I J C T A, 9(3), 2016, pp. 1687-1693

A Survey on Various Ontology Matching Systems

S. Ranjini* and K. Saruladha**

Abstract: Due to the abundant growth of the semantic web massive amount of heterogeneous data's are growing enormously. This problem of heterogeneity among the web resources can be resolved using various ontology matching techniques and systems. Ontology matching is a great way of solving the heterogeneity and interoperability issues among the web data's that use different but related ontologies. Ontology matching creates a sharable semantic space for the ontologies to access and associate the similar elements across different ontologies. In this paper various existing ontology matching techniques, its limitations and future works are discussed and a brief comparison of all the ontology matching systems is tabulated and presented.

Keywords: Abundant, semantic web, heterogeneity, ontology matching, semantic space, ontology matching systems.

1. INTRODUCTION

Ontology is a formal, explicit specification of a shared conceptualization [1]. The "conceptualization" refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. "Explicit" means that the type of concepts used, and the constraints on their use are explicitly defined. For example, in medical domain, the relation between the concepts diseases and symptoms are causal and a constraint is that a disease cannot cause itself. "Formal" refers to the fact that the ontology should be machine readable, which excludes natural language. "Shared" reflects the notion that ontology captures consensual knowledge that is, it is not private to some individual, but accepted by a group. Depending on the precision of this specification, the notion of ontology encompasses several data or conceptual models. Ontologies are used in many applications, such as database integration, peer-to-peer systems, E-commerce, semantic web services, social networks. However, in open or evolving systems, such as the semantic web, different parties would, adopt different ontologies and many of these ontologies contain overlapping information. Thus using ontologies does not reduce heterogeneity, it raises heterogeneity problems to a higher level.

Ontology matching is a solution to the semantic heterogeneity. It aims at finding correspondences between semantically related entities of different ontologies. These correspondences may stand for equivalence, subsumption, or disjointness, between ontology entities. Ontology entities, in turn, usually denote the named entities of ontologies, such as classes, properties or individuals. However, these entities can also be more complex expressions, such as formulas, concept definitions, queries or term building expressions. Ontology matching results, called alignments, can thus express with various degrees of precision the relations between the ontologies under consideration. As of today, there are quite a lot of applications that require matching large ontologies, such as medicine [2], biology domains, large life-science ontologies [3], E-business [4], web directory [5], web data [6], etc. Therefore, these emerging demands on matching large ontologies bring a new challenge [7] for ontology matching technique.

* M. Tech, Pondicherry Engineering College, *Email: ranjini21691@gmail.com*

** Assistant Professor, Pondicherry Engineering College, *Email: charuladha@pec.edu*

The remainder of this paper is organized as follows: Section 2 deals with the notations and terminologies used in the paper. Section 3 discusses the literature survey made on the various ontology matching systems. Section 4 discusses on the survey summary. Finally, Section 5 concludes the paper.

2. PRELIMINARIES

Definition 1: (Ontology) An ontology is a quintuple $O(C, P, I, A, T)$ where:

- C is the set of classes, i.e., the set of concepts that populates the domain of interest.
- P is the set of properties, i.e., the set of relations existing between the concepts of domain.
- I is the set of individuals, i.e., the set of objects of the real world representing the instances of a concept.
- A is the set of axioms, i.e., statements that say what is true about the modeled domain. Examples a subclass, equivalent classes and disjoint classes axioms.
- T is the set of annotations, i.e., Meta data used to provide some human friendly information. Examples are labels and comments.

