IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

REVIEW OF COMPUTER ENGINEERING STUDIES
Vol.2, No.3, 2015, pp.13-16
http://dx.doi.org/10.18280/rces.020303

SEMANTIC RETRIEVAL OF KEYWORD BASED ON ONTOLOGY REASONING

Li Jinyu[1], Gao Yi[1], Li Kun[1] and Yan Hongcan[2*]

[1] Yi Sheng Innovation education College, North China University of Science and Technology,
46 Xinhua Road, Tangshan, Hebei, China;
[*2] College of Science, North China University of Science and Technology,
46 Xinhua Road, Tangshan, Hebei, China.

E-mail: yanhongcan@ncst.edu.cn

ABSTRACT

Semantic retrieval is based on the semantic level which is expressed by key phrases to recognize and deal with the user's retrieval request, the knowledge base is critical and foundation for reasoning and knowledge accumulation of semantic search, and the ontology is the basis of the knowledge base. As for the user's demand for natural language search, a method of the natural language keyword analysis and extension that based on domain ontology is proposed. Using the maximum mechanical matching method to break out the key phrases form the user's retrieved language. And identify the bit word, the upper word as hypernym, and the next word as hyponym of keywords by ontology reasoning, namely synonyms and synonyms, so as to realize the semantic retrieval that based on keywords. Go through the empirical study for the ontology and the knowledge base of Computer Ethics, it shows that this method can greatly improve the retrieval efficiency of fuzzy query.

Keywords: Domain ontology, Semantic retrieval, Chinese word segmentation, Ontology reasoning.

1. CONSTRUCTION OF DOMAIN ONTOLOGY

In the traditional search engine based on keyword matching, semantic retrieval can greatly improve the retrieval recall [1]. At present, the Web search engine can realize the fuzzy and intelligent retrieval by keyword split and combination. But it still cannot solve the problem of synonyms and antonyms. It can search all things and the quality is not high. It is difficult to meet the requirements of people's search in the network information age. Then people use a dictionary, thesaurus and a dictionary of intelligent retrieval. But semantic retrieval is still not achieved. TimBerners-Lee proposed the semantic web which can be a good solution to this problem. Knowledge representation and semantic extension of intelligent retrieval is the key to realize. In this paper, the ontology can provide the knowledge representation and semantic extension in the retrieval. So the application of the concept of the semantic web can improve recall [2-3].

Domain ontology is the basis for the semantic retrieval. Domain ontology construction is a professional complex process that requires the help of professional knowledge in the field, in the case of full knowledge of core knowledge and core concepts in the field, it can adopt a bottom-up approach to achieve constructed. At the same time, many researchers have proposed some s that are beneficial to the structure of the ontology. One of the principles is influential, which is put forward by Gruber in 1995:

(1) Clarity: Ontology must be effectively defined the meaning of the term;

(2) Coherence: Ontology should be consistent. It must support the reasoning that the definitions are consistent with others;

(3) Extendibility: Ontology can provide a Shared vocabulary. The task of the share can be expected to provide the corresponding concept foundation;

(4) Minimal Encoding Bias: Ontology should be at the level of knowledge. It has nothing to do with certain symbols and codes;

(5) Minimal Ontological Commitment: Ontology should be Minimal. It can satisfy the specific knowledge sharing [4].The principle should be tried to follow in the process of building. Its building process is basically the following:

First, clarifying the scope. The purpose of the structure of domain ontology is based on the sharing of knowledge in a field. At the build time, this ontology should be clearly targeted for future applications in the fields, scope and purpose of its use, the situation of comprehensive understanding in the field in favor of domain ontology; second, learn from the existed ontology. With the development of information retrieval, the research of ontology is also deepening and expanding, so when constructing a domain ontology, the study should learn from the previous portion of the ontology that has been studied. Based on the existed domain ontology, doing the transformation can accelerate the work progress; third, list the concept and establish the frame of ontology according to their relationships. According to the areas which come down to find the related ontology concepts. In order to get more accurate key concepts, domain ontology should confirm many times. And domain ontology should continue to be improved [5]. Fourth, based on the characteristics of the concept and property to be divide and clear the logical relationships between concepts, to get a hierarchy concept map with clear layer of structure, and establish a frame for ontology; fifth, coding and the situation. Select the appropriate language to describe ontology, encoding and formalization constructed domain ontology; Sixth, further improved. The knowledge of domain boundaries are uncertain, there is often a cross and general fields, so it is need to constantly expand in the process of structure and to be the new ontology [6].

In the build process ,domain ontology will go through these stages substantially, but the build process is in the dynamic, sometimes back and forth constantly modified in several stages, so the study should be based on users' own accumulated and extended knowledge to be modified and responded in a timely manner during the build process .

2. CHINESE WORD SEGMENTATION

Word segmentation of natural language is the most basic in Chinese information processing, whether in machine translation or information retrieval or other applications, if it involves Chinese, they are inseparable from the Chinese word segmentation, so Chinese word segmentation is the intelligent retrieval first problem [7].

2.1 Chinese words segmentation method

Currently the most mature Chinese words mainly positive maximum matching method; reverse maximum matching; word segmentation based on statistical methods. This paper use maximum matching method for the user's query language word processing [8].

2.2 Java classes and core code for words segmentation

Maximum matching refers to a dictionary as the basis, take the longest word in the dictionary for the first time to scan the text, then successive reduce the number of scanning string in the dictionary. To improve sweep efficiency, can also according to count how many design multiple dictionary, and then according to the number of words from different dictionaries. For example: the longest word in the dictionary as "the People's Republic of China" a total of seven Chinese characters, the maximum matching initial word for seven Chinese characters. Word for word, and then, the corresponding word to look up in the dictionary.

The following is the code for Maximum Matching.

```
importjeasy.analysis.MMAnalyzer;
importjava.io.StringReader;
//The maximum length of word matching window to match the code 12
public void word Segment(String Sentence)
//Input a string to be treated as an object
{
intsenLen = Sentence.length() ;
  //Calculate their characters in length
int i=0, j=0;
//i control the initial position of the variable, j control the end position of the variable
int M=12;   //pString word;
 //The word that need to do comparison in thesaurus
booleanbFind = false;
  //Determine whether it is variable lexicon of words
While (i <senLen)  //If the length of i less than the length of this sentence ,i will enter circulation
{
int N= i+M<senLen ? i+M: senLen;
//If i'M <senLen is true on the implementation of i'M, if it is false to perform senLen assign the result to N
bFind=false;    //Suppose the phrase is not taken out of the lexicon of the words natural language processing
for(j=N; j>i; j--)     //Forward Maximum Matching
{
word = Sentence.substring(i, j);
   //Interception string. If there are matched to the appropriate word, it will be in this loop until the loop is over
if(dic.Find(word))
{
System.out.print(word + " ") ;
  //If the word forms the lexicon, print out
bFind=true;
i=j;   //If you find the word in the lexicon interception t the end of the assigned position j to i
//System.out.println(i);
break;    //Out of the for loop
...}}
if(bFind == false)
 //If word is not form the lexicon, do this in the following way
{
word = Sentence.substring(i, i+1);
System.out.print(word + " ");
++i;  //Control reading order
}}
System.out.println();
}
```

Such as "we play in a wildlife park" the natural language, positive maximum matching method, the final segmentation result is: "we/in/vivid/content/garden/play", among them, 2 words dictionary words, not a dictionary word is 1.

3. WORD MEANING EXPANSION BASED ON ONTOLOGY REASONING

In order to realize the semantic extension of keywords, we must find out the key words with a word, hypernym, and under a word (hyponym) through the ontology reasoning.

3.1 Rules of ontology reasoning: SWRL（Semantic Web Rule Language） rules

Realizing participle extension must follow certain rules in the process of ontology building [9]. Although the OWL has strong ability of knowledge description, the ability mainly come from the inference based on class and relationship. If the knowledge is difficult to express by class, the OWL is hard to describe.

In order to solve these problems, SWRL rules are used to describe the user-defined rules. SWRL is based on OWL. It can be integrated with OWL definition of classes and properties. Then the operation will be conducted on the basis of description logic rules. SWRL rules as follows:

Rule1: is same as (? X, ? y) ∧has label (? X, ? z) →has label (? Y,? z)

An instance of domain ontology: Computer crime, computer intellectual property and computer society problem three are the same subclass relations of computer ethics.

Rule2: is same as (? X, ? y) ∧is part of (? X , ? z) →is part of (? Y,? z)

An instance of domain ontology: Public privacy and personal privacy is a subclass of same level. If privacy is a subclass of computer privacy protection, it infers that public privacy is also a subclass of computer privacy protection.

Rule3: A (? X) ∧ B (? y) ∧is same as (A, B) →is same as (? Y,? z)

An instance of domain ontology: The net has similarities with Internet addiction. Then social problems for the computer can be deduced the parent class.

Rule4: subclass of (? X, ? y) ∧ is part of(? X, ? z) → is part of (? Y,? z) …

An instance of domain ontology: Software piracy problem is the subclass of intellectual property rights. Network intellectual property and intellectual property are the subclass parent class relations. So it can infer that there is a common parent class relationship between intellectual property rights and software piracy problem [10].

3.2 The knowledge reasoning in the ontology

This paper choose OWLDL based on description logic language to build ontology. So the semantic retrieval is a kind of semantic reasoning which is based on description logic knowledge. The knowledge base is composed of two parts: Tbox and Abox. Tbox is a set of terms of the related concepts and relations. It is used to describe the general properties of the concepts and relationships in the application domain. Abox is a collection of related individuals. It can describe the relation between individual and individual in application domain extension knowledge [11].

The function of the logic inference engine is to provide a reasoning service for the knowledge base (KB). Then the implicit information inside the surface information is discovered.

The problem of reasoning in Tbox mainly includes four aspects: Satisfiability, Subsumption, Equivalence and Equivalence.

The consistency detection is a condition in ABox reasoning. If an A is consistent with the T in ABox, then there is a A and T common model I. Reasoning in ABox also includes an example test, that is, the given individual and the concept of matching, test whether there is an instance relationship between them.

In this paper, we first match the keyword parameters with the concept of ontology. TBox is a collection of concepts in the description logic, which usually contains a hierarchy of concepts. Terminology axiom consists of two forms: inclusion axiom and equivalent axiom. In order to realize the concept classification, the inclusion axioms are used to define inclusion relations. We apply equivalence axioms to define complex concepts in the field.

Then the SWRL rules and the SWRL merge that already exists in the knowledge base are merged. The rule will be loaded in the inference engine, and it is also required to load the information from the ABox to complete the reasoning.

4. REALIZATION OF SEMANTIC RETRIEVAL

Semantic retrieval is based on domain ontology and segmentation, and the query process based on keyword expansion is as follows:

Step 1: According to the query request, using the forward maximum matching method to resolve the key words;

Step 2: Get the appositive, superordinate and subordinate words query of keywords with Jean-bit words;

Step 3: Extract extended term from the knowledge base ,and return search results.

In this paper, we will take the computer ethics as an example to carry out the semantic retrieval.

First build domain ontology of Computer Ethics.

According to the contents of the research, the concept of computer ethics is listed, and the semantic tree of concept hierarchy is sorted out, as shown in Figure 1.

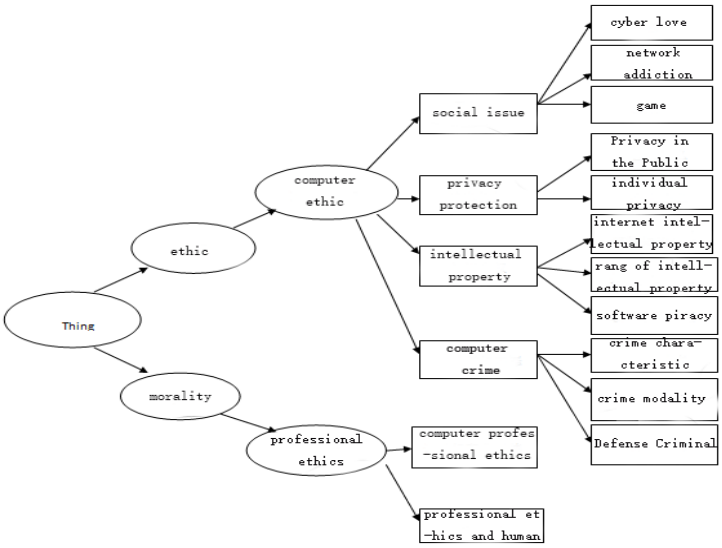
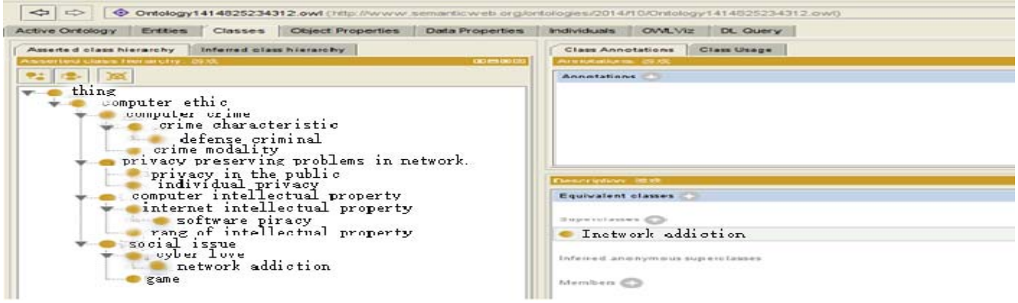


Figure 1. Semantic tree of background knowledge

Using protégé ontology software to build ontology, there are four file types in protégé, this paper chooses OWL RDF files, The exported ontology files support OWL ontology file viewing and saved, read and modify in editing tools, and provide the knowledge resources described for knowledge reasoning and retrieval. The ontology of computer ethics created by protégé is shown in Figure 2.



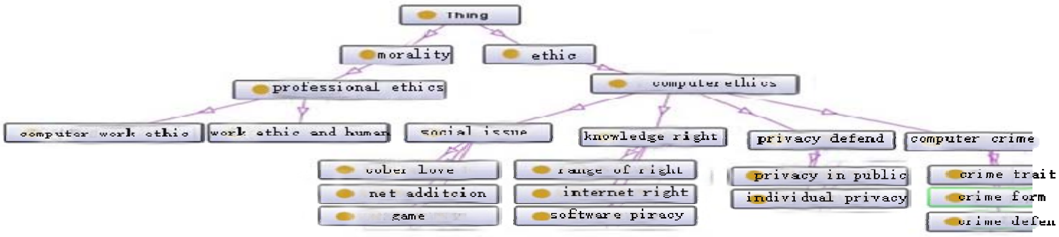


Figure 2. The ontology construction of Computer Ethics

Then the keyword semantic expansion: For example, the need to query the statement is "With the rapid development of computer networks, computer crimes are also increasing in China.", According to the maximum mechanical matching method, the query sentence is segmented as: "With", "the rapid development", "computer networks", "computer crimes", "increasing" and "n China". Extracting key words for "computer crime". Reasoning by Jean inference engine. According to the information of the established of the related concepts and relationships of the terms of the TBox and the description of the relevant individual instantiation set ABox deduce the apposition "privacy" and the upper word " computer ethics" and the next word" Forms of crime".

5. CONCLUSIONS

This paper used seven steps to construct the domain ontology of computer ethics as the example, used the maximum mechanical matching method to decompose the key words in the language. And then, by using the software of protégé and, to construct the domain ontology of computer ethics. To find out the key words, affix superordinate and subordinate verbs by ontology reasoning, so as to realize the semantic retrieval of key words. In this paper ,the retrieval of word segmentation and semantic language extension, laid the technical foundation for the realization of intelligent retrieval, the next key step is the problem of ranking search results, will be and the Rank algorithm to calculate the similarity between extended words and key words combination, realize the relevance ranking search results.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (61370168) and the National Natural Science Foundation of Hebei (F2014209238).

REFERENCES

1. Ma Sen, Zhao Wen, Yuan Chongyi, Zhang Shikun, Wang Lifu, Research on Key Technologies of Semantic Retrieval Based On Rule Reasoning [J], *Electronic Journal*, 2013, 41 (5): 977-981.
2. Meng Hongwei, Zhang Zhiping, Zhang Xiaodan, Research on the Model of the Intelligent Retrieval Model Based on Domain Ontology [J], *Information Journal*, 2013, 32 (9):180-184.
3. Song Wei, Zhang Ming, Concise Tutorial [M], Beijing: Higher Education Press, 2004.
4. Yu Jing, Wu Guoquan, Based On Domain Ontology Of Government Information Retrieval System [J], *Chinese Information Science*, 2010, 30 (6): 1664-1667.
5. Dong Hui, Tang Min, The Application of Semantic Retrieval in the Environment of Web 2.0, 2011, [J], *Chinese Journal of Library*, 2, 37 (192): 115-119.
6. Da Kangta, Orbost, Smith, The Semantic WEB, XML, WEB Services, the Future of the Knowledge Management Guide Yue GAO Fengyi [M], Science and Technology of China Press, 2009.3.
7. The King Peak, Hua-Fang Wang, Intelligent Development Trend of Digital Library Information Retrieval Technology [J], *Modern Intelligence*, 2008.28 (11), 93-93-13.
8. Li Fei, Zhao Shixia, Semantic Information Retrieval Based on Ontology Technology Research [J], *Information with Computer*, 2010 (6): 106-107.
9. Nicola Guarino, Daniel Oberle, Steffen Staab, What Is an Ontology? [J], *Handbook on Ontologies*, 2009.05 (22): 1-17.
10. Liu Chunmao, Mi Guowei, We B 2. 0 for Society under the Environment of Semantics Information Constructing Knowledge [J], *The New Recognition of Affection to the Theory and Practice*, 2010 (2): 89-92.
11. Xu Zheng, Zheng Quan, Video Semantic Retrieval System Based on Ontology Research [J], *Journal Of Computer Applications*, 2012 (3), 836-837.