워드넷 온톨로지를 이용한 토픽맵 매핑

유우종 김진우 권주흠
고려대학교 컴퓨터학과 소프트웨어시스템 연구실
{milkybell, pkm311, jkwon}@software.korea.ac.kr

Mapping Topic Maps using WordNet ontology

Woojong Yoo, Jinwoo Kim, Juhum Kwon
Software System Lab., Dept. of Computer Science and Engineering, Korea University

Abstract

In this paper, we provide a mechanism to map topics between similar Topic Map ontologies, which enlarge the capacities of current Topic Maps. That is, there is no explicit way to map topics across multiple Topic Maps in current technology. Therefore, our approach provide semi-automatic or automatic mapping of topics between Topic Maps. In order to map the topics, we provide precise definitions of mapping properties and mapping rules.

1. Introduction

Topic Map has emerged as a mean to provide machine understandable semantics [1, 2, 3, 4]. Thus, the idea of Topic Map can be enriched with precisely defined metadata scheme (i.e., Topic Map ontology) in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications [5]. Currently, related to the Topic Map, several important researches are being under considered to enhance the capability of Topic Map such as formalizing knowledge, reasoning mechanism, knowledge representation, etc [1]. However, the current Topic Map requires more than that. It amounts to agree on mapping of topics for similar Topic Maps constructed by different domain experts. If not, tedious manual work is required to integrate Topic Maps when users navigate Topic Maps. In this paper, we provide a semi-automatic or possibly automatic mechanism to map the topics. As shown in figure 1, the main idea of mapping topics across multiple Topic Maps is as follows:

- First, we precisely define the mapping properties and mapping rules to map the different Topic Map ontologies.
- Second, find the places in the ontologies where they overlap and relate concepts that are semantically close via equivalence and subsumption relations
- Third, when topics are going to be mapped, we use the WordNet hierarchy, which is a linguistic ontology [6, 7, 8], as a reference model to map topics. Because, the WordNet knowledge taxonomy provides the general framework of real world knowledge representation. In this paper, we assumed that the topics in the Topic Maps were constituted with the natural language as in the WordNet concept.

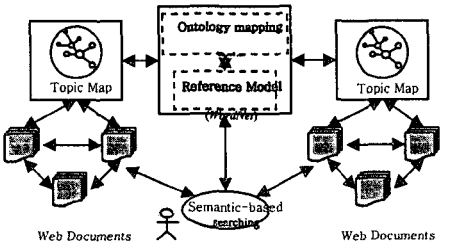

Fig. 1. A Framework of mapping topic map

The novel contribution of this paper is that we provide a precise mapping properties and rules for Topic Map ontologies.

The organization of this paper is as follows: the next section discusses related work. In section 3, we provide mapping technique. Finally, in section 4, we give a conclusion and address future work.

2. Related Work

In this section, we briefly discuss the concept of WordNet and Topic Map in subsection 2.1. In subsection 2.2, we explain the basic concept on the ontology integration, which is directly related to our research.

2.1 WordNet and Topic Maps

WordNet is an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets. Each concept, topic in this paper, is interrelated with semantic relation thereby a knowledge taxonomy can be generated. Also, the synonym set, called Synset, is the collection of synonymous concept set. Therefore, we can traverse the taxonomy via the semantic relation such as IS-A, Part-of, etc. Also, we can find the synonymous word from corresponding synset.

Topic map, as defined in ISO/IEC 13250, is used to organize information in a way that it can be optimized for navigation. Therefore, the main goal of creating Topic map is designed to link large quantities of unorganized information. That is because information is not useful if it cannot be found or linked. For example, in the paper publishing world, there are several mechanisms to organize and index the information contained within a book or document. Indexes allow readers to go directly to the portion of the document that is relevant to their information needs. Similarly, Topic map can be thought of as the online equivalent of printed indexes. Topic Map allows users to create a large quantity of metadata and tightly interconnected data, where they constitute a kind of semantic network above the data themselves. Therefore, a Topic map is considered as an aggregate of topic characteristics, including Topic, Association, and Occurrence [3, 4].

2.2 Ontology integration

The need to integrate number of ontologies is that ontologies are existed in the same area but cover different aspects, contain overlapping information [9, 10]. Ontology integeration can be divided into several sub-part such as mapping, aligning, merging, and integrating. Mapping relates similar concepts or relations from different sources to each other by an equivalence relation. Aligning brings two or more ontologies into mutual agreement, making them consistent and coherent. Merging merges different ontologies about the same subject into a single one that unifies all of them. And, integrating builds a new ontology reusing other available ontologies (i.e., assemble, extend, specialize). Therefore, ontology integration is a complicated process mostly done by hand. That is because it is difficult to find the terms that need to be aligned and the consequences of a specific mapping possibly unforeseen implications are difficult to see.

