International Journal of Engineering &Technology, 7 (3.12) (2018) 481-484

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Heterogeneity Management Using OAEI Benchmark Dataset

KaladeviRamar

Vel Tech Multi Tech Dr.RR and Dr.SR Engineering College
Chennai, Tamilnadu, India.

Abstract

The evolution of ontologies and itsapplications are in various fields like artificial intelligence, reasoning, philosophy, biological science, and medical field. The components of ontologiesare concepts, instance, relationships, constraints, axioms and inference mechanism. Ontology is a main source for enabling interoperability in the semantic web. In this paper heterogeneities are identified between information systems and the possible rectification are carried out using OAEI benchmark datasets. Proposed method is compared with S-Match algorithm. The evaluation results shows that proposed method is performed better and structure changes of input ontologies not affect the results.

Keywords: *Ontology, OAEI, Reference ontology.*

1. Introduction

Semantic heterogeneity is considered as same information beingrepresented using different terminologies or structures. Based on the previousworks in databases, semantic heterogeneities are categories into attribute level, entity level, Abstraction level and Data value incompatibilities [Kashyap&sheth, 1996].

Semantic web addresses the semantic heterogeneity issue byproviding solution through ontologies. Ontology is defined as a formal and explicitrepresentationof concepts [Gruber, 1993]. Increase in the awareness of semanticrepresentation of information leads huge availability of ontologies forthe same domain. The ontologies even have mismatches between semantic ofthe structure or concepts called ontological heterogeneities. Identifying relation between the heterogeneous ontologies is difficult foriniitiating semantic knowledge retrieval and information exchange.

The objective of ontology mapping to find identify the relationship between theconcepts of input ontologies thatis useful to achieve interoperabilitybetween variousontologies [Patel et al. 2005]. OAEI bibliographic dataset are considered to find heterogeneities among different representation of same domain (bibliographic) ontologies. Some ontologies are considered as reference ontology in order to identify equal, less general and more general relationships among the concepts of ontologies.

The remaining paper is organized as follows; in chapter 2 ontology mapping using background or reference ontology is discussed. Chapter 3 explains the proposed work. Results are analyzed in chapter 4. Finally chapter 5 concludes the entire work.

2. Related works

Most of the existing ontology matching methods are exploiting the contents of ontologies like concepts, property, instance, axioms. Other views of the methods use external knowledge (reference ontology) in the mapping. This knowledge is derived in distinct ways and from different knowledge sources. When referring background source two aspects are to be considered. First one is relating the input ontologies to the background knowledge and the second is extraction of knowledge from background source.

Semantic web is used as knowledge source that explores the method of using many background ontologies in the mapping task (Sabou et al. 2008). In the experimental setup semantic web is used as a source of these ontologies. In Zhang &Bodenreider (2005), matching is done through reference ontology. Interoperability among various ontologies is identified using matching to reference ontology of the domain. Instead of mapping every entity of ontologies to other, all entitiesare matched to specified reference domain ontology, producing the mutual matches via the reference. An algorithm to identify and use missing background knowledge automatically during the mapping process is discussed in (Giunchiglia et al. 2006) on heuristic based. Iteratively potential background knowledge which is missed is discovered and a pair of matched concepts is called as candidate match and if the entities are not matched, then the major of their sub concepts in the hierarchy below are mapped. Detected missing knowledge is added to the background knowledge and it can be reused in the future.

BLOOMS method, using LOD as background knowledge is based on the aim of using information already available on the Linked Open Data cloud (Jain et al. 2010). BLOOM accepts two ontologies as inputs, which contains schema details.

BLOOM continues with the below mentioned steps. It constructs a forest for each concept name using information from Wikipedia and forests are compared to yield decision on which concept names are to be aligned. By using the alignment API and a reasoner, post processing has been done. The algorithm in (Mascardi et al. 2010) used top level ontologies used as semantic bridges in the ontology mapping process. It follows a systematic procedure of the relationships between features of mapped ontologies (no. of simple and composite concepts, stems, top level concepts, common english prefixes and suffixes, ontology depth), matching algorithms, upper ontologies, and experiment results. Mora et al. (2013) extended the active learning framework for ontology matching proposed in (Shi et al.2009). The existing method is improved by correct graph propagation algorithm, user feedback and by using upper ontologies as semantic bridges. The main limitation related with this method is the lack of reference alignment for further computation of performance metrics.

