Large-Scale Ontology Matching: a Review of the Literature

Samira Babalou, Mohammad Javad Kargar*, Seyyed Hashem Davarpanah
Department of Computer Engineering, Faculty of Engineering, University of Science and Culture, Tehran, Iran,
samira_babalou@yahoo.com, kargar@usc.ac.ir, davarpanah@usc.ac.ir

***Abstract*—Ontology as the base of semantic web is used in many applications. Different ontologies in the same domain lead some heterogeneities and ontology matching systems are developed for resolved them. Heterogeneities have arisen owing to the fact that these ontologies have been created by various people through diverse methods. Nowadays, using large-scale ontologies in some applications such as medical fields seems inevitable. By using large-scale ontologies, some problems like the shortage of memory consumption and long duration of execution appeared in ontology matching systems. In this paper, large-scale ontology matching systems are studied and proposed a general architecture for them. Then large-scale ontology matching systems classified based on the partitioning large ontologies into several sub-ontologies, as known as the modularization, decomposition, summarization, clustering, and divide and conquer categories. This new classification will be useful for future research works in this field. In order to find out the efficiency of the ontology matching systems the results of OAEI (Ontology Alignment Evaluation Initiative) for the period 2011 to 2015 are compared. In spite of great progress, increasing accuracy is required in some section such as conference and benchmark sections.**

***Index Terms*—Ontology, ontology matching, large-scale ontology matching, literature review.**

I. Introduction

Ontology as the conceptual framework can model the description of one domain. They use in various research fields such as knowledge representation, natural language processing, information retrieval, data base, knowledge management, database integration, information transformation, digital libraries, geographic information systems, visual information retrieval, or multi-agent systems [1].

Today, we witness the growth of data explosion in scientific and commercial domains. WEB is faced with databases by huge amount of information with diverse representations. Existing of a huge heterogeneous of data has turned into one of the most noticeable challenges in the some areas like data integration[2]. For this reason, the matching techniques attempt to develop automatic methods to be able to finding the correspondence between these volumes of data in order to obtain useful information in a variety of applications. In the other words, the ontology matching is a research area that focuses on knowledge discovery using matching ontologies.

Owing to the decentralized nature of the Web, there are numerous ontologies in the domains with overlapping applications and even in the identical domains. For interaction between the web applications, ontology matching has been always recommended for controlling the web heterogeneity [3]. In fact, the matching operation is one of the vital operations in many applicable domains such as ontology integration, semantic web, data warehouse, e-commerce, sensor networks, peer-to-peer systems, semantic web services, social networks [1, 4]. By now, ontology matching has drawn the attention of many researchers. Notwithstanding the accomplished advancements in this field, the issue of ontology matching still remains as the real challenge [5].

Once we work with small ontologies, they can be easily matched by using the existing matching tools, but available ontology matching tools have some problems to process large-scale ontologies. It is not only the tools but also the existing memories and systems are unable to process this large information volume. In addition, this will take a very long time. It needs to be highlighted that increasing the speed of such systems would still have a long way to go [6]. The problem of lack of scalability of ontology matching systems has been given a rise as a substantial challenge for years. The efficiency of matchers is of prime importance, especially, when a user cannot wait too long for the system to respond or when memory is limited. Current ontology matchers are mostly design time tools, which are usually not optimized for resource consumption [6].

One of the challenges of large ontology matching systems is the method of partitioning them. Therefore, this research, will show the variety of methods for partitioning the large-scale ontologies and according them classified the ontology matching systems in modulation, decomposition, summarization, clustering, and divide and conquer categories.

Also, in our current work, in order to find out the efficiency of the ontology matching systems, the results of OAEI (Ontology Alignment Evaluation Initiative) for the period 2011 to 2015 are compared. We investigate the issue of evaluating ontology matching by the OAEI result, since they have a large impact on the development of matchers in recent years and it shows their practical usefulness.

Using the ontology in computer science dates back to early 1990s [7]. Many ontology matching techniques have been investigated for years and many review papers such as [6, 8-11] have discussed them. A comparative analysis of partitioning-based ontology matching has done by Algergawy et.al. [12]. Rahm [13] provided an overview of selected approaches and current implementations for large-scale schema and ontology matching. The main goal of this paper is discussion about large-scale ontology matching.

Today, many applications need matching the large-scale ontologies. For example, in library management, the thesauri need to remove the redundant books for integration[3]. Moreover, in medical and biological fields, large ontologies such as GALEN2 and FMA[14] need matching in order to provide a similar for access and manipulation [15].

The current matching tools can efficiently deal with small ontologies (smaller than 500 classes) [16] but matching the large-scale ontologies is still a serious challenge. For example, in OAEI 2011 in the section for large ontologies with 2000-3000 classes, only 6 out of 16 participating systems could process such ontologies [17]. Similarly, in OAEI 2010 only 50% of the matching ontologies systems could match large ontologies in one hour. Nonetheless, these emerging demands on matching large ontologies have resulted in new challenges for the matching ontology technology. Therefore, in matching large ontologies, the partitioning techniques for ontologies have been proposed with the aim of reducing the space and time complexities.

