific domain which causes heterogeneity problem. The heterogeneity problem leads to inability to get accurate result in semantic web. All entities of first ontology matches with all entities of second ontology. It checks the classes, object properties, data properties. Matching techniques depend on some feature of ontologies such as RDF statement and class hierarchies [7].

The matching process is dependent on similarity values. Structure matching consists of two stages; the first one involves matching of RDF statements and the second one involves matching of class hierarchies[8].Mapping and merging of multiple ontologies to produce consistent, coherent and correct merged global ontology is an essential process to enable heterogeneous multi-vendors semantic-based systems to communicate with each other[10].

III. CONCLUSION

This paper presented a scope of ontology merging to reduce the heterogeneous problem. Global ontology is created by measuring the lexical, semantic and to detect homonyms conflict using set of SWRL rule in knowledge base. The similar classes and instance are combined as a single ontology and dissimilarities are added directly in to the global ontology. The contribution presented in this paper minimizes human involvement during ontology merging. Ontology merging approach is suggested that semantic heterogeneity can be resolved with the help of ontology [9].

IV. REFERENCES

[1]. Saeed Hassanpour, Martin J.o'Connor and Amar K.Das,"Visualizing Logical Dependencies in SWRL rule Bases", International Conference, Vol.35.pp, (2012).

[2]. Cho,Kim,Kim.P,"A New Method for Ontology Merging based on Concept using Word net", in advanced communication technology, Vol 3,(2006).

[3]. C.R. Rene Robin, G.V. Uma, "A novel algorithm for a fully automated ontology merging using hybrid strategy", Journal of scientific research, vol.47 no.1, (2010).

[4]. Siham Amrouch, Sihem Mostefai,"Syntatico- Semantic algorithm for an automatic ontology merging", International Conference information technology and e-service, vol.5.pp, (2012).

[5]. Prasenjit Mitra, Gio Wiederhold, "Resolving Heterogeneity in Ontologies", IEEE TKDE, Vol.99 (2010).

[6]. Nora Maiz, Muhammad Fahad ,Omar Boussaid, Fadlia Bentayeb, "Automatic Ontology Merging by Hierarchical Clustering and Inference Mechanisms" , proceeding of I-KNOW (2010),pp.81-93

[7]. Mohammed Maree, Mohammed Belkhatir,"A Coupled Statistical and Semantic Framework for Merging Heterogeneous Domain Specific Ontologies", International Conference on tool with Artifical Intelligence, Vol 3 (2010).

[8]. Kamel Hussein Shafa'amri, Jalal Omer Atoum, "A framework for improving the performance of ontology matching techniques in semantic web", International journal of advanced computer science and application vol.3.no.1 (2012).

[9]. Michel Klein, "Combining and relating ontologies: an analysis of problem And solution", International conference,(2009).

[10]. Muhammad Fahad, Muhammad Adbul Qadir,"A Framework for Ontology Evaluation", pp, (2010).