An Approach of Concept Similarity Computation in Taxonomy Ontology

A. Selvakumar
P.G. Student, Department of Computer Science & Engg, Sri Manakula Vinayagar Engineering College, Puducherry, India.
igaselva@yahoo.com

R. Sudha
P.G. Student, Department of Computer Science & Engg, Sri Manakula Vinayagar Engineering College, Puducherry, India.
jsudha16@yahoo.com

Abstract—This paper describe a method to calculate the concept similarity in taxonomy ontology and help searching system of E-commerce to answer users query more effectively. Instead of using the current ontology mapping methods which are focus on analyzing the terms relationship based on some thesauruses and special construction definition of terms, in this paper we firstly show the creation of taxonomy ontology, and then based on it, we describe a method to calculate the similarity between terms in the taxonomy ontology and terms in source ontology, Following with this method, the similarity can be calculated and used in answering users' queries. Finally, this paper supplies an instance to validate this method.

I. INTRODUCTION

The purpose to synthesize knowledge from different bottom information sources impulses the research on searching method. Most information of these sources is unified by bottom ontology based on taxonomy. These taxonomy ontologies are composed of terms which are structured by subsumption relationships. What is more, each term will be mapped to one or more objects which are stored in databases. Many on-line business websites such as ebay, Taobao, Amazon all have this kind of structure. Obviously, they put the things belong to the same taxonomy into one category, such as automotive, books and so on. What is more, one category can be separated into some subdirectory when it is necessary. For example, "books" can be composed of "Computer books", "Fiction" and so on. When consumers are choosing products, how to make them feel comfortable about this process is important to any E-commerce companies. This paper describes a method which can make the computation of concept similarity easier and help the users of E-commerce find the products more effectively. Following with this method, users only need to choose a term they want in taxonomy ontology, then all the terms in source ontologies that are similar to it can be found out totally. In last part of this paper, there is an instance to validate this method. The remaining of the paper is organized as follows:. Section 2 discusses the constitution of taxonomy ontology. Section 3 gives out the process of calculating the concept similarity and Section 4 conclude the whole paper.

II THE CONSTITUTION OF TAXONOMY ONTOLOGY

A. *The domain taxonomy*

The domain taxonomy can also be called classifier which is used to annotate the terms in both taxonomy ontology and information source ontology. It is composed of all attributes which belong to the entity of the domain. Moreover, each attribute can contain many different values. For example, an automobile may have three (sometimes it is more than three) attributes, they are brand, use and the number of box. And as to each attribute, it can include a variety of values. For instance, in automobile markets, we consider Benz, Volkswagen and Honda as the values of brand. This taxonomy method above is a widest taxonomy which does not always conform to the real condition, that is to say, maybe Benz does not produce a car with two boxes, but you still need to add the two boxes to attribute "box", for Honda makes two boxes cars. That is why we say it is a widest taxonomy. We consider each attribute as an element, and when the entity of this domain has N ($0<N<+\infty$) attributes, these attributes can be grouped as (A1 , A2 , …AN) , what is more, the different values of each attribute should be signed by an Arabian number started from 1, such as 1,2,3…and so on. Fig. 2 shows the automobile domain taxonomy with three attributes. In this Figure, all the values of attribute are signed by Arabian numbers.

B. *The taxonomy ontology*

Drawing the values from every attribute and aggregating them together can describe a meaningful entity in real world, and it can be called a term. These terms construct a tree on which the child node instantiates one more attributes than its father node. Definitely speaking, this tree is taxonomy ontology. This paper builds a group for each term in taxonomy ontology based on the attributes of domain taxonomy and use the Arabian number of values in domain taxonomy to annotate the term in taxonomy ontology. In addition, the order of term's attributes should conform to ones in domain taxonomy

978-1-4244-4711-4/09/$25.00 ©2009 IEEE

and the unknown attribute in terms is annotated by zero. That means group Benz|SportsCar) can be annotated by (1,2,0). For instance, the whole automobile taxonomy