The rest of the paper is organized as follows. In section II, the overall workflow of large-scale ontology matching discuss, while in section III categorizing of dividing techniques will be shown. The evaluation of OAEI sections show in the section IV, experimental summary shows in the section V, and finally section VI conclude the paper.

II. Large-Scale Ontology Matching Systems

The increasing knowledge on the advantages of ontologies for data processing has led to creating ontologies for the real world domains. Yet, the real world ontologies in medicine, electronic, and business fields would have very large sizes. Ontologies with large-scales can be considered as a type of ontology made for describing the complex real world domains. While dealing with large-scale ontologies, the input ontologies are divided into several sub-ontologies. Afterwards, all the sub-ontologies are matched with each other and then the result of them combined in order to obtain the overall result of matching [18]. Babalou et.al. [18] shows the general architecture for these large-scale ontologies which it has shown in the Fig. 1.

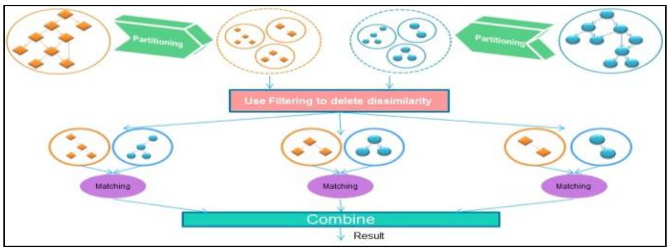

Fig. 1. Overall architecture for matching large ontologies [18]

One of the most challenging parts of these systems is partitioning phase that we categorized large-scale ontology matching systems based on partitioning phase. Before that, we explained the most prominent large-scale ontology matching systems as whole.

III. Techniques of Dividing Large-scale Ontologies

According to the conducted researches, the partitioning methods of large ontologies into small ontologies can be categorized as follows: (1) Modulation, (2) Decomposition, (3) Summarization, (4) Clustering, (5) Divide and conquer, and (6) Other approaches.

A. Modularization

The integrated large-scale ontologies have some problems for maintenance, scalability and conducting. Hence, the modulation of an ontology including the identification of the components (modules) of that ontology which can be considered as a discrete part while these modules are linked and related to other modules. Modules are extracted according to the sub-categories or applications [19]. A module is a minimum set axioms (sub-class, instantiation, equivalence, etc.) which maintain their entire entities and relations. For example, the relations of a class or a concept are within that module, not within another module i.e. it has encapsulating characteristics [20]. There exist many methods such as [4, 21-25] which deal with extracting the modules from the ontologies.

Grau et al. [24] conducted the modulation for large-scale ontologies on the basis of E-connections [26]. The E-connection is a set of partitioned knowledge bases that has been made up of ontology by repeating the analysis of the concepts and roles by using the description logic. Garcia et al. [21] used the partitioning techniques of the graphs from the iGraph library (http://igraph.sourceforge.net/) for modulation of the large ontologies. MOM (Modularization-based Ontology Matching) [4] is a modularization-based approach which decompose a matching problem of large ontology into several small problems by using E-connections which it is similar to the approach proposed by Grau et al. [23]. Grau et al. [23] has provided on formal definition of module so that each entity within ontology has been semantic encapsulation, and it detects the axioms associated with each entity in the ontology. Similarly, Grau et al. [22, 25] and Jimenez-Ruiz [22, 25] uses logic-based approaches for extracting the modules. Moreover, the LogMap [16] has done the ontology matching by using reasoning and the module extraction technique [22].

Due to having the encapsulation characteristic, modules can be used as independent units. Nonetheless, modulation suffers from weakness for ontology matching in fields such as anatomy [20]. Indeed, due to the fact that one concept cannot be as the sub-classes of other concepts in another module, these approaches might create modules with large sizes while ontology matching tools cannot process them. In addition, most of the concepts have "part-of" relations in ontologies such as anatomy. Modulation of the ontologies using E-connections such as [24] maintains the encapsulation property of each module but using the obtained modules by E-connections for refining the ontologies is impossible [27]. Furthermore, E-connections cannot describe the subclass and sub-property relations [28].

B. Decomposition

Description logic is used for designing, merging, and developing the ontologies. Decomposition the ontologies will be used instead of merging them for large ontologies[29]. Several approaches, among which [29-32], have dealt with decomposing the ontologies. With regards to the research on the semantic web, the "Description Logic" (DL) has turned into the main language for describing the majority of terminological knowledge in the Ontology Web Language (OWL) [32].

Stuckenschmidt and Klein [32] tried to make use of Distributed Description Logic (DDL) and local reasoning by compiling implied axioms for decomposing the ontologies into several sub-ontologies. Thi-Anh-Le and Nhan [29] could decompose particular ontology into several sub-ontologies by using description logic on the basis of the graphs partitioning algorithms. They employed two methods based on the ontologies displaying approaches: one approach was to display the ontology with one symbol graph using the minimal separator method based on partitioning and the other approach was axiom graph using the eigenvectors and eigenvalues based on segmenting method. Koney et al. [30] decomposed the ontology independent of syntax forms; in other words, the ontology partitioning results depend on the terminologies' meaning. In addition, Pham et al. [31] used the technique of ontologies overlapping decomposition in order to divide the ontology into several sub-ontologies. Employing the description logic for decomposing the ontologies has complexities. The DL reasoners will not be appropriately scalable by being merged in large ontologies through mapping [33].

C. Summarization

Summarization or extraction approach provides a summary of that ontology as a smaller or more compacted ontology. Indeed, the summarized ontology covered all the main concepts. While defining the ontology summarization in [34], it is admitted that this is an automatic process for generating a summarized version of the ontology in which the important information has been provided for the user. Summarizing the ontology helps in the users' rapid understanding and effortless perception, and also beneficial in facilitating the engineering works of the ontologies [35] as well as being used in other works such as reasoning applications [36].

Li and Motta [35] have categorized the ontologies summarization methods into three categories, namely the driving force, the working unit of summarization, and extractive or abstractive. They additionally investigated the criteria for assessment of the ontologies summarization. Readers are referred to [35] for further studies and information.

Peroni et al. [37] extracted key concepts such as ontologies summarization. They employed criteria such as the name, density, coverage surface, and popularity. In addition to defining the ontology summarization, Li et al. [34] studied the most important properties of the ontologies which needs to be included in the summarized ontology. Zhang et al. [38] considered the RDF sentences as the main unit of summarization while extracting the sentences was considered as the results of summarization. They built RDF sentences graph and considers the RDF sentences as the nodes and the link between them as the edges. Afterwards, for each node they calculate the centrality measure as the proportional importance of them.

The ontology summarization approaches are useful in query answering but are not efficient in the ontology matching, Because of the sub-ontologies obtained from this approach have large sizes and the existing tools for ontology matching cannot easily process those large-scale ontologies.

D. Clustering

The simplest but the most useful approach for executing the matching on large-scale ontologies is clustering [13] in which the large ontology is divided into several clusters using different techniques.

Algerygawy et al. [39] use graph clustering method so that the nodes in one cluster are similar to each other regarding their structures while the nodes in different clusters differ from each other. In fact, the criteria for clustering are the structural similarity of the nodes and their connections. In this approach, the idea of structural similarity has been derived from the AHSCAN algorithm [40] emphasizing that the nodes in one network will have very high structural similarity if they have similar connections. Also, The AHSCAN algorithm has been adopted from the SCAN algorithm (A Structural Clustering Algorithm for Networks) [41] whose operation trend is bottom-up i.e. it initially considers every node as a separate cluster and then merges the clusters on the basis of their structural similarity . Indeed, it uses the relations connected to one element of its neighbors for computing the similarity measure. Noteworthy, the neighborhood similarity has been used in [3, 42-45].

SeeCOnt [46] is a seeding based clustering method which propose a rank function to determine the seed of each cluster

(CH). In the membership function, it uses the string and structural similarity measures between CHs and concepts. Also, Ahmed et.al. [47] propose a revision and an enhancement of K-means clustering algorithm based on a new semantic similarity measure.

In ontology clustering, Saruladha et al. [44] used TNSP (Tversky-based Neighbour Structural Proximity) and DNSP (Dice based Neighbour Structural Proximity). Furthermore, in another research Saruladha et al. [43] could enhance AHSCAN [40] algorithm by limiting the calculations to the neighboring node. Pei et al. [48] clustered the schemas on the basis of their textual similarities. Then they cluster the schemas properties which are equal in the schema clusters for finding the corresponding characteristics among these schemas. Ultimately, the properties are clustered according to different schema clusters by using the statistical information collected from the existing attribute clusters for finding the corresponding properties among the schemas.

By making use of the hierarchical clustering algorithm, Tran et al. [20] divided the large ontologies semantically into groups called clusters. In order to calculate the semantic similarity between the concepts, the information content [49] of each entity is used so that the semantic closeness of the entities will be assessed. The information content is a fundamental criterion which measures some information such as generality grade of correctness by using the concepts which appear in their texts.

The clustering results depend on the type of clustering algorithm and similarity measures. In addition, the hierarchical clustering algorithms are not appropriate for large datasets while the partitioning algorithms would outperform them in large datasets [50]. Moreover, due to the largeness of the data, using the clustering methods with high orders would lead increase calculation complexities and harder implementations.

E. Divide and Conquer

The “Divide and Conquer” techniques divides one problem into several sub-problems. These techniques have used in large-scale ontology matching, too. Hu et al. [3] used the “Divide and Conquer” technique for the scalability problem of the ontologies in order to match large-scale ontologies. They use of a structure-based partitioning algorithm through a bottom-up approach called Rock [51].

