Global Journal of Computer Science and Technology
Software & Data Engineering
Volume 13 Issue 4 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Ontology Mapping for Cross Domain Knowledge Transfer

By Santosh Kumar Banbhrani, Xu DeZhi & Mir Sajjad Hussain Talpur
Central South University

Abstract - The proliferation of domain specific ontologies has improved the ability to represent process and store information in regard to highly specialized domains. However, adhoc transfer of information between domain specific ontologies is not possible. Consequently, multiple solutions have been pro-posed and evaluated as means of facilitating the adhoc transfer of information between another. These range from, structural approaches, which attempt to match knowledge structures between ontologies; lexicographical approaches, that use high level reasoning to match concepts between related ontologies and finally, local structure approaches which look for similar local structures between ontologies to facilitate the transfer of information. To date, the success rate of the published algorithms has been relatively poor. Some of the most successful algorithms, at best are able to match around 50% of the concepts between related ontologies. In this paper we propose a novel global-local hybrid approach to improve the success and accuracy of adhoc information transfer between domain specific ontologies. We demonstrate the efficiency of the proposed algorithm by matching the nodes of three inter-related medical domain ontologies. This demonstrates a significant improvement over existing lexicographical and structural approaches.

Keywords : domain knowledge, heterogeneity, ontology mapping, semantic web.

GJCST-C Classification : D.2.11, D.2.13

ONTOLOGY MAPPING FOR CROSS DOMAIN KNOWLEDGE TRANSFER

Ontology Mapping for Cross Domain Knowledge Transfer

Santosh Kumar Banbhrani [α], Xu DeZhi [σ] & Mir Sajjad Hussain Talpur [ρ]

Abstract - The proliferation of domain specific ontologies has improved the ability to represent process and store information in regard to highly specialized domains. However, adhoc transfer of information between domain specific ontologies is not possible. Consequently, multiple solutions have been proposed and evaluated as means of facilitating the adhoc transfer of information between another. These range from, structural approaches, which attempt to match knowledge structures between ontologies; lexicographical approaches, that use high level reasoning to match concepts between related ontologies and finally, local structure approaches which look for similar local structures between ontologies to facilitate the transfer of information. To date, the success rate of the published algorithms has been relatively poor. Some of the most successful algorithms, at best are able to match around 50% of the concepts between related ontologies. In this paper we propose a novel global-local hybrid approach to improve the success and accuracy of adhoc information transfer between domain specific ontologies. We demonstrate the efficiency of the proposed algorithm by matching the nodes of three inter-related medical domain ontologies. This demonstrates a significant improvement over existing lexicographical and structural approaches.

Keywords : domain knowledge, heterogeneity, ontology mapping, semantic web.

I. Introduction

Ontology's have become a valuable tool to help quantify and process information for decision support systems in highly specialist knowledge domains. Consequently large amounts of both qualitative and quantitative data are processed and stored [3] in various expert systems. The drawback of using these ontologies is that, automated transfer between the systems requires extensive operator intervention in the form of specialist data transfer tools. These tools require the designer to manually map the common information concepts between the two ontologies. As the complexity of the data stored an ontology increases, the complexity of the mapping task and the probability of an error increases. One recent study published by Oellrich et al.[2] found that formal mapping between two ontology's representing the same knowledge domain (Human Pheno-type Ontology and Mammalian Phenotype Ontology defined using the phenomeblast software) was successful at mapping only 48% of the concepts between ontology's. This lack of success at mapping between ontology's has multiple underlying factors such as, knowledge conceptualizations by the developers with implicit assumptions and/or conflicting knowledge structures due on developer assumptions. The assumptions underlying the development of ontology definition and structure arise out of a lack of external standards for the knowledge domain being modeled. External standard setting bodies represent a specific expression of the nature of the information being classified and are able to establish formal relationships for information stored in an ontology can only mitigate this challenge. Thus at the instance of definition, an ontology can at best represent a subset of the scientific world-view in regard to that knowledge domain. This problem is further exacerbated by the presence of multiple standard setting bodies. For example when developing an ontology for medical diagnosis support systems, the developers have a choice of at-least five medical terminology thesauri when using the English language. Individually these controlled vocabularies have well defined application areas with little or no overlap. However, when used to develop an ontology for a specific purpose (clinical diagnosis) the underlying assumptions and world-views of the thesaurus chosen guide and inform the structure of the ontology. This acts as an impediment to the transfer of information between ontologies based on different thesauri. Additionally, when developing an ontology for a specific application area, by choice, only a small subset of concepts in a domain will be used to create the ontology. Due to this, translating all the concepts between from one ontology to another is extremely unlikely to succeed. Therefore, success in concept translation will rely on being able to map all relevant concepts.

This document will report the results of a lexicographical and structural hybrid approach that has been found effective at mapping relevant concepts between related ontologies, developed using a well-defined and restricted vocabulary. This document is organized as follows, the next section will review existing literature for inter-ontology data transfer, following this the next section will present the results for when mapping between three medical domain algorithms with

Author α : Masters of Engineering in Computer Application Technology. E-mail : santosh.banbhrani@gmail.com
Author σ : Professor School of Information Science and Engineering. E-mail : hunan.xu@mail.csu.edu.cn
Author ρ : Ph.D. (Computer Science) Research Scholar. E-mail : mirsajjadhussain@gmail.com
Address : School of Information Science and Engineering, Central South University, Changsha Hunan, China.

techniques identified in literature. After this the next section will describe the novel algorithm proposed in this paper will be described. Finally the results after re-mapping the same three ontologies are presented, after this the conclusion identifies further work that is needed to validate this technique.

II. Related Work

As noted in the previous section, mapping between ontologies developed for limited vocabularies is an extremely active research area. Multiple techniques have been proposed and demonstrated as being effective at mapping between related ontologies; one comprehensive survey of ontology mapping tools published in 2006 by Choi, Song and Han [1] proposed that the terms "*ontology mapping*", "*ontology alignment*" and "*ontology merging*" refer to and indicate different approaches to solving a common subset of challenges.

The paper segments ontology mapping into the following subsets:

A global ontology and local ontologies Here the mapping between ontologies is used to query information from other ontologies, or to map a concept from one ontology into a view.

Mapping between local ontologies This is used to transform entities in one or more source ontology into entities in the target ontology.

Ontology merge and alignment Used to identify unique concepts found in one or more source ontologies being considered for merging or to identify redundant or overlapping concepts.

From the tools described in the paper, semantic matching was common to all the tools described in addition to semantic matching the following approaches were been used for mapping entities and concepts between ontologies. These include but are not limited to, hierarchical mapping, [4-5] probability distribution mapping [6].

Table 1 : Entity mapping success rate between the ontologoies

	Ontology 1	Ontology 2	Ontology 3
Ontology 1	34	12	10
Ontology 2	12	40	13
Ontology 3	10	13	53

Lexical mapping [7] and probabilistic pair matching [8-9]. The tools evaluated in the survey were mainly semi-automated and were designed to be used as support tools for human decision making when mapping entities and concepts between ontologies. Only one of the surveyed tools "CTX Match" [2] is a complexly automated algorithm. As the survey is now 10 years old, the need for a completely automated ontology mapping algorithm has become imperative.

The need for an automated tool has primarily grown due to two reasons, the increasing size and complexity of ontologies used in expert and decision support systems. Secondly, the need to migrate large amounts of accumulated data from obsolete systems to updated ontologies. Invariably for obsolete systems, underlying documentation may be missing, incomplete or unavailable due to various factors.

This work proposes a completely automated ontology mapping algorithm, therefore from the tools evaluated in the survey, the CTX Match algorithm proposed by Bouquet, Serafini and Zanobini8 will be directly comparable.

CTXMATCH is a hierarchical logical reasoning tool that uses the hierarchical relationship between the entities in both the target and source ontologies. The mapping between the source and target inputs H, and H1 in HCs, and for each pair of concepts
(a node with relevant knowledge including meaning in Hierarchical classifications), returns their semantic relation . For example, k is more general than , k is less general than , k is equivalent to , k is compatible with and k is incompatible with .

a) CTXMATCH Results

After processing the three ontologies with the CTXMATCH algorithm the following results were obtained.

