SOMREP: Mapped Meta Information (MMI) for Semantic Web using Ontology Mapping Patterns

S. Chitra
Lecturer
Manakula Vinayagar Institute of Technology, Puducherry

G. Aghila
Professor, Dept. of CS
Pondicherry University, Puducherry

ABSTRACT
Ontologies are currently emerging as representation techniques for overlapping complimentary context domains. A single ontology is no longer enough to support the tasks predicted by a distributed environment like the Semantic Web. Multiple ontologies need to be accessed from several applications. A system incorporating the semantics either implicitly or explicitly to form a formal specification of ontology mapping is vital to achieve inter-operability between the existing ontologies in both homogenous and heterogeneous environment. As a base for the above purpose, this work describes the various patterns that can be used to find the similarity between ontologies through the components Concepts, Relations, Attributes and Values. These patterns play a major role in the specification of ontology mappings, which in turn reduces the number of errors, makes the mapping more concise and understandable. Concept-to-concept mapping pattern has been implemented and analysis has been done using OAEI benchmark test dataset (Google and Yahoo directories). The results and analysis of the application for scenarios of the mapping patterns also have been discussed.

Keywords
Ontology, Ontology Mapping, Ontology Integration, Mapping Patterns.

1. INTRODUCTION
The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries[1]. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. Ontologies have become common on the World-Wide Web. According to dictionary meaning Ontology is defined as [3] "a shared vocabulary, which can be used to model a domain, that is, the type of objects and/or concepts that exist, and their properties and relations".

Ontologies have been used popularly in many fields such as knowledge representation, information retrieval, natural language understanding, biology-Science and Web Services. In recent years, the Semantic Web [1], which aims at providing high-quality intelligent services on the Web, exploits ontologies to model the knowledge of various semantic web applications. In turn, the Semantic Web promotes the researches of ontology greatly. Usually, ontologies are distributedly used and built by different communities.

The Semantic Web contains many distributed ontologies with overlapping domains [5]. In order to allow for interoperation between applications on the Web, these ontologies need to be related to each other through ontology mapping. Of course, ontology integration is necessary [8]. Regarding to [17] "ontology integration is the process of building ontology in one subject reusing one or more ontologies in different subjects". Ontology integration can be spited into various integration types, depending on the degree of integration. Those types are usually referred as ontology mapping, aligning or merging. Different levels of integration can be distinguished: alignment, partial compatibility and unification[2].

Mapping of ontologies refers to an identification of identical concepts or relations between different ontologies [2]. Related to ontology mapping is ontology aligning, which brings two into a mutual agreement and makes them consistent and coherent [8]. In the case of ontology merging a new ontology is built at the base of two or more existing ontologies. This new ontology combines the existing ones.

Ontology mapping is a very challenging topic. It can be done either manually or using semi-automated/automated tools. Manual mapping becomes impractical as the size and complexity of ontologies increases. This work proposes 16 different mapping patterns for mapping two ontologies which can be thought of as building block for various levels of integration process.

Organization of this paper is as follows: Section 2 describes the related work that has been done on Ontology Mapping and mapping patterns. Section 3 briefly explains the proposed Mapping patterns and architecture of the framework. Section 4 depicts analysis done with OAEI benchmark test dataset. Section 5 shows the simulation graph and Section 6 concludes this work with Future Enhancement.

2. RELATED WORK
Ontology mapping is an open problem. Even though ontology mapping solutions for all domains have been proposed in recent years, this section reviews only the specific work related to utilization of mapping patterns for ontology mapping.

Using ontologies in a dynamic environment, such as a Grid, to share some common concepts is still a challenge. It is difficult to keep a static mapping between ontologies. Nelson et al., [11] have adopted the concept of Tuple Space and propose a flexible approach for managing ontologies in a Grid. This approach simplifies the communication process and provides flexibility of participation of all participants.

Ondrej Svab [12] describes a mapping pattern as a graph structure. This paper examined about three simple patterns. Also paper is particularly interested in two types of ontology

[1] http://www.w3c.org/rdf

[2] http://www-ksl.stanford.edu/onto-std/

design patterns: naming conventions and structural patterns. Naming conventions are related to naming classes, properties and/or instances. Structural patterns concern the modeling choices in using certain ontology entities and connecting them together.