The Partition-based Block Matching method (PBM) [52] has utilized the idea of “divide and conquer” for partitioning the large ontologies into small blocks. It divide every ontology by the ROCK algorithm [51] into independent blocks on the basis of structural and linguistic similarity. Then it defines the weighty links according to a structural proximity. Then, according to these weighty links, it uses two criteria, namely the Cohesiveness within the blocks and Coupling between the categories so that it will be able to divide them into blocks. Coma++ [53] uses a fragment-based matching method for matching the large ontologies.

F. Other Approaches

There exist other approaches which do not include in our previous categories. GOMMA [54] match the large-scale ontologies using parallelizing and implementing on several machines. Yam++ [55] has also performed the ontologies matching very well by employing machine learning and information retrieval techniques. Furthermore, for large ontologies matching, the MapReduce technique [56] has been used by employing the partitioning method which was based on the words weight in the V-Doc+ [57].

IV. Evaluation

OAEI (Ontology Alignment Evaluation Initiative) is an international initiative that began operation since 2004. The aim of OAEI is to compare systems on the same basis and to allow anyone for drawing conclusions about the best matching strategies. OAEI is considered as a comprehensive exam for the existing matching systems. This initiative has provided a systematic approach for assessing the ontology matching algorithms and for identifying their strengths and weaknesses. It also introduces new sections and challenges annually while the participants attempt to enhance their systems every year. In fact, the participating at the OAEI is shown to efficiency and practicality of the ontology matching systems.

The number of participating systems in the OAEI is increasing every year. Fig. 2. illustrates the number of the participants in the OEAI. It has shown that the participants in the OAEI ascending increased. However, 2014 has suffered a significant decrease with only 14 systems. The systems participating in the OAEI would execute their systems on standard datasets. Therefore, the results of them can be able to compare in different years due to equality of their datasets. For this reason, this paper surveyed the results for years 2011-2015 (http://oaei.ontologymatching.org). Noteworthy, the comparison and evaluation of the results of the OAEI have been done until 2010 in some papers such as [6, 8]. Measures such as the Precision, Recall, and the F-measure, specialized for mapping the ontologies [58], are used for determining the goodness of a matching. We studied three oldest test cases i.e. Benchmark, Anatomy, and Conference. Because of this paper focus on large-scale ontologies we studied Largebio section, too.

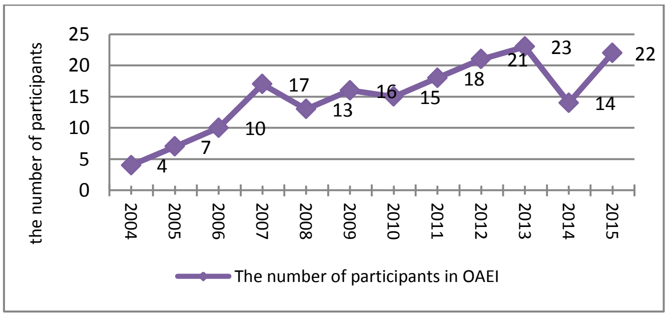

Fig. 2. The number of participants in OAEI

A. The Benchmark Section

The aim of the Benchmark data set is to provide a stable and detailed picture from each algorithm. To do so, the algorithms are run systematically on the produced test cases. Test sets of the Benchmark have been built around the seed ontology and many variations of it. For this purpose, the usual bibliography ontology has been use. Fig. 3. evaluates and compares the three top systems in years 2011-2015 on the basis of the highest F-measure.

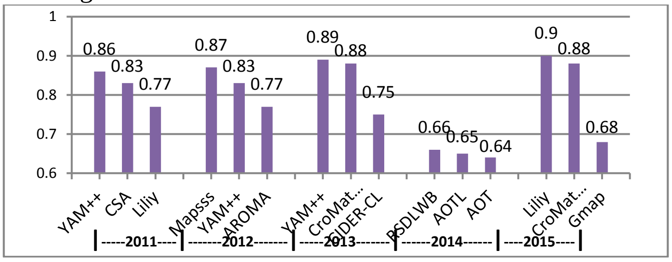

Fig. 3. Three superior systems in the Benchmark section in 2011-2015

As shown by Figure 6, the highest F-measure of the systems in this section was below 0.9. Yet, this value is significantly higher than the years before 2011. YAM++ is the top system in 2011-2013, while Liliy has remarkable advancement in comparison with the results of the previous years in 2015.

B. The Anatomy Sections

The section of Anatomy test includes ontologies from the medical fields. The two datasets used in this section are the human anatomy and anatomy of the mouse. These two datasets have been in use since 2007 while they have undergone some trivial changes in recent years. Three top systems in 2011-2015 have been shown in Fig. 4. on the basis of the highest F-measure.

