A Domain Information Retrieval Method based on Pension Insurance Ontology Concept Similarity

Qingsheng Wan, Shaobin Huang*, Kai Sun, Lifu Feng, and Ya Li
College of Computer Science and Technology, Harbin engineering University, Harbin, China
Email: wanqingsheng@yeah.net
*Corresponding author

***Abstract*—Matching key terms is the main method for most engines for information retrieval, but this method cannot meet the needs of people who want to retrieve useful domain information. This is because that it doesn't consider the semantic relation of resources in the context. Ontology is an effective method that can solve the problem of domain knowledge sharing and semantic understanding. In this paper, an optimization strategy is presented for information retrieval domain, and an algorithm is also proposed for ontology concept similarity calculation. Experiments are carried out on the audit guidance system, and the results show that this method significantly improves the accuracy of information retrieval based on domain ontology.**

***Index Terms*—Domain Ontology; Concept Similarity; Information Retrieval**

I. INTRODUCTION

With the rapid development of Internet, web has become a global information source. Search engine is a general way for users to gain information. However using the traditional search engine, it's usually quite difficult for the user to find the needed information accurately. There are following several reasons: The first reason is that the understanding of the user's problem is not accurate enough, which leads to a lot of noise in returned results so that the user can't easily find the information they need; The second reason is that the processing of information content mostly uses the coding-procedure-based pretreatment technology or text analysis technology, which reflect only one side of the content; The third reason is that the problem proposed by the users and information source content may not be completely consistent, and it is difficult to guarantee correctly matching between the content and user questions, and the correct rate is very low. To improve the existing retrieval accuracy of the system, we must solve the problems mentioned above. Some of the existing researches show that technology based on Ontology is one of the methods to solve these problems.

By the establishment of a Ontology based on the WordNet [5], an approach in [1] has solved the problem of information retrieval from the yellow pages and product catalogues, but in general the yellow pages and product catalog information have a certain structure, therefore the technique used in [1] could not be well applied to retrieve in natural language text. References [2-4] focus on the universal Ontology that is applied to information retrieval and natural language understanding. But it is difficult or even impossible to building a universal Ontology which can cover all areas of knowledge. So a more realistic method is to establish a field of Ontology, and to use it to solve the specific problems of information retrieval on that field.

Traditional information retrieval methods are mainly divided into two categories. The one is the method based on keyword matching. This method first allows users to present a retrieval request in the form of key words, then performs a matching between keywords submitted by users and documents from a document library, finally returns to user the documents where the keywords submitted by the user appeared. At present, many search engines in Web are using this search method, such as Google [6], Baidu [7], etc. But a disadvantage of this method is that any semantic information is not contained in the retrieval process, which is a very important reason for the not-high retrieval accuracy. The other one is called conceptual information retrieval [8], which extracts various conceptual information by processing information in the document on semantic level, thus forming a concept base, and then retrieves relative information in the concept base, and provide search results according to the understanding of user's question. This method overcomes the limitation that the semantic information retrieval is not considered in retrieval based on keywords, and this method has better natural language interface. But a disadvantage of the conceptual information retrieval is that the description of relations between concepts is not contained in the concept base, thus it is unable to deal with problems regarding relations between concepts.

Ontology is a clear description of the concept, which abstracts some applied field in the real world into a set of concepts and relations between concepts [9]. We put the Ontology into the traditional information retrieval technologies, which can not only inherit the advantages for conceptual information retrieval, but also can overcome the limitation that conceptual information retrieval cannot deal with relations between concepts.

doi:10.4304/jnw.8.12.2908-2914

This paper discusses how to improve accuracy of retrieval in Pension Insurance domain, and it focuses on Ontology concept similarity calculation which plays an important role in information retrieval optimization [15-17]. Based on Ontology concept similarity, we propose an Information Retrieval model based on Ontology. Comparing with common retrieval models based on keywords, our model can improve the accuracy of retrieval efficiently.

II. STRATEGY OF DOMAIN INFORMATION RETRIEVAL OPTIMIZATION

It is difficult for machines to understand the keywords inputted by users. Mechanical matching is still the main strategy of current Search Engine for finding and sorting. Retrieval results cannot cover the synonyms for the words in the keyword list, which causes unsatisfactory accuracy and completeness of the search results [18].

We present an optimized method to solve this problem. We add ontology into Search Engine, use semantic character of Ontology to optimize user inputs and then send the inputs to Search Engine.

According to Pension Insurance Ontology, our optimization methods are given as follows.

A. Optimization based on "subClassOf" Relation

"subClassOf" Relation represents the relationship layer of Ontology concepts, just like the Parent-Child Relationships in a tree structure [19-21]. We can build an Ontology concept tree structure with the help of "subClassOf" Relation. In this tree, every concept in the lower layer is the embodiment of those in the higher layer, and every concept in the higher layer is the generalization of those in the lower layer. Thus, we match keywords which we want to retrieve in the ontology tree. If such keyword exists, we can use sub-concepts instead of the origin word to improve the accuracy of retrieval process. If the desired keyword cannot be found, we can use the father concept to get more results. This method also can be used to generate a keyword set which can be used to generate more useful results.

B. Optimization based on Equivalent Relation and the Similarity of Ontology Concepts

Equivalent relation means that there exist concepts that have the same semantic. Compared with the method of retrieval by matching keywords, it has higher search efficiency, accuracy and comprehensive. We also increase or replace keywords by judging the similarity between concepts. An algorithm for the ontology concept similarity calculation is given in the next section.

C. Optimization based on Custom Relation

The pension insurance ontology which is used in the optimization strategy contains 11 relation attributes. These attributes reflect the constraint relation and the process of pension insurance business. Take "handing" as an example, we can get the concepts "issue", "payment", "declare" through the pension insurance ontology, because these business processes all need the process "handing", and the concept "operator" is the performer of "handing", and the "handing agency" is the agency of "handing". Continue finding, we can find the concept "examine and approve". In other words, using the ontology can get the relation concepts which reflect pension insurance business by a keyword. This strategy is also very useful in the question-answering system.

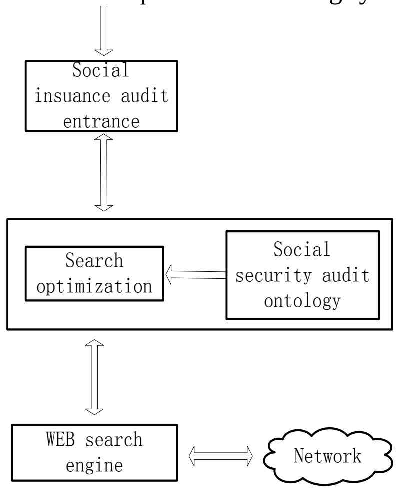

Figure 1. Optimized retrieval process

As shown in Fig. 1, it is a solution that the pension insurance ontology has been applied to social security domain information retrieval. The function of each module in Fig. 1 is as follows: The module of "social security entrance" is the interface to users, it is used to input search terms and check the retrieval results. The module of "social security audit ontology" is based on the pension insurance ontology in the format of OWL. The module of "search optimization" uses optimization strategies which are presented above to increase, replace and delete the retrieval keyword, then remove the same results in the many groups of results which are based on many groups of retrieval terms. The module of "web search engine" uses search engine to retrieve.

The goal of audit guidance system is to guide audit methods, policies and regulations. The core technology of the audit guidance system is retrieval process. The retrieval efficiency determines directly that whether the system can meet the needs of clients. This paper judges the efficiency of the optimization strategy by comparing the retrieval results according to the keywords matching and the optimization strategy presented above.

III. ONTOLOGY CONCEPT SIMILARITY

A. Related Research of Ontology Concept Similarity

The basis of concept similarity calculation using ontology is that the two concepts have a semantic correlation, and there exists a path in hierarchy network diagram. Resnik measures the semantic similarity of two words according to the maximum amount of information of their common ancestor node [22]. Agirre and Rigau take semantic distance, depth and area density of concept tree into consideration while calculating similarity between words [23]. Leacock and Chodorow propose a semantic similarity calculation model based on distance, which is simple and intuitive, but is heavily dependent on pre-established ontology hierarchical network [24]. Li et al calculate the similarity using China National Knowledge Infrastructure and Synonyms Dictionary, and it can be used in semantic disambiguation [25]. Liu and Li use HowNet to transform similarity calculation between concepts into calculation between sememes, and determine concept similarity by calculating the distance between sememes [26]. Zhu et al introduce the idea of semantic distance in Computational Linguistics to calculate concept similarity, but they take little consideration about the influencing factors of similarity [27]; Huang and Zhou optimize the five decision factors of semantic similarity, and provide a new semantic similarity calculation model [28].

B. Factors Affecting the Similarity

There are some principles to be considered when calculating concept similarity [29]. First of all, we need to quantify the similarity. The similarity is a numerical value which is generally between 0 and 1. If a similarity is not in this range, it should be normalized. Second, the similarity calculation should be simple, convenient and the computational complexity should be as low as possible. Third, take full advantage of the ontology characters, and take into account various impact factors as many as possible. Finally, the concept similarity should be symmetrical, that is $Sim(C_1, C_2)=Sim(C_2, C_1)$.

Definition 1 Semantic coincidence degree means the number of the same superior concepts of inner concepts in the ontology. Semantic coincidence degree shows the degree of similarity [15].

Use OL (C_1, C_2) to denote the semantic coincidence degree of the two concepts C_1 and C_2. $O(C_1)$ denotes the set of nodes while node C_1 up going to root node, $|O(C_1)|$ denotes the length of set $O(C_1)$, then the semantic coincidence degree of C_1 and C_2 is

$$OL(C_1, C_2) = |O(C_1) \cap O(C_2)| \tag{1}$$

The greater the semantic coincidence degree is, the greater the similarity; with the increase of coincidence degree between concepts, their similarity also increases.

Definition 2 Semantic distance means the number of edges that are crossed by the shortest path which connects two concept nodes in ontology diagram, that is, the number of directed edges of the shortest distance between two concepts.

The greater the semantic distance is, the lower the similarity. It is a commonly used measurement to represent semantic distance using the number of directed edges of the shortest path. However, in the hierarchical tree of ontology, the organization of concept is top-down, and the classification is from big to small, from coarse to fine, the similarity between concepts that far away from the root is greater than between concepts that near the root.

Definition 3 The width of concept means the number of nodes that are the direct child nodes of concept C. It is denoted as Wid (C).

Definition 4 (The weight of concept) Using Weight(C) to denote the weight of concept C, using Parent(C) to denote the father node of concept C, and the weight of concept C is formulated by

$$Weight(C)=\begin{cases}\dfrac{1}{Wid(C)} & C=Root\\ \dfrac{1}{2}\times\dfrac{1}{Wid(C)}\times Weight(Parent(C)) & C\neq Root\end{cases} \tag{2}$$

Definition 5 (The weight of edge) the edges introduced from concept C have the same weight, using the weight of concept C to present the edge weight. Denoting the edge from C_1 to C_2 as edge (C_1, C_2), then

$$Weight\left(edge\left(C_1, C_2\right)\right) = Weight\left(C_1\right) \tag{3}$$

Then, the edge weight can denote the semantic distance of two concepts, that is, the sum of weight of edges that are crossed by the shortest path of two concepts. Assume that C_1, C_2… C_n is the shortest path between concept C_1 and C_n, then:

$$Dist(C_1, C_n)=\sum_{i=1}^{n-1} Weight(C_i) \tag{4}$$

Definition 6 (The depth of concept) In an ontology tree, assume that the depth of root node Root is 1, that is, Dep(Root)=1, then the depth of other nodes is

$$Dep(C)=Dep(Parent(C))+1 \tag{5}$$

Definition 7 (The depth of tree) The depth of tree Dep(Tree) is the biggest depth of nodes in the tree, that is:

$$Dep(Tree)=Max(Dep(C)) \tag{6}$$

In the hierarchical tree of ontology, the similarity increases when the two concepts get near to the root. The similarity of two concepts with the same semantic distance increases with the growth of the sum of their depth, and decreases with the growth of the depth difference.

If the two concepts have a larger amount of shared information, then their similarity is larger. Each concept can be regarded as a refinement of its ancestor node, and every child node contains the information of all its ancestor nodes. In order to make full use of information theory and the theory of probability and statistics, the paper introduces statistic characters of ontology to similarity calculation.

Definition 8 (The information amount of concept C) Denoting P(C) as

$$P(C)=\frac{a}{total\ amout\ of\ training\ data} \tag{7}$$

a=number of times that concept C appears in training data. Then the information amount of concept C, namely Info(C) is

$$Info(C)=-lg(P(C)) \tag{8}$$

Definition 9 (The edge strength) The strength of directed edge, namely edge (C_1, C_2) is

$$Int\left(edge\left(C_1,\ C_2\right)\right)=\left|Info\left(C_1\right)-Info\left(C_2\right)\right| \quad (9)$$

Definition 10 (The strength between two concepts) The strength between two concepts C_1 and C_n is the sum of strength of edges in the shortest path of C_1 and C_n. Assume that C_1, C_2 … C_n is the shortest path between concept C_1 and C_n, then:

$$\ln t(C_1,C_n)=\sum_{i=1}^{n-1}\ln t(edge(C_i,C_{i+1})) \quad (10)$$

C. The Formula of Concept Similarity

Considering the factors affecting the similarity mentioned above, we propose the formula of concept similarity:

$$Sim(C_1,C_2)=1-\sqrt{\frac{1}{OL(C_1,C_2)}\times\frac{|Dep(C_1)-Dep(C_2)|+1}{Dep(C_1)+Dep(C_2)}\times\frac{1}{\ln t(C_1,C_2)}\times Dist(C_1,C_2)} \quad (11)$$

As we can see, the concept similarity is positively correlated with semantic coincidence degree and the concept strength is negatively correlated with semantic distance and depth difference.

D. Realization of Algorithm

In order to facilitate the computing of the ontology concept similarity, this paper defines the ontology in the form of O : < C , R , Cr>, in the definition, O means ontology, C means concept, R means the relation between concepts, Cr means taxonomic concept which is the relate hierarchy concept of the concept “C”. And the Cr can be defined in the form of < C, R , C>. Then we can get the index table which makes the concept as a key value in the form of <<C，<C , R , C>>>. By the concept index table, we can organize the ontology as a concept syntax tree. At last this paper presents a concept similarity computing algorithm defined on concept tree [30].

Definition 1: Concept syntax tree is a group that consists of four elements in the form of “concept Tree:< C , D , CNs , PN >”, where C means concept, D means the depth value which is the depth of concept in the ontology concept syntax tree, CNs means the children concept set of “C”, PN means the parent concept of “C”.

Our recursive algorithm which generates the concept syntax tree is as follows.

Algorithm 1:

Input: Concept index table “conceptIndexTable<cpt , < cpt1 , relation , cpt2 >>”, root node “root< Thing , 0 , null , null >”

Output: Concept syntax tree “conceptTree< cpt , depth , children , parent>”

Algorithm name: creatCptTree(conceptIndexTable, node)

```
Crs ← conceptIndexTable.getValue(node. concept);
if(Crs != null)
 while(Crs.hasNext())
 inf ← Crs.nextElement();
 newNode.cpt ← inf.cpt2;
 newNode.depth ← node.depth + 1;
 newNode.parent ← node;
 creatCptTree(conceptIndexTable, newNode)
 node.children.addElement(newNode);
 end while
end if
```

After generating the concept syntax tree, this paper presents some formulas and definitions about the concept similarity computing. The concept similarity value between two concepts generally is expressed as num 0 to 1, and the greater num means the greater similarity. For example, formula 1 means the relation between the similarities of concept “c1” and “c2”.

$$sim(c_1,c_2)=1-dc(c_1,c_2) \quad (12)$$

In formula 12, $dc(c_1,c_2)$ means the distance of two concepts.

In order to compute the distance of two concepts, this paper takes “milestone” as the distance between node and the root node in the concept syntax tree. The formula of computing milestone is as follows.

$$milestone(n)=\frac{1}{2k^{l(n)}} \quad (13)$$

In formula (13), $l(n)$ means the distance between node and the root node. (l (root) = 0)

$$dc(c_1,c_2)=dc(c_1,ccp)+dc(c_2,ccp) \quad (14)$$

$$dc(c,ccp)=milestone(ccp)-milestone(c) \quad (15)$$

Formula (13) and formula (14) show how to compute the distance between two concepts. In formula (13) and (14), “ccp” means the parent node of node “c1” and “c2”.

The above calculation model is based on such an idea. In the concept syntax tree, the similarity between two concepts is greater if the depth between the two concepts is greater. And the similarity between two concepts which are parent-child relation is greater than the similarity between two concepts which are brother’s relation.

Algorithm 2 describes the computation process of concept similarity through concept syntax tree.

Algorithm 2:

Input: concept syntax tree: conceptTree<cpt, depth, children, parent>, Pension Insurance ontology: ontoLib

Output: concept similarity matrix: cptSimMatrix

Algorithm name: simCompute(conceptTree,ontoLib)

```
conceptLib ← ontoLib;
for each cpt1 in conceptLib;
 for each cpt2 in conceptLib;
 if(find(conceptTree,cpt1)&&find
(conceptTree, cpt2))
 then partNode ← findParent(conceptTree ,
cpt1,cpt2);
 simValue ← 1–dc(cpt1, partNode) –dc
 (cpt2, partNode);
 end if
 end for
end for
return cptSimMatrix
```

IV. EXAMPLES OF PENSION INSURANCE ONTOLOGY

The research method of this paper is based on Pension Insurance ontology which is constructed by our research group. It contains 363 concepts and 11 attribute relations. Let us take Fig. 2 as an example, there are 3 concepts under the top level concept “thing”, such as “person”, “unit”, and “basic pension insurance”. And the concept “basic pension insurance” contains 16 atomic concepts; they include all the business in the pension insurance domain. Constructing of the pension insurance ontology is favor of the ontology integration of medical insurance ontology, maternity insurance ontology and so on.

Figure 2. Fension insurance ontology

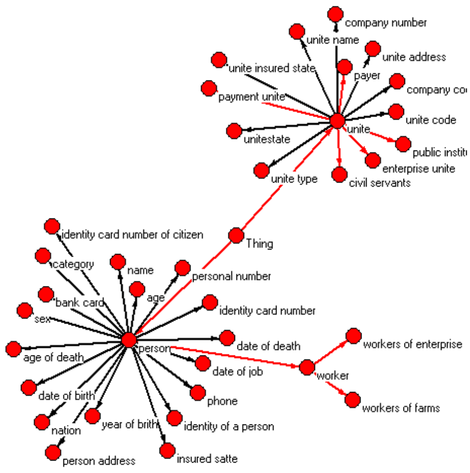

Figure 3. Part of the ontology visualization

Fig. 3 shows the visualization of some ontology about the concept “person” and “unit”, where black line means the object attributes which are contained by the concept, such as concept “person” contains object attribute “name”, “sex”, “residence”, “ID num”, “nation” and so on, and red line means the taxonomic relation between concepts, they are likely to be the parent class and child class relations in the object-oriented programming, such as concept “person” has a child concept “workers”, concept “workers” has concepts “enterprise’s employee”, “farm worker” and so on. These children concepts all inherit the object attribute of their parent concept.

V. Experimental Results

This chapter first makes an experiment which computes ontology concept similarity using algorithm 2. We get a matrix of 163 dimensions, as shown in Fig. 4. After domain experts’ checking, the similarity matrix reflects the real similarity between pension insurance ontology concepts. It also reflects the validity of the pension insurance ontology constructed by our laboratory.

Figure 4. Part of the ontology concept similarity matrix

In order to verify whether the strategies of domain information retrieval optimization work or not, we carried out an experiment to analyze the retrieval effectiveness. The experimental environment is Google search engine, Baidu search engine and the audit guidance system which is developed by our laboratory. Take retrieve “person” as an example, in pension insurance domain, concept “person” has some special meanings such as “insured people”, “insured workers” and so on. But concept “person” is a generalized concept in daily life. Fig. 5 shows the retrieval results of “person” by Google. We can find that the retrieval result retrieved by Google is unsatisfactory.

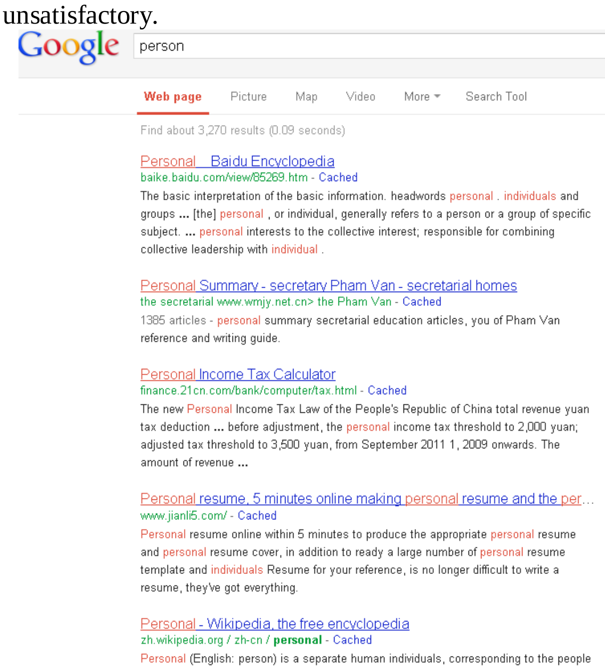

Figure 5. Retrieval results by Google

We then take “person” as a retrieval word, and use the optimization strategy to optimize the retrieve process. Through the pension insurance ontology, it can get concept “insured people”, “insured workers” by the relation of equivalence, and it also gets concept “enjoying treatment people”, “insured state” by relation “subClassOf”. Through the concepts and relations, we can get 4 groups of retrieval words, such as “person + insured people”, “person + insured workers”, “person + insured state”, “person + enjoying treatment people”.

Figure. 6 shows the retrieval results retrieved by 4 groups' retrieval words. As shown in Fig. 6, the first 40 results are related to pension insurance domain, and it is in line with retrieval intentions of the people.

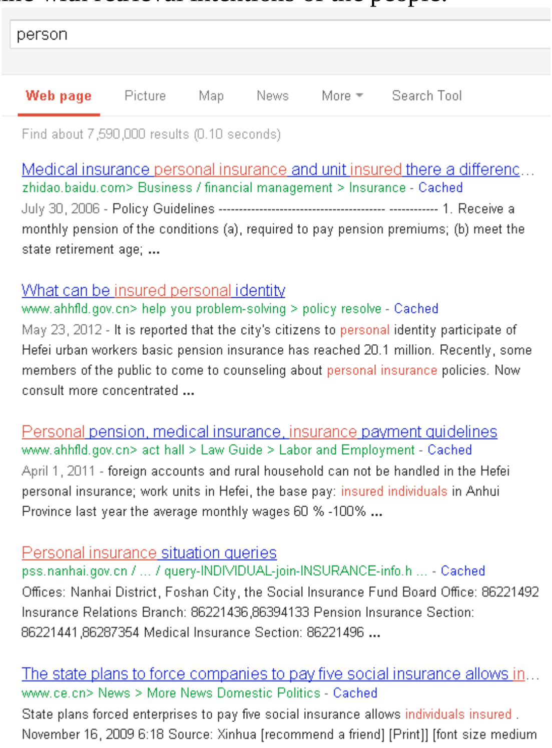

Figure 6. Retrieval results after optimization

TABLE I. COMPARISION OF THE RETRIEVAL RESULTS BY DIFFERENT SEARCH ENGINE

	person	units
Google	0/3,760,000,000	0/ 1,150,000,000
Baidu	0/ 100,000,000	0/100,000,000
before optimization	17/40	19/40
after optimization	40/40	40/40

Table 1 compares the experimental results between "person" and "units" which are taken as retrieval word individually through different search engines and retrieval systems. The result is in the form of "num1/num2", num1 means the num of useful results at the first 40 results; num2 means the num of all the retrieval results. The audit guidance system only shows the first 40 results, so the num2 is 40 at the audit guidance system. Because the common search engine is for general knowledge, it is unsatisfactory in order to get domain knowledge. Experimental results show that the strategy of domain information retrieval optimization is more effective than before.

ACKNOWLEDGMENT

This research was supposed by the "National Science & Technology Pillar Program" under the grant number of 2009BAH42B02 and 2012BAH08B02.

REFERENCES

[1] N. Guarino, C. Masolo, and G. Vetere. OntoSeek. "Content Based Access to the Web". *IEEE Intelligent System*, vol. 14, no.3, pp. 70- 80, 1999.

[2] C. G. Wu, W.P. Jiao, Q. J. Tian, et al. "Information Retrieval Server based on Ontology and Multi-Object". *Research and Development of Computer*, vol. 38, no. 6, pp. 641- 647, 2001.

[3] Y. B. Pan, Y. X. Chen. "Natural Language Understanding based on Ontology". *Computer Technology and Automation*, vol. 22, no.4, pp .71- 74, 2003.

[4] L. J. Liao, Y. D. Cao, X. Y. Li. "Information Extraction based on Ontology". *Engineering and application of computer*, vol. 23, no.4, pp. 110- 113, 2002.

[5] G. A. Miller. "WORDNET: A Lexical Database for English". *Communications of the ACM*, vol. 38, no.11, pp. 39- 41, 1995.

[6] Google. http: / / www. google. com.

[7] Baidu. http: / / www. baidu. com.

[8] S. Y. He. "The theory and practice of the conceptual information retrieval". *Journal of the China Society for Scientific and Technical Information*, vol. 14, no.2, pp.134- 141, 1995.

[9] T. R. Gruber. "A Translation approach portable ontology specifications". *Knowledge Acquisition*, vol. 5, no.2, pp. 199- 220, 1993.

[10] Graciela Brusa, M. Laura Caliusco and Omar Chiotti. "Towards ontological engineering: a process for building domain ontology from scratch in public administration". *Expert Systems*, vol. 25, no.5, pp. 484-503, 2008.

[11] M. Li, L. Yuan, and W. Du, "An Adaptive Motion Model and Multi-feature Cues Based on Particle Filter for Object Tracking," *Journal of Multimedia*, vol. 7, no. 5, pp. 364–371, 2012.

[12] N. Mallesh and M. Wright, "An analysis of the statistical disclosure attack and receiver-bound cover," *Computers & Security*, vol. 30, no. 8, pp. 597–612, 2011.

[13] Jairo Francisco de Souza, Sean W.M. Siqueira and Rubens N. Melo, "Applying Ontology Similarity Functions to improve software agent communication", *Models for Capitalizing on Web Engineering Advancements: Trends and Discoveries*, 12, pp. 43-57, 2012 .

[14] Alexander Maedche and Steffen Staab, "Learning Ontologies for the Semantic Web", *Intelligent Systems*, IEEE, vol. 16, no. 2, pp. 72- 79, 2001.

[15] C. Su and H. Li, "An Approach to Interactive Affective Learning Algorithms," *Journal of Computers*, vol. 7, no.10, pp. 2425–2431, 2012.

[16] J. Lin, J. Zhao, and L. Xue, "Non-linear Multi-attribute Based Online Auction Bidding Model and Platform," *Journal of Computers*, vol. 7, no. 10, pp. 2390–2396, 2012.

[17] J. Zhang, F. Xiong, and D. Zhang, "Steganalysis for LSB Matching Based on the Dependences Between Neighboring Pixels," *Journal of Multimedia*, vol. 7, no. 5, pp. 380–385, 2012.

[18] Anna Formica, "Ontology-based concept similarity in Formal Concept Analysis", *Information Sciences*, 2006, vol. 176, pp. 2624–2641.

[19] Christos Andronis, Anuj Sharma and Vassilis Virvilis, "Literature mining, ontologies and information visualization for drug repurposing", *Briefings in Bioinformatics*, vol.12, no. 4, pp. 357-368, 2011.

[20] E. M. N. Goncalves, "Specifying Knowledge in Cognitive Multiagent Systems Using a Class of Hierarchical Petri Nets," *Journal of Software*, vol. 7, no. 11, pp. 2405–2414, 2012.

[21] M.-T. Chu and R. Khosla, "Strategic Knowledge Workers Features in the Context of Communities of Practice," *Journal of Software*, vol. 7, no. 11, pp. 2389–2399, 2012.

[22] P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a taxonomy", *Proceedings of the*

14th International Joint Conference on Artificial Intelligence. Montereal, Canada, pp. 448-453, 1995.
[23] E. Agirre, G. Rigau. "A Proposal for Word Sense Disambiguation Using Conceptual Distance", Proceedings. of the 1st International Conference on Recent Advances in NLP. Tzigov Chark, Bulgaria, 1995.
[24] C. Leacock, M. Chodorow. "Combining Local Context and WordNet Similarity for Word Sense Identification", *Computational Linguistics*, vol. 24, no. 1, pp. 147-165. 1998.
[25] S. J. Li, J. Zhang, X. Huang. "Semantic Computation in Chinese Question-answering System", *Journal of Computer Science and Technology*, vol. 17, no.6, pp. 933-939, 2002.
[26] Q. Liu, S. J. Li. "Semantic Similarity Calculation Based On China National Knowledge Infrastructure". *International Journal of Computational Linguistics & Chinese Language Processing*, vol. 7, no.2, pp. 59-76, 2002.
[27] L. J. Zhu, L. Tao, H. Liu. "Concept Similarity Calculation in Domain Ontology", *Journal of South China University of Technology* (Natural Science Edition), vol. 32(supplement), pp. 147-150, 2004.
[28] G. Huang, Z. R. Zhou. "Study on Concept Similarity Calculation Based on Domain Ontology", *Computer Engineering and Design*, vol. 28, no. 10 pp. 2460-2463, 2007.
[29] Z. P. Zhang, H. L. Zhao, Z. H. Zhang. "Concept Similarity Computation Based on Ontology". *Computer Engineering*, vol.35, no. 7, pp. 17-19, 2009.
[30] X. Y. Du, M. Li and S. Wang, "A Survey on Ontology Learning Research", *Journal of Software*, vol. 17, pp. 1837-1847, 2006.

Qingsheng Wan, male, born in 14 February 1975, received the B.S degree in Agricultural Mechanization and Electrization from Northeast Agricultural University, Harbin, China, in 2000, and received the Master degree in Agricultural Mechanization and Electrization from Northeast Agricultural University, Harbin, China, in 2007. He is currently working towards the Ph. D in Harbin engineering University, Harbin, China. His research direction interests include artificial intelligence, natural language process and so on.