Ondrej Svab [13] considered three categories of patterns: content patterns, logical patterns and frequent errors. Content patterns use specific non-logical vocabulary and describe a recurring, often domain-independent state of affairs. Logical patterns, in turn, capture the typical ways of modeling problems, which can be tackled in a specific ontological language. Frequent errors describe inadequate constructions that are often used by inexperienced modelers.

Ming Mao [9] proposes a new generic and scalable ontology mapping approach, it takes advantage of propagation theory, information retrieval technique and artificial intelligence model to solve ontology mapping problem. It utilizes both linguistic and structural information of ontologies, measures the similarity of different elements of ontologies in a vector space model, and integrates interactive activation network to deal with constraints.

Namyoun et al.,[10] have reported about the tools, systems, and related work of ontology mapping. They explain about three ontology mapping categories as 1) mapping between an integrated global ontology and local ontologies, 2) mapping between local ontologies and 3) mapping on ontology merging and alignment. In their work comparison has been done on the evaluation criteria, which are input requirements, level of user interaction, type of output, content of output, and the five dimensions: structural, lexical, domain, instance based knowledge, and type of result.

Yves et al., [20] proposed a new methodology by supporting the uncertainty modeling for ontology mapping using Naïve Bayes theorem. Their methodology works with manual validation, without taking the degree of uncertainty into consideration.

An Ontology-based approach for semantic service selection which takes into account the heterogeneity of service descriptions has been proposed in [18]. The approach is based on an application ontology, which is merged by different service ontologies and constructed as a semantic net with multiple concept relations. Yannis et al., [19] focuses the survey on current state of the art in ontology mapping. Ravi et al.,[15] does the analysis on ontology mediation tools.

As the survey indicates the potential of the ontology mapping to various domains, this work proposes 16 different various patterns that can be used to find the similarity between ontologies through the components Concepts, Relations, Attributes and Values.

3. PROPOSED MAPPING PATTERNS AND ARCHITECTURE

Ontology Mapping is found to be the basic operation for all ontology management operations. The general definition of ontology mapping is as given below:

Ontology Mapping Definition: Given two ontologies O1 and O2, mapping one ontology onto another means that for each entity (Concept C, Relation R, or Instance I) in ontology O1, try to find a corresponding entity, which has the same intended meaning in ontology O2 [14].

During the process of ontology mapping the mapping pattern plays an important role. Mapping patterns reflect the internal structure of ontologies as well as mappings between elements of ontologies. Different mapping patterns can be formulated for ontology mapping as: concept-concept (P1), attribute-relation and relation-attribute (P2), concept-relation and relation-concept (P3), relation-relation (P4), attribute-attribute (P5), concept-attribute and attribute-concept (P6), concept-value and value-concept (P7), relation-value and value-relation (P8), attribute-value and value-attribute (P9). Based on the level of requirement of the task to be completed, the patterns may be used (refer Annex I).

This work limits its perspective, only to the concept-to-concept mapping pattern for ontology mapping. Fig.1 shows the pictorial representation of mapping two concepts from two ontologies.

3.1 Concept-to-concept Mapping Pattern

Concept A in ontology1 is at level 0 in Fig.1, whereas the same concept is at level 1 in ontology2. The grey arrow shows the mapping between concepts from both ontologies.

3.2 Architecture of Proposed Framework

SOMRep is a Standard Ontology Mapping Representation Framework. Its architectural design is shown in Fig.2. It is assumed that ontologies are available in library in XML format. Similarly data structures to be followed for different mapping patterns are stored in repository. Mapping of two ontologies is done through mapping process and an XML output file MMI is produced. With this MMI user's query can be answered. Similarity measures that are used to map two ontologies are discussed below.

