VOL. 10, NO. 5, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

HETEROGENEOUS INFORMATION MANAGEMENT USING ONTOLOGY MAPPING

Kaladevi Ramar and T. T. Mirnalinee
SSN College of Engineering, Anna University, Tamilnadu, India

ABSTRACT

Increase in web and information technologies has made available to large number of independently created and managed information systems. These systems include similar information from disparate sources cause information heterogeneity. To achieve interoperability between heterogeneous information systems and unified integration of those systems heterogeneities between information systems needs to be reduced. Mostly information heterogeneity occurs in three levels: syntactic, structural and semantic. The semantic heterogeneity issue is not completely addressed yet. In this research syntactic, structural, data and semantic heterogeneities between information systems is considered and a novel ontology mapping technique is developed to resolve semantic heterogeneity achieving semantic interoperability between ontologies. Background knowledge has been taken as reference ontology as a part of this work. The Ontology Mapping For Information Management (OMFIM) algorithm is evaluated with OAEI (Ontology alignment Evaluation Initiative) benchmark dataset and the performance is compared against S-match algorithm. Result shows that our proposed method outperforms the S-match algorithm for solving semantic heterogeneity and also best suitable for the systems with insufficient lexical overlap and poor structural correspondence.

Keywords: interoperability, ontology mapping, semantic heterogeneity, integration, semantic web.

1. INTRODUCTION

Managing huge availability of similar information from disparate systems is a serious challenge. This information leads to heterogeneity among systems. This information heterogeneity creates problem with interoperability and integration of different information systems. Heterogeneity occurs at three levels [1]. Syntactic heterogeneity is handled by converting different formats of information to standardized formats such as XML, RDF and OWL. Besides the manually encoded transformation rules, few other middleware components are used to solve the structural heterogeneity problems [2]. However the existing techniques are not sufficient to solve the problem of semantic heterogeneity.

Due to an increased awareness of ontology [3] applications and the availability of multiple ontologies over same domain leads to semantic heterogeneities between ontologies. Ontology mapping has been the suggested solution to find semantic correspondences between similar elements of different ontologies thereby enabling semantic interoperability between them [4].

Ontology mapping is primarily classified into four types such as terminological, structural, instance based and semantic based matching methods. Poor lexical overlap, less structural similarities, and unavailability of instance between ontologies, directs the researchers to another possibility for mapping i.e based on background knowledge. Background ontology is a common vocabulary for a domain to share information and support information integration. The available background knowledge is from web, Linked Open Data (LOD), wordnet, domain ontologies and upper ontologies [5].

In this work we have proposed a system called Ontology Mapping for Information Management (OMFIM) which is based on background ontology. In OMFIM, matching between concepts of input ontologies are found based on their correspondence with the concepts of background ontology.

This paper is organized as follows; in Section 2 some related works are reported to assess the state of the art methodologies and techniques for interoperability among heterogeneous information systems using semantic web technologies. Section 3 discusses our proposed system to resolve semantic heterogeneity using ontology mapping, results and comparison with state of the art technique. Section 4 deals with conclusion and gives directions for future use.

2. RELATED WORK

To guarantee interoperability among several information systems and integration of heterogeneous information, ontology mapping is the suggested solution. We have been discussed some of the ontology mapping systems here.

PROMPT system is developed to support ontology mapping and merging [6]. PROMPT uses a measure of linguistic similarity among concept names and mixes it with the structure of the ontology and users actions. Anchor-PROMPT is, an extension of PROMPT, with a sophisticated prompt mechanism for term matching. It treats ontology as a directed labeled graph [7].

PRIOR+ is an integrated approach based on information retrieval and artificial intelligence techniques [8]. It has string similarity, structure similarity and profile based similarity matchers. PRIOR+ used a harmony based adaptive aggregation method to aggregate multiple similarities. S-Match is a schema and ontology mapping system that uses reasoning and theorem proving methods to find mappings [9]. It starts with a combination of matchers using lexical information and external resources. Then it uses a SAT solver to find semantic relations. COMA++ is a matching prototype which uses several

VOL. 10, NO. 5, MARCH 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

characteristics such as the names and data types of the schema elements and structural information of schemas to determine similarities between them [10]. Two instance-based matchers such as constraint and content based are proposed for in COMA++ to gain a further quality improvement. Falcon-AO is an automatic ontology matching tool [11]. Falcon-AO considers ontologies as graph-like structures, and then generates mappings between elements in the two graphs. There are two matchers integrated in Falcon-AO: LMO for syntactic comparison based on edit distance, and GMO for graph-based comparison.

YAM++ is an ontology mapping approach to deal with both terminological and conceptual heterogeneity of ontologies [12]. It is a combination of machine learning and graph matching techniques. LogMap ontology matching tool addressed scalability and logical inconsistency issues [13]. The core of LogMap has an iterative process starting from the initial anchors, alternates mapping repair and mapping discovery steps. In order to detect and repair unsatisfiable classes, it has built-in reasoning and diagnosis capabilities. WikiMatch, an ontology matching approach based on Wikipedia as a large knowledge external resource [14]. The knowledge in Wikipedia covers almost all possible domains at least to a certain depth. This approach is exploited Wikipedia’s search functionality and inter-language links for finding mappings between ontologies. RiMOM is a systematic approach to quantitatively estimate similar characteristics for each alignment task between ontologies and propose a strategy selection method to automatically combine the matching strategies based on two estimated factors(linguistic and structural) [15].

The above literature concludes that different methods have been developed to assure the interoperability and integration of heterogeneous information systems through ontology mapping. Still information retrieval, querying, integration of information is inaccurate and incomplete because most of the existing systems have the assumptions that the input ontologies are either lexically or structurally similar. This paper deals with considering various heterogeneities in information systems and resolving semantic heterogeneity using ontology mapping.

3. PROPOSED SYSTEM FOR SOLVING SEMANTIC HETEROGENEITIES

Our proposed work classifies heterogeneity between information systems and provides a system to resolve semantic heterogeneities between input ontologies using OMFIM algorithm.

3.1 Heterogeneous in information systems

We classify information heterogeneities into, entity level, attribute level, abstraction level and data value incompatibilities based on [16, 17]. Entity level incompatibilities arise, because of using different descriptions for semantically similar entities. It includes naming, schema isomorphism and database identifier conflicts. Attribute level incompatibilities arises when semantically similar attributes are modeled using different descriptions. These include naming, data precision, data representation and data scaling conflicts. Abstraction level incompatibility arises when two semantically similar entities or attributes are represented at different levels of abstraction. These include type conflicts, dependency conflict and generalization conflicts. Data value incompatibilities arise due to the values of the data present in different databases. Temporal inconsistency, noisy data are of this kind.

In this paper, our proposed algorithm OMFIM resolves naming, schema isomorphism, generalization, and dependency conflicts.

3.2 Heterogeneous ontologies from OAEI benchmark dataset

Ontology describes the knowledge of the domain. For instance bibliographic system from OAEI benchmark dataset is considered [18]. Snippets of the test cases 103 and 205 are shown in fig 1 (a) and (b). Similar concepts are represented with different terminologies. For example the concepts 'MSCthesis', 'Doctoralthesis', ‘nonformal’ in testcase 103 are semantically similar to the concept 'Masterthesis', 'PHDthesis', ’informal’ in testcase 205. Similarly the concept ‘Booklet’ of 103 is semantically related (i.e less general) with 'Nonformal' of 205. ‘TechReport’ of 103 and ‘TechnicalReport’ of 205 are belongs to same parent ‘Report’. Discovering correspondences between heterogeneous ontologies is crucial for enabling efficient semantic based knowledge integration and retrieval.

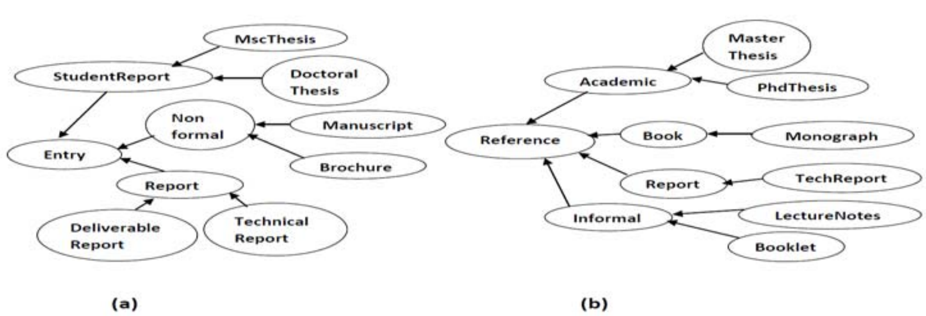

Figure-1. Ontological representation for two bibliographic systems.

To resolve the above heterogeneities OMFIM algorithm is proposed. The correspondence between the concept of ontology1 C_i(S1) and concept of ontology2 C_j(S2) with background ontology is called anchors. Relation identified between C_i(S1) of ontology1 and C_j(S2) of ontology2 using anchors is called anchoring matching. The correspondence is computed using, Levenshtein distance function [19]. OMFIM algorithm is given below.

Algorithm
Input: Ontologies of two input systems S1, S2 and background ontology B.
Output: Mapping relations
C_i(S1), C_j(S2) and C_k(B) represents the concepts of ontologies S1, S2 and B respectively, i=1 to m, j= 1 to n, k=1 to p (m, n, p are number of concepts in S1, S2, B)
Perform lexical matching between C_i(S1) and C_j(S2) $\forall_{i,j}$
if C_i(S1) is equivalent to C_j(S2)
then R → "="
else
- Compute lexical matching between C_i(**S1**) with C_k(B) and C_j(S2) with C_k(B) $\forall_{i,j,k}$. The matched background concepts are called anchors.
- Find the Position of anchors in background.
- Position of anchors belongs to S1 is represented as PBC_i(S1)
- Position of anchors belongs to S2 is represented as PBC_j(S2)

Identify the relation between concepts based on the position of anchors
{
Case 1: PBC_i(S1) is equivalent to PBC_j(S2)
R → "=";
Case 2: PBC_i(S1)) is parent of PBC_j(S2)
R → ">";
Case 3: PBC_i(S1) is child of PBC_j(S2)
R → "<";
Case 4: parent of PBC_i(S1) = parent of PBC_j(S2)
R → "||";
Case 5: otherwise
R → "⊥";
}

4. RESULTS AND DISCUSSIONS

To evaluate our approach, benchmark test cases from OAEI 2007 ontology matching campaign is used. OAEI has 54 test cases. Benchmark tests can further be divided into 5 groups. Test cases 101 to 104 contain ontologies have same label description and hierarchy structure. Test cases 201 to 210 contain ontologies that have similar hierarchy structure. Test cases 221 to 247 contain ontologies which have similar label description. Test cases 248 to 266 have ontologies where both label description and hierarchy structure are different. Test cases 301 to 304 have ontologies which are real world cases defined by different institutions.

From each group one test cases such as 104, 201, 228, 261 and 302 are selected for evaluation which can be used as background ontologies. Two randomly chosen test sets from the selected test cases are considered as input ontologies. Then OMFIM algorithm is used to identify the semantic relationship between the input ontologies using background ontology.

The standard metrics, such as Precision, Recall, and F -measure, are used to evaluate the OMFIM mapping algorithm. OMFIM system is compared with S-match system because OMFIM system is more aligned with S-match system. Both systems represent semantic relationships between concepts as equivalence (=); more general (>); less general (<) and disjointness (⊥). S-match considers elements in same level as (<) relation. But OMFIM takes this as sibling relation (||). S-match uses WordNet as background source along with element level matchers. Similarly OMFIM uses background ontology as reference source with edit distance string matching method. S-match relations are based on synonyms, hyponym or meronym, hypernym or holonym, antonyms between concepts using wordnet and the hierarchy of nodes. OMFIM finds the relations based on input concept positions in background ontology that is either in same position or in same level or in parent child hierarchy.

The performance of S-match and OMFIM is shown in Table-1 in terms of precision, recall and F-measure.

Table-1. Performance of OMFIM and S-match.

