

Research on Method of Concept Similarity Based on Ontology

Xiquan Yang¹, Ye Zhang^{1,2}, Na Sun¹, Deran Kong¹

¹ School of Computer Science, Northeast Normal University, Changchun, China

Email: yangxq375@nenu.edu.cn

² School of Business, Bohai University, Jinzhou, China

Email: zhangy782@gmail.com

Abstract—This paper brought forward ontology-based information retrieval model and constructed a prototype model. By combing information retrieval approach with ontology, this paper used domain ontology to extend the original keywords input by users and calculates the concept similarity. The experiment results prove that this method is feasible and the computational results are reasonable.

Index Terms—Ontology, semantic web, concept similarity, semantic characteristics

I. INTRODUCTION

With the rapid development and extensive application of ontology in the information retrieval and artificial intelligence fields, ontology-based concepts similarity calculation has become one of the hottest topics. Ontology is playing an increasingly important role in the areas of software engineering, artificial intelligence, information retrieval and web services found [1]. Knowledge can be effectively organized in the field of ontology areas in order to be better shared and reused. However, the similarity between in the concepts of quantitative calculation in ontology is the core issue of search term expansion and the sort of search results in the realization process of concept or semantic retrieval system. Existing ontology inter-entities semantic similarity algorithm researches are including the following areas: Massimo Paolucci proposed DAML-S description-based matching algorithm [2] [3] in 2002. The algorithm essentially makes use of the class levels and type description mechanism provided in the ontology model and simplifies the relations between class levels for four matches, the match results are shown in a discrete level, and the method neglects many other element semantics in OWL. Linguistics think word distances are closely related to word similarity [4], the smaller the distance between two words is, the higher the similarity is; conversely, the greater the distance between two words is, the lower its similarity is.

From the 1970s, psychologists put forward a number of non-geometric similar models, in which the similarity models, proposed by Tversky model in 1977 [5], has been widely concerned and applied in the fields of cognitive psychology and artificial intelligence. The Reference [6], on the basis of Tversky's work, gives a more detailed calculation model. That is the similarity between concepts i and h is defined as a weighted

function of similarity and dissimilarity measures between two concepts, this algorithm can describe a flexible similarity between the concepts in the ontology. Meanwhile, it depends on the quantities of the sharing characteristics and differences.

II. ONTOLOGY KNOWLEDGE

Ontology as a structural knowledge representation can provide a modeling tool to describe knowledge model at the level of the semantic and knowledge, and also provide a semantic framework for text information. It includes vocabularies which are generally acknowledged by experts and scholars in specific domain, and these vocabularies are organized as a formal of directed acyclic graph to describe concept, property and the relationship between them [7]. The more upward level, the more abstract concepts are, and the more downward level, the more specific concepts are. We can form a normalized or standardized architecture by ontology description, and make the machines to realize the knowledge share and reuse according to the restrictions in ontology.

There are many definitions about ontology, with the increasing understanding continuously deepening, the definitions are developing more and more perfect. Neches thought the definition of ontology is that "A ontology defines the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary." [8]. But the most acknowledged definition is proposed by Studer in 1998, which is "Ontology is a formal and normalized explanation of a shared conceptualization." [9] In fact, ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. Ontology which provides semantic information for the classification plays an important role during classifying process. We construct category vector using the concepts extracted from domain ontology, the category vector is defined as $C = \{C_1, C_2, \dots, C_i, \dots, C_n\}$, C is the predefined category label and C_i is concepts in the ontology.

