An Ontology Mapping Method Based on Tree Structure

Sheng Li
School of Computer Science and Technology, Huazhong University of Science and Technology
kinglisheng@163.com

Heping Hu
School of Computer Science and Technology, Huazhong University of Science and Technology
hepinghu@andin.net

Xian Hu
School of Mathematics and Computer Science, Hubei University
xuanziyishi@163.com

Abstract

This paper provides an approach for ontology mapping based on tree structure grammar. Its excellence is that it combines both the similarity of the inner structure of concepts in different ontologies and the language similarity of concepts. In this way, more potential relationship between concepts could be found, and the validity of the calculation of similarity could be improved.

1. Introduction

In the methods of ontology mapping, the ones based on the structure of concepts are that considering about the hiberarchy of the concepts such as the relationship of concepts and semantic neighborhood when processing mapping. Since the former mapping method had rarely adequately consider about the influence that is evocable by the difference of the inner structure of ontologies, the veracity of the similarity computing will be reduced. For this reason and inspired by some idea about the tree grammar deduction in syntax pattern recognition, we provide a new ontology mapping method based on the tree structure. Its can combine both the similarity of the inner structure of concepts in different ontologies and the language similarity of concepts, so that the validity of the calculation of similarity will be improved.

2. Ontology Mapping Based on Index Tree

Since most ontology designed with hiberarchy, we can simply represent ontology with a tree hiberarchy structure. Now, we provide an ontology mapping method base on tree index structure. Its excellence is that it can combine both the similarity of the inner structure of concepts in different ontologies and the language similarity of concepts.

2.1 Creating Index Tree for Ontology

In order to indicate that the concepts in different floor have different effect to the similarity, we use index tree to number each concept (node) in the ontology. Here are the approaches for creating of index tree:

(1) To a domain ontology O, it has one and only node C_0 so-called main concept, we take it as the root of index tree. Except for C_0, all of other nodes have a and only a parent node;

(2) Suppose that there are totally n concepts in ontology O, the max number of branches in the tree is x. Therefore, we create a full X-tree (FXT), and there are x^k nodes in the k floor of a FXT. If a concept is not exist in some position, we define it as an empty node, and label it as Φ;

(3) If a node is number as C_k (k is a serial number like 1.1.2…), then, its children nodes will be numbered as $C_{k.1}$, $C_{k.2}$ …. Fig 1 shows the structure of index tree created from ontology:

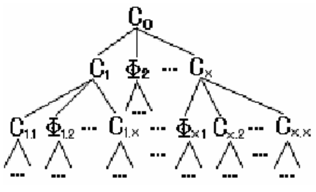

Fig 1: Index tree corresponding to the concepts of ontology

In fig 1, a node represents a concept. The symbol Φ represent that in this position there is no corresponding concept. Moreover, in order to use index tree expediently, we add an accessional process:

(4) If the max branches or the max layer of two mapping

Proceedings of the Second International
Conference on Semantics, Knowledge, and Grid (SKG'06)
0-7695-2673-X/06 $20.00 © 2006 IEEE

ontology are different, we must make them uniform in shape by adding several Φ to the lesser one.

2.2 The Similarity Algorithm Based on Tree-type Index

In order to carry out ontology mapping between two trees created through approaches hereinbefore, we firstly make definitions as follows:

Definition 1: Floor(C_i) is the layer of node C_i in the index tree created from ontology O, if the number of total concepts in O is n, then 0≤level(C_i)≤n-1;

Definition 2: word_affinity (C_1, C_2) is the language similarity of leaf node C_1 and C_2, and this calculation could be finished by the semantic similarity calculating software from HowNet system[3].

Definition 3: sim_tree(C_1, C_2) is the similarity of the trees rooting in C_1 and C_2;

Definition 4: In the index tree rooting in node C_i, the distance between C_j and its root is distance(j) = level(C_j) - level(C_i);

Then, mapping algorithm is provided as follows:

(1) Suppose there are k layers in the ontology O and O'. From the highest floor m=0 to the lowest floor m = k-1, comparing each concept C_j in the m floor of ontology O to the concept C'_j in the same floor of ontology O' one by one;

(2) If m = k-1, that is to say the two concepts lie in the leaf nodes, sim(C_j , C'_j) = word_affinity(C_j , C'_j) ;

(3) If m ≠ k-1 , sim(C_j , C'_j) = sim_tree(C_j , C'_j);

The similarity between concept C_i in ontology O and C'_i in O' equals to the similarity of the trees that respectively rooting by them:

$$\text{sim\_tree}(C_i , C'_i) = \sum(p(j) * \text{sim\_tree}(C_j , C'_j))$$

where the similarity coefficient p(j) = $2^{-(\text{distance(j)}+1)}$.

In this algorithm, when the distance between a concepts and its root is closer, the effect coursed by its language similarity is greater.

2.3 Deciding the Threshold

During the initialization of threshold, we consider that the higher layer the concepts lie in, the higher similar degree with them. So, we set the initial threshold of the root is 0.8, after the root, each floor reduces 20% from their upper. Then, we use the method of machine learning and manual estimation, adjusting the threshold of each floor. Repeat this process until the thresholds fluctuate in a steady narrow filed.

3. The Experiment Results

In order to test the validity of this method, we chose a group of corresponding ontologies Bio Tutorial Ontologies provided by Protégé[2], and comparing the result with similarity calculating method provided by[3]. Table 1 shows the result data of experiment.

Layer	Index tree structure algorithm			Simple language similarity algorithm		
	Threshold	Relation	Accuracy	Threshold	Relation	Accuracy
1	0.9427	1	100%	0.8000	1	100%
2	0.7188	2	100%	0.6400	2	100%
3	0.5.317	7	77.8%	0.5120	6	66.7%
4	0.3797	8	72.7%	0.4096	6	54.5%
5	0.2891	2	50.0%	0.3277	0	0

Table 1: the result of experiment

6. Conclusions and the Future Works

This mapping method combines the similarity of the inner structure of concepts in different ontologies and the language similarity of concepts, so that the accuracy of the calculation of similarity has been improved. However, since this method need to develop two ontologies into full index trees, in some cases, a lot of empty concepts (Φ) need to be added, so the calculation efficiency could be affected. This problem need to be further studied.

Conference

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The SemanticWeb. Scientific American, 2001.

[2] http://protege.stanford.edu/plugins/owl/

[3] Pun Liu, Sujian Li. Word semantic similarity calculation based on WordNet [EB/OL]. http://www.keenage.com, 2004