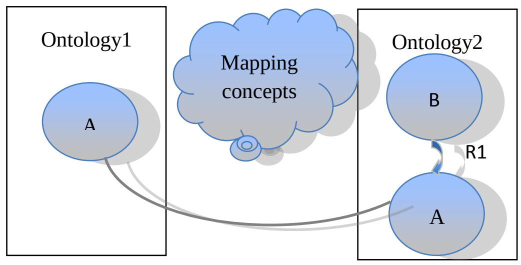


Fig. 1 Concept-to-Concept Mapping Pattern

3.3 Mapping Process

Many similarity measures have been proposed for concept similarities, including the string-based similarity, graph-based similarity, instance classification similarity and knowledge resource similarity [16]. The string-based similarity is widely used for ontology mapping. The graph-based similarity utilizes the similarity of the structures of ontologies. Instance classification similarity says that, if the classification of instances is similar to the concepts in different ontologies, the concepts are similar. The knowledge-based similarity utilizes other knowledge resources, such as dictionary and Word-Net to calculate the similarity. Although there are many similarity measures, we discuss three similarity measures for use in our framework. The similarities measures are Syntactic, Semantic and Structural similarity. We discuss them in order.

3.3.1. Syntactic Similarity

Syntactic similarity is string-based similarity used for concept pairs. We utilize the following similarities: prefix, suffix [7], n-grams and edit distance.

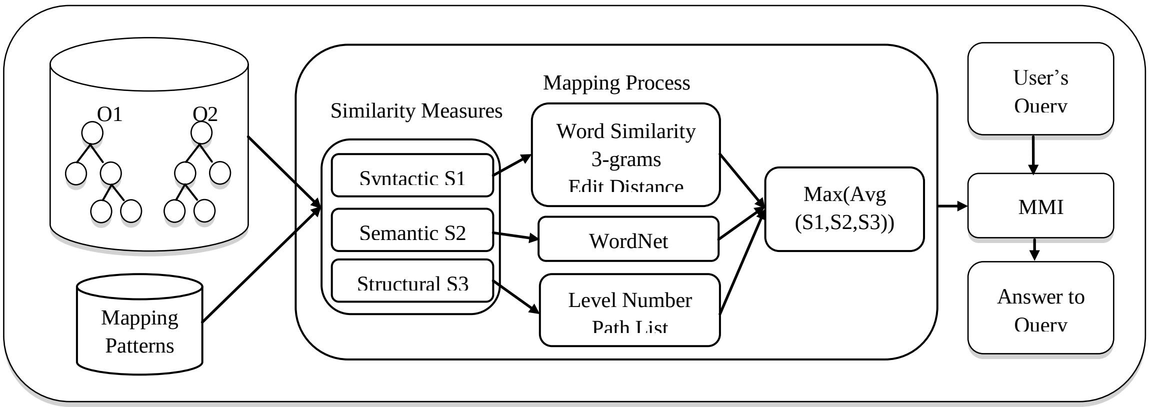


Fig 2: Architecture of Proposed Framework

The prefix similarity measure is for the similarity of concept prefixes such as Eng. and England. The suffix similarity measure is for the similarity of word suffixes such as phone and telephone. This similarity is measured using (1)

$$sim_{presuf} v1 = \frac{\max(\max(prefix(e1,e2)), \max(suffix(e1,e2)))}{\mathrm{Max}(|e1|, |e2|)}$$

where e1 and e2 are the entities from ontologies 1 and 2 respectively. Edit distance can calculate the similarity as a count of the string substitutions, deletions and additions. The formula used is (2)

$$sim_{edit} v2 = \frac{1}{1 + \mathrm{edit}(e1, e1)}$$

For n-gram, the word is divided into n number of strings, and the similarity is calculated by the number of same string sets. In our system, we utilize 3-gram for calculating the similarity (3)

$$sim_{gram} v3 = \frac{1}{1 + (|g(e1)| + |g(e2)| - 2 * (g(e1) \cap g(e2)))}$$

Here, g (e1) and g (e2) are number of 3-grams obtained for concepts e1 and e2 from ontology 1 and 2 respectively. g (e1) ∩g (e2) is number of common terms between the concept pair e1 and e2.

3.3.2. Semantic Similarity

The knowledge resource, which is semantic based, is calculated for words. We use Word Net [4] as the knowledge resource for calculating the semantic similarity. We utilize only synset from Word Net.

$$sim_{word} v4 = \begin{Bmatrix} 1.00 \ if \ (e1 = e1) \\ 0.75 \ if \ (e1 \in synset(e2)) \\ 0.00 \ if \ (e1 \notin sysnset(e2)) \end{Bmatrix} \quad (4)$$

The similarity measure using synset is set slightly lower than the measure for actual string equality matches, in order to privilege exact matching between terms.

