Reuse of Chinese Domain Ontology for the Restricted Domain Question Answering System

Jie Liu, Yun Ma and Liming Luo
College of Information and Engineering, Capital Normal University, Beijing, China
Email: liujxxxy@126.com, 195307728@qq.com, luolm@cnu.edu.cn

Zhengtao Yu
School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming Yunnan Province, China
Email: ztyu@hotmail.com

***Abstract*—Expressing knowledge by ontology is favorable for reuse and reasoning of knowledge. The structure of domain ontology for question answering system (QA) was formally defined, and the relations between ontology elements and semantic analysis of question and answer extraction were illustrated. The method of reusing Chinese ontology was proposed in the same domain, and the method can construct new initial ontology after extracting and classifying new instances based on the existing ontology using support vector machine (SVM) and reuse the other elements of the existing ontology. By the experiment of ontology reuse in certain medical domain, the average F-value of extraction and classification of instances reached 82.8%, and it verified validity of the method. The proposed method has significance for fast constructing Chinese ontology for QA in same domain.**

***Index Terms*—question answering, ontology, ontology reuse, knowledge representation**

I. INTRODUCTION

QA not only can let askers use the natural language to question, but also can return succinct, accurate answers to askers, not using some relevant WebPages. QA can be divided into restricted domain QA (RDQA) and common domain QA according to the range of question and answer. With the support from TREC (Text Retrieval Conference), great progress have been made in common domain QA [1]; In RDQA, QA systems in English, Japanese and German have already been applied. In China, many research institutions have put considerable effort into the research of Chinese QA, such as the expert system about relationship of Institute of Computing Technology of the Chinese Academy of Sciences [2], the campus navigation system of Tsing-Hua University [3], the QA system based on frequently asked questions of Harbin Industry University [4], the bank domain QA system of Beijing Institute of Technology [5], etc.

Ontology has been widely used in the field of information, such as the references [6-8]. In addition, ontology has been widely applied in QA systems of western language [9]. At present, Chinese QA systems based on ontology also become more and more, such as the references [10-12]. But in these studies, only the reference [12] explicitly gave the structure of ontology suit for QA.

With the extensive use of ontology, automatically constructing ontology is becoming more and more important. In western language, there have existed some systems and tools of ontology reuse. The reference [13] achieved efficient ontology reuse through the process of ontology module extraction in order to reuse those concepts and relations; the reference [14] focused on how vocabularies (concepts and relations) can be extracted and integrated into the target ontology in the domains of e-Recruitment and medicine; the reference [15] proposed a set of tasks which are relevant to ontology reuse and formalized them as reasoned problems aiming at reusing the general properties and relationships between the notions, and so on. In China, there are few studies about the method of ontology reuse, let alone ontology reuse for QA. Although some methods of ontology reuse have existed, they all focus on the reuse of concepts, relations and properties of conventional ontology, not on that of Chinese QA. So it is necessary that researching Chinese ontology reuse for QA. The proposed method in the paper focuses on the reuse of special ontology used in QA; besides, the method not only can reuse concepts, relations and properties of ontology, but also can extract instances from the Chinese domain texts and integrate these instances into the new ontology. In this paper, ontology reuse means that extracting new ontology elements from the Chinese domain texts using the existing ontology and inheriting part of elements of the existing ontology. In same domain, ontologies are different mainly because their instances are different from the elements associated

Manuscript received December 17, 2011; revised February 10, 2012; accepted February 20, 2012.

This work is supported by The Natural Science Foundation of Beijing City (No.4123094), National Nature Science Foundation of china (No.60863011), The Science and Technology Project of Beijing Municipal commission of Education (No.KM201110028020, No.KM20101028019). Beijing Key Construction Discipline "Computer Application Technology.

Corresponding author is Jie Liu. Email:liujxxxy@126.com

doi:10.4304/jcp.7.11.2684-2690

with the instances. For example, different hospital ontologies consist of different instances such as doctor's name, therapeutic disease, medical cost and instrumental name; but they consist of same concepts such as doctors, diseases, equipments etc., same relations and attributes. Therefore it is important that studying the method of rapidly constructing ontology for QA.

In this paper, we will introduce the ontology structure proposed in the reference [12]. Besides, in the same domain, it will be proposed that a method of quickly constructing new ontology for QA using the existing resources.

II. QA AND ONTOLOGY REUSE

Fig.1 shows the relationship among QA , ontology and ontology reuse.

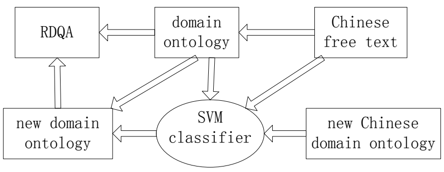

Figure 1 Relationship among QA , ontology and ontology reuse

In Fig.1, domain ontology has be acquired from Chinese free text. SVM classifier is trained by domain ontology and Chinese free text, and then SVM Classifier acquires new instances from new Chinese free text and allocates new instances to corresponding ontology concepts, and lastly new instances and part of inherited ontology elements form the new domain ontology.

It should be noted that the ontologies and free texts belong to same domain.

The structure of this paper is as follows. The ontology structure for RDQA is introduced; the relations between the defined ontology structure and RDQA are analyzed in detail. Then the methods about how to extract features will be introduced in detail. Later experiments and results will be presented. Last conclusions will be given. In this paper, the resources of examples and experiments are from the medical field.

III. ONTOLOGY OF RDQA

A. Ontology Structures of RDQA

Ontology structures may be different for different purposes. Ontology structure suit for RDQA is defined according to the demand for semantic analysis of question and answer extraction. Ontology structure of RDQA is given as follows [12]:

Definition 1 Ontology $O=(C,\ H^C,R,\ rel,I,\ p,\ p_c,v,r,\iota_C,\iota_R,\iota_p,\iota_v,\iota_{Iv},\mathrm{A})$, here

(1) C denotes the set of concept names;

(2) H^C denotes conceptual hierarchy or taxonomy, $H^C \subseteq C\times C$, and $H^C(C_1,C_2)$ means that C_1 is the sub-concept of C_2 ;

(3) R denotes the set of non-hierarchical relation (or object relation) names;

(4) $rel: R\rightarrow C\times C$, is a function and denotes the non-hierarchical relation between concepts;

(5) I denotes the set of instances;

(6) P denotes the set of attribute names of concept;

(7) p_c denotes the set of concrete content of P;

(8) v denotes the set of domain event names;

(9) r denotes the set of role names of domain events;

(10) r_c denotes the set of concrete content of r.

(11) Function: $\iota_C: C\rightarrow \kappa_1(I)$ means corresponding relation between the concept C and the instance I;

(12) Function $\iota_R: R\rightarrow \kappa_2(I^+)$ means embodiment of rel in I and denotes a concrete non-hierarchical relation;

(13) Function $\iota_p: p\rightarrow \kappa_3(I)$ means corresponding relation between the instance I and the attribute P;

(14) Function: $\iota_{vr}: vr\rightarrow \kappa_4(I)$ means corresponding relation between the role r, the event v and the instance I;

(15) A is called as the set of axioms, and is propositions about the elements of the set $C\cup R\cup I\cup P\cup v\cup r$, and is the correct hypothesis without proving, commonly according with certain logic language such as description logic.

Definition 2 Let Γ be logic language, and ontology O expressed by Γ, which is denoted by (O, Γ), and the set of axioms A of O accord with axiom system of Γ. In this paper, the ontology KB of QA system is expressed by OWL language based on description logic.

B. Ontology Segment

Fig.2 is the segment of medical domain ontology which is translated from Chinese ontology.

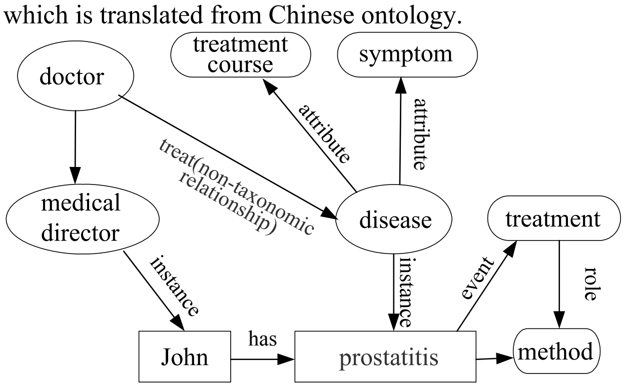

Figure 2 Ontology fragment of medical domain

In Fig.2, "John" and "prostatitis" are two instances, and "doctor" and "disease" are two classes, and "medical director" is the subclass of "doctor", and "treatment" is a non-taxonomic relationship between "doctor" and "disease", and "symptoms" and "treatment course" are two properties of the class. The triple (John has prostatitis) is embodiment of an instance of the non-taxonomic relationship (patient has disease). The triple (prostatitis event treatment) shows that "treatment" is an event of "prostatitis". The triple (treatment role method) shows that "method" is a role of "treatment".

C. The Relation Between Ontology Structure and QA System

In QA, ontology structure mainly serves semantic analysis of question and answer extraction. The three are inseparable. The relations among them can be abstractly exemplified by the following examples. Here, QAR denotes the results of question analysis, and QT denotes focus of question, and C denotes classes, and I denotes instances, and AE denotes answer extraction. The concrete methods of question analysis and answer extraction will be not described here.

Question one: Which famous doctors is in your hospital?
QAR={QT=I, C=doctor}
AE: lookup the instances I by the function $\iota_{\text{doctor}} \rightarrow \kappa_1(I)$.

Question two: Which diseases can hysteroscopy treat?)
QAR= {QT=I, C=disease, I=hysteroscopy, R=treat}
AE: lookup the instances I of the class "disease" by the function $\iota_{\text{treat}}:\text{treat} \rightarrow \kappa_2(\text{hysteroscopy}, I)$.

Question three: What are the symptoms of prostatitis?
QAR={QT=pc, I=prostatitis, P=symptom}
AE: pc can be got by the function $\iota_{\text{symptom}}:\text{symptom} \rightarrow \kappa_3(\text{prostatitis})$

Question four: What is the method of treatment of prostatitis?
QAR={QT=rc, I=prostatitis, v= treatment, r= method }
AE: rc can be got by the function $\iota_{\text{treatment, method}}:(\text{treatment} \cdot \text{method}) \rightarrow \kappa_4(\text{prostatitis})$.

From analysis above, we can see that the defined ontology structure is in line with the requirements of RDQA. In reference [12], the experimental results have also verified the structural role.

IV. Ontology Reuse in Same Domain

From the definition 1 we can see that the ontology elements such as C, HC, R, rel, P, v, r, l_p, l_{vr} can be reused, I, l_C, p_c, r_c, l_R need to be learned. This paper only studies the method of extracting I,l_C. This paper presents a new method of extracting I,l_C using SVM [16] based on the existing Chinese ontology and texts of resource in same domain.

In the proposed method, the class of SVM classifier means the concept of ontology; the sentences including instances are used as the training resources of SVM classifier; the sentences including candidates are used as the studying resources of SVM classifier; the new instances will be acquired from the new free texts and be classified. In this paper, instance classification means allocating new instances to corresponding ontology concepts.

The performance of SVM application is mainly determined by the quality of selected features if algorithm is given. Appropriate features mean those features related to classification and can make SVM have good discerning ability. This paper selects some features which include shallow features such as word, POS and chunk, and semantic features such as named entity, synonym and semantic representation of ontology in sentences. These features can fully express semantics of domain texts.

A. Selection of Syntax Features

(1) Word and POS: Word and POS are the basic units of sentence. Therefore word and POS are chosen as basic features of classification. This paper adopts a tool called ICTCLAS, which is developed by Chinese Academic Science (CAS), so as to segment word and annotate POS. The standard used for annotation of POS is established by Peking University.

(2) Chunk: Analysis of chunk is based on the theory of shallow semantic analysis, which is becoming a significant part of the semantic analysis. Chunk is an important element in a sentence. Therefore, chunk is chosen to be added in features of classification. The definition of chunk class refers to Chunking Shared Task proposed by Conference on Computational Natural Language Learning-2000 (CoNLL-2000), in which 12 chunk categories are defenined such as noun chunks (NC), verb chunks (VC), adverb chunks (ADVC), adjective chunks (ADJC), quantity chunks (QC), "的"(de) noun chunk (DNC), preposition chunk (PC) etc. Identification of chunk uses the chunk model based on maximum entropy set up by CAS [17].

Here are examples about annotation of word segmentation, POS and chunk:

Chinese sentence S1:前列腺炎是难以彻底治愈的泌尿生殖系统疾病.(Prostatitis is a disease in urogenital system, which is treatable but difficult to get completely healed.)

Word segmentation and POS annotation: 前列腺炎/n 是/v难以/d彻底/a治愈/v的/u泌尿/v生殖/v系统/n疾病/n /w.

Chunk annotation: [NC前列腺炎/n][VCC是/v][ADVC难以/d][VCC彻底/a治愈/v][DNC的/u][NC泌尿/v生殖/v系统/n疾病/n] /w.

B. Selection of Semantic Features

Semantic features include the semantics of words and ontology. The semantics of words includes named entity (NE) and the synonyms of words (SC). The semantics of ontology includes class of instance, object relation, property of class and the synonyms of ontology vocabularies (SO).

(1) Named entity (NE): NE means noun phrase having exact meaning, which plays an important role in distinguishing the class of instance. For example, the instances of class "doctor" commonly are some person's name. In the sentence S2 "约翰教授擅长治疗前列腺炎 (Professor John is good at treating prostatitis)", person's name is important for judging whether "约翰" is the instance of class "doctor". After NE is identified, the result of annotation S2 is as follow:

S2: [NAME 约翰]教授擅长治疗不孕症.

Algorithm of named entity identification is Chinese named entity identification based on HMM model, which is specified in reference [18]. After being trained by corpus of Chinese People's Daily published in Jan. 1998, the algorithm showed high accuracy of identification.

(2) Synonyms of words (SW): Many words have synonyms, so the synonyms of words in RS and CS are used as one feature of classification. The TongYiCiCiLin (A Chinese Dictionary of Synonyms) [19] is use as the tool of synonym extension. For example, the word "治疗" (treatment) in S2 has eleven synonyms in TongYiCiCiLin such as "医治", "医","治", "看", "疗", "治病", "看病", "医疗", "诊疗", "诊治" and "临床". These synonyms will be added in feature representation of the candidate "John" if "John" is a candidate.

(3) Semantic features of ontology

In ontology, the elements related to an instance commonly include class (concept) of instance (CI), object relation (OR), class (concept) attributes (CA), event of instance (EI), and role of event (ER). Above Fig.2 is a fragment of certain hospital ontology which shows the various elements have the close semantic relationship with "prostatitis". If these elements appear along with the candidate instance in the same sentence, then it plays a strong role to determine whether the candidate is a real instance.

(4) Synonyms of ontology vocabularies (SO): Ontology vocabularies are critical semantics of classification, so the synonyms of ontology vocabularies will specially be added in feature representation.

C SVM Feature Coding and Extraction of Feature Value

The resources for classifier training are the existing ontology and the resource texts from which the existing ontology was obtained. Sentence in the resource texts involving with instance is called resource sentence (RS). The resource for extracting new instances is the new Chinese free texts. Unlisted word or noun in new texts which is not included in the existing ontology is called candidate. Sentence including candidate is called candidate sentence (CS). Every instance and candidate respectively corresponds to a RS and CS.

To extract instance by SVM, for every instance and candidate, corresponding RS and CS have to be annotated for SVM training and learning. The standards of annotation and the definitions of various features have been introduced before. All features in SVM classifier must been digitalized, so features extracted should be coded, and then transferred into digital form. The format of feature vector input is as follow: <label> <index1>:<value1> <index2>:<value2>…<indexn>:<valuen>, in which label is the code of corresponding ontology class, index is the code of defined feature, and value is value of corresponding feature item. When feature codes are determined, it needs for every RS and CS to determine the values which correspond to these feature indices.

Encoding feature is supported from the recourses including table of ontology class, dictionary, POS marked set, table of chunk class, TongYiCiCiLin and so on. Related coding resources are defined as: there are *W* words in dictionary, *P* POSs in POS set, *C* chunk classes, *N* named entity classes, *CI* ontology classes, *R* object relations, *G* properties of class, and *M* words in TongYiCiCiLin. The eigenvector codes are showed as Tab.1.

When feature codes are determined, it needs for every RS and CS to determine the values which correspond to these feature indices. Every index is determined by only two values, 0 or 1. For RSs and CSs, after segmenting word, and annotating POS, chunk, named entity, ontological object relation, property of class, it is checked whether relevant index exists or not. If yes, the corresponding indexical value is 1. And if not, that is 0. For words and ontology vocabularies, if they can be found in TongYiCiCiLin, corresponding index is set as 1.

TABLE I.
DIMENSION OF EIGENVECTOR OF MEDICAL DOMAIN

Feature items	Initial number	Last number	Amount
word	1	W	W
POS	W+1	W+P	P
chunk	W+P+1	W+P+C	C
NE	W+P+C+1	W+P+C+N	N
SW	W+P+C+N+1	W+P+C+N+M	M
CI	W+P+C+N+M+1	W+P+C+N+M+CI	CI
OR	W+P+C+N+M+CI+1	W+P+C+N+M+CI+R	R
CA	W+P+C+N+M+CI+R+1	W+P+C+N+M+CI+R+G	G
EI	W+P+C+N+M+CI+R+G+1	W+P+C+N+M+CI+R+G+E	E
ER	W+P+C+N+M+CI+R+G+E+1	W+P+C+N+M+CI+R+G+E+ER	ER
SO	W+P+C+N+M+CI+R+G+E+ER+1	W+P+C+N+2M+CI+R+G+E+ER	M

There are 105716 words in word list by CAS and 77343 words in TongYiCiCiLin. In latter experiments, dimensions of eigenvector are shown as Tab.2. So feature vector by the method above is up to 260513 dimensions. Obviously this figure is large, SVM can process multi-dimension features, but training time is a bit long.

D SVM Training and Instance Extraction Strategy

A method of one-to-one classification is based on the assumption that all classes are independent, not related with each other, treated equally, and all hierarchies are flattened in this method.

While ontology has good hierarchical structure, it should be taken advantage of in classification. Therefore

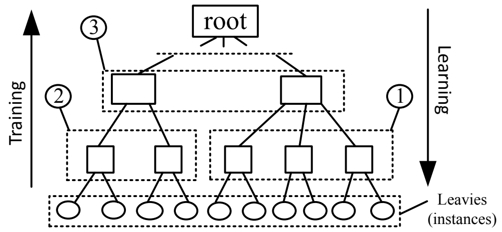

Figure 3 A virtual ontology tree. Square stands for concept, and ellipse for instance

ontology can be transformed into a tree having hierarchical structure by adding virtual nodes in which every node participates in classification except leaf and root nodes.

Training of SVM classifier is started from the concept node at the bottom up to the root in sequence. After current hierarchy is trained, instances of classes belonging to the same parent class are integrated into instances of their parent class. Then the hierarchy containing the parent class is trained. For example, in Fig.3, SVM is trained in sequence:

a) train concepts ① and ②;

b) integrate respectively instances of groups ① and ② as instances of parent classes, and then train ③;

c) train the upper hierarchy in sequence up to root.

Classification of a candidate sample is started from top hierarchy to lowermost concept hierarchy.

TABLE II.
DIMENSION OF EIGENVECTOR OF MEDICAL DOMAIN

Word	Pos	Chunk	NE	SW	CI	OR	CA	EI	ER	SO
105716	25	12	9	77343	22	18	45	11	31	77343

V. EXPERIMENTS AND RESULTS

In order to evaluate the effect of various features on instance extraction and classification, 4 feature sets are defined, including feature set 1 (word), feature set 2 (word, POS and chunk), feature set 3 (Word, POS, Chunk, NE and SW), and feature set 4 (Word, POS, Chunk, NE, SW, CI, RI, CA, EI, ER and SO). Feature sets are based on word firstly, and then gradually they added in POS, chunk, named entity class, synonyms of words, ontology elements and Synonyms of ontology vocabularies.

Since there is no unified training and testing data group for Chinese ontology learning so far, the existing ontology and resource text about one hospital in medical field are taken as training resource, and the resource texts occupy 1.5M and contain 3125 RSs. The texts about other hospitals are taken as learning resource, which occupy 1.2M and contain 2563 CSs.

According to described methods above, the training and learning eigenvectors of SVM classifier are generated.

The accuracies and recall rates of some classes are high in experiments while that of other classes are low, so average accuracies and recall rates are taken to evaluate the performance of the system.

In order to evaluate the effect of different methods of machine learning about instance extraction and classification, four methods of machine learning are chosen, i.e. KNN, NB, DT and SVM. The feature of word is trained and tested alone using different methods and then compared with each other. For the SVM, we use LIBSVM as the tool for classification. RBF function is adopted as kernel function of SVM, which is better than kernel functions of Linea, Polynomial and Sigmoid. Tab.3 shows the results.

TABLE III.
RESULTS OF CLASSIFYING INSTANCE BY WORD FEATURE BY DIFFERENT MACHINE LEARNING ALGORITHMS

Machine learning method(%)	KNN	NB	DT	SVM
Average accuracy(%)	58.1	56.2	59.3	65.1

It can be seen from Tab.3 that the effect of SVM is obviously better than that of other learning algorithms when choosing word as feature.

In order to evaluate the effect of different features on classification, instances are extracted and classified respectively according to 4 feature sets using SVM classifier. Tab.4 gives average accuracies, recall rates and F-scores of instance extraction and classification for different feature sets. The computation formula of F-score is as follow:

$$F\text{-}score = \frac{\left(\beta^2+1\right)\times precision \times recall}{\beta^2 \times precision + recall} \tag{1}$$

Here, β=1.

TABLE IV.
AVERAGE ACCURACY AND RECALL RATE OF DIFFERENT FEATURE SETS

Evaluating standards	Feature set1	Feature set2	