Definition 2: (Ontology Alignment Process) The ontology alignment process can be seen as a function f which, from a pair of ontologies $O1$, $O2$ to align, an input alignment A', a set of parameters p, a set of resources r, returns a new alignment A between these ontologies.

$$A = f(O1, O2, A', p, r)$$

Graphically, the ontology alignment process can be represented as in Figure.1 shown below:

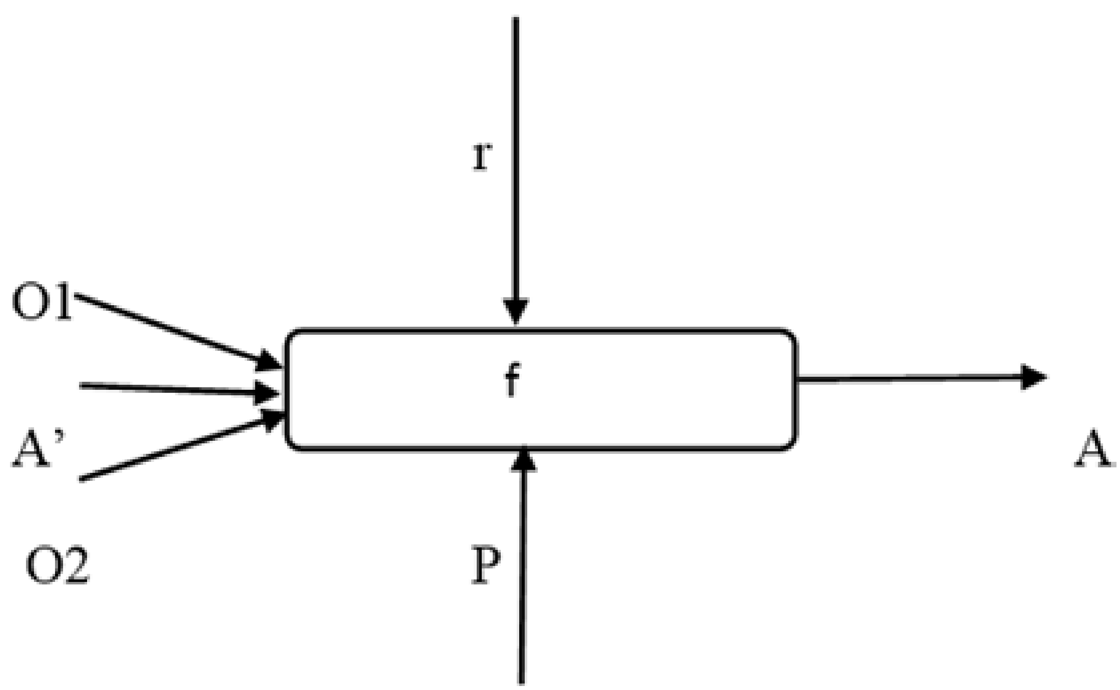

Figure 1: Ontology Alignment Process

Definition 3: (Alignment)

An alignment A is a set of k correspondences, where each correspondence L(with $L = 1, 2, \ldots k$) represents a triple defined as follows:

$$L = (e_i, {}_{ej}, \ddot{Y}) \text{ with } i \varepsilon \{1, 2, \ldots, |O1|\}$$
$$j \varepsilon \{1, 2, \ldots, |O2|\}$$
$$\ddot{Y} \; \varepsilon [0, 1]$$
$$\ddot{Y} > t \; \varepsilon [0, 1]$$

Where e_i is the ith entity of ontology *O*1, e_j is the jth entity of ontology *O*2,Ÿ is the confidence value, *t* is the threshold value used to filter valid correspondences. The next section discusses about the literature survey made on the various existing ontology matching systems.

3. LITERATURE SURVEY

Several techniques have been reported in the literature for matching ontologies. This section in general discusses all the related researches carried out on various existing ontology matching systems.

FALCON-AO [8] is used to align ontologies and is an ontology matching algorithm. It uses two matchers namely LMO(Lexical Comparison Approach) and GMO(Graph Based Approach) and follows partitioning technique. Falcon-AO flows three phases namely partitioning, matching and discovering alignments. It uses the OAEI 2005 data set and provides an automatic divide and conquer method for matching. Falcon-AO is mainly implemented to handle large RDFS and OWL ontologies but has a very minimum common vocabulary and gives poor performance if the structure of the ontologies are different.