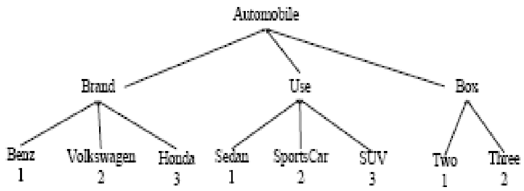

Figure 2. The automobile domain taxonomy

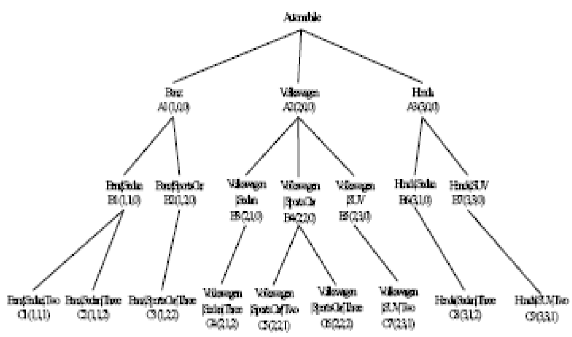

Figure 3. The automobile taxonomy ontology

C. The source ontology annotation

Similar to taxonomy ontology, a group T1(X1, X2… X|N|) can be defined to describe each term in source ontology, in which Xn(n=1,2…|N|) is a value of attribute, and |N| equals to the number of attributes in domain taxonomy. What is more, even (n=1,2…|N|) can also be defined as a group when it is necessary, such as X1=(X11, X12…X1n). In conclusion, a term T1 can be written as T1{(X11, X12…X 1n1), (X21, X22…X2n2)…}. As to the unknown attribute, use zero to denote it. An example of such a defined information source is showed in Fig. 4.

III THE PROCESS OF CALCULATING CONCEPT SIMILARITY

A. Create the small ontology tree

In order to reduce the complexity of the computation, we need to create the small ontology trees from the original ontology trees instead of using the original ones directly. In the above discussion, we know that a term either in taxonomy ontology or in source ontology can be described by a group such as T1(X1, X2… X|N|). Based on it, we can create the small ontology trees for original ontologies by the following steps:

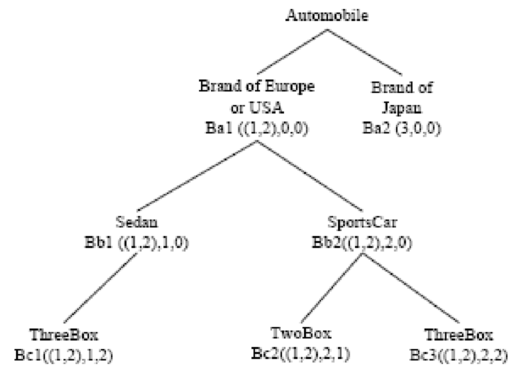

Figure 4. The annotation of source ontology

Step 1: Get the term we want to query in the taxonomy ontology(usually it will be selected by users). And then register its group. Take the taxonomy ontology in . Users may want to query "Volkswagen|Sedan". And we register the group (2,1,0). Step 2: Find the first element in the group, here in this example, it is "2". Next, use breadth-first search to each source ontologies. When we find out a term whose first element is or include "2", this term can be a root of a small ontology tree. Take the three source ontologies in Fig. 5 as an example. Following the above method, we can create four different small ontology trees which are shown in Fig. 6. What is more, tree "A" in Fig. 6 comes from ontology "1" in Fig. 5; Tree "B" in Fig. 6 comes from ontology "2" in Figure5; Trees "C" and "D" comes from ontology "3" in Fig. 5. Obviously, there are two kinds of small ontology tree. First one's root only has one element which is instantiated, and it will always be the first element because the instantiation of the attributes is ordered. Second one's root not only has the first element instantiated, but also includes other instantiated elements, Such as Tree "D" in Fig. 6. Therefore, both two kinds of small ontology trees are justified.

B. Calculate the similarity of concept

We use S(T1, T2) to describe the similarity between concept T1 and T2(and in ontology, concept is equal to term, so in the following discussion we use term instead of concept). Thus, we will have the below formula to define the concept similarity.

Def 4.2.1 Concept Similarity

$$S(T1,T2)=\frac{\sum_{i=1}^{n}[\frac{1}{m}\delta_i(T1,T2)]}{n}$$

In the above formula, i represents the serial number of attributes and n represents the whole number of attributes which are instantiated in the term queried. m means the possible values a attribute can have, when an attribute only have one value, m is equal to one. Usually, as for terms in source ontologies, an attribute can have more than one value,

Figure 5. Three information source ontologies

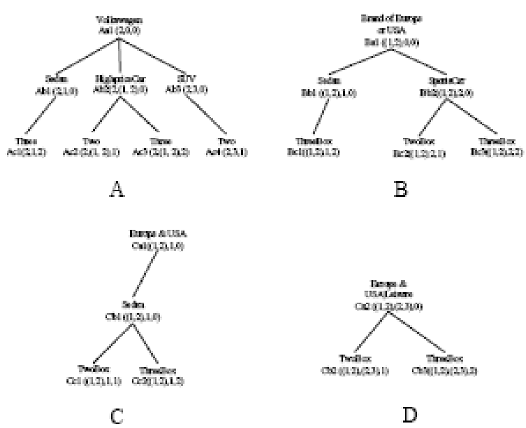

Figure 6. Four small ontology trees

so m is more than one. Moreover, we should define in the following way:

Def 4.2.2 The definition of .

$$\delta_i(T1,T2)=\begin{cases}1 & \text{when } T1 \text{ and } T2\text{'s first } i \\ & \text{attributes are similar} \\ 0 & \text{when } T1 \text{ and } T2\text{'s first } i \\ & \text{attributes are not similar}\end{cases}$$

Now, using the (1) and (2) together, we can calculate the concept similarity of two terms.

C. An instance of calculating the similarity of concept

According to the taxonomy ontology in Fig. 3, if the user's query is (2, 1, 0), we should choose the terms which have two attributes instantiated in small ontology trees in Fig. 6. And then use (2, 1, 0) as T1 and each term chosen from small ontology trees as T2, and put T1 and T2 into (1) and (2). Then we have the computation result as shown in Table1. In Table 1 we can see that the point "Ab1" has one hundred percent similarity with T1, so it is the best term we want to find. And then "Ab2", "Bb1", "Cb1" have seventyfive percent similarity with T1, "Ab3" has fifty percent similarity with T1 and "Bb2", "Ca2" have twenty-five percent similarity with T1. The more vivid picture about the similarity among these terms has been shown in Fig. 7.

TABLE I. THE RESULT OF SIMILARITY COMPUTATION

TABLE I. THE RESULT OF SIMILARITY COMPUTATION

T1	T2	Similarity(%)
(2, 1, 0)	Ab1(2, 1, 0)	100
	Ab2[2, (1, 2), 0]	75
	Ab3(2, 3, 0)	50
	Bb1[(1, 2), 1, 0]	75
	Bb2[(1, 2), 2, 0]	25
	Cb1[(1, 2), 1, 0]	75
	Ca2[(1, 2), (2, 3), 0]	25

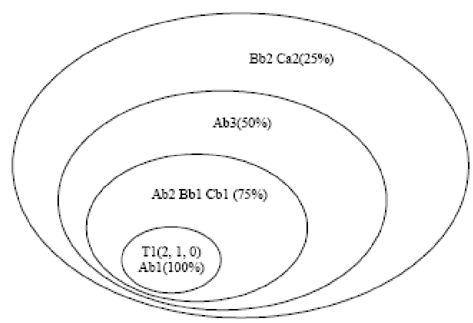

Figure 7. Equal-value illustration of similarity with T1 as the center

As being described in introduction, when the similar terms in source ontologies been found out, next task is to search the objects which belong to each term. If the system is set to choose objects whose similarity are more than sixty percent, then according to Fig. 7, "Ab1", "Ab2", "Bb1" and "Cb1" will be chosen. Fig. 8 is the same as picture 2 in Fig. 5, but each term of it has object. We can find "object 1" belongs to "Bb1", so "object 1" is chosen. The choice of other objects in picture 1 and 3 of Fig. 5 is based on the same method and we won't discuss here.

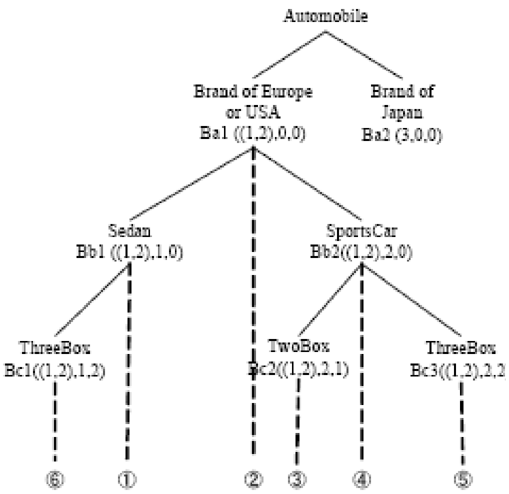

Figure 8. The source ontology with objects

IV. CONCLUSION

This paper describes a method to calculate the concept similarity in taxonomy ontology. It can help users to find the right products in on-line business websites. He or she only need to submit a query about the term in taxonomy ontology, the system will map this term to all the similar terms in source ontologies automatically. Based on this original model system, we should do further research and make it adaptable to real application.

REFERENCES

[1] Yannis Tzitzikas, Nicolas Spyratos, and Panos Constantopoulos, "Mediators over Ontology-based Information Sources," Proc. IEEE Symp. Second International Conference on Web Information Systems.

[2] Yannis Tzitzikas, Anastasia Analyti, Nicolas Spyratos, and Panos Constantopoulos, "An Algebra for Specifying Compound Terms for Faceted Taxonomies," Technical Report TR-314, Institute of Computer Science-FORTH,October 2002, pp. 2-6.

[3] Nicolas Spyratos, Yannis Tzitzikas, and Vassilis Christophides, "On Personalizing the Catalogs of Web Portals," 15th International FLAIRS Conference, FLAIRS'02, May 2002, pp. 1-10.

[4] Gao Kening, Ma Anxiang, and Zhang Bin, "Web integration based on classification ontology," Journal of Southeast University.

[5] Huang Guo, Zhou Zhu-rong, "Research on domain ontologybased concept semantic similarity computation," Computer Engineering and Design, Editorial Department of Computer Engineering and Design, 2007