3.3.3. Structural Similarity

Since the ontologies are organized as concept hierarchies, we defined the similarity using the structure of ontologies.

Shortest path is calculated from the root to concept pair and level number is obtained. With this level number, level similarity between the concept pairs is calculated as

$$sim_{level} v5 = \frac{2}{\mathrm{level}(e1) + \mathrm{level}(e2)}$$

Finally, all the similarity measures are applied to summation to obtain the most correspondence for each concept in ontology 1.

$$sim_{map}(e1, e2) = \max\left(\mathrm{avg} \sum_{i=1}^{5} v_i\right) \quad (6)$$

After obtaining the concept pair with maximum similarity measure, the path is extracted from root to the current concept in both the ontologies. One level parent and children information is also taken from ontologies. Along with this information all the similarity measures for that particular concept pair is written into output XML file which is mapped Meta information (MMI) file as shown in Table 1.

MMI is used for answering the query posted by the users. It is evident that ontology mapping is the basic operation for all ontology management operations like ontology alignment, ontology integration/merging. Different stages of integration process are: Syntactic Integration based on syntactic mapping, Structural Integration based on organization of ontological concepts and Semantic Integration based on the concept meaning considering that a meaning depends on the context is applied. For each stage of integration process, ontologies have to be mapped. Although mapping patterns are applicable for all stage of integration process, some mapping patterns are found to be more applicable or appropriate in some stage of integration.

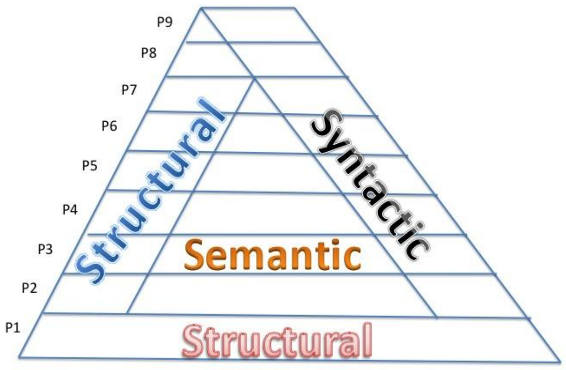
The usage level of mapping patterns for semantic, syntactic and structural integration process is clearly mentioned in the Fig. 3. Thus, mapping patterns can be used more effectively for all ontology management operations.

Table 1. Mapped Meta Structure in XML Format(Output)

```
<?xml version="1.0" encoding="UTF-8"?>
<Result>
<RecordInfo>
<Entity1>Arts and Humanities</Entity1>
<Entity2>Industrial Goods and Services</Entity2>
<PreSuf>0.0</PreSuf>
<N-Grams>0.02702702702702703</N-Grams>
<Edit-Distance>21</Edit-Distance>
<E1-Level>1</E1-Level>
<E2-Level>2</E2-Level>
<E1-Path>root- Arts and Humanities</E1-Path>
<E2-Path>root- Business- Industrial Goods and Services</E2-Path>
<Max-Value>5.256756756756757</Max-Value>
</RecordInfo>
…..
</Result>
</xml>
```

4. ANALYSIS

A prototype of SOMRep has been implemented as a Java application. We used Word Net, which provides synset to calculate the semantic similarity. Using the prototype, experiment was carried out using the 2008 bench mark test



Three stages of Ontology Integration

Fig. 3: Mapping patterns used at 3 levels of integration

given by the Ontology Alignment Evaluation Initiative (OAEI) [6], in order to determine the accuracy of the Mapping process algorithm. We have chosen directory dataset to evaluate Mapping process algorithm. The directory data consists of real world Web site directories Google's and Yahoo's. Each directory's ontology is organized as taxonomy, with concept names in a hierarchical structure.

5. APPLICATION RESULTS

We have done two sets of application analysis using the proposed mapping patterns. The first application is to derive equations for next level mapping. The second application is to discover the most appropriate web service for same domain.