III. SEMANTIC CHARACTERISTICS-BASED ONTOLOGY MATCH ALGORITHM

This paper presents a concept similarity computing approach in domain ontology retrieval, which combines the semantic features of concepts in ontology, analyzes the Semantic Web standard description language OWL, and gives a similarity to each match results. The percentage of similarity can represent any precision degree of similarity. The similarity of ontology concept should not only consider these entities in ontology, but also take the attributes and relations between concepts into account, which have a certain effect on the similarity measure of concepts. Assume that C_1 and C_2 are in the entities defined in ontology, similarity calculation method of the concepts should have the following two characteristics:

Characteristics 1: The results of similarity calculation should be a continuous change value, and can reflect the similarity measure of concepts with fine particle size

Characteristics 2: $MD(C_1, C_2)$ is the similarity of two inter-entities in ontology.

It not only depends on the public characteristics of concepts C_1 and C_2 , but also need to consider the difference between their features. These features include the level of features and its attribute characteristics.

Assume that C_1 and C_2 are two concepts in ontology, $MD(C_1, C_2)$ is the similarity between the two concepts, the definition as in (1) ($\alpha, \beta \in [0,1]$):

$$MD(C_1, C_2) = \frac{\sum(C_1 \cap C_2) - \alpha \sum(C_1 - C_2) - \beta \sum(C_2 - C_1)}{\sum(C_1 \cup C_2)} \quad (1)$$

Given the similarity algorithm $MD(C_1, C_2)$, the algorithm will transform the keywords entered by users into ontology inter-concept class, and calculate the various types amount of information collection in ontology, and get the similarity results. The value of similarity is from 0 to 1. And similarity algorithm is described as follows:

Step 1: pretreatment, use ontology inter-entities to build a similarity measure matrix A , where $A_{ij} = MD(C_1, C_2)$, which can provide any similarity between the two concepts of measurement, and the concept elements of all

the entities defined explicitly between equivalent class and DisjointWith class are assigned to 1 and 0 respectively.

Step 2: for all entities in ontology, calculate the concept differences between the entities. According to actual needs, we can set the difference parameters between the concepts, and calculate the various features vector computing the amount of information collection.

Step 3: calculate the amount of information collection between intersection and union of semantic feature vectors, and set value 1 for the amount of information of each feature.

Step 4: calculate the degree of similarity metric.

IV. EXPERIMENT

This experiment takes wine domain ontology provided by the W3C as an example, it reflects the ability of the description of OWL better. For domain ontology retrieval reasoning, we adopt Jena API to parse wine ontology, and build retrieval prototype system to retrieval the wine ontology.

The statistical graph of semantic similarity between WineTaste class and other classes in domain ontology. From Figure 1, we can see the changes about different values of WineTaste when α and β were respectively assigned 0 and 1. With the increasingly stringent requirements of results, that is, as the proportion of differences increases, the concept similarity is getting smaller and smaller, more and more similarity approaches zero.

Figure 2 and Figure 3 respectively show the values of similarity measure when $\alpha = 0$ and $\beta = 0$, correspondingly β are 0.2, 0.5, 0.8 and α are 0.2, 0.5, 0.8. The experiment results show that when biased towards the parameters α , it will return more results but the parameters β requires more stringent similarity results. In practical application process, the parameters can be adjusted according to users' requirements, and users can obtain different sets of results, which has a good ability of personalized customization.