SAMBO [9] is a system for aligning and merging biomedical ontologies. Alignment process in SAMBO is performed by computing the similarity values between the terms in the source ontologies. It uses several strategies for performing the alignments and uses two techniques namely relationship alignment and concept alignment. But in SAMBO evaluating alignments manually consumes a lot time and does not have any visualization technique.

DSSIM [10] is a system for mapping ontologies with uncertainty and is based on multiagent based matching framework. It follows the dempster shafter theory of evidence to solve the problem of uncertainty. DSSIM handles large OWL and SKOS ontologies. Instead of human interaction for the mapping process DSSIM uses the iterative closed loop technique. It uses the OAEI 2006 dataset and uses classes and properties that have flat hierarchy due to which all mappings are not fully utilized by the semantic similarity component and also the system does not consider individuals of classes for matching.

RIMOM [11] is a dynamic multistratergy ontology alignment framework and uses the minimization of Bayesian decision. It automatically determines the alignment methods and the information's to be used for similarity calculation. It uses the OAEI 2006, 2007 benchmark and directory datasets. But RIMOM produces unsatisfactory results during alignment process and is inefficient in handling large ontologies for matching.

ASMOV [12] is a system that performs alignments of ontologies with semantic validation. ASMOV automatically adjust weights based on features within the ontologies and evaluates those features within the alignment process. It uses three similarity measures namely lexical, external and internal similarity measures. ASMOV uses the pruning technique to solve the semantic issues at the final stage by removing the incorrect and invalid mappings. But in ASMOV the weights chosen for computing the similarity must be the same for any domain and the convergence technique forces the alignment process to end prematurely leading to minimum alignment generation.

ANCHOR-FLOOD [13] This system handles large ontologies of RDFS and OWL efficiently. It uses two techniques namely schema matching and instance matching and uses the OAEI 2009 datasets to produce the results. The system compares by considering similar concepts from two ontologies and terms it as anchors. Using these anchors it finds the neighbors of the ontologies namely sub concepts, super concepts and matches the corresponding neighbors. But in this system ignores the aligning of distantly placed anchor pairs and its structural transformation is not good for instance matching.

AGREEMENTMAKER [14] This system is used for matching large real world schemas and uses the OAEI 2007 biomedical datasets for matching. It uses a wide range of automatic matchers, user interfaces, user feedback and visual comparison for matching. It handles large scale ontologies and uses two techniques namely similarity computation and alignment selection method. But agreement maker system consumes more memory and the internal data structures used are time consuming which decreases the efficiency of the ontology matching system. The next section briefs the various existing ontology systems.

4. COMPARITIVE ANALYSIS OF THE EXISTING ONTOLOGY MATCHING SYSTEMS

This section discusses and tabulates a comparative analysis of all the existing systems techniques, drawbacks and future works as shown in below table 1.

Table 1
Comparitive Analysis Of The Existing Ontology Matching Systems

Paper Title	*Techniques*	*Limitations*	*Futurework*
(1) FalconAo:Aligning ontologies with falcon[8]-2005 dataset	1. Lexical Comparison (LMO) 2. Graph Matching technique (GMO)	1. Common vocabularies between ontologies is different.	1. Structural and Linguistic comparability needs to be improved.
(2) The results of Falcon Ao in OAEI 2006 campaign[15] -2005 dataset	3. Matchers: V-DOC, I-SUB and GMO.PBM Technique used & Central controller	1. PBM does not perform well with relations that are Complicated.	1. Machine learning approaches to be developed for automatic adjustment of parameters.
(3) Falcon Ao: results for OAEI 2007[16]-2007 dataset	5. components: Central controller, model pool, matcher library, alignment set, repository	1. prefix “rdfs” is not bound in environment task.	1. To develop a stable system that overcomes the drawbacks and builds a comprehensive ontology matching system
(4) Object coref & falcon ao: resultsfor OAEI 2010[17] -2007 data set	Semantics owl: sameas, owl: inverse functional property, owl: functional property, owl: max cardinality	1. Using SPARQL endpoint is time consuming for large scale ontologies.	1. A system to divide the objects into different domains
(5) Falcon Ao++:an improved ontology alignment system -2014[18]-conference track dataset	1. Divide and conquer technique 2. String Similarity technique	1. Input information is a bottleneck 2. supports only one to one mapping.	1. v-doc can be extended to consider further neighbors
(6) Asmov: ontology alignment with semanticvalidation[12] -2006, 2004 data set	similarity measures used, rule and validation techniques used, pruning techniques used	1. weights chosen should be same for all domains and convergence causes to produce less than optimum alignment	1. more tests and turnings of weights to be done 2. convergence has to be improved
(7) Asmov results for OAEI 2010[19]-2009 data set	1. unique process of semantic verification 2. text matching algorithm	1. entire content of ontologies should be stored into memory prior matching.	1. Increase in storage memory. 2. use of permanent storage
(8) Agreement maker:efficient matching for real world schemas & ontology[14] -2007 dataset	1. Similarity computation technique 2. Alignment selection technique	1. more memory is needed 2. construction of internal data sets is time consuming	1. A system that consumes less memory space and a system that uses internal data structure to save time
(9) Agreement maker light results for OAEI 2014[20] -2012 dataset	1. Lexical matching techniques 3. improved alignment repair module	1. does not handle instance matching	1. To reduce erroneous mappings partial reference alignments are needed
(10) Anchor prompt:using non local context for semantic matching[21]-2001 dataset	1. Anchor pair techniques used 2. ontologies are treated as graphs 3. merging of ontologies of overlapping domains. 4. Analyzing of non local concept	1. The system does not perform well when two ontologies differ in connecting with its classes and properties in depth.	1 A system that is capable of equally connecting in depth to its classes. 2. A system capable of working with any type of ontologies

(*contd*...)

(*Table 1 contd...*)

Paper Title	*Techniques*	*Limitations*	*Futurework*
(11) Sambo a system for aligning and merging biomedical ontologies [9]-2004 data set	1. aligning ontologies 2. merging ontologies 3. single filtering technique	1. evaluating alignment strategies manually takes time.	1. developing tool kit for aligning and providing best suited alignments
(12) Sambo results for the OAEI 2007[22]-2006 data set	1. Interacts with user to decide on final alignments 2. conflict checker used	problem of choosing which alignment best suits the given task	1. developing recommendation methods for alignment strategies.
(13) Sambo and Sambo dtf results for the OAEI 2008 [23]-2007 data set	1. sambo uses string matching and domain knowledge techniques 2. Sambo dtf uses advanced filtering techniques (double filtering).	1. second and third results for anatomy is not produced	1. developing recommendation systems suggest best matchers filters that increases precision of the system
(14) Dssim:ontology mapping with uncertainty[10]-2006 dataset	1. Dempster Shafer theory of evidence 2. iterative closed loop technique	1. flat hierarchy of classes and properties so semantic similarity of all mappings not found of classes.	1. a system that can handle complex comparison operation 2. scalability of uncertainty should be handled.
(15) Dssim:managing uncertainty on the semantic web[24]-2006 dataset	1. Dempster Shafer theory of evidence 2. Dempster Shafer rule of combination	1. Dempster combination rule is computationally expensive	1. precision can be increased by considering natural language descriptors
(16) Dssim results for OAEI 2009[25]-2006, 2007, 2008 dataset	1. multiagent agentArchitecture 2. Demster Shafer theory	1. translation is needed to convert different language ontologies in library track	1. A system that provides multilingual and domain specific background knowledge to improve the precision of the system
(17) Rimom: results for oaei 2010[26]-2019 dataset	1. Inverted index technique used 2. (a) interface layer (b) task layer (c) component layer	1. evaluation of IMEI track is not performed for OAEI 2010	1. IMEI track to be developed which provides good platform to test instance matching algorithm
(18) Rimom 2013 results for OAEI 2013[27]-2010 dataset	4. modules:data preprocess, unique subject Matching, One left object matching, score matching	1. for checking anatomy results more biological information is needed	1. perform multifarm experiment on the conference
(19) Rimom-im: results for OAEI 2014[28]-2013 dataset	1. two techniques used: (a) blocking method (b) similarity aggregation method	aligned predicates is not present	1. an algorithm that automatically aligns predicates
(20) Rimom 2:a flexible ontology matching framework [29]-2010 dataset	1. flexible framework 2. user input strategies	1. for large ontologies data set user input becomes an bottleneck	1. a system to be developed to prove its efficiency over large ontologies with many number of data sets