As depicted by Fig. 4. a considerable advance is observed in 2013 in this section. All the F-measures of the three top systems on this year were above 0.9. Also, the best F-measure belonged to AML in 2013-2015. The YAM++ system also experienced a significant growth in these years. In 2012 and 2013, this system has been chosen as the most top system while in 2011 it encountered the lack of memory error. CODI and AgeMaker, participating in the year 2010, could be among the three top systems in this section in the following year by enhancing their systems. Yet, it is indispensible to highlight that the LogMap is 25 times quicker than AgreementMaker and 75 times rapider than CODI.

This section has been executed on systems with various powers in recent years. Therefore, it is impossible to establish a comparison between the results of the execution time of these systems due to the inequality of the execution environment For instance, in 2013, the three top systems on the basis of execution time were LogMap, GOMMA, and IAMA but the three top systems based on the F-measure were AML, GOMMA and YAM++. This indicates that there is no direct

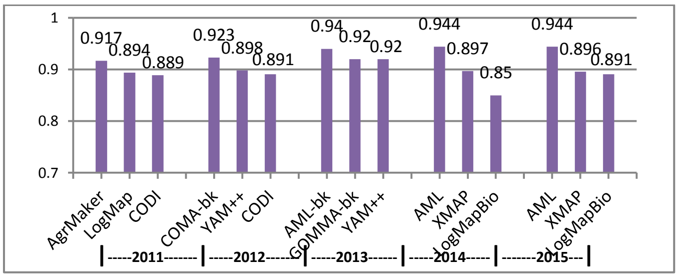

Fig. 4. Three superior systems in the anatomy section through 2011-2015

relationship between the qualities of the matching results with execution time.

C. The Conference Section

This section introduces matching several moderately expressive ontologies, which the results of the participants are evaluated with the reference alignments using the logic reasoning. The assessments are performed by the SEALS (Semantic Evaluation At Large-scale) infrastructure which is a software infrastructure for assessing the automatic running. The dataset of this section entailed 16 ontologies in the field of organizing conferences. Fig. 5. shows the three systems with the highest average of the F_1-measure in years 2011-2015. According to Fig. 5, the 2015 result is significantly better than the results of other years, also YAM++ in 2011-2013 ranked the first.

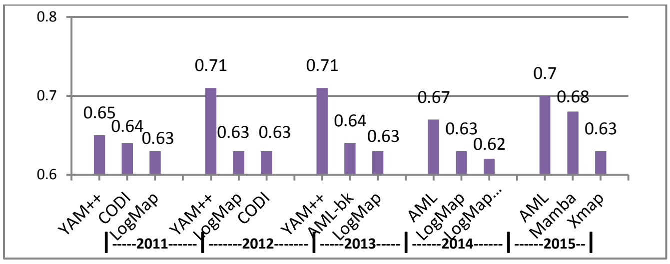

Fig. 5. Three superior systems of OAEI in 2011-2015 in the conference section

D. The Large biomedical ontologies Section (largebio)

The Largebio is one of the most challenging sections of scalability and complexity. The ontologies existing in this dataset are semantically stronger while entailing hundred thousands of classes. In fact, the aim of this section is to detect the alignments among large medical ontologies FMA, SNOMED-CT, and NCI that they have 78989, 306591, and 66724 classes, respectively.

TABLE I. shows a summary of the results for the three top systems in 2011-2015 in the Largebio section. This table has been sorted on the basis of the average of F-measures. For similarity of the experiments' tests in these years, only the FMA section with NCI (Test1), FMA with SNOMED (Test2), and NCI with SNOMED (Test3) have been compared. For 2011, only the FMA-NCI section has been tested.

TABLE I. THREE SUPERIOR SYSTEMS OF OAEI IN 2011-2015 IN THE LARGEBIO SECTION

	Test1		Test2		Test3	
	Matcher	F-m.	Matcher	F-m.	Matcher	F-m
2011	LogMap	0.83	-	-	-	-
	GOMMAbk	0.82	-	-	-	-
	GOMMA	0.81	-	-	-	-
2012	YAM++	0.86	ServOMapL	0.77	YAM++	0.68
	GOMMA	0.84	ServOMap	0.75	ServOMapL	0.67
	ServOMapL	0.84	YAM++	0.74	LogMap	0.67
2013	YAM++	0.87	YAM++	0.82	ServOMap	0.72
	LogMap	0.83	AML-BK	0.77	YAM++	0.71
	LogMapBK	0.83	AML	0.76	AML-BK	0.70
2014	AML	0.84	AML	0.75	AML	0.76
	LogMap	0.83	LogMap	0.71	LogMapBio	0.70
	LogMapBio	0.79	LogMapBio	0.70	LogMap	0.70
2015	XMAP-BK	0.86	XMAP-BK	0.81	AML	0.76
	AML	0.84	AML	0.75	LogMapBio	0.71
	LogMap	0.83	LogMap	0.72	LogMap	0.70

V. EXPERIMENTAL SUMMARY

As the overall, in these five years, there is an evident progression observed in comparison with the preceding years. The average of the F-measure of three top systems in 2011-2015 have been displayed in Fig. 6. to show advancing of these years in the benchmark, conference, anatomy and LargeBio sections. Notwithstanding the mentioned advances, the average of the results of these systems in sections such as conference and benchmark are still low and they require further accuracy.