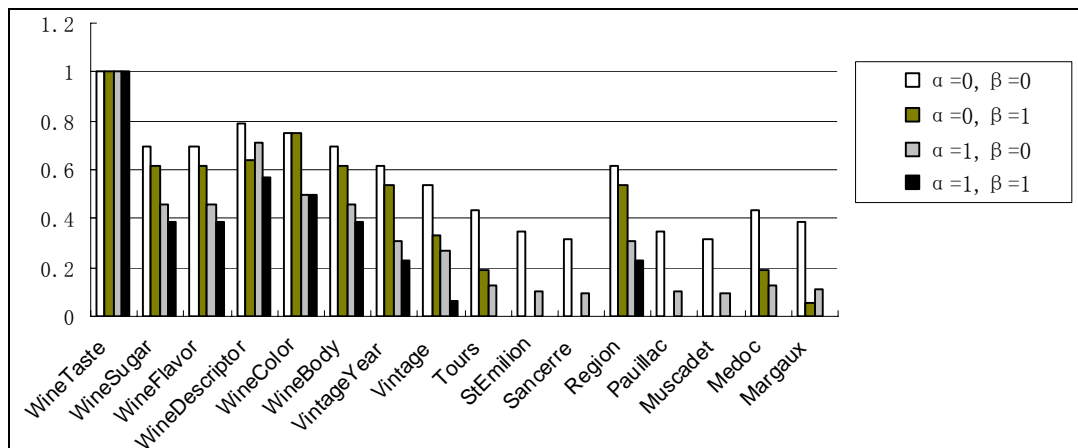


Figure 1. the similarity of wine taste and other concepts.

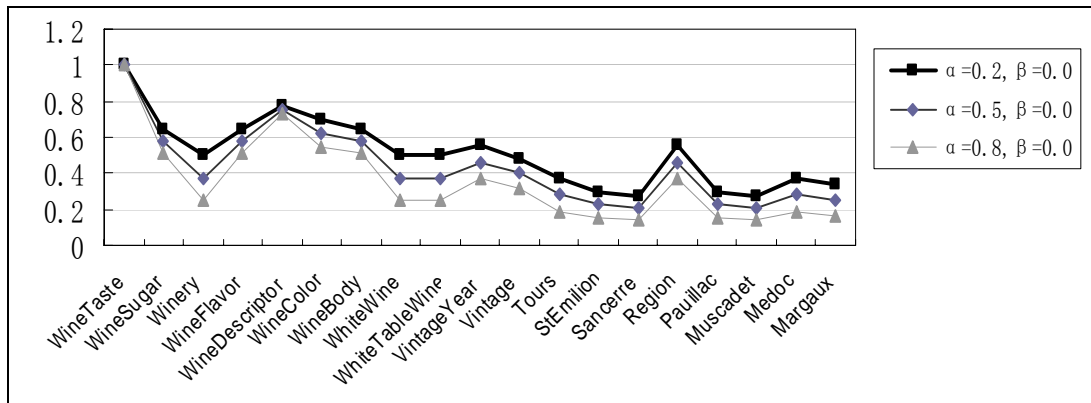


Figure 2. the changing situations of Concept Similarity in Ontology with the changes of α .

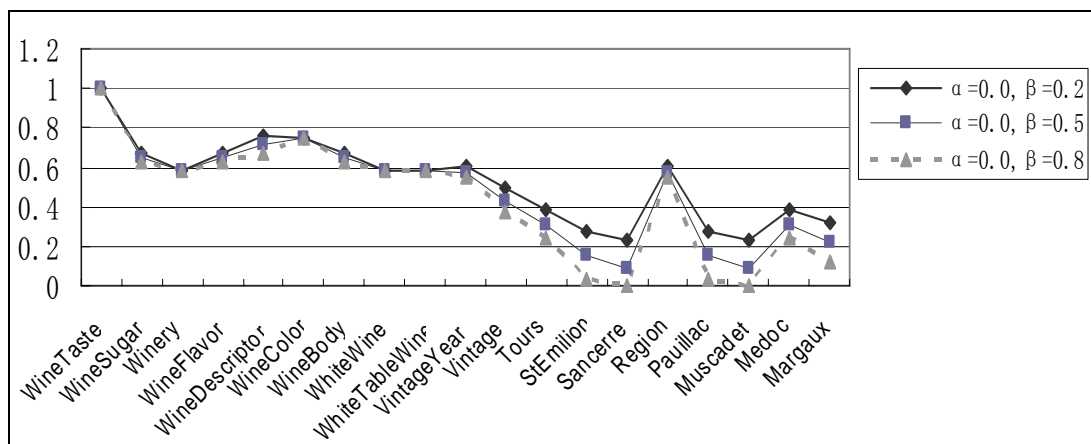


Figure 3. the changing situations of Concept Similarity in Ontology with the changes of β .