Corpus is used as background knowledge in (Madhavan et al. 2005). This approach exploits corpus of schemas and mappings in a specific domain in order to increase the robustness of mapping algorithms. Corpus is utilized in two ways. First, increase the evidence of each element to be matched by considering evidence from related elements in the corpus. Second, learnt statistics about elements and their relationships and use them to infer constraints that are to prune candidate mappings. Domain ontology is used as background knowledge to map source and target ontologies in (Aleksovski et al. 2006). Anchoring matching method connect an input ontology concept to related concepts in the reference ontology. The concept of input ontologies can be anchored to concepts of background knowledge can produce relationships on anchors which represent various properties.

3. Proposed System

The standard benchmark dataset from OAEIis taken for identifying mapping between related ontologies. Bibliographic dataset from Ontology Alignment Evaluation Initiative dataset is taken asinput ontologies. Sample test cases 205 and 103 are illustrated in Figure 3.1 (a) and (b). Similar concepts are represented with different terminologies. For example the concepts ('MSCthesis', 'Masterthesis'), ('Doctoralthesis', 'PhDthesis'), (‘Nonformal’, ’Informal’) are semantically equal which are from in test case 103, 205. In the same way, the concept‘Booklet’ of 103 is less general relation with ‘Nonformal’ of205. Likewise ‘TechReport’ of 103 and ‘TechnicalReport’ of 205 belong to the sameparent ‘Report’. In order to achieve efficient semantic based knowledgeintegration and retrieval, the heterogeneities needs to be redeemed.

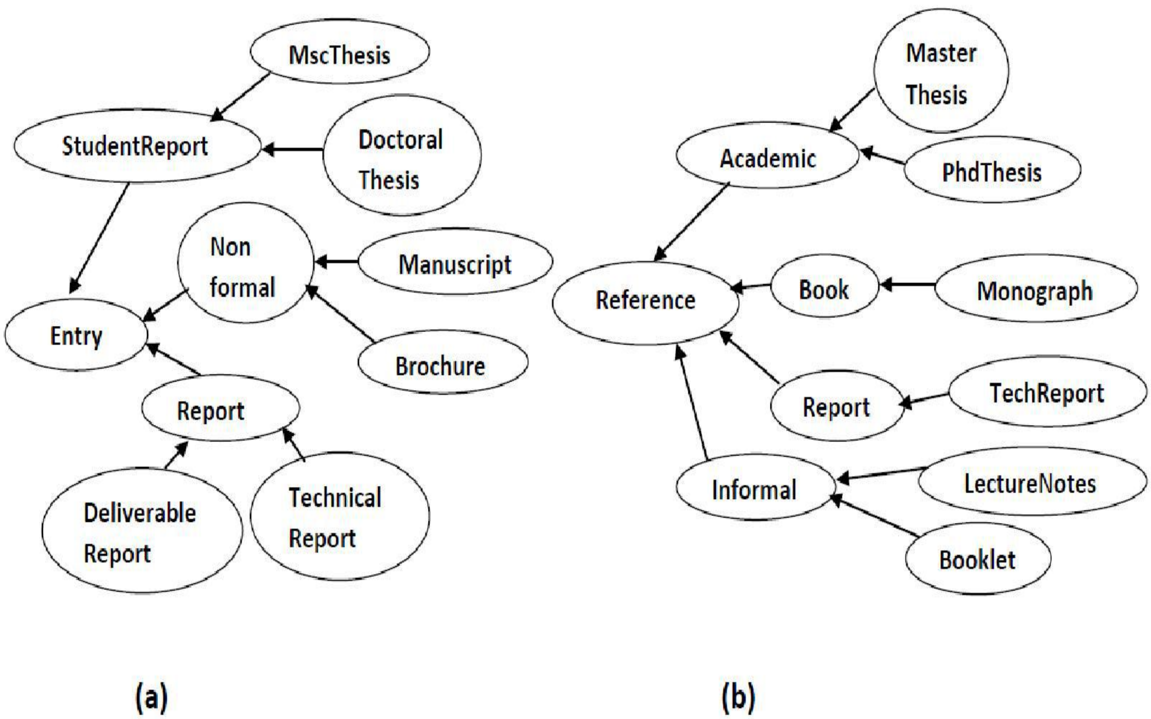

Fig. 3.1: Ontological Representations for Two Bibliographic Systems

The concept 'DoctoralThesis' from Figure 3.1(a) and 'PhdThesis' from Figure 3.1(b) is not lexically similar. Then anchoring matching is performed for 'DoctoralThesis' with background Ontology to derive its positioni.e '1.7.1.1'. Similarly anchoring matching is performed for'PhdThesis' with background ontology to derive its position i.e '1.7.1.1'. Therelation estimator finds the relation between 'DoctoralThesis' of ontology 1and 'PhdThesis' of ontology 2 as equal using the positions found throughanchoring matching which is represented in Figure 3.2.

Similarly the concepts 'Collection' and 'Book' are not relatedsemantically. Using anchoring matching from background bibliographicontology 'Collection' is in position '1.4.1 ' and 'Book' is in position '1.4' . Thenthe relation estimator finds the relation between 'Lecture notes' of ontology 1and 'Reference' of ontology 2 as less general based on the positions foundthrough anchoring matching which is depicted in Figure 3.2.

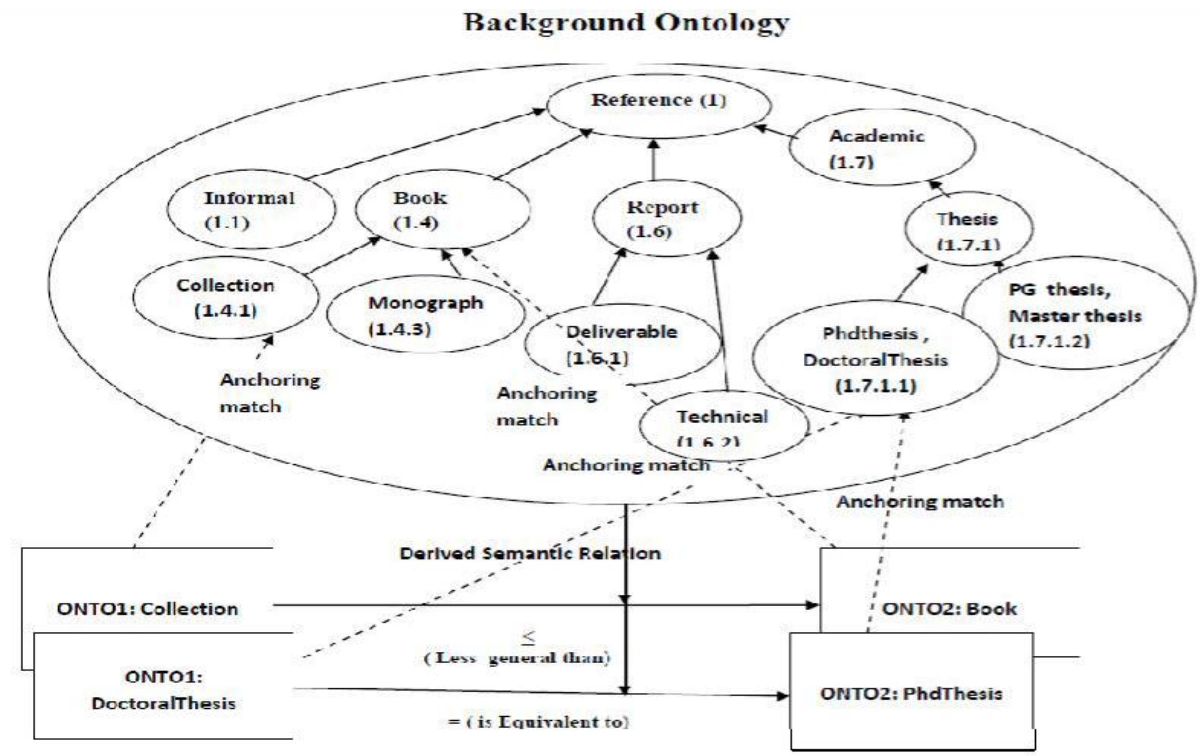

Fig. 3.2: Ontology Mapping Scenarios

4. Results and Discussion for OAEI Test Cases

To evaluate the proposed approach [Kaladevi&Mirnalini, 2015], OAEI benchmark data set are used. It has totally 54 test cases. Thesebenchmark test casescan begrouped into 5 categories .Test cases 101 to 104 have ontologies with same label andhierarchy structure. Ontologies belongs to 201 to 210 test cases have similarhierarchy structure. Test cases are from221 to 247 that havesimilar label representation. Both labelrepresentation and hierarchy structure is different for test cases from 248 to 266. Test cases from 301 to 304 are real time cases given fromvarious institutions.

From each test case category one test case such as 104, 201, 228, 261, 302are considered as background/ reference ontologies. Twotestsets randomly selected from the above test cases are given as source ontologies. Then proposed algorithm is used to identify the semanticrelationship among the source ontologies using background ontology.The performance metrics, such as Precision, Recall, and F - measure, areconsidered to analyze theproposed mapping algorithm. Proposed system is compared with S-match algorithm because both are more aligned. These systems define semantic relationships among concepts as equivalence (=), less general (<),more general (>); and not equal ($\perp$). S-match evaluates entities insame hierarchylevel as (<) relation.