5.1. Scenario 1

The reported mapping patterns are more applicable to binary mapping, since they consider the binary relation property while doing the mapping process. For example, consider one-to-one entity for mapping. If these basic mapping patterns are to be utilized for the next level consideration of ternary and quandary, the need for application of these mapping patterns to be explored. Therefore, it is necessary to devise rules for high level mapping patterns. Ternary and quandary mapping patterns are combination of three and four proposed patterns. For example, the one of the ternary mapping pattern is shown below:

$$Rule\ 1 = P_{17}: \{P_1, P_7, P_{16}\}$$

Where $P_{17} \rightarrow$ new ternary mapping pattern
$P_1 \rightarrow$ Concept-to-concept mapping pattern
$P_7 \rightarrow$ Attribute-to-attribute mapping pattern
$P_{16} \rightarrow$ Value-to-value mapping pattern

Example for quandary mapping pattern is shown as:

$$Rule\ 2 = P_{19}: \{P_1, P_7, P_{16}, P_6\}$$

Where $P_{18} \rightarrow$ new ternary mapping pattern
$P_1 \rightarrow$ Concept-to-concept mapping pattern
$P_7 \rightarrow$ Attribute-to-attribute mapping pattern
$P_{16} \rightarrow$ Value-to-value mapping pattern
$P_6 \rightarrow$ Relation-to-relation mapping pattern

Using these mapping patterns ontology mapping can be done more effectively.

5.2. Scenario 2

The mapping can be classified as Heavy weight and Light weight mapping. Heavy weight mapping between concepts of two ontologies shows that there exists more similar semantic correspondence between those two concepts. Light weight mapping indicates that least similar semantic correspondence between concepts of two ontologies.

During the mapping of two ontologies, if parent and more than 50% of the children are same then it is concluded as Heavy weight mapping. This determination is based on the percentage of children for each parent as shown in Table 2.

Light weight mapping is controversy of Heavy weight mapping. If the number of children to any parent is less than 50% then it is decided to be Light weight mapping. Table 2 shows this clearly. Based on Table 2, mapping is done for Concept-to-Concept mapping pattern and classification of mapping is obtained. Heavy and Light weight classification of www.Amazon.com and www.Ebay.com has been done and similarity matrix for the concepts of level1 is obtained as shown in Fig. 4a. Since similarity matrix for this scenario is sparse, it is reduced into matrix as in Fig. 4b.

Table 2: Values for Heavy/Light weight mapping based on %

S. No	Parent Level	Children Level	Value	Weight
1	F	F	1.00	H
2	F / P	90% - 80%	0.9	H
3	F / P	80% - 70%	0.8	H
4	F / P	70% - 60%	0.7	H
5	F / P	60% - 50%	0.6	H
6	F / P	50% - 40%	0.5	L
7	F / P	40% - 30%	0.4	L
8	F / P	30% - 20%	0.3	L
9	F / P	20% - 10%	0.2	L
10	None	None	0.00	L

F – Full, P – Partial, H – Heavy, L – Light

This scenario dealt with one level of mapping between web services. It can be extended in depth to obtain more accurate mapping results.

Ebay / Amazon	A2 Computers & Networking	B2 Electronics	C2 Travel
A1 (Computers & Office)	0.8	0.00	0.1
B1 (Electronics)	0.00	0.8	0.00
C1 (Travel)	0.00	0.2	0.4

Fig. 4a Similarity Matrix

Ebay / Amazon		Similarity Value
A1	A2	0.8
A1	C2	0.1
B1	B2	0.8
C1	B2	0.2
C1	C2	0.4

Fig. 4b Reduced Similarity Matrix

6. CONCLUSION

Ontology mapping is found to be the basic process for all type of operations as far as ontology is concerned. In this paper, we have proposed a standard ontology mapping representation framework, SOMRep to semi-automatically perform mapping between two ontologies. We have designed different possible mapping patterns for various ontology management operations. The patterns identified may be applied on quite different ontologies depending on the requirement or need, of the application on hand. We have considered syntactic, semantic and structural similarities in ontologies for mapping. Application results show how concept-to-concept mapping pattern is applicable for devising ternary, quandary rules and for accurate web service discovery.

The work can be extended for other mapping patterns in depth to obtain more accurate mapping results. In future, this framework can be evaluated with different domain and the mapping results may be obtained.

7. REFERENCES

[1] Berners-Lee .T,Hendler .J and Lassila, O., "The Semantic Web", Scientific American, Vol. 284, No. 5, pp. 34-43, 2001.

[2] Doan, A.; Madhavan, J.; Domingos, P.; Halevy, "A Learning to Map between Ontologies on the Semantic Web", Proceedings of the Eleventh International WWW Conference, 2002.

[3] Dnyanesh Rajpathak and Enrico Motta, "An Ontological Formalisation of the Planning Task**",** Proceedings of the Formal Ontologies in Information Systems, 2004.

[4] C. Fellbaum, "Word Net: An Electronic Lexical Database", MIT Press, 1998.

[5] Hendrik Thomas, Declan O' Sullivan, Rob Brennan, "Evaluation of Ontology Mapping Representations", Knowledge and Data Engineering Group, 2009.

[6] Juanzi Li, Jie Tang, Yi Li and Qiong Luo, "RiMOM: A Dynamic Multistrategy Ontology Alignment Framework", IEEE Transactions on Knowledge and Data Engineering, vol: 21, p: 1218-1232, 2009.

[7] Judith Wilde, "Building Vocabulary: Prefixes, Roots, and Suffixes", Printed: NCELA, 2006.

[8] Klein, M, "Combining and relating ontologies: An Analysis of Problems and Solutions", Workshop on Ontologies and Information Sharing, 2001.

[9] Ming Mao,"Ontology Mapping: An Information Retrieval and Interactive Activation Network Based Approach", Springer LNCS 4825, pp. 931–935, 2007.

[10] Namyoun Choi, Il-Yeol Song, and Hyoil Han, "A Survey on Ontology Mapping", SIGMOD, Vol. 35, No. 3, 2006.

[11] Nelson C.N. Chu, Quang M. Trinh, Kwn E. Barker and Reda S. Alhajj, "A Dynamic Ontology Mapping Architecture for a Grid Database System", IEEE Fourth International Conference on Semantics, Knowledge and Grid, p : 343 – 346, 2008.

[12] Ondrej Svab "Exploiting patterns in Ontology Mapping",6th International and 2nd Asian Semantic Web Conference, 2008.

[13] Ondrej Svab, Vojtech Svatek and Heiner Stuckenschmidt2, "A Study in Empirical and 'Casuistic' Analysis of Ontology Mapping Results", Springer LNCS, Vol: 4519, p: 655-669, 2007.

[14] PengWang, Baowen Xu, "Debugging Ontology Mappings: A Static Approach", Computing and Informatics, Vol. 22, p: 1001–1015, 2007.

[15] Ravi Loudrusamy and Gopinath Ganapathy, "Feature Analysis of Ontology Mediation Tools", Journal of Computer Science, 4(6), p: 437-446, 2008.

[16] Ryutaro Ichise, "Machine Learning Approach for Ontology Mapping using Multiple Concept Similarity Measures", Proceedings of the 7th IEEE International Conference on Computer and Information Science, p: 340-346, 2008.

[17] Sofia, H. Martins, J. P.," A Methodology for Ontology Integration", Proceedings of the International Conference on Knowledge Capture, ACM SIGART, 2001.

[18] Xia Wang, Manfred Hauswirth, Tomas Vitvar, and Maciej Zaremba, "Semantic Web Services Selection Improved by Application Ontology with Multiple Concept Relations", ACM, p: 2237-2242, 2008.

[19] Yannis, K. and S. Marco, "Ontology mapping: The state of the art", The Knowledge Engineering Review, 18: p: 1-31, 2003.

[20] Yves R. Jean-Marya, E. Patrick Shironoshitaa, Mansur R. Kabuk "Ontology Matching with Semantic Verification" Web Semantics: Science, Services and Agents on the WWW, vol: 7, p: 235-251, 2009.

[21] Zhaohui Wu, Yuxin Mao, and Huajun Chen, "Subontology-Based Resource Management for Web-Based e-Learning ", IEEE Transactions on Knowledge and Data Engineering, vol. 21, p: 867-879, 2009.