Test case	OMFIM			S-match		
	Precision	Recall	F-measure	Precision	Recall	F-measure
104	1	.83	.91	.86	.96	.91
201	1	.94	.97	1	.83	.91
228	1	.81	.9	.92	.81	.86
261	1	1	1	1	.69	.82
302	.91	1	.95	.88	.83	.85

Table-1 shows for all the test cases precision of OMFIM is better than or on par with S-match. Recall of OMFIM is better than S-match for three test cases. For remaining two test cases OMFIM obtained less recall because OMFIM not captures hierarchical information for more than two levels. For the test case 104 S-match obtained better recall because its performance is depends on the semantics relation holding between concepts of the input nodes and its structure. On the basis of F-measure, it is apparent that OMFIM outperforms S-match. Performance of OMFIM is consistent even with the change in the structure of input ontologies. Since OMFIM is relying on the background structure and not on the structure of input ontology. But changes of the structure in the input ontologies have a great effect in S-match.

5. CONCLUSIONS

Ontology mapping is essential for ontology evolution, ontology integration, web service composition, search, and query processing. Since heterogeneities is a crucial roadblock to achieve integration and interoperability between systems. In our system types of syntactic, structural, semantic and data conflicts were considered for an information system. Semantic heterogeneity between the information systems is addressed by a novel ontology mapping algorithm OMFIM using concept positions of the input ontologies in background ontology. The evaluation results show OMFIM is performed well for testcases with exactly same names, same graphs, same graph and different linguistics, different in label description and graph structure and real world cases. OMFIM is consistent for change in input structures. This algorithm needs to be evaluated for more benchmark datasets i.e. all combinations of testcases are evaluated against the reference ontology and our system would be extended to create a common knowledge source.

REFERENCES

Stuckenschmidt H, Harmelen VF. 2005. Information sharing on the semantic web, Springer.

Wiederhold G. 1992. Mediators in the architecture of future information systems, IEEE Computer. 25: 38-49.

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

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

Shvaiko P, Euzenat J. 2013. Ontology matching: state of the art and future challenges, IEEE Trans on Knowledge and Data Engineering. 25(1): 158-176.

Noy N, Musen M. 2000. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment, Proc. Of the National Conference of Arti_cial Intelligence (AAAI). pp. 450-455.

Noy N, M. Musen. 2001. Anchor-PROMPT: Using non-local context for semantic matching, in Proceedings of the IJCAI 2001 Workshop on Ontology and Information Sharing. pp. 63-70, Seattle, WA.

Mao M. 2008. Ontology Mapping: Towards Semantic Interoperability in Distributed and Hetergeneous Environments, Ph.D. dissertion, Pittsburgh Univ., Pittsburgh, PA.

Giunchiglia F, Yatskevich M, Shvaiko P. 2007. Semantic matching: Algorithms and Implementation, Journal on Data Semantics. 1: 1- 38.

Do HH, Rahm E. 2002. COMA- A system for exible combination of schema matching approaches, in Proc. Very Large Data Bases Conf. (VLDB). pp. 610-621.

Jian N, Hu W, Cheng G, Qu Y. 2005. Falcon-AO: Aligning Ontologies with Falcon, In K-Cap 2005 Workshop on Integrating Ontologies. pp. 87- 93.

Ngo D, Bellahsene Z. 2012. YAM++: A Multi-strategy Based Approach for Ontology Matching Task, Lecture Notes in Computer Science. 7603: 421-425.

Jimenez-Ruiz E, Cuenca Grau B. 2011. LogMap: Logic-based and scalable ontology matching. In: Proc. 10th International Semantic Web Conference (ISWC), Bonn (DE). pp. 273288,

Hertling S, Paulheim H. WikiMatch. 2012. Using Wikipedia for Ontology Matching, Seventh International Workshop on Ontology Matching (OM 2012).

Li J, Tang J, Li Y Luo Q. RiMOM. 2009. A Dynamic Multistrategy Ontology Alignment Framework, IEEE Trans. Knowledge and Data Engineering. 21(8): 1218-1232.

Kashyap V, Sheth. 1996. A Semantic and schematic similarities between database objects: A context-based approach, The VLDB Journal. 5: 276-304.

Meenakshi N, Verma K Sheth A, J.A. Miller JA. 2007. Ontology Driven Data Mediation in Web Service, International Journal of Web Services Research. 4(4): 104-126.

OAEI Test Cases:
http://oaei.ontologymatching.org/2007/benchmarks/.

Levenstein I. 1966. Binary codes capable of correcting deletions, insertions and reversals, Cybernetics and Control Theory. 10: 707- 710.