However, proposed systemconsiders the as sibling relation (||). Wordnet is used as background source in addition to element level matchers. Likewise, proposed system exploits reference ontology as background source with lexical matching techniques. S-match mapping relations derived from synonyms, antonyms, meronym or hyponym, holonym or hypernym of Wordnet and their hierarchy levels. Proposed systemidentifies the relations using concept locations in background ontology that is either in same level or in same node or in less or more general relation.The performance of S-match and proposed one is shown in figure 4.1.

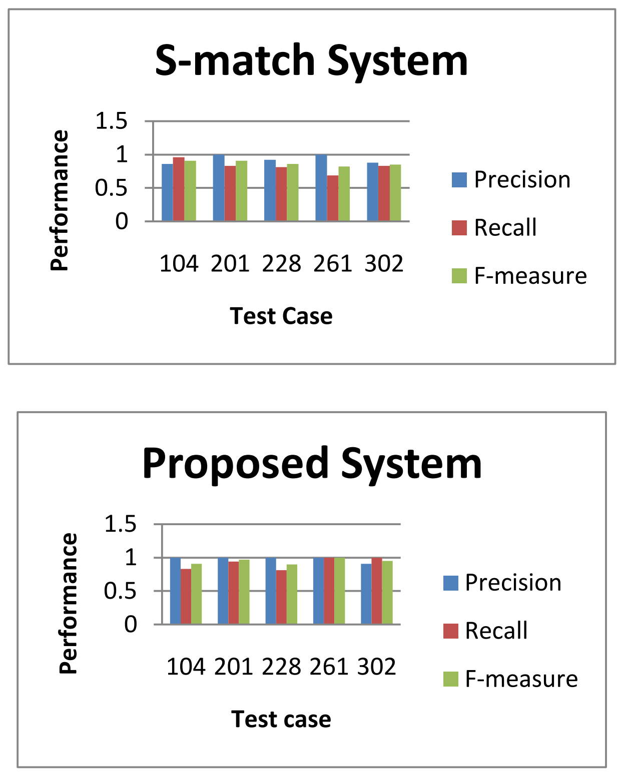

Fig. 4.1: Performance of S-match and proposed systems for various OAEI test cases

Figure 4.1 shows precision of all test cases of proposed system is better thanor on par with S-match. For two test cases, proposed oneresults less recallbecause it does not capture nested information for more than twolevels. S-match recall rate is betterfor test case 104 l because itdepends on the semantic and hierarchical relation

between concepts ofthe tree nodes and structure. With respect to of F-measure, it is apparent that proposed system outperforms S-match. Proposed system performance isnot affected even after the change in the structure of source ontologies, since it is dependson the reference ontology structure and not on the input structure. But structural changes in the source ontologies have a greater impact in S-match.
Figure 4.2 (a),(b) shows the two test sets from test case 104 withslight changes in structure but no change in label. For the first test set, precision of S-match is 57% and proposed is 89%. For the second test set, 47%precision is obtained from S-match and 89% precision is obtained from proposed system. S-match shows lower precision, because the algorithm is based onthe senses of Wordnet as well as the structure of concept as nodes. Therefore, the change in the structure of the above test sets of 104 yields less precision.But, proposed system's performance is not changed by the change of input structurebecause it depends on the background structure.