3. Mapping Technique

This section discusses the mapping properties and mapping rules to map the Topic Map ontologies.

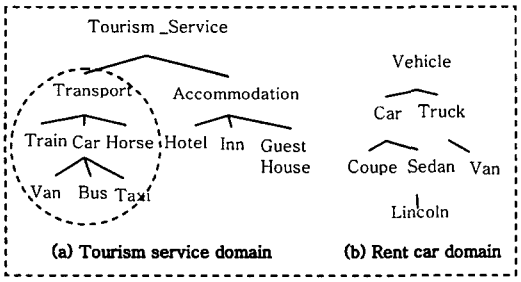

Fig. 2. Knowledge Taxonomy

As shown in figure 2, two Topic Map knowledge taxonomies are used in order to demonstrate the proposed mapping technique. The figure 2 (a) is a Topic Map ontology derived from tourism domain and figure 2 (b) is derived from rent car domain.

3.1 Mapping Properties

As discussed in the section 2, the WordNet uses sets of synonyms, called *synsets*, to represent word senses. For example, the semantically synonymous word of *car* is *auto, automobile, machine, and motorcar* in the *synsets*.
Then, let λ be a mapping function of topic T in the Topic Map ontology to corresponding synonym set, *synsets*, in the WordNet, then λ is defined as $\lambda: T \rightarrow Synset$. From the mapping function, we can define the mapping properties of topics as follows:

Definition 3.1 (Equal) Two topics t_i and t_j, $t \in T$,under ontology p and q, ${}^p t_i$, ${}^q t_j$, respectably are equal if and only if the topics satisfy the following condition:
$(\{\lambda({}^p t_i)\} \wedge \{\lambda({}^q t_j)\} \neq \varnothing) \Rightarrow ({}^p t_i = {}^q t_j)$

Let $t_1 \leq t_2$ be represented as the implication form of subsumption hierarchy such that t_1 is the subtopic of t_2, then we can define the following as the specialization of topic :

Definition 3.2 (specialized topic) Two topics t_m and t_n under ontology p and q, ${}^p t_m$, ${}^q t_n$, respectably are specialized topics of ${}^p t_i$ and ${}^q t_j$ if and only if the topics are in the form of ${}^p t_m \leq {}^p t_i$, ${}^q t_n \leq {}^q t_j$ and satisfy the following cases:
1. $({}^p t_i = {}^q t_j) \Rightarrow \forall({}^p t_m, {}^q t_n) \leq ({}^p t_i)$ and $\forall({}^p t_m, {}^q t_n) \leq ({}^q t_j)$
2. $({}^p t_i \neq {}^q t_j) \Rightarrow \forall({}^p t_m) \leq ({}^p t_i)$ and $\forall({}^q t_n) \leq ({}^q t_j)$

The new topic indicates that a topic in the Topic Map ontology does not have corresponding topic in the WordNet. Therefore, the following definition justifies the notion of new topic:

Definition 3.3 (New topic) Given a topic t_i under ontology p or q respectably, a topic t_i is a new topic t_n if and only if the topic satisfy the following condition:
$(\{\lambda(t_i)\} = \varnothing) \Rightarrow (t_i = t_n)$

The irregular topic is the case that the mapping from a topic in the Topic Map ontology to *Synsets* in WordNet satisfies $\{\lambda(T)\} \neq \varnothing$. However, the topic does not follow the WordNet taxonomy.

Definition 3.4 (Irregular topic) Let t_i be a topic under ontology p or q and $t_j \in Synsets$ be a topic in WordNet. Then, t_i is a irregular topic t_m if and only if the topic satisfy the following condition:
$(\{\lambda(t_i)\} \neq \varnothing)$ and $(t_i \nleq t_j)) \Rightarrow (t_i = t_m)$

3.2 Mapping Rules

The method for mapping of ontology varies according to the purpose of a system. In the proposed approach, the mapping is based on the mapping properties defined in sub section 3.1, where we presented four different constructs of mapping rules used to map the separate Topic Map ontologies as summarized in table 1 below.

Table 1: Mapping Rules.

Operation	Symbol	Construct of Mapping Rule
Equal	$\Leftrightarrow_{Equal}$	$(T1@A \Leftrightarrow_{Equal} T1@B) = T1@\{A,B\}$
Specialized Concept	$\Rightarrow_{Specialized}$	$T1@A \Rightarrow_{Specialized} T2@B$
New Concept	$\Rightarrow_{New}$	$T1@A \Rightarrow_{New} T2@A$
Irregular	$\Phi_{Irregular}$	$T1@A\ \Phi_{Irregular}\ T2@A$

The equal operation takes the form of $(T_1@A \Leftrightarrow_{Equal} T_1@B) = T_1@\{A,B\}$ where $T_1@A$ and $T_1@B$ indicate the source topic T_1 in the domain A and B respectably from the Topic Map ontologies. That means, the two topics of T_1 are mapped into one topic if and only if both topics are semantically equal. For example, the topics *car* in figure 3 can be represented as $(Car@a \Leftrightarrow_{Equal} Car@b)=Car@\{a, b\}$. The specialized topic takes the form $T_1@A \Rightarrow_{Specialized} T_2@B$ meaning that the topic T_1 is the specialized topic of the topic T_2. Further, the T_2 should be the shortest upper topic in the mapped Topic Map ontology. Identifying the shortest upper topic can be obtained by referencing WordNet hierarchy thereby the topic T_1 can be attached to T_2. For instance, the rule $Vehicle@b \Rightarrow_{Specialized} Transport@b$ indicates that the topic vehicle is specialized topic of the topic transport although there is no explicit specification between them. Again, this implication is derived from WordNet taxonomy. The form of new topic $T_1@A \Rightarrow_{New} T_2@A$ implies that the topic T_1 is the specialized topic of the topic T_2. However, the topic T_1 does not exist in the WordNet vocabulary. The mapping rule $Lincoln@b \Rightarrow_{New} Sedan@b$ can be the example for the new topic. Therefore, in this case the topic T_1 moves together with the topic T_2 while being attached to T_2. Finally, the irregular topic is the case that two topics are equal but one or both of the topics does not follow WordNet taxonomy. For instance, from the mapping rule $Horse@a\ \Phi_{Irregular}\ Transport@a$, we know that both topics exist in WordNet but they do not follow the hierarchy. The figure 3 shows the results of mapping Topic Map ontologies where the topics are mapped according to mapping properties and mapping rules. In figure 3, the thick solid lines between topics *transport* and *vehicle*, and between topic *sedan* and *Lincoln* indicate the new link generated from the mapping rule of specialized topic and new topic respectably.

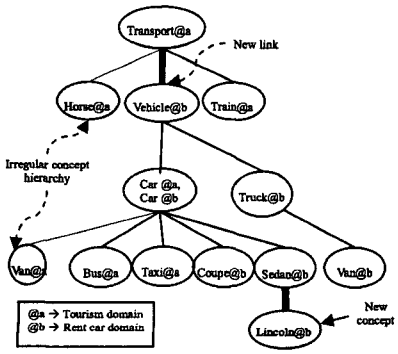

Fig. 3. A mapped Topic Map ontology constructed from the mapping rules.

4. Conclusion

Topic Map mapping technique can be used in many different areas such as ontology integration, Semantic web searching, etc. In this paper, we proposed a technique to map topics across multiple Topic Map ontologies. The outstanding feature of the proposed approach was that our method resolved interoperability issue across multiple Topic Map ontologies. Because, the current Topic Map has no mechanism to interrelate topics across multiple domains of topics. The strongpoint of this paper is that we provided precise mapping properties and mapping rules. We are planning to implement prototype system for the future work.

References

[1] Buchheit, M., Donini, F. M., and Schaerf, A. (1993). Decidable reasoning in terminological knowledge representation system, Journal of Artificial Intelligence Research, 1, 109-138.

[2] Heflin, J. and Hendler, J. (2000). Semantic Interoperability on the Web, Proc. Extreme Markup Languages, Graphic Communications Assoc. (pp. 111–120). Montreal.

[3] International Organization for Standardization, ISO/IEC 13250, *Information Technology-SGML Applications-Topic Maps*, Geneva: ISO, 1998.

[4] S. Pepper, "The TAO of Topic Maps", XML 2000 Conference & Exposition

[5] Lassila, O. (1998). Web Metadata: A Matter of Semantics, IEEE Internet Computing, 2, 4, 30–37.

[6] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1993). Five papers on WordNet, CSL Report 43, Princeton University.

[7] Miller, G. A. (1995). WordNet: A lexical database for English, Communications of the ACM, 38, 11.

[8] Miller, G. A. (1998). Nouns in WordNet, WordNet: An Electronic Lexical Database (pp. 23-46), C. Fellbaum, ed., Cambridge, Mass.: The MIT press.

[9] Sofia Pinto H. and Martins, J. P. (2001). A methodology for ontology integration, ACM, Int. Conf. On K-CAP, 22-23.

[10] Sofia Pinto, H. (1999). Some Issues on Ontology Integration, Proc. of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden.