Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 85 (2016) 301 – 308

Procedia Computer Science

International Conference on Computational Modeling and Security (CMS 2016)

COGOM:COgnitive Theory Based Ontology Matching System

K.Saruladha[a*],S.Ranjini[b]

[3a]Pondicherry Engineering College,Senior Assistant Professor,Department Of Computer Science,Pondicherry-605008

[b]Pondicherry Engineering College,M.Tech,Department Of Computer Science,Pondicherry-605008

Abstract

Ontology matching systems take a prominent position in solving semantic heterogeneity problems to facilitate sharing and reuse of ontologies. The process of generating ontology alignments through ontology matching techniques purely lies on how the concepts and relationships are modeled. This paper focuses on designing an ontology matching system in which concepts are modeled based on cognitive units of knowledge comprising of objects, attributes and relationships. The proposed cognitive based ontology matching system(COGOM) identifies semantically related concepts by aggregating the attribute similarity degree, structural similarity degree and semantic conception degree. The similarity computation is adapted from the Tversky psychological model of similarity. The proposed ontology matching system is adaptive in nature because of the cognitive based knowledge expression and the computational overhead of generating alignments is improved by forming quality clusters of semantically correlating concepts thus reducing the concept match space. The precision and recall metrics are used for evaluation of the proposed system using the benchmark data sets of OAEI 2015.

Keywords: Semantic heterogeneity; COGOM; Tversky; Quality clusters;Precision;Recall.

1.Introduction

Despite the abundant data and information available in the semantic web the heterogeneity nature of the data leads to lot of problems. The realization of the goals of semantic web is possible only when the heterogeneous nature of knowledge sources are handled efficiently. For this purpose ontologies provide vocabulary and the

Peer-review under responsibility of the Organizing Committee of CMS 2016
doi:10.1016/j.procs.2016.05.237

synonyms of the terms used in the vocabulary regarding a particular domain of interest. Most of these ontologies are constructed and maintained by different knowledge engineers with different backgrounds thereby constructing several ontologies with different terminologies for the same domain. This is where ontology matching techniques gains importance. Ontology matching is a solution to the semantic heterogeneity problem. Ontology matching aims at determining the correspondences between similar concepts present in different ontologies which are developed for the same domain using semantic similarity measures and creates a sharable semantic space. For the past decade, many ontology matching techniques(systems) were proposed and were tested using OAEI benchmark data sets. But still there are certain challenges to be addressed to improve the effectiveness and efficiency of ontology matching. Despite many approaches believed to improve effectiveness of system, ontology matching systems do produce incorrect alignments which reduces the precision. When the size of ontology is increasing the match space becomes very large and hence results in lesser efficiency(execution time).Hence this paper proposes a cognitive based model which aims to improve efficiency without compromising on effectiveness of the ontology matching system.

This proposed ontology matching system(COGOM) models concepts as cognitive units of knowledge comprising of objects, attributes and relationships.In order to evaluate the proposed system and its metrics this paper uses two evaluation criteria namely precision and recall. The number of input parameters considered in this paper is significantly more than the existing system ,it will hopefully provide higher results and better solutions and hence proves that the proposed work will develop an scalable and efficient ontology matching system.

The remainder of this paper is organized into the following subsections: Section 2 reports the existing ontology matching systems. Section 3 focuses on limitation details of the existing ontology matching systems. Section 4 focuses on the motivations and principles underlying the design of the proposed (COGOM)system. Section 5 deals with theoretical foundations of concept analysis methods. Section 6 explains about the proposed COGOM architecture. Section 7 explains the algorithm for the proposed ontology matching system .Section 8 deals with the Implementation and experimental results. Section 9 deals with the evaluation methodology. Section 10 concludes the paper and also discusses the possible future work that can be performed with this proposed system as a base.

2.Related researches

This section tabulates all the related researches carried out and the following table summarizes few of the ontology matching systems based on their data sets, techniques and limitations used. Though many ontology matching systems have been developed there is very few system based on cognitive theory. As this paper aims at proposing a cognitive based ontology matching system the formal theory of “Concept Algebra” is used. Concept Algebra is a formal theory for abstract concepts and knowledge manipulation. The mathematical representation of concepts are developed based on concepts object-attribute-relation called OAR theory. Concept Algebra has a higher formalization degree for the sake of being an abstract mathematical structure, and its cognitive relations between concepts are conductive to knowledge reasoning. Hence by determining similarity between concepts, relations and their correlation degree, concepts with rich semantics are determined which helps in forming better alignments.

Table1.Literature survey of all of Ontology Matching Systems

Ontology Matching System	Data Sets	Techniques	Limitations
FalconAO: Aligning Ontologies With Falcon[7]-2005	OAEI 2005 data set	(1)Lexical comparison(LMO) (2)Graph matching technique(GMO)	Common vocabularies between ontologies is different and GMO cannot perform very large ontologies.
SAMBO: A System For Aligning And Merging Biomedical Ontologies[9]-2006	OAEI 2004 data set	Single filtering technique	Virtualization is not there and evaluating alignment strategies manually takes time.
DSSIM: Ontology Mapping With Uncertainty[10]-2006	OAEI 2006 data set	Dempster Shafer theory of evidence	Consideration of flat hierarchy of classes and properties for matching and hence semantic similarity of all mappings are not found
RIMOM:A Dynamic Multistrategy Ontology Alignment Framework[11]-2009	OAEI 2006,2007 data sets	1,Multistratergy ontology alignment framework 2.Minimization of Bayesian decision	Alignments based on background knowledge produces unsatisfactory results
FalconAO++:An Improved Ontology Alignment System[8]-2014	OAEI conference track data set	1.Divide and conquer technique. 2.String similarity technique	Input information is a bottleneck and.supports only one to one mapping

In the next section the limitations of existing ontology matching systems is discussed.

3.Limitations of the existing ontology matching system

Apart from Ontology Matching system tabulated in Table1, many other ontology matching systems like:ASMOV,Agreement Maker, Agreement Maker Light Ontology Matching System, Anchor Flood, Anchor Prompt,SAMBO and SAMBO DTF,RIMOM2 and RIMOM-IM are reported in the literature[1].Based on the literature survey made, the existing ontology matching system suffer from the following limitations:

- Improper matching with background knowledge leading to incorrect alignments increases recall and decreases precision.
- No techniques are used to consider only the relevant properties for matching that reduces the processing time, computation time and cost.
- Due to little common vocabulary the existing ontology matching system finds difficult to form approximately relative clusters leading to the formation of multiple clusters.
- The existing system does not filter out the good quality clusters after the clusters are formed..

Having studied the limitations, the motivations and principles underlying the design of COGOM system is discussed in the next section.

4.Motivations and principles underlying the design of the proposed COGOM system

4.1.Motivations and Principles: The proposed COGOM system is aimed at:

- Improving the efficiency of ontology matching system in terms of precision and recall and,
- Decreasing the execution time by reducing the concept match space by generating alignments using quality

clusters alone.

The principles kept in mind in designing the proposed COGOM system is narrated below:

- The COGOM system is based on cognitive theory. As the concept is based on cognitive theory it closely resembles the human perspective of categorizing related concepts and forms clusters.
- The proposed attribute similarity and structural similarity are correlated against human judgments and the quality of the clusters thus formed include highly correlating concepts and hence reduces concept match space.

5.Theoretical foundations of concept analysis methods

In information processing the perception of modeling a concept plays a vital role. In this direction many theoretical philosophical foundations have been reported in the literature. Out of which in information and knowledge processing two important techniques are found to be used in most of research work. One is formal concept analysis and the other is concept algebra. As the proposed COGOM system is based on cognitive model(concept algebra) this section enlightens and discusses the pros and cons of formal concept analysis(FCA) and concept algebra(CA).

5.1.Formal concept analysis(FCA)

Formal Concept analysis[5] provides a robust method for data inquiry, knowledge representation and efficient management of information.FCA is the definite representation of entire lattices and its properties which are expressed in terms of formal contexts. FCA- based approach gives the information that a particular concept contains a particular attribute/property or not. This information is not enough in order to find out how strongly that particular attribute/property is related to that concept. So Fuzzy Formal Concept Analysis(FFCA) was introduced.

5.2.Fuzzy formal concept Analysis(FFCA)

Fuzzy formal context[6] contains the list of concepts, attributes and a membership value. This captures how strong the particular attribute is related to the concept. But still the similarity measure value was constrained with only objects and attributes. In order to overcome this dis-advantage Concept Algebra approach is used.

5.3.Why concept algebra?(CA)

Concept Algebra[3] is an extension of formal concept analysis(FCA). Concept algebra is a cognitive theory based knowledge representation method and includes a set of formal treatments in order to find how concepts in ontologies are semantically rich and correlated with each other. Concepts in concept algebra are expressed with composition operations and relation operations. Unlike FCA and FFCA the concept similarities are found with help of depth analysis of properties, relationship types and semantic correlation degree between the concepts. Also quality clusters are generated by means of using the concept algebraic relationship formulas of super-concepts, sub-concepts and intention relations between concepts. Using these concept algebraic rules and formulas concepts and relationships are better conceptualized than the existing methods and helps in better formation of similar domain clusters. Hence in the proposed COGOM instead of modeling concepts using FCA or FFCA cognitive theory based concept algebra concept analysis method is used.

6.Proposed COGOM architecture

This section discusses in detail the proposed COGOM system.Firstly,the idea of designing COGOM based on cognitive theory is conceived from the work reported by Rodriguez[4].In Rodriguez, the mapping of entity classes was done by considering two ontologies SDTS and WRDNET.The inter-concept similarity was computed using Tversky psychological model. But Rodriguez has not addressed the issue of reduction in concept match space.COGOM takes this into consideration. Secondly, in Yingxu wang[2],Guanyu Li[3] the concept network formation is done and it is based on cognitive theory. But the asymmetric property and the relative importance of concepts are not considered in computing structural similarity and attribute similarity. Keeping these two conceptions in mind the COGOM system

- Models concepts as cognitive units of knowledge and reduces the concept match space by splitting ontologies into clusters using the notion of concept network. The clusters include semantically correlating concepts and hence the entire ontology need not be matched. The inter cluster matching would aid in generating alignments. Because of this the concept match space is considerably reduced.
- The Intra Similarity measure is Tversky psychological model based and hence do not ignore the asymmetric properties and relative importance of concepts in the ontology.
- Further the commonality among concepts is identified based on he attribute commonality, structural commonality(super concept and sub concept commonality).This section describes the overall architecture of the COGOM system.

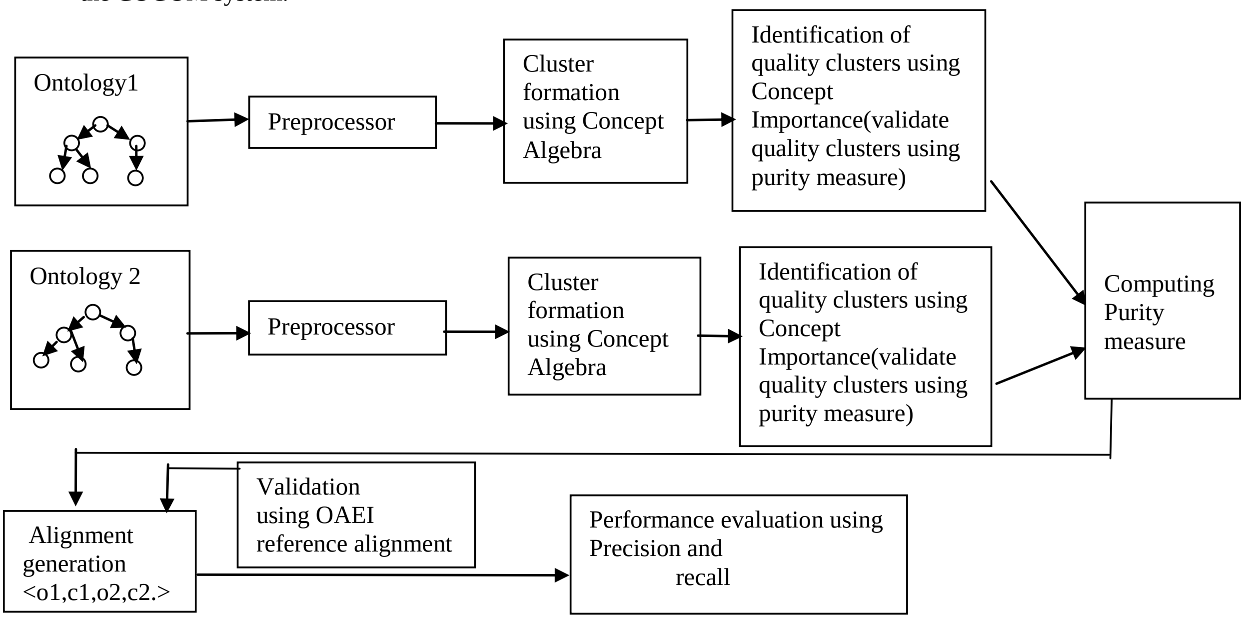

Fig1.Proposed COGOM architecture diagram

6.1Explanation

The proposed ontology matching system COGOM architecture involves the following process: Initially two ontologies are taken as input and are pre-processed and corresponding classes, attributes are extracted and is given as input to the cluster formation phase and the attribute and relation correlation ship degree is determined and the quality cluster are generated and validated using purity measure. The alignments that are generated are cross verified against the OAEI reference alignments and finally the precision and recall of the system are computed and proves to be highly efficient and scalable system compared to the existing ontology matching systems.

7.Alignment (A) Algorithm COGOM(O1(Ontology),O2(Ontology))

Step1:(Preprocessing)

//E is set of entities and L is set of links and O is ontology

//Input Russian ontologies of OAEI benchmark data set in RDF form

//Output alignments {<O1,C1>,<O2,C2>,SV} where SV is the similarity value that ranges from 0-1.

Graph1(E1,L1)=Preprocessor(O1);

Graph2(E2,L2)=Preprocessor(O2);

Step2:(Similarity measure)

Compute:

(i)Attsim(C_i,C_j) //Compute attribute similarity degree(Attsim) using Tversky Ratio model

(ii)Rsim(C_i,C_j) //Compute relationship similarity degree(Rsim) using Rodriguez method.

(iii)SCD(C_i,C_j)//Compute semantic conception degree(SCD)

Step3:(Cluster formation using concept algebra)

//Formation of cluster using similarity measure

Generate clusters for ontology O1 using concept algebra by aggregation of Attsim and Rsim using SCD computed in step 2.

Generate clusters for ontology O2 using concept algebra by aggregation of Attsim and Rsim using SCD computed in step 2.

Step4:(Quality clusters generation)

//Two clusters C1 and C2

Important concepts are identified using Concept Importance. For those important concepts alone

Compute :

(i)Sup(C_i,C_j) // Compute the ratio of commonalities against distinctiveness of the super-concepts.

(ii)Sub(C_i,C_j) //Compute the ratio of commonalities against distinctiveness of the sub-concepts.

(iii)Int(C_i,C_j) //Compute the ratio of commonalities against distinctiveness of the Intention concepts.

(iv)CCD(C_i,C_j)//Compute cluster correlation degree between concepts using sup, sub and int weights

IF(CCD>ε) // where ε is a threshold which is experimentally determined

Include cluster as member of quality cluster set

ELSE(CCD<ε)

Ignore cluster

Step5:(Validation of clusters)

//Quality cluster set

Check and validate the quality of clusters using purity measure

Purity $_j$ = max P_{ij} //If purity > Threshold clusters are of good quality clusters.

Compute Inter cluster similarity of quality clusters generated for the two ontologies(O1,O2)

Step6:(Generation and validation of clusters)

Generate alignments in the form {<O1,C1>,<O2,C1>,SV}

8.Implementation and experimental results

8.1Implementaion Library ------ JENA

Jena is a Java framework for writing Semantic Web applications. It is used in pre-processing step to convert OWL ontology to using two Russian ontologies(Russian O1,Russian O2) OWL files .

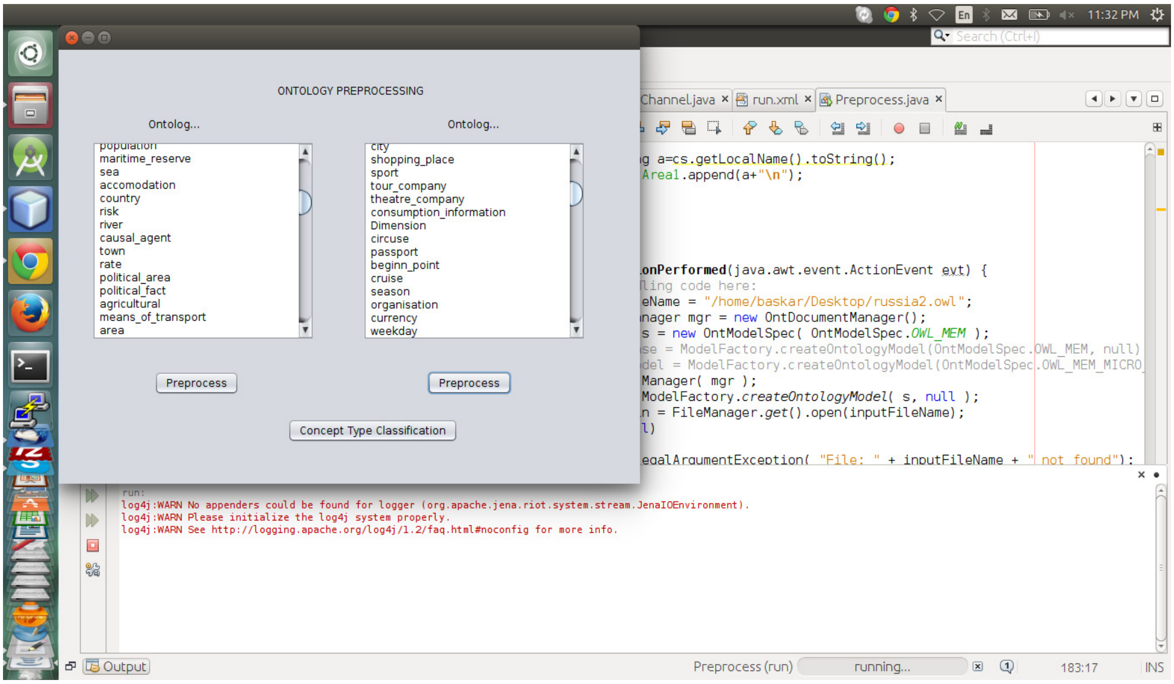

Fig.2.(a)Snapshot of Preprocessing

9.Evaluation metrics

The proposed ontology matching system is assessed and compared with the existing ontology matching system. There are two evaluation metrics computed namely precision and recall of the system which are formulated bellow:

(i)Precision = Number of correctly found correspondences / Number of all found correspondences (1) (ii)Recall = Number of correctly found correspondences / Number of all reference alignments (2)

10.Conclusion and future work

As various computing technologies are growing day to day it has lead to massive amount of disparate information's which results in increasing difficulty of managing these heterogeneous resources across various domains.Thus the proposed system will handle the heterogeneity issues by developing an algorithm which forms robust clusters with proper concept depth explanation styles and uses purity measures to filter good quality clusters thereby decreasing the match space by considering the potentially important concepts thereby enhancing the efficiency and effectiveness of the proposed ontology matching system.The future work will focus on developing parallel matcher work flow which will optimize and will increase the scalability of the system along with developing ontology repairing alignment strategies which will correct the misalignment's occurred during the matching process and increases the precision of the system.

References

[1] Pavel Shvaiko and Jerome Euzenat.Ontology matching:State of the art and future challenges.IEEE Transactions On Knowledge and Data Engineering ;January 2013.vol.25. no.1.

[2] Yingxu Wang.On concept algebra:A denotational mathematical structure for knowledge and software modeling.Int'I Journal of Cognitive Informatics and Natural Intelligence;April-June 2008.2(2),1-19.

[3] Guanyu Li,Yunhao Sun,Jiahong Yan.Concept algebra based ontology merging method.Journal of Information and Computational Science;May 2015.12:7,2665-2679.

[4] M.Andrea Rodriguez,Max J.Egenhofer.Determining semantic similarity among entity classes from different ontologies.IEEE Transactions On Knowledge And Data Engineering,Department Of Computer Science;March-April2003.vol.15.no.2,442-456.

[5] Anna Formica.Concept similarity in formal concept analysis:An information content approach.Elsevier Journal;February 2007.

[6] Kyoung-Mo Yang,Eung-Hee,Suk-Hyung Hwang,and Sung-Hee Choi.Fuzzy concept mining based on formal concept analysis.International Journal Of Computers;2008.Issue 3,Vol2.

[7] Ningheng Jian, Wei Hu, Gong Cheng,Yuzhong Qu. Falcon- AO:Aligning ontologies with falcon.Southeast University;2014.

[8] Fatsuma Jauro,S.B.Junaidu,S.E.Abdullahi.Falcon-AO++:An improved ontology alignment system.Department of Mathematics;2014.

[9] P. Lambrix and H. Tan.SAMBO - A system for aligning and merging biomedical ontologies.J. Web Semantics; 2006.vol. 4, no.1,pp. 196-206.

[10] Miklos Nagy,Maria Vargas-Vera,Enrico Motta.DSSIM-Ontology mapping with uncertainty.2006.

[11] J. Li, J. Tang, Y. Li, and Q. Luo.Rimom: A dynamic multistrategy ontology alignment framework. IEEE Trans. Knowledge and Data Eng; Aug. 2009.vol. 21, no. 8, pp. 1218-1232.