Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2015, 7(3):476-481

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Research on concept similarity calculation method based on semantic grid

Zhao Guozeng

Department of Computer and information Engineering, Luoyang Institute of Science and Technology, Luoyang, China

---

ABSTRACT

Study the concept of similarity is the basis to achieve semantic interoperability in the Semantic Grid，but also knowledge representation and information retrieval in an important research content. In this paper, the ISA hierarchy ontology relationship, considering the degree of semantic overlap, semantic distance, level, depth, level, based on the gradient to other factors, presents a conceptual ontology similarity calculation. Experimental results show that the method and the method and the subjective judgments compared to traditional computing methods to improve the accuracy of computer similarity.

Key words: ontology; semantic similarity; semantic distance; level depth

---

INTRODUCTION

Semantic Grid is based on the semantics of Semantic Web technologies to represent and represented by grid computing architecture technologies. It is described by standardized and articulate including computing, storage, databases, services, and other semantic content of information resources, provide an open, safe, orderly, and scalable management architecture to solve complex network computing environment, and information distributed collaborative sharing problems.[1] The key point is that all resources, including services, all with a machine-process able way to describe, its goal is to achieve semantic interoperability.

Since 2001, Semantic Grid research began to receive attention, but is still in its infancy, there are many key issues must be resolved. In the semantic grid system, there are a lot of resources can be shared and heterogeneous grid services, semantic similarity calculation is to establish the semantics of these heterogeneous data association in important ways.

The concept of ontology-based semantic similarity calculation, there are many industry-related research. According to the definition of ontology model, ontology mapping methods can be divided into: Based on the concept of grammar, based on the concept definition, based on the concept instances, and based on the concept structures. [2] These methods are reflected to some extent the relationship between the concepts in the mapping process can reference information includes these aspects, but each individual information cannot fully reflect the relationship between concepts. The definition of the concept of information concept main consideration synonym set and feature set, the string comparison by calculating the similarity of the element name, this method only considers the definition of the concept of information, when the name of the entity is not given in accordance with the concept of meaning, and only when a mere symbol, concept similarity is calculated using the definition of the concept of information would not achieve very good results. Examples of information distribution concept embodies the concept of an instance in the case of a complete set of instances, instance information can be used to calculate the similarity concept concepts. When an instance of the concept is not sufficient, the result instance information with the concept similarity computing concept is not accurate. The concept of structural information reflects the structure of the main body, namely the relationship between concepts semantic level, if only the concept of structural information to calculate the concept similarity, then the calculation result is inaccurate.

In this paper, the combined effects of various factors on the semantic similarity, the traditional algorithm is improved, the use of the ISA on the relationship between ontology structure, according to the concept of semantic overlap degree of internal ontology, semantic distance, and poor levels of depth within the ontology a variety of factors to calculate the semantic similarity between the concepts.

COMPUTATIONAL MODEL OF SEMANTIC SIMILARITY

1. Ontology and semantic similarity

In the ontology, knowledge is the use of classes, instances, relations, functions, and axioms to the standardization of the five primitives. Therefore, ontology can be expressed by a group of five dollars: $O = (C, I, R, F, A)$. Where, C, I, R, F, A denote concepts, instances, relations, functions, and axioms. Classes in the ontology are often classified organization to the concept of form, as the five basic primitives to describe the concept [3].

(1) Class or concept: that is a collection of objects, abstract objects of reality. Used to define the general framework, which includes the name of the concept, and the relationship between concepts of collection, and describe them using natural language.
(2) Relation: concepts and that concept in the field of interaction between the n-dimensional forms can be a subset of the Cartesian product defined: $R: C_1 \times C_2 \times \cdots \times C_n$ 。
(3) Function: refers to the special relationship. This relationship can be in front of all the elements to uniquely determine the next element..
(4) Axiom: a theoretical system that is recognized as the true proposition.
(5) Instance: instance of a specific element that is the object.

Concept of similarity is the existence of semantic concepts in the ontology association, the concept of semantic similarity is defined as follows:

Definition 1 When the ontology of the two concepts has some common characteristics, then they have similarities, similarity is expressed as:

$$Sim(e_{1i}, e_{1j}) = \frac{\alpha}{d + \alpha} (\in (0,1]) \quad (1)$$

Definition 1: When the ontology of the two concepts have some common characteristics, then they have similarities, similarity is expressed as:1, α is the adjustable parameter, d is an integer. If the two concepts are identical, the shape is 1, then $Sim(e_{1i}, e_{1j}) > s$, the two concepts have a certain degree of similarity, s is the set similarity threshold

The concept of semantic similarity calculation is to establish an important method of semantic association. In data mining, artificial intelligence, information retrieval, WEB services and other areas found to play an important role. Concept belongs to the ontology in accordance with the classification, the concept of similarity into the same ontology in the concept of similarity and conceptual similarity between heterogeneous ontologies. This concept involved is from the same ontology.

2. The calculation of semantic similarity

"Semantic similarity" There are two related concepts, namely "semantic relevance" and "semantic distance." Semantic similarity is the degree of word and meaning can be replaced by the degree of compliance; and semantic relatedness refers to the degree of association between the words.

In the ontology, all the concepts are organized according to the relationship in a classification level network, by calculating the level of any two concepts in the semantic network to measure the distance between the concepts of semantic similarity.

For the calculation of semantic similarity, there are two main ways: First, based on two concepts of the most informative common ancestor node to measure the similarity of two concepts, the concept of information known as the volumetric method; the second is a computer based on the concept of semantic distance the model. Currently, many semantic distance calculation, but most are based on WorldNet [4] The main idea of the classification tree method is the use of tree structure, the concept of hierarchical classification of the ontology, which typically is based on the recommended upper shared ontology (Suggested Upper Merged Ontology, SUMO) concept hierarchy tree

algorithm [5-6]. The algorithm need only consider the ontology in the relationship between the concepts of upper and lower (SubClassOf), regardless of the ontology in other binary relations, but the algorithm of the ontology in the concept of structural information and semantic information with less, making the semantic similarity rough calculation. To be able to use the ontology in other rich binary relation, proposed an algorithm based on How Net. The essence of the algorithm HowNet lexical semantics is the soul of the relationship, but the algorithm ignores some of the semantic distance of factors, such as the concept of node depth, breadth, and so on.

The concept of domain ontology in the current study is the similarity of the concept of using the relationship between the upper and lower bit calculation, ignoring the other semantic information, affecting the relationship between the concepts of similarity calculation. Similarly, in the use of a large number of other relationships is how to influence the synthesis of similarity. Constituted by the relationship between ontology levels of classification, the root node is an abstract concept, with the level of preferences, the more specific meaning of the concept. Clearly, the semantic distance of two pairs of the same concept, the more at the top, the similarity also too small, so the level of similarity with the concept which related.

3. Semantic similarity of the factors considered in the calculation

Ontology and semantic similarity of concepts in the concept of semantic overlap between the degrees of semantic distance, there is a certain level of depth of impact. Therefore, the calculation of similarity in the ontology should be considered in the above three factors.

Definition 2 Degree of semantic overlap is the concept within the ontology contain the same number of concepts, it shows the degree of similarity between the two concepts.. If using $a(n_i)$ said the number traced back to the root node, then $s(n_i, n_j)$, said the concept of n_i, n_j degree of semantic overlap.

$$s(n_i, n_j) = a(n_i) \bigcap a(n_j) \tag{2}$$

As in Figure 1, $s(n_7, n_9) = 2$, $s(n_7, n_8) = 3$. Can be seen, the concept of semantic similarity and semantic overlap is proportional to.

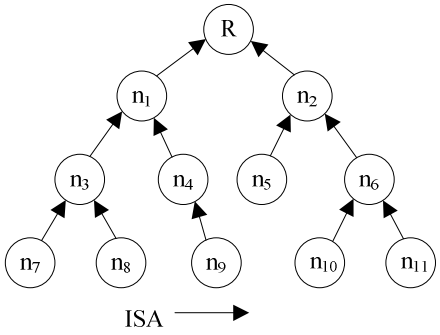

Fig.1An ontology concept hierarchy tree

Definition 3 Semantic distance refers to the ontology in the path connecting two nodes in the shortest path across the edges, denoted by $dist(n_i, n_j)$.