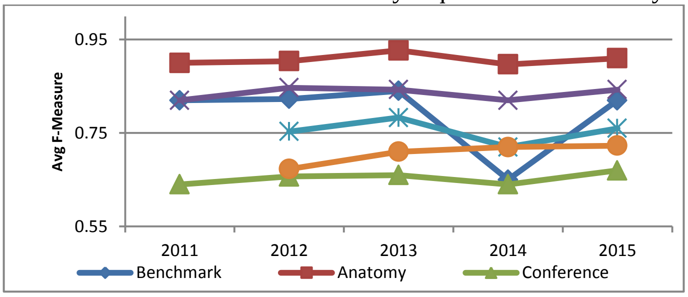

Fig. 6. Evaluating the Mean of F-measures for the years 2011-2015 in various datasets.

It is ostensibly evident that the LargeBio section has experienced a remarkable progress, but the lack of high accuracy has been counted as one of the unsolvable challenges of these systems. The best result is belonging to anatomy section (around 0.9), and the worst one is related to conference section (lower than 0.7). Also, we can see from the anatomy section, that after a few years progress, the improvement rate is slowing down in 2014 and 2015. Hence, in the forthcoming years, some specific actions have to be taken in this field.

VI. CONCLUSIONS

Variety of ontologies in the same domain in semantic web have led to heterogeneity and therefore led to development of ontology matching systems. It is more than one decade that the ontology matching systems have attempted to solve the problems of heterogeneity and ontologies matching. Today, in many real applications like medical domain, size of ontologies is very large and ontology matching systems, dealing with this large size encounter many challenges like shortage of memory and long processing times.

For this purpose, this paper is studied the ontology matching systems with a focus on large-scale ontologies. Then new classification proposed based on according to partitioning large ontology in several small sub-ontologies. These categorizations include modulation, decomposition, summarization, clustering, and divide and conquer. While clustering, modulation, and divide and conquer are used for large ontologies matching systems, the other two methods, namely the decomposition and summarization have not been applied in ontology matching applications due to their low efficiency.

The international OAEI initiative is prominent in ontology matching systems by providing systematic methods for evaluating the ontology matching algorithms as well as identifying their strengths and weaknesses. Various sets of systems participate in various OAEI test cases, which we report the F-measure obtained from these systems and the respective progress or regress made. Taking into account the results of the systems participating in the OAEI show the grade of them.

Comparing the results of OAEI in the last five years reveal the capability of these systems in dealing with large ontologies, Nevertheless, increasing accuracy is required in some section such as conference and benchmark sections. Our investigations also indicate that there is no direct relationship between the time of execution of matching and the quality of the obtained results.

REFERENCES

[1] R. Kolli, "Scalable matching of ontology graphs using partitioning," University of Georgia, 2008.

[2] S. Sellami, A.-N. Benharkat, Y. Amghar, and R. Rifaieh, "Study of Challenges and Techniques in Large Scale Matching," in *ICEIS (1)*, 2008, pp. 355-361.

[3] W. Hu, Y. Qu, and G. Cheng, "Matching large ontologies: A divide-and-conquer approach," *Data & Knowledge Engineering*, vol. 67, pp. 140-160, 2008.

[4] Z. Wang, Y. Wang, S. Zhang, G. Shen, and T. Du, "Matching large scale ontology effectively," in *The Semantic Web–ASWC 2006*, ed: Springer, 2006, pp. 99-105.

[5] J. Euzenat, C. Meilicke, H. Stuckenschmidt, P. Shvaiko, and C. Trojahn, "Ontology alignment evaluation initiative: six years of experience," in *Journal on data semantics XV*, ed: Springer, 2011, pp. 158-192.

[6] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, pp. 158-176, 2013.

[7] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge acquisition*, vol. 5, pp. 199-220, 1993.

[8] S. Liu, D. Shaw, and C. Brewster, "Ontologies for Crisis Management: A Review of State of the Art in Ontology

Design and Usability," in *Proceedings of the Information Systems for Crisis Response and Management conference (ISCRAM 2013 12-15 May, 2013)*, 2013.

[9] J. Zhu, "Survey on Ontology Mapping," *Physics Procedia*, vol. 24, pp. 1857-1862, 2012.

[10] D. FAROOQUI NK, M. F. NOORDIN, A. MUHAMMAD, M. M. NADZIR, S. MURUGANANDAM, D. SRINIVASAN, *et al.*, "ONTOLOGY MATCHING: IN SEARCH OF CHALLENGES AHEAD," *Journal of Theoretical and Applied Information Technology*, vol. 74, 2015.

[11] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, "Ontology matching: A literature review," *Expert Systems with Applications*, vol. 42, pp. 949-971, 2015.

[12] A. Algergawy, F. Klan, and B. Konig-Ries, "Partitioning-based Ontology Matching Approaches: A Comparative Analysis," presented at the The Ninth International Workshop on Ontology Matching collocated with the 13th International Semantic Web Conference, Riva del Garda, Trentino, Italy 2014.

[13] E. Rahm, "Towards large-scale schema and ontology matching," in *Schema matching and mapping*, ed: Springer, 2011, pp. 3-27.

[14] C. Rosse and J. L. Mejino Jr, "A reference ontology for biomedical informatics: the Foundational Model of Anatomy," *Journal of biomedical informatics*, vol. 36, pp. 478-500, 2003.

[15] S. Zhang and O. Bodenreider, "Hybrid Alignment Strategy for Anatomical Ontologies: Results of the 2007 Ontology Alignment Contest," in *OM*, 2007.

[16] E. Jiménez-Ruiz and B. C. Grau, "Logmap: Logic-based and scalable ontology matching," in *The Semantic Web–ISWC 2011*, ed: Springer, 2011, pp. 273-288.

[17] J. Euzenat, A. Ferrara, W. van Hage, L. Hollink, C. Meilicke, A. Nikolov, *et al.*, "Results of the Ontology Alignment Evaluation Initiative 2011," in *6th OM workshop*, 2011.

[18] S. Babalou, M. J. Kargar, S. H. Davarpanah, and A. Alsayed Algergawy, "Centralized Clustering Method to Increase Accuracy in Ontology Matching Systems," *Amirkabir International Journal of Modeling, Identification, Simulation & Control*, 2015.

[19] M. d'Aquin, "Modularizing ontologies," in *Ontology Engineering in a Networked World*, ed: Springer, 2012, pp. 213-233.

[20] D.-T. Tran, D.-H. Ngo, and P.-T. Do, "An information content based partitioning method for the anatomical ontology matching task," in *Proceedings of the Third Symposium on Information and Communication Technology*, 2012, pp. 272-281.

[21] A. C. Garcia, L. Tiveron, C. Justel, and M. C. Cavalcanti, "Applying Graph Partitioning Techniques to Modularize Large Ontologies," in *ONTOBRAS-MOST*, 2012, pp. 72-83.

[22] B. C. Grau, I. Horrocks, Y. Kazakov, and U. Sattler, "Just the right amount: extracting modules from ontologies," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 717-726.

[23] B. C. Grau, B. Parsia, E. Sirin, and A. Kalyanpur, "Modularizing OWL ontologies," in *Proc. KCAP-2005 Workshop on Ontology Management, Banff, Canada*, 2005.

[24] B. C. Grau, B. Parsia, E. Sirin, and A. Kalyanpur, "Automatic Partitioning of OWL Ontologies Using E-Connections," *In International Workshop on Description Logics*, vol. 147, 2005.

[25] E. Jiménez-Ruiz, B. C. Grau, U. Sattler, T. Schneider, and R. Berlanga, "Safe and economic re-use of ontologies: A logic-based methodology and tool support," in *The Semantic Web: Research and Applications*, ed: Springer Berlin Heidelberg, 2008, pp. 185-199.

[26] O. Kutz, C. Lutz, F. Wolter, and M. Zakharyaschev, "E-connections of abstract description systems," *Artificial intelligence*, vol. 156, pp. 1-73, 2004.

[27] M. Grüninger, T. Hahmann, A. Hashemi, and D. Ong, "Ontology Verification with Repositories," in *FOIS*, 2010, pp. 317-330.

[28] P. Doran, "Ontology reuse via ontology modularisation," in *KnowledgeWeb PhD Symposium*, 2006.

[29] P. Thi-Anh-Le and L.-T. Nhan, "Decomposing ontology in Description Logics by graph partitioning," in *The 8 th International Conference on Computing and Information Technology (IC2IT 2012)*, 2012.

[30] B. Konev, C. Lutz, D. Ponomaryov, and F. Wolter, "Decomposing Description Logic Ontologies," in *KR*, 2010.

[31] T. A. L. Pham, N. Le-Thanh, and P. Sander, "Decomposition-based reasoning for large knowledge bases in description logics," *Integrated Computer-Aided Engineering*, vol. 15, pp. 53-70, 2008.

[32] H. Stuckenschmidt and M. Klein, "Reasoning and change management in modular ontologies," *Data & Knowledge Engineering*, vol. 63, pp. 200-223, 2007.

[33] M. Ba and G. Diallo, "Large-scale biomedical ontology matching with ServOMap," *IRBM*, vol. 34, pp. 56-59, 2013.