Subsequent research into algorithmic ontology mapping has improved on hierarchical mapping by using techniques derived from directed graph matching. To illustrate the potential of a generic directed graph node matching technique, the same three ontologies were re-mapped. Comparing the results in 1 and 2, we can see that even without semantic matching the directed graph entity matching technique is more effective when mapping between ontology 1 and ontology 3. The next section consist of the following the problem statement, introduction to directed graph matching and finally the algorithm description.

Table 2 : Directed graph entity matching

	Ontology 1	Ontology 2	Ontology 3
Ontology 1	34	07	17
Ontology 2	07	40	10
Ontology 3	17	10	53

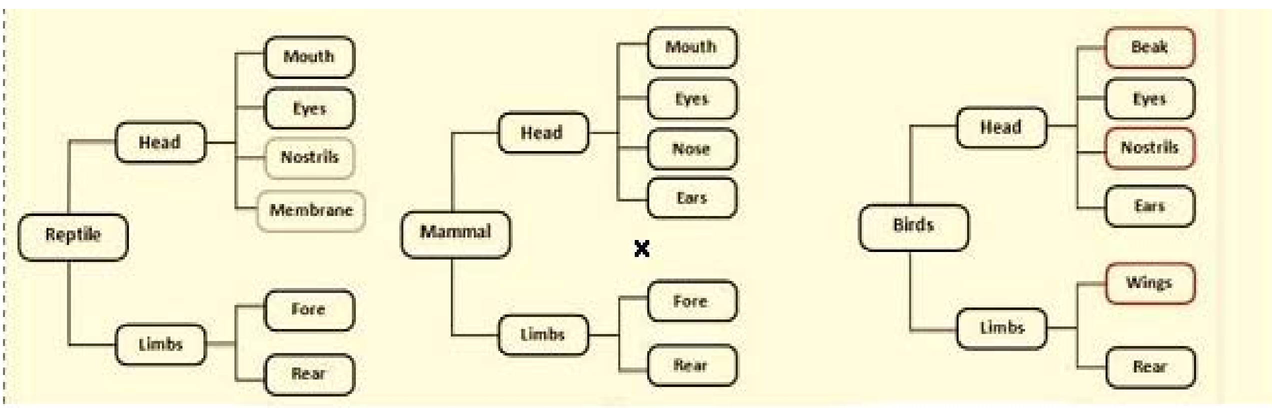
III. Problem Statement

The work in this document is based from a study comparing the performance of sophisticated algorithms evaluated as part of The Ontology Alignment Evaluation Initiative to a simple lexicographical ontology mapping algorithm. From these documents the following common themes can be identified:

- Mapping between domain specific ontologies is a challenging problem, for which currently manual concept mapping is the only effective solution.
- Pattern matching and machine learning algorithms are reasonably successful at ontology mapping. As the assumptions and world views that are a factor into ontology development are di cult to quantify, there is likely to be an upper limit to the concept mapping accuracy.
- Mapping between ontologies using limited vocabularies for similar use-cases is more likely to be effective and accurate.

a) Directed Graph Matching Example

As identified in existing literature lexicographical ontology mapping techniques are extremely effective especially for limited vocabularies. Using a zoological reference textbook as a sample vocabulary for the ontologies in figure 1. A reasoner success-fully maps the *beak* class to the mouth class. The vocabulary used will introduce uncertainty in the mapping of the classes *nostrils* and *membrane* classes of the rep-tile ontology and the *wings* class of the bird ontology. Therefore, the number of mapped classes will range from 18 to 22, i.e. 81% to 100%.



(a) (b) (c)

Figure 1 : Sample Ontologies

On the other hand for example, using a directed graph technique to map in-formation between the classes is extremely simple as the ontologies have the same structure.

Table 3 : Simulated lexical ontology entity matching for figure 1

	Reptile	Mammal	Birds
Reptile	09	07	08
Mammal	07	09	06
Birds	08	06	09

Table 4 : Simulated directed graph matching for figure 1

	Reptile	Mammal	Birds
Reptile	09	09	09
Mammal	09	09	09
Birds	09	09	09

Therefore this map will be 100% successful at class mapping. If on the other hand the structure of the ontologies were to be modified to reflect a different world view, as illustrated in figure 2. For this sample, the graph based method would fail when mapping the world view, as illustrated in figure 2. For this sample, the graph based method would fail when mapping the "*head*" information between the bird and the other two ontologies. Therefore of the 22 classes only 19 classes are successfully mapped, 86% success rate.

These results are summarized in tables 3, 4 and 5.

IV. Proposed Work

To reduce the uncertainty in the lexicographical approach, a novel combined ontology mapping algorithm was proposed. The algorithm combines the lexicographical mapping with the directed graph approach to reduce mapping uncertainty.

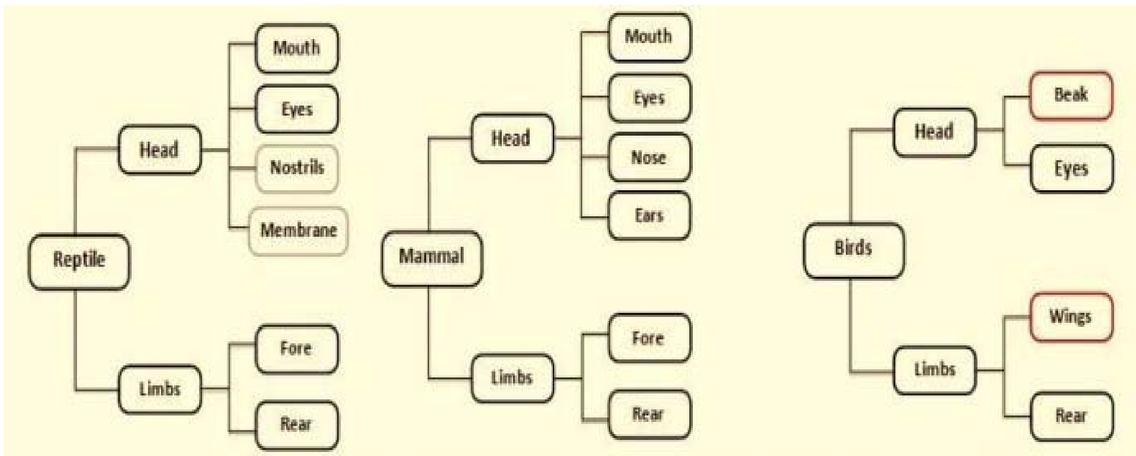
1. Use a thesaurus based synonym (lexicographical) search to identify concept commonality and term networks in the two domains.
2. Read class structure for source and target ontologies to generate node-edge graphs to identify common class structures.
3. Use value matching to bootstrap and validate structural mapping.
4. Use word networks to map areas that do not match structurally and re-evaluate parent nodes.
5. Repeat steps three and four to find any updated root nodes that:

Table 5 : Simulated lexical ontology entity matching for figure 2

	Reptile	Mammal	Birds
Reptile	09	09	07
Mammal	09	09	07
Birds	07	07	09

Table 6 : Results of processing with proposed algorithm

	Ontology 1	Ontology 2	Ontology 3
Ontology 1	34	24	23
Ontology 2	24	40	32
Ontology 3	23	32	53



(a) (b) (c)

Figure 2 : Sample Ontologies

The limited vocabulary dictionaries that were used for the lexicographical matching were obtained from the nih.gov website. To generate the lexicographical map, the Apache-NLP libraries [10] in java were used to generate word associations between the three dictionaries. To generate the `terms of interest' networks of descriptor terms are generated for words common to all three directories.

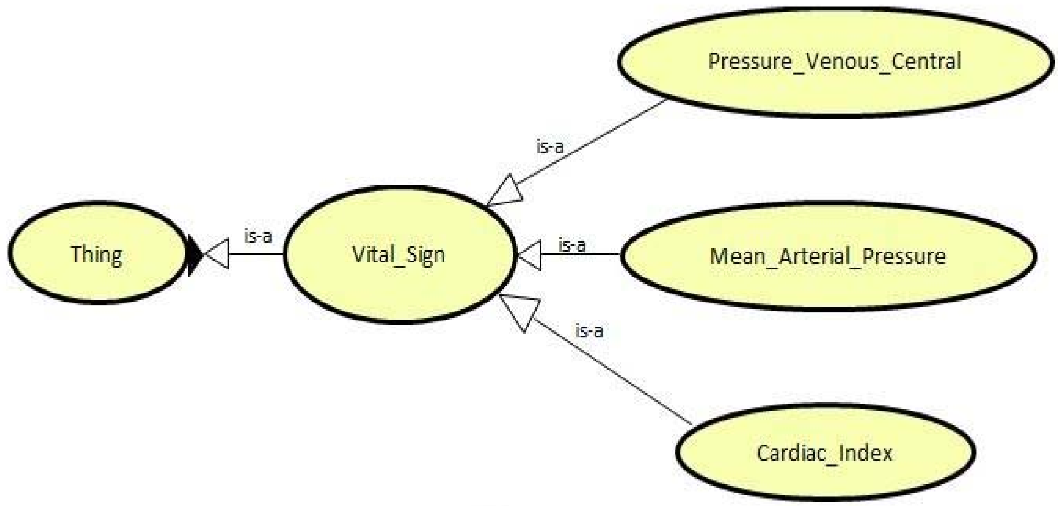
Using the direct graph matching technique illustrated in the previous section (section??) is used as follows. For a root node pair $(R—N)$ in ontology 1 and a similar root-node pair $(R^{'}—N^{'})$ in ontology 2. Node pairs that have the same structure i.e. same properties such as scientific units (physical, chemical or biological), data types are considered matches. For root's, with more than one nodes, an arbitrary value (experimentally determined to be .75) is used as a threshold. That is, if more than 75% of the child nodes of a node match the child nodes of a root node of the target ontology then the root nodes are considered a match.

Following this, any nodes in the source and target ontologies that do not match. NLP network search is used to find matches any nodes that initially were not found to have any corresponding matches. Any subsequently matching nodes are marked as such and the nodes are revaluated to identify any root nodes that may now meet the threshold for matching child nodes. The results of processing the three ontologies and with the proposed algorithm are detailed next.

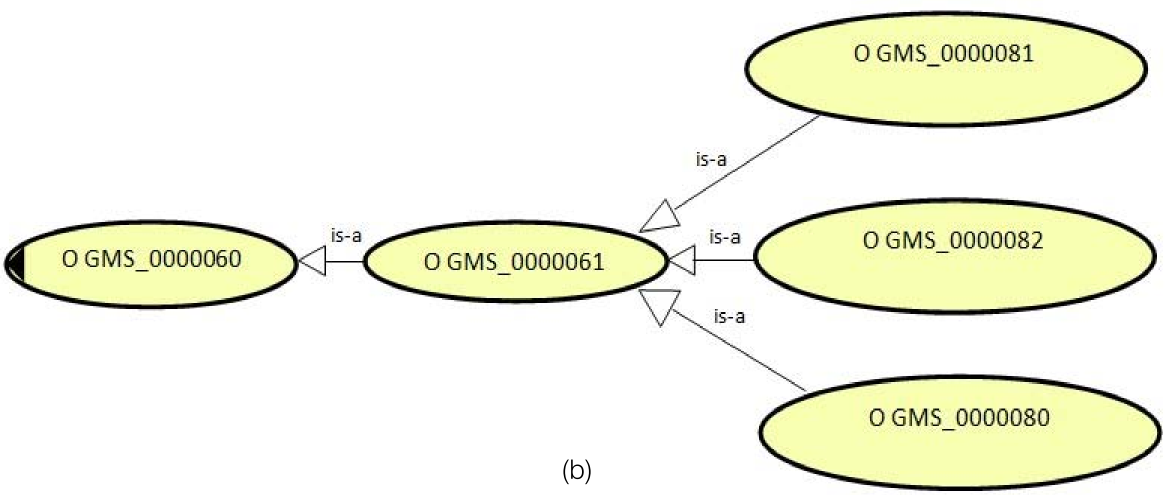
V. RESULTS

As we can see in the table 6, the proposed algorithm improves significantly when compared to the hierarchical or directed graph matching as illustrated in tables 1 and 2. One reason for this could be because all three ontologies are medical support ontologies that use a significantly constrained vocabulary. This and the availability of comprehensive dictionaries that the JAVA NLP toolkit has been designed to process, probably make these ontologies non exemplars when identifying drawbacks to this approach.

Global Journal of Computer Science and Technology (C) Volume XIII Issue IV Version I Year 2013



(a)



(b)

Figure 3 : Successful Combined Match

VI. Conclusion

Since many ontology mapping algorithms have been proposed, a group of criteria are urgently needed to evaluate and compare the results of different algorithms. However, former measures have their own limitations, and none of them can guarantee that semantically equivalent alignments always score the same, which should be a basic character of a real semantic evaluation. By this we have demonstrated that the proposed algorithm can be very effective then existing algorithms. Their performance is equivalent to the performance of the more innovative algorithms. Our evaluation has validated that most of the progressive algorithms are either not freely available or do not scale to the size of biomedical ontologies. We have tested this algorithm and got result 2 which is better than the result 1 which is based on existing algorithms which I used as part of our algorithm. Now that we have some preliminary significant results demonstrating the effectiveness of this approach for use with medical support ontologies. The effectiveness of this algorithm needs to be evaluated with larger and more complex ontologies. In future work we will focus upon testing with ontologies of greater size. Those tests will provide for solider proof whether this method can be successfully applied to the ontology integration problem.

Acknowledgment

This work is supported by National Natural Science Foundation of China under grant No.60970096 and 90818004, the Postdoctoral Science Foundation of China under Grant No.20080440988 and the Natural Science Foundation of Hunan Province, China Under Grant No.09JJ4030.

References Références Referencias

1. Namyoun Choi, Il-Yeol Song, and Hyoil Han. A survey on ontology mapping. ACM Sigmod Record, 35(3):34 {41, 2006}.
2. Anika Oellrich, Georgios Gkoutos, Robert Hoehndorf, and Dietrich Rebholz-Schuhmann. Quantitative comparison of mapping methods between Human and Mammalian Phenotype Ontology. Journal of Biomedical Semantics, 3 (Suppl 2):S1, 2012.
3. L. Yao, A. Divoli, I. Mayzus, J.A. Evans, and A. Rzhetsky. Benchmarking ontologies: bigger or better? PLoS Computational Biology, 7(1): e1001055, 2011.
4. Doan, A., Domingos, P., & Halevy, A. (2003). Learning to match the schemas of data sources: A multistrategy approach. Machine Learning, 50(3), 279-301.
5. Paolo Bouquet, Luciano Sera ni, Stefano Zanobini, Semantic Coordination: A New Approach and an Application", ISWC 2003, LNCS 2870, pp. 130-145, 2003.
6. AnHai Doan, Jayant Madhavan, Pedro Domingos, Alon Halevy\Learning to Map between Ontologies on the Semantic Web", VLDB Journal, Special Issue on the Semantic Web, 2003.
7. John Li, \LOM: A Lexicon-based Ontology Mapping Tool", Proceedings of the Performance Metrics for Intelligent Systems (PerMIS. '04), 2004.
8. Mitra, P and Wiederhold, G, \Resolving Terminological Heterogeneity in Ontologies", Proceedings of the ECAI'02 workshop on Ontologies and Semantic Interoperability, 2002.
9. Prasenjit Mitra, Natasha F. Noy, Anju Jaiswals\OMEN: A Probabilistic Ontology Mapping Tool" International Semantic Web Conference 2005: 537-547
10. opennlp.apache.org