And semantic similarity as a concept in linguistics semantic distance is often mentioned, it refers to the closeness of the two concepts. [7] In general, the semantic distance between two concepts of the smaller, the more similar their semantics, whereas farther. In the field of information retrieval, the value of semantic distance is smaller, the text with user queries closer, when the distance is zero, the text in full compliance with the user's request, when the distance is greater than a certain value, the text is not associated with the user's query, cannot be set back as a result.

$$dist(n_i, n_j) = |(a(n_i) \bigcup a(n_j)) - (a(n_i) \bigcap a(n_j))| \tag{3}$$

For example, $dist(n_7, n_9) = 4$, $dist(n_7, n_8) = 2$, shows that the smaller the semantic distance, the higher the similarity, on the contrary, the lower the similarity.

In SUMO [8], all based on semantic concepts are organized in a hierarchical tree. Therefore, to measure semantic similarity between any two concepts which, of course, the first thought in SUMO, all based on semantic concepts are organized in a hierarchical tree. Therefore, to measure semantic similarity among any two concepts naturally think first calculate the semantic distance between these two concepts in a hierarchical tree, and then converted to a semantic similarity semantic distance. SUMO hierarchical tree structure determines the use of the two concepts in the shortest path tree to represent their semantic distance is a natural distance measure. Semantic distance between two concepts of equal indicate their semantic similarity is the same, however, SUMO-level tree, from top to bottom, the classification concept is descending, the concept of similarity between categories are generally smaller than small classes. Therefore, in the same semantic distance (as defined above) under the circumstances, in a hierarchical tree from the concept of the similarity between the roots farther away than the similarity between the concepts of the root near large. Thus, the concept of the depth in the tree which is a factor to be considered, namely: in the tree edges should be given different depths different weights. In addition, the parent node classification level of detail the concept of semantic distance is also a factor to be considered in the calculation.

Definition4: Level hierarchy in which the concept of depth is the sum. The same two concepts semantic distance, the similarity with the sum of the level where they are increases, with the level difference between them increases and decreases [9-10].

In the ontology of the tree, set owl: thing for the root concept, known as the Root, $h(Root)=1$, then any non-root node n_i in the ontology tree depth $h(n_i)$ is:

$$h(n_i)=h(parent(n_i))+1 \quad (4)$$

Where, $parent(n_i)$ is the parent node n_i .

General concept of ontology tree is the maximum depth of the tree depth. Constituted by the relationship between the depths of the ontology classification tree, each layer is a layer on the concept of refinement. Greater depth, the more specific meaning of the concept, whereas the more abstract. Obviously, the concept of similarity is calculated and the depth. In addition, when calculating the similarity of two concepts, the concept of similarity of the depth difference reflects an important factor. Deviation increase with depth, the similarity is between the concepts of diminish. Figure 1, $Sim(n_3,n_8)>Sim(n_1,n_8)$ similarity.

Considering the above factors, the concept of semantic similarity is calculated as follows:

$$Sim'(n_i,n_j)=\frac{\alpha * s(n_i,n_j)*(h(n_i)+h(n_j))}{dist(n_i,n_j)*(|h(n_i)-h(n_j)|+1)} \quad (5)$$

Semantic similarity in the range [0, 1], the next step needs to be normalized similarity, the following formula:

$$Sim(n_i,n_j)=1-\mu^{Sim'(n_i,n_j)} \quad (6)$$

Formula， μ is the normalization factor.

Program implementation, the concept of similarity between the calculations is as follows:
Step1: Initialize the concept, set the adjustment factor;
Step2: calculate the degree of conceptual overlap between the semantic, semantic distance, level, depth and conceptual difference between the levels of depth;
Step3: the concept of similarity of the computer;
Step4: normalized.

RESULTS

Experiment based on the traditional algorithm, considered in the calculation of similarity between the concept of degree of semantic overlap, semantic distance, the depth of the conceptual level and other factors. In this paper, SUMO randomly select 15 of the concept, the experimental results were compared. First, let 10 people based on experience to determine the similarity between a given concepts, the concept of similarity for these subjective data,

the average value as a subjective standard. In addition, according to the method of calculating this similarity, taking into account the degree of semantic overlap, semantic distance and poor level of depth factors, through appropriate adjustment factor to calculate given the similarity between the concepts of. Meanwhile, in order more clearly, according to the similarity based on semantic distance calculation, only consider the relationship between the upper and lower, the same relationship, regardless of the direction and relevance, to calculate the semantic similarity. The results shown in Figure 2.

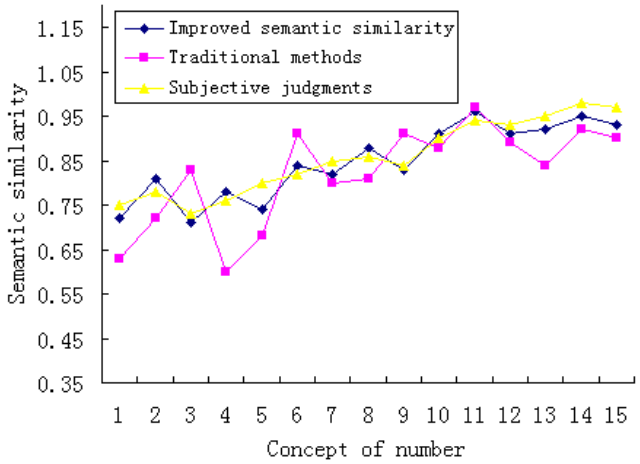

Fig.2 Computer to calculate and compare the subjective judgments of similarity

As can be seen from the chart, the algorithm calculated closer to the concept of semantic similarity of human subjective judgments, therefore, this method is effective. In addition, compared with traditional algorithms, but also has obvious advantages.

CONCLUSION

In the Semantic, Web as the representative of semantic technology, through clear, standardized description of the semantics of information resources trying to solve the Internet resources in the system automatic discovery, seamless integration of direct exchange of data and services. Meanwhile, as the representative of Grid technology architecture is defined by the different levels of standards and specifications, providing more responsive to change for the next generation Internet, independent fault-tolerant, dynamic and scalable distributed computing architecture.

Currently, the Semantic Web and Grid computing is still in the stage of rapid development, the relevant theory and architecture is not mature, still in constant development and improvement. As a combination of the two, semantic grid architecture and application technology is not perfect, there are a number of key issues that must be resolved. Therefore, research in this area has great innovative. Semantic grid semantic mapping and query reasoning there is no practical and proven algorithms and acceptable framework, is still in the primary stage of research and development. Exist in a variety of heterogeneous semantic grid heterogeneous information resources, semantic mapping is an important way to establish these semantic relationship of heterogeneous data, and based on this proposed on the basis of research on ontology mapping algorithm based on semantic mapping linguistics algorithms. Currently, the Semantic Web and Grid computing is still in rapid development stage. As a combination of both, the semantic grid architecture and application technology also far from perfect, there are many key issues must be resolved. The calculation of semantic similarity is an important method to solve the semantic grid semantics between heterogeneous data resources associated. This took into account the degree of semantic overlap, semantic distance, and level gradient to other factors, given the similarity of concepts in ontology computational methods, and verified by experiment the validity of the method. This study is just the beginning; there are many issues to be further research to better the expected results.

Acknowledgements
This work is supported by the Natural Science and Technology Research of Foundation Major Project of Henan Province under Grant (No.142102210471, 142100220568)

REFERENCES

[1] Besnard Ph,Cordier M O,Moinard Y. *Integrated Computer-Aided Engineering*,**2008**,15(4):351-367.
[2] EhrigM,SureY.Ontology mapping-an integrated approach[C]//Proe of the 1st ESWS,Lecture Notes in Computer

Science.[s.I]:Spring Verlag,**2004**(1):76-91.
[3] Sheth.Changing focus on interoperability in information system: Form system, syntax, structure to semantics [C].Interoperating Geographic Information Systems, Boston: Kluwer Academic Publisher, 1999:13-14.
[4] ZHANG Min,SONGRui-hua,MA Shao-ping. *Computer Science*,**2004**,27(10):1395-1401
[5] Budanitsky A, HirstG. *Computational Linguistics*, **2004**, 32(1):13-47.
[6] Niles I,PeaseA.Towards a Standard Upper Ontology //Proc of theInternational Conference on Formal Ontology in Information Systems.Ogunquit,USA,**2001**:2-9.
[7] XU De- zhi,WANGHuai-*Computer Engineering and Applications*,**2007**,43(8):154- 156.
[8] Adam Pease.Standard upper ontology knowledge interchange format [EB/OL].http://suo.ieee.org/SUO/KIF /suo-kif.htm.
[9] WU Jian,WU Zhao-Hui,LIYing,DENGShui-Guang. *Chinese Journal of Computers*,**2005**,28(4):595-602.
[10] ZHANG Yan-xia,ZHANG Ying-jun, PAN Li-hu,XIE Bin-hong,CHEN Li-chao. *Computer Engineering*, **2012**,38(12):176-178.