[34] N. Li, E. Motta, and M. d'Aquin, "Ontology summarization: an analysis and an evaluation," *International Workshop on Evaluation of Semantic Technologies (IWEST 2010)*, vol. Shangai, China, 2010.

[35] N. Li and E. Motta, "Evaluations of user-driven ontology summarization," in *Knowledge Engineering and Management by the Masses*, ed: Springer, 2010, pp. 544-553.

[36] A. Fokoue, A. Kershenbaum, L. Ma, E. Schonberg, and K. Srinivas, "The summary abox: Cutting ontologies down to size," in *The Semantic Web-ISWC 2006*, ed: Springer, 2006, pp. 343-356.

[37] S. Peroni, E. Motta, and M. d'Aquin, "Identifying key concepts in an ontology, through the integration of cognitive principles with statistical and topological measures," in *The Semantic Web*, ed: Springer, 2008, pp. 242-256.

[38] X. Zhang, G. Cheng, and Y. Qu, "Ontology summarization based on rdf sentence graph," in *Proceedings of the 16th international conference on World Wide Web*, 2007, pp. 707-716.

[39] A. Algergawy, S. Massmann, and E. Rahm, "A clustering-based approach for large-scale ontology matching," in *Advances in Databases and Information Systems*, 2011, pp. 415-428.

[40] N. Yuruk, M. Mete, X. Xu, and T. A. Schweiger, "AHSCAN: Agglomerative hierarchical structural

clustering algorithm for networks," in *Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in*, 2009, pp. 72-77.

[41] X. Xu, N. Yuruk, Z. Feng, and T. A. Schweiger, "SCAN: a structural clustering algorithm for networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007, pp. 824-833.

[42] K. Saruladha, G. Aghila, and B. Sathiy, "LOMPT: An efficient and Scalable Ontology Matching Algorithm," *Procedia Engineering*, vol. 38, pp. 2272-2287, 2012.

[43] K. Saruladha, G. Aghila, and B. Sathiya, "A partitioning algorithm for large scale ontologies," in *Recent Trends In Information Technology (ICRTIT), 2012 International Conference on*, 2012, pp. 526-530.

[44] K. Saruladha, G. Aghila, and B. Sathiya, "Neighbour based structural proximity measures for ontology matching systems," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, 2012, pp. 1079-1085.

[45] N. Schlitter, T. Falkowski, and J. Lässig, "DenGraph ‐ HO: a density ‐ based hierarchical graph clustering algorithm," *Expert Systems*, 2013.

[46] A. S. Alsayed, Babalou; Mohammad Javad, Kargar; Seyyed Hashem, Davarpanah., "Seecont: A new seeding-based clustering approach for ontology matching.," *In Advances in Databases and Information Systems - 19th East European Conference, ADBIS 2015*, pp. 245–258, 2015.

[47] S. S. Ahmed, M. Malki, and S. M. Benslimane, "Ontology Partitioning: Clustering Based Approach," *International Journal of Information Technology and Computer Science(IJITCS)*, 2015.

[48] J. Pei, J. Hong, and D. Bell, "A novel clustering-based approach to schema matching," in *Advances in Information Systems*, ed: Springer, 2006, pp. 60-69.

[49] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *arXiv preprint cmp-lg/9511007*, 1995.

[50] O. A. Abbas, "Comparisons Between Data Clustering Algorithms," *International Arab Journal of Information Technology (IAJIT)*, vol. 5, pp. 320-325, 2008.

[51] S. Guha, R. Rastogi, and K. Shim, "ROCK: A robust clustering algorithm for categorical attributes," *Information systems*, vol. 25, pp. 345-366, 2000.

[52] W. Hu, Y. Zhao, and Y. Qu, "Partition-based block matching of large class hierarchies," in *The Semantic Web–ASWC 2006*, ed: Springer, 2006, pp. 72-83.

[53] H.-H. Do and E. Rahm, "Matching large schemas: Approaches and evaluation," *Information Systems*, vol. 32, pp. 857-885, 2007.

[54] T. Kirsten, A. Gross, M. Hartung, and E. Rahm, "GOMMA: a component-based infrastructure for managing and analyzing life science ontologies and their evolution," *J. Biomedical Semantics*, vol. 2, p. 6, 2011.

[55] D. Ngo and Z. Bellahsene, "YAM++: a multi-strategy based approach for ontology matching task," in *Knowledge Engineering and Knowledge Management*, ed: Springer, 2012, pp. 421-425.

[56] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, pp. 107-113, 2008.

[57] H. Zhang, W. Hu, and Y.-z. Qu, "VDoc+: a virtual document based approach for matching large ontologies using MapReduce," *Journal of Zhejiang University SCIENCE C*, vol. 13, pp. 257-267, 2012.

[58] H.-H. Do, S. Melnik, and E. Rahm, "Comparison of schema matching evaluations," in *Web, Web-Services, and Database Systems*. vol. 2593, ed: Springer, 2003, pp. 221-237.