Research on a Method of Semantic Query Expansion Based on Civil Aviation Emergency Domain Ontology

Wang Hong [1,2], Fan Hongjie [1], Li Jian [1], Wang Jing [1]
[1]College of Computer Science and Technology
Civil Aviation University of China
Tianjin 300300, China
[2]College of Management
Tianjin University
Tianjin 300072, China
hwang@cauc.edu.cn, 1259786745@qq.com, ljian7115@sina.com, j_wang@cauc.edu.cn

ABSTRACT: *In order to solve the problem of semantic query based on civil aviation emergency domain ontology (CAEDO), a method of semantic query expansion is proposed considering the taxonomic and non-taxonomic relationships in the domain ontology comprehensively. In this method, the calculation of semantic similarity is improved based on the distance and semantic sharing degree for the taxonomic relationships, and the calculation of semantic relevancy on the basis of object properties is proposed for the non-taxonomic relationships. Furthermore, the feasibility of the method is verified by some experiments. The experimental results show that compared with the traditional method, the precision and recall are improved. Moreover, a methodological support is provided for semantic retrieval and generation of rescue plan based on CAEDO.*

Keywords: CAEDO, Semantic Query Expansion, Semantic Similarity, Semantic Relevancy

Received: 23 December 2015, Revised 26 January 2016, Accepted 1 February 2016

1. Introduction

Domain ontology is used to describe and express the relationships among concepts in a specific field. Semantic query expansion based on domain ontology, which adds more relevant concepts to original query set, is used to clearly express user's query intention. Currently, the research of ontology-based semantic query expansion method mainly focus on the calculation of semantic similarity based on the distance [1]-[4], the concept depth [5]-[7], the path weights [8]-[9] and the semantic sharing degree [10]-[12]. These methods are implemented on the taxonomic relationships in the ontology. Research of semantic query on the non-taxonomic relationships is few. Therefore, a method of semantic query expansion is proposed to comprehensively consider the taxonomic and non-taxonomic relationships in the CAEDO.

2. Analysis of Semantic Query Based on CAEDO

In the field of civil aviation emergency management, we've done some research and have made some achievements before. By improving a hierarchical clustering algorithm, we have obtained the taxonomic relationships among domain concepts. Meanwhile, by presenting a method of relation extraction based on NNV association rules [13], we have extracted the non-taxonomic relationships among concepts and successfully constructed the CAEDO. We have implemented the semantic query for rescue resources on the taxonomic relationships. However, this method ignores the impact of non-taxonomic relationships on query performance. Therefore, by comprehensively considering the relationships between concepts, a method of semantic query expansion is proposed based on CAEDO.

For the keyword set entered by a user, based on the relation tree of CAEDO, judge the relationships between keyword (a concept in the CAEDO) and any other ontology concept. If there is only taxonomic relationship between keyword and concept, calculate the semantic similarity on distance and semantic sharing degree. If keyword has taxonomic and non-taxonomic relationships with concept, calculate the semantic relevancy on non-taxonomic relationship and the semantic similarity on taxonomic relationship. When detecting, choose the concept whose similarity value is higher than the threshold to add into the extended set. This set can be used in semantic retrieval to assist the generation of rescue plan. The model of semantic query expansion is shown in figure1.

As the figure 1 shows, the model includes four modules:

CAEDO: The civil aviation emergency domain ontology we have constructed.

Definition Layer: Define semantic similarity, semantic distance, semantic sharing degree and relationship weights.

Similarity Calculation Layer: Calculate the semantic similarity on taxonomic relationships and the semantic relevancy on non-taxonomic relationships based on the CAEDO.

Application Layer: Achieve the application of extended set that use to semantic retrieval and generate rescue plan.

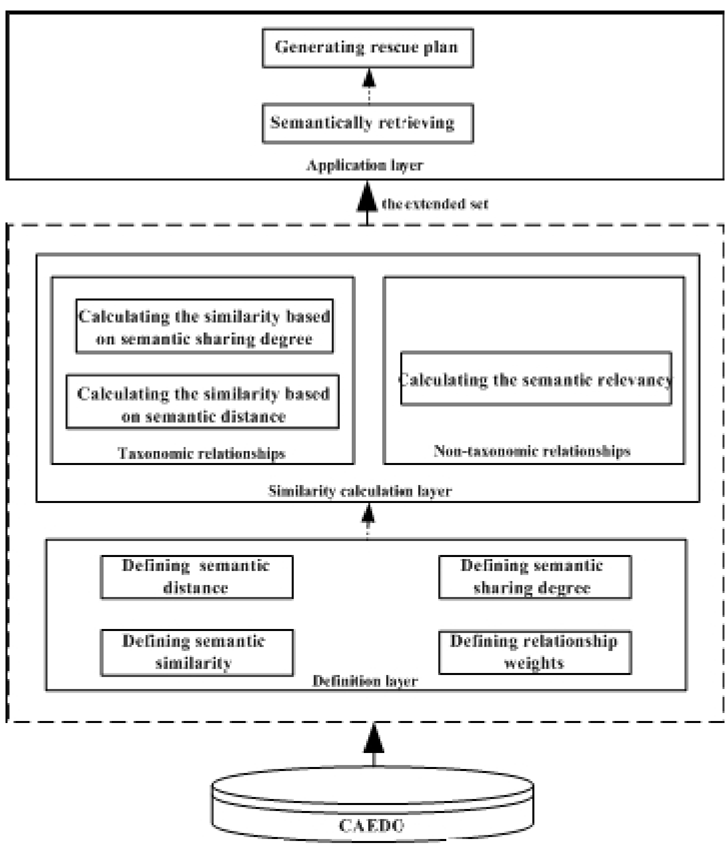

Figure 1. The model of semantic query expansion based on CAEDO

3. Method of Semantic Query Expansion Based on CAEDO

3.1 The Definition of Semantic Similarity

Semantic similarity refers to the coincidence degree of meaning between two concepts [6]. Suppose $Sim(c_i,c_j)$ is the semantic similarity between concepts c_i and c_j. It must meet the following conditions:

(1) The value of semantic similarity is a real number in the interval [0,1]. That is $Sim(c_i, c_j) \subseteq [0.1]$.

(2) If two concepts are completely similar, their semantic similarity is 1. That is $Sim(c_i, c_j) = 1$ if and only if $c_i = c_j$.

(3) If two concepts have no same property, their semantic similarity is 0.That is $Sim\ (c_i, c_j) = 0$.

(4) Semantic similarity between concepts is symmetric. That is $Sim(c_i,c_j) = Sim(c_j,c_i)$.

3.2 The Method of Semantic Query Expansion

When calculating the semantic similarity between keyword (a concept in the ontology) and any other ontology concept in the semantic query expansion, if merely considering the taxonomic relationship, it is very difficult to access the concepts whose similarity with keyword is very small but relevancy is large. In order to avoid this situation, a method of semantic query expansion is proposed considering the taxonomic and non-taxonomic relationships between concepts.

In the constructed domain ontology, if there is only taxonomic relationship between concepts, calculate the semantic similarity on the distance and semantic sharing degree. If there are taxonomic and non-taxonomic relationships between them, they are regarded as directly connected by an edge, calculate their semantic relevancy on object properties and semantic similarity on semantic sharing degree.

Suppose keyword c_i is a concept in the domain ontology and c_j is any other concept in the domain ontology. Their semantic similarity is calculated as follows:If c_i and c_j only have taxonomic relationship, $Sim(c_i, c_j)$ is:

$$Sim(c_i,c_j) = \vartheta_1 Sim_{Dist}(c_i,c_j) + \vartheta_2 Sim_{Con}(c_i,c_j) \tag{1}$$

If c_i and c_j have taxonomic and non-taxonomic relationships, $Sim(c_{i,}c_j)$ is:

$$Sim(c_i,c_j) = \varphi_1 Sim_{NI}(c_i,c_j) + \varphi_2 Sim_{Con}(c_i,c_j) \tag{2}$$

$\vartheta_i, \varphi_i (i = 1,2)$ are variable parameters, which reflect the proportion of various factors, $\vartheta_1 + \vartheta_2 = 1$ and $\varphi_1 + \varphi_2 = 1$.

Choose the concept whose similarity value is higher than the threshold which we set to add into the extended set.

3.3 Calculation of Semantic Similarity on Taxonomic Relationships

3.3.1 Calculation of Similarity Based on Semantic Distance

Semantic distance is the length of shortest path connecting two concepts in the ontology relation tree [6]. Suppose c_i, c_j are two concepts in the tree of CAEDO, the traditional similarity calculation based on semantic distance[1] is:

$$Sim(c_i,c_j) = \frac{\alpha}{Dist(c_i,c_j) + \alpha} \tag{3}$$

$Dist(c_i,c_j)$ is the length of shortest path which connects c_i and c_j in the ontology relation tree. α is a variable parameter.

This calculation only considers the inverse relation between semantic distance and similarity ignoring the effect of relationship types on similarity in the ontology. Therefore, the calculation of similarity on semantic distance is improved by considering relationship weights.

Relationship weight is used to measure the influence degree of relationship type on similarity. In the CAEDO, taxonomic relationships have *Kind-of* relationship and *Instance-of* relationship. Since the refinement degree of instance is greater than all of its ancestor nodes, relationship weight of *Instance-of* is larger than *Kind-of*. Combining with the experience of experts to constantly test the value of relationship weight, the final value is shown as follows:

$$weight(c_i, c_j) = \begin{cases} 0.74, type(c_i, c_j) \text{ is } Instance\text{-}of \text{ relationship;} \\ 0.7, type(c_i, c_j) \text{ is } Kind\text{-}of \text{ relationship;} \end{cases}$$

$type(c_i, c_j)$ means the relationship type between concepts c_i and c_j. $weight(c_i, c_j)$ is the relationship weight of concepts c_i and c_j.

Suppose c_i, c_j are any two concepts in the relation tree of CAEDO, their semantic distance based on relationship weights is:

$$Dist(c_i, c_j) = \frac{\eta}{\prod_{k=1}^{n} weight(k)} - \eta \quad (4)$$

n means the total number of edges in the shortest path from c_i to c_j. $weight(k)$ is the relationship weight of two concepts that are directly connected by the k-edge. η is a adjustable parameter and usually set to 1.

The calculation of similarity based on semantic distance between concepts c_I and c_j is:

$$Sim_{Dist}(c_i, c_j) = \frac{1}{(Dist(c_i, c_j))^2 + \alpha} \quad (5)$$

α is a variable parameter, which reflects the relationship between semantic distance and semantic similarity.

(2) Calculation of Similarity based on Semantic Sharing Degree

Semantic sharing degree refers to the number of same upper concepts which any two concepts share in the relation tree of ontology [2]. Semantic sharing degree expresses the same degree of two concepts. If there are two concept pairs having the same semantic distance, the concept pair that shares more same upper nodes should be more similar. Therefore, the calculation of similarity based on semantic sharing degree is introduced.

Suppose c_i, c_j are two concepts in the CAEDO. Combining with the structure of domain ontology relation tree, their similarity based on semantic sharing degree [10] is:

$$Sim_{Con}(c_i, c_j) = \frac{count[Up(c_i) \cap Up(c_j)]}{count[Up(c_i) \cup Up(c_j)]} \quad (6)$$

$Up(c_i)$ is the set of nodes(upwards) reachable from c_i, $Up(c_j)$ is the set of nodes (upwards) reachable from c_j, $Up(c_i) \cap Up(c_j)$ are the reachable nodes sharing by c_i and c_j. $Up(c_i) \cup Up(c_j)$ are all the nodes(upwards) reachable from c_i and c_j. *count* refers to the number of nodes.

3.4 Calculation of Semantic Relevancy on Non-taxonomic Relationships

Semantic relevancy refers to the degree of correlation between concepts. Some concepts may not be similar, but they can be associated by some other relationships. Therefore, the calculation of semantic relativity is proposed based on object property non-taxonomic relationships.

In the CAEDO, object property non-taxonomic relationships mainly include *Synonymy-of* relationship, *Own* relationship, *Located-in* relationship, *Adjacent-to* relationship, *Associated-with* relationship and *Invoke* relationship. *Synonymy-of* relationship represents that the semantic between concepts is exactly same, so in all relations, it has the largest weight. In the refinement degree, instance has the greatest refinement degree, so its weight is larger than other object property non-taxonomic relationships. Combining with the relevant degree between concepts and the experience of experts to constantly test the value of relationship weight, the final value is shown as follows:

$$weight(r)=\begin{cases}1, \text{if } r \text{ is } \textit{Synonymy - of} \text{ relationship;}\\ 0.72, \text{if } r \text{ is } \textit{Own / Located - in} \text{ relationship;}\\ 0.68, \text{if } r \text{ is } \textit{Adjacent - to / Associated - with} \text{ relationship;}\\ 0.66, \text{if } r \text{ is } \textit{Invoke} \text{ relationship;}\end{cases}$$

weight(r) means the weight of the relationship r in the CAEDO.

Suppose $R = \{r_1, r_2, r_3 \dots r_m\}$ is a relationship set and r_i ($i = 1, 2, 3\dots\dots m$) is a defined object property non-taxonomic relationship in the CAEDO.

Suppose c_i,c_j are two concepts in the domain ontology, their relevance value on the relationship r_i is shown as follows:

If c_i and c_j have the relationship r_i, their relevance value on the relationship r_i is $V_{ri}(c_i, c_j) = weight(r_i)$;

If c_i and c_j don't have the relationship r_i, their relevance value on the relationship r_i is $V_{ri}(c_i, c_j) = 0$;

Due to there may be various object property non-taxonomic relationships between concepts, so the relationship weight of c_i and c_j is:

$$w(c_i, c_j) = \frac{\sum_{k=1}^{m} V_{r_k}(c_i, c_j)}{n} \tag{7}$$

m means the number of elements in the set R, n is the number of non-duplicate object property non-taxonomic relationships between concepts c_i and c_j.

Two semantically relevant concepts based on object properties are regarded as directly connected by an edge. The calculation of semantic relevancy between c_i and c_j is:

$$Sim_{NI}(c_i, c_j) = \frac{1}{\left(\frac{\varepsilon}{w(c_i, c_j)} - \varepsilon\right)^2 + \sigma} \tag{8}$$

ε , σ are adjustment factors and usually set to 1.

4. Process and Implementation Effect of Semantic Query Expansion

4.1 The Process of Semantic Query Expansion

The basic idea of proposed method is to get and store the quantitative values of similarity between concepts based on CAEDO. There is a threshold set to limit the number of extended concepts. When detecting, choose the concept whose similarity value with keyword is higher than the threshold to add into the extended set. The process of semantic query expansion is described figure 2.

4.2 Implementation Effect of Semantic Query Expansion

According to the characteristics of civil aviation emergency field, we have constructed domain ontology. CAEDO is a kind of normalized description of concepts and their relationships involved in emergency management of civil aviation field. Part of the CAEDO is shown in the figure 3.

As shown in figure 3, oval represents a domain concept. Solid line and empty arrow represent the *Kind-of* relationship. Solid line and solid arrow represent the *Instance-of* relationship. Dotted line and solid arrow represent an object property non-taxonomic relationship.

According to the proposed model and the semantic query expansion method, we design a contrastive experiment. In general, the effect of semantic distance and object property on similarity is larger, so ϑ_1 values 0.9, ϑ_2 values 0.1, φ_1 values 0.85, φ_2 values 0.15, α values 1 and γ values 0.65. We take 5 group queries for test; the following table 1 is three of the queries.

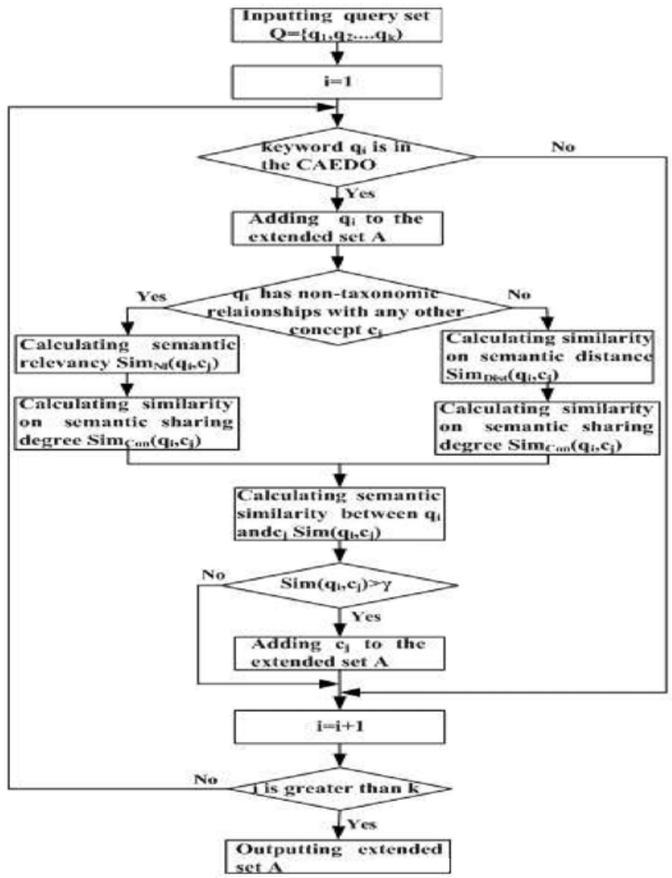

Figure 2. The process of semantic query expansion based on CAEDO

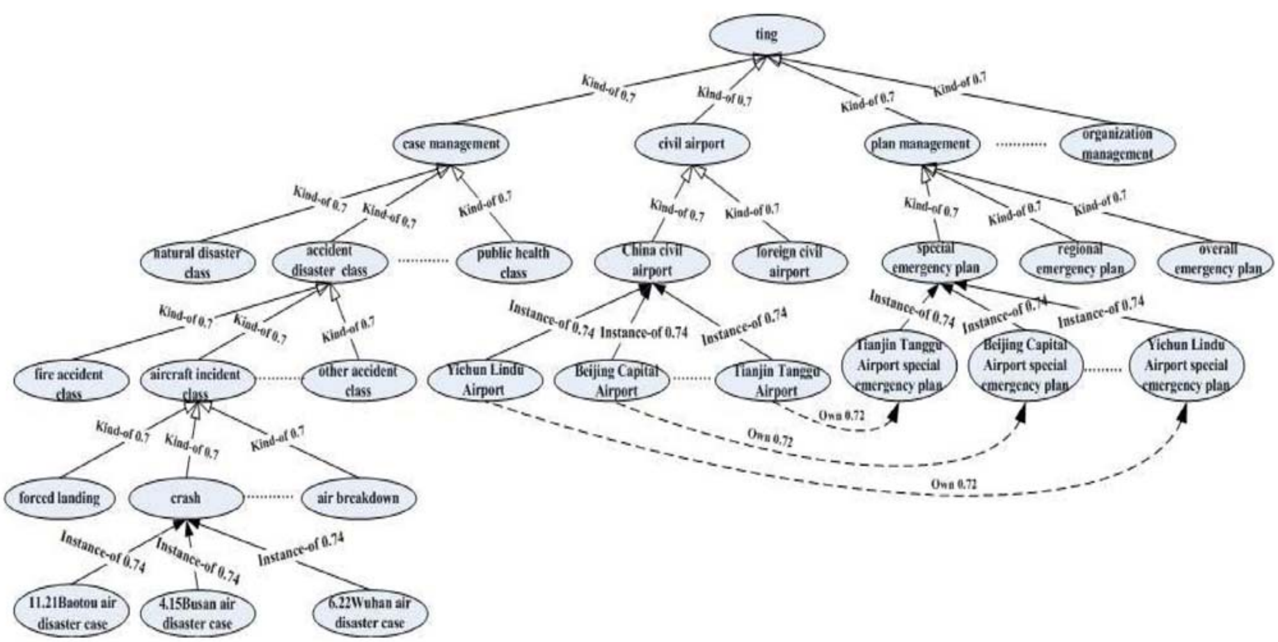

Figure 3. Part of the CAEDO

User's query	Extended set(the proposed method)	Extended set(the traditional method)
Yichun Lindu Airport,crash	{Yichun Lindu Airport, China civil airport, Yichun Chinese Medicine Hospital, Yichun Fourth People's Hospital, medical rescue equipments, medical rescue vehicles, medical rescue supplies, Yichun Fire Brigade, fire fighting equipments, fire engines, Yichun Lindu Airport special emergency plan}{crash, aircraft incident class, 11.21 Baotou air disaster case, 4.15 Busan air disaster case, 6.22 Wuhan air disaster case}	{Yichun Lindu Airport, China civil airport}{crash, aircraft incident class, 11.21 Baotou air disaster case, 4.15 Busan air disaster case, 6.22 Wuhan air disaster case}
Pingyang district, disintegrate	{ Pingyang district, Ruian city, Pingyang People's Hospital, Tengjiao Central Hospital, Pingyang Maternal and Child Health Hospital, Pingyang Chinese Medicine Hospital, Xiaojiang Hospital, medical rescue equipments, medical rescue vehicles,medical rescue supplies, Pingyang Fire Squadron, Aojiang Fire Squadron, Shuitou Fire Squadron, fire fighting equipments, fire engines}{disintegrate, aircraft incident class, 5.8 Shenzhen air disaster case, 6.6 Xi'an air disaster case, 7.31 Nanjing air disaster case}	{Pingyang district, Ruian city, Pingyang People's Hospital, Tengjiao Central Hospital, Pingyang Maternal and Child Health Hospital, Pingyang Chinese Medicine Hospital, Xiaojiang Hospital, Pingyang Fire Squadron, Aojiang Fire Squadron, Shuitou Fire Squadron} {disintegrate, aircraft incident class, 5.8 Shenzhen air disaster case, 6.6 Xi'an air disaster case, 7.31 Nanjing air disaster case}
Tianjin special emergency plan	{Tianjin special emergency plan, regional emergency plan,Tianjin Binhai Airport special emergency plan, Tianjin Tanggu Airport special emergency plan, China civil aviation overall emergency plan, national disposal of civil aircraft accident emergency plan}	{Tianjin special emergency plan,regional emergency plan}

Table 1. The extended results

Use recall and precision to measure the effect of two methods. The calculation formulas are shown as follows:
Recall = all the related concepts which are detected / total domain ontology concepts.
Precision = all the related concepts which are detected / the entire related concept.

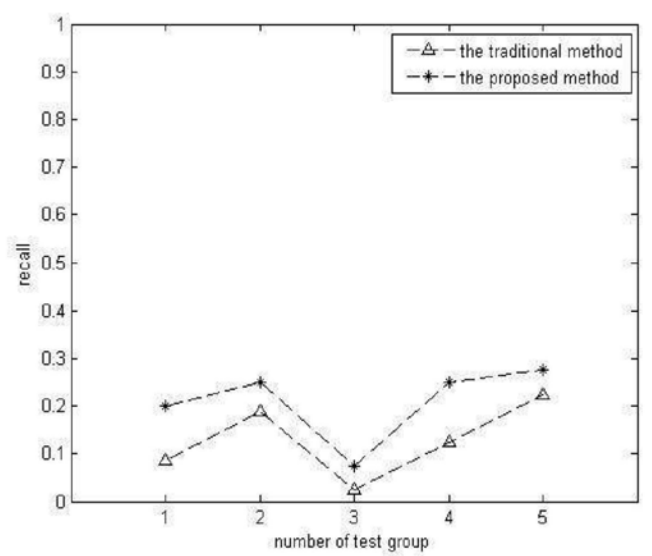

Figure 4. Comparison of recall

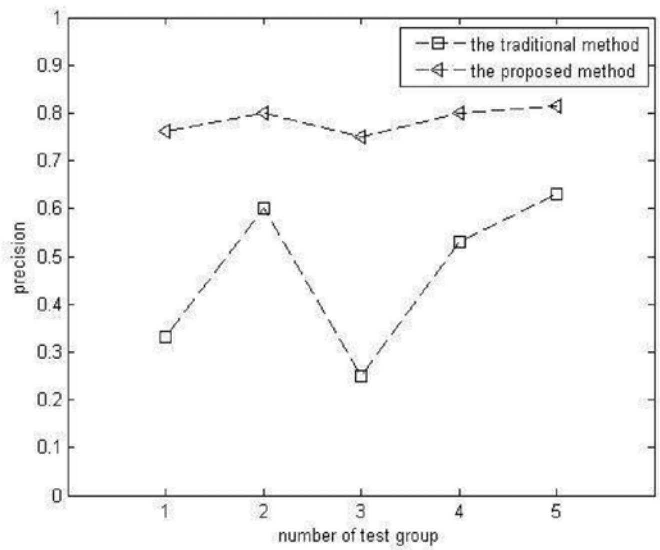

Figure 5. Comparison of precision

Experimental results show that the traditional method only can expand the concepts that have larger semantic similarity with keyword on the taxonomic relationships. Compared with the traditional method, the proposed method also can obtain the concepts that are related to keyword on the non-taxonomic relationships. It can effectively improve the recall and precision in the semantic query expansion.

5. Conclusions

We have proposed a method of semantic query expansion which comprehensively considers the taxonomic and non-taxonomic relationships between concepts through the calculation of semantic similarity and semantic relevancy based on the CAEDO. The experimental results prove that this method has higher recall and precision than the traditional method and can expresses user's intention more clearly. In the civil aviation emergency rescue system, the semantic query expansion method that we proposed has been used in semantic retrieval to assist the generation of rescue plan. Future work is to solve the sequencing problem of extended results.

6. Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant No. 61079007, the Scientific Research Foundation of Civil Aviation University of China under Grant No. 09QD04X, the Central University Foundation of Civil Aviation University of China under Grant No. 3122014C016.

References

[1] Zhang, Fan., Zhong, Jinhong., Huang, Ling. (2010). A Improved Calculation Method of Semantic Similarity between Domain Ontology Concepts. *Computer Engineering*, 36 (23) 66-68

[2] Li, Wenjie., Zhao, Yan. (2010). A Algorithm of Semantic Similarity Based on Ontology Structure. *Computer Engineering*, 36 (23) 4-6

[3] Liu, Dexi ., Wu, Shihan. (2011). A Query-Oriented Summarization System for XML Elements. *International Conference on Management of e-Commerce and e-Government*, (22) 222-225

[4] Li, Rong, Cao, Shunliang., Li, Yuanyuan. (2006). A Measure of Semantic Similarity between Gene Ontology Terms Based on Semantic Pathway Covering. *Process in Natural Science*, 16 (7) 721-715

[5] Cui. Qiwen., Xie, Fu . (2012). A Improved Calculation Method of Semantic Similarity between Domain Ontology Concept.*Computer Applications and Software*, 29 (2) 173-175.

[6] He, Yuanxiang., Shi, Baoming., Zhang, Yong. (2013). Research on a Algorithm of Semantic Similarity Based on Ontology. *Computer Applications and Software*, 30 (11) 31 2-315

[7] Xu, Xianghua., Huang, Jialai., Wan, Jian. (2008). A Method for Measuring Semantic Similarity of Concepts in the Same Ontology. *International Mut symposiums on Computer and Computational Sciences*, 22 (8) 207-213

[8] Zhou, Jianfeng. (2010). Improved Semantic Similarity Calculation Model Based on Ontology. *Journal of Information*, 29 (12) 144-147.

[9] Huang, Guo., Zhou, Zhurong . (2007). Research on Semantic Similarity Computation Based on Domain Ontology. *Computer Engineering and Design*, 28 (10) 2460-2463.

[10]Li, Zhaolong., Du, Junping. (2012). Semantic Similarity Calculation Model Based on Tourism Domain Ontology. *Proceedings of the 31st Chinese Control Conference*, 27 (25) 3863-3868.

[11] Chen, Huacheng., DU, Xuehui., Chen, Xingyuan. (2012). Query Expansion Model Based on Interest Ontology. International Conference on Information Management, *Innovation Management and Industrial Engineering*, 12 (7) 475-478.

[12] Zhang, Wei., Duan, Liguo., Chen, Liguo(2010). The Research on Concept Semantic Similarity Computing Based on Semantic Tree. *International Symposium on Intelligence Information Processing and Trusted Computing*, 43 (10) 324-327.

[13] Wang, Hong, Gao, Siting, (2012). Application and Research of Non-taxonomic Relation Extraction Method Based on NNV Association Rule. *Computer Applications and Research*, 12 (9), 3665 -3668.