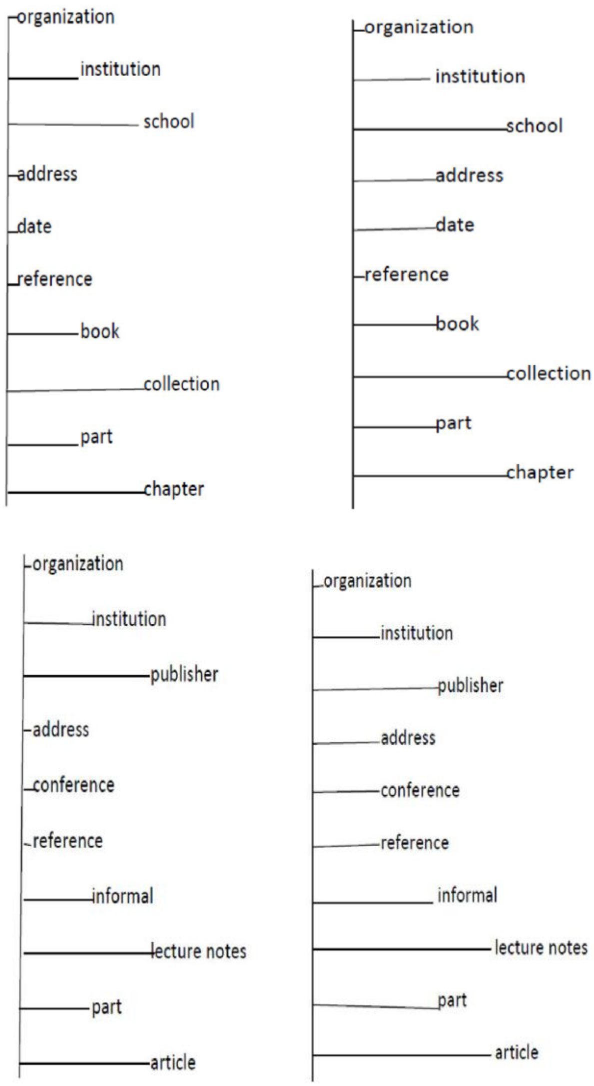

Fig. 4.2: (a), (b) Test Set1, 2 from Test Case 104 with Change in Structure

5. Conclusion

Information heterogeneities are a difficult point to achieve interoperability, integration and achieve a common understanding between information systems. Therefore proposed ontology mapping algorithm is suggested as a solution to address semantic heterogeneities. Since background ontology has rich knowledge, more semantic relations are identified efficiently between the input ontologies. Ontology mapping solutions are essential for the areas such as ontology and data integration, ontology evolution, web service composition data warehouses, information sharing, , search, and query answering. Proposed algorithm is tested with OAEI benchmark dataset and the results are compared with S-match algorithm. Results are comparatively higher and structural changes do not affect performance in proposed systems.

References

[1] Aleksovski, Z, Klein, M, ten Kate, W & van Harmelen, F 2006,'Matching unstructured vocabularies using a background Ontology', InProceedings of Knowledge Engineering and Knowledge Management(EKAW), pp. 182-197.

[2] Giunchiglia, F, Shvaiko, P &Yatskevich, M 2006, 'Discovering missing background knowledge in ontology matching', In ECAI,pp. 382-386.

[3] Gruber T. 1993. A Translation Approach to Portable Ontology Specifications, Knowledge Acquisition. 5: 199-220.

[4] Jain, P, Hitzler, P, Sheth, A, Verma, K&Yeh, P 2010, 'Ontologyalignment for linked open data', in Proc. 9th International SemanticWeb Conference (ISWC), pp. 402-417.

[5] Kaladevi, R &Mirnalini, TT 2015, 'A framework for heterogeneous information management using ontology mapping', ARPN Journal of Engineering and Applied Sciences, ISSN. 1819-6608, vol. 10, no. 5, pp. 2078-2081.

[6] Kashyap, V &Sheth, A 1996, 'Semantic and schematic similaritiesbetween database objects: A context-based approach', The VLDBJournal, vol. 5, pp. 276-304.

[7] Madhavan, J, Bernstein, P, Doan, A &Halevy, A 2005, 'Corpus basedschema matching', in Proc. 21st International Conference on DataEngineering (ICDE), pp. 57-68.

[8] Mascardi, V, Locoro, A & Rosso, P 2010, 'Automatic ontologymatching via upper ontologies: A systematic evaluation', IEEETransactions on Knowledge and Data Engineering, vol. 22, no. 5,pp. 609-623.

[9] Mora, M, Ernesto, R&Ichise, R 2013, 'Ontology matching by activelypropagating user feedbacks through upper ontologies', AcademicJournal, vol. 10, issue 2, pp. 85-93.

[10] Patel M, Koch T, Doerr M, Tsinaraki,C. 2005. Semantic Interoperability in Digital Library Systems, UKOLN, University of Bath. pp. 1-73.

[11] Sabou, M, d'Aquin, M& Motta, E. 2008, 'Exploring the semantic webas background knowledge for ontology matching', Journal on DataSemantics, vol. XI, pp. 156-190.

[12] Shi, F, Li, J, Tang, J, Xie, G & Li, H 2009, 'Actively learning ontologymatching via user interaction', In Proc. 8th Int. Semantic Web Conf,pp. 585-600.

[13] Zhang, S &Bodenreider, O 2005, 'Alignment of multiple ontologies ofanatomy: deriving indirect mappings from direct mappings to areference', In Proceedings of AMIA Annual Symposium, pp. 864-868.