The next sections deals with the jist of the limitations faced by the above discussed ontology matching systems.

5. SURVEY SUMMARY

Based on the ontology matching systems discussed so far this section briefs a short survey summary of the general limitations suffered by the existing ontology matching systems.

- Improper matching techniques are used which decreases efficiency and increases the execution time of the system.
- The matchers that are used for matching ontologies produce high execution time and decreases efficiency.

- No parallelization of matching tasks are done.
- No modularization of ontologies.
- Improper matching with background knowledge leading to incorrect alignments increases recall and decreases precision.
- Improper matcher selection leads to inefficiency of the system.
- The proposed algorithm for instance matching takes extra time for prepossessing.
- 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 ontology matching does not support many to many mapping.
- The existing system does not filter out the good quality clusters after the clusters are formed.
- There are no ontology repairing techniques used to repair the final resulted incorrect alignments produced by the matchers.

6. CONCLUSION

Ontology matching algorithm plays an important role in traditional applications, such as ontology integration, schema integration and data warehouses. Recently emerged new applications, such as peer-to-peer information sharing, web service composition, autonomous communication system, mobile devices communication, navigation and web query answering make use of the ontology matching algorithms. This paper discusses in detail about the various existing ontology matching systems, techniques used, its limitations and future work. The main aim of this paper is to present a detailed overview of the existing ontology matching techniques which are used to increase the efficiency and effectiveness of the ontology matching process by choosing the proper ontology matching system that fits in properly so as to improve the precision and recall of the ontology matching system which in turn reduces the heterogeneity issues. Therefore, we are working towards discovering a better ontology matching system which incorporates an efficient technique that inculcates the advantages and discards the disadvantages produced by the ontology matching systems and its techniques discussed in this paper.

References

[1] T.R. Gruber, “A translation approach to portable ontology specifications”, Knowledge Acquisition, vol. 5, no. 2, pp. 199-220, 1993.

[2] Zhang S., Mork P., Bodenreider O. and Bernstein P.A., “Comparing two approaches for aligning representations of anatomy”, Artif Intell Med, vol. 39, no. 3, pp. 227-236, 2007.

[3] Kirsten T., Thor A. and Rahm E., “Instance-based matching of large life science ontologies”, in Proceedings of the data integration in the life sciences (DILS), LNCS, Springer, Heidelberg, vol. 4544, pp. 172-187, 2007.

[4] Smith K., Morse M. and Mork P., “The role of schema matching in large enterprises”, in Proceedings of the CIDR, 2009.

[5] Avesani P., Giunchiglia F. and Yatskevich M., “A large scale taxonomy mapping evaluation”, in Proceedings of the international conference on semantic web (ICSW), LNCS, Springer, Heidelberg, vol. 3729, pp. 67-81, 2005.

[6] Su W., Wang J. and Lochovsky F.H., “Holistic schema matching for web query interfaces”, in Proceedings of the international conference on extending database technology (EDBT), Springer, Heidelberg, vol. 3896, pp. 77-94, 2006.

[7] Shvaiko P. and Euzenat J., “Ten challenges for ontology matching”, in Proceedings of the Move to Meaningful Internet Systems, pp. 1164-1182, 2008.

[8] Ningheng Jian, Wei Hu, Gong Cheng, Yuzhong Qu, “Falcon-AO:Aligning ontologies with falcon”, Southeast University, 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.

[12] Yves R. Jean-Mary, Mansur R. Kabuka, “ASMOV:Ontology Alignment With Semantic Validation”, University of Miami, 2007.

[13] M. H. Seddiqui, and M. Aono, “Anchor-flood: results for OAEI-2009”, Proceedings of ontology matching workshop of the 8th international Semantic Web conference, Chantilly, VA, USA, 2009.

[14] I.F. Cruz, F.P. Antonelli, and C. Stroe, “Agreementmaker: Efficient Matching for Large Real-World Schemas and Ontologies”, Proceedings of VLDB, vol. 2, no. 2, pp. 1586-1589, 2009.

[15] Wei Hu, Gong Cheng, Dongdong Zheng, Xinyu Zhong, Yuzhong Qu, “The Results Of Falcon-AO in the OAEI 2006 Campaign”, ISWC Ontology matching workshop, 2006.

[16] Wei Hu, Yuanyuan Zhao, Dan Li, Gong Cheng, Honghan Wu,”FALCON-AO:Results for OAEI 2007", ISWC Ontology matching workshop, 2007.

[17] Wei Hu, Jianfeng Chen, Gong Cheng, YuzhongQu,”Object Coref & Falcon-AO:Results for OAEI 2010", ISWC Ontology matching workshop, 2010.

[18] Fatsuma Jauro, S.B. Junaidu, S.E. Abdullahi, “Falcon-AO++: An Improved Ontology Alignment System”, International Journal of Computer Applications, Vol. 94, No. 2, pp. 0975–8887, 2014.

[19] Y. R. Jean-Marry, E.P. Shironoshita, M.R. Kabuka, “ASMOV: Results for OAEI 2010,” Proceedings of ontology matching workshop of the 9th international semantic web conference, 2010.

[20] Daniel Faria, Catarina Martins, Amruta Nanavaty, “Agreement Maker Light Results for OAEI 2014”, University of Illinois At Chicago, 2014.

[21] Natalya F. Noy, Mark A. Musen, “Anchor-Prompt:Using Non-Local Context For Semantic Matching”, vol. 47, Stansford University, 2012.

[22] He Tan and Patrick Lambrix, “SAMBO Results for the ontology alignment evaluation” intiative 2007, Department of computer and information science, 2007.

[23] Patrick Lambrix, He Tan, and Qiang Liu, “SAMBO AND SAMBODTF Results for the ontology alignment evaluation initiative 2008, Department of Computer Science and Information Science, 2008.

[24] Miklos Nagy, Maria Vargas-Vera, Enrico Motta, “DSSIM-Managing uncertatinity on the semantic web”, The open university, 2007.

[25] M. Nagy, M. Vargas-Vera, and P. Stolarski, “Dssim Results for OAEI 2009,” Proceedings of Fourth Int’l Workshop Ontology Matching (OM) at the Int’l Semantic Web Conf.(ISWC), pp. 160-169, 2009.

[26] Zhichun Wang, Xiao Zhang, Lei Hou, “RIMOM results for OAEI 2010”, Tsinghua University, 2010.

[27] Qian Zheng, Chao Shao, Juanzi Li, Zhichum Wang, Linmei Hu, “RIMOM 2013 results for OAEI 2013”, Beijing Normal University, 2013.

[28] Chao Shao, Linmei Hu, Juanzi Li, “RIMOM-IM results for OAEI 2014”, TsinghuaUniversity, 2014.

[29] Zhichun Wang, Xiao Zhang, Lei Hou, Juanzi Li, “RIMOM2: A Flexible Ontology Matching Framework”, Proceedings of the ACM WebSci’11, pp. 1-2, 2011.