As there are no accurate scales to measure the absolute similarity of two concepts, to a large extent, the calculation results are dependent on the definitions of the concept level defined by experts in the field. If the level of concept adjusts, the concept similarity will correspondingly change a lot. The purpose of similarity calculation is to get the match order of one concept and other concepts in order to achieve the best matches. So we can consider the reasonable similarity algorithm in two aspects concerned above, and select the appropriate parameters. The concept WineTaste's matching order proposed by experts is the concept itself, his father and son nodes, the closer attribute node and his brother nodes. From Figure 1, Figure 2 and Figure 3, we can see that although the parameters are different, the matching sequences acquired by similarity are basically consistent.

TABLE I. THE SIMILARITY BETWEEN WINE TASTE AND WINE DESCRIPTOR

	WineTaste	WineDescriptor
$\alpha=0.2, \beta=0.2$	1.000	0.742
$\alpha=0.3, \beta=0.5$	1.000	0.692
$\alpha=0.5, \beta=0.5$	1.000	0.678
$\alpha=0.5, \beta=0.8$	1.000	0.635

According to

the views of experts in the field, the similarity of father and son's relationship should be from 0.7 to 0.8. Through a lot of comparison, the algorithm selects $\alpha = 0.2$, $\beta = 0.2$ as proposed parameters, their minor amendments such as $\alpha = 0.23$ or $\beta = 0.27$ will not affect the outcome of nature. But other circumstances such as $\alpha = 0.5$, $\beta = 0.5$ will lead to the results that similarity value is out of the intuitive scope Table 1 data also illustrates this point, not to list one by one.

V. CONCLUSION

The quantification representation of similarity measure between concepts is to make machines understood the meanings of concepts, and it is the important foundation of intelligent reasoning and judgments. This paper uses similarity algorithm to abstract the levels of ontology concepts and attributes, and gets quantified similarity measure results which changed continuously by making good use of the interaction and union of ontology concepts. Combining semantic distance in ontology with OWL semantic features, we realize to compute the concept similarity, which shows more information about the category of concepts classified by experts. In addition, by adjusting the value of different parameters, users can get a flexible similarity expression of different services, and custom their similarity results according to their own needs.

Further study includes expanding ontology retrieval algorithms to support more OWL semantic elements, adding the calculation of importance of different semantic features and adding the uncertainty information reasoning and so on.

REFERENCES

- [1] Klein M., Bernstein A. Searching for Services on the Semantic Web using Process Ontologies. The First Semantic Web Working Symposium ,2001
- [2] Paolucci M, Kawamura T. Semantic Matching of Web Services Capabilities. In Proc. Of the International Semantic Web Conference. 2002
- [3] Li Kuang, Jian Wu, Shuiguang Deng, Ying Li, Wei Shi, Zhaohui Wu. Exploring Semantic Technologies in Service Matchmaking, Proceedings of the Third European Conference on Web Services ,2005
- [4] Luigi Lombardi, Giuseppe Sartori. Concept similarity: An Abstract Relevance Classes Approach. The 7th International Conference on Cognitive Modeling, 2006.
- [5] TVERSKEY A. Features of Similarity. Psychological Review. 1977. 84(4):327
- [6] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press..
- [7] T.R. Gruber, "Toward principals for the design of Ontologies used for knowledge sharing". International Journal of Human computer Studies, 1995,43 (5/6):907
- [8] R. F. Neches, T. R. Finin, T. Gruber, R. Patil, T. Senator, W.R. Swartout, "Enabling Technology for Knowledge Sharing", AI Magazine, 1991, pp: 36-56.
- [9] R. Studer, V. R.Benjamins, D. Fensel, "Knowledge Engineering, Principles and Methods". Data and Knowledge Engineering, 1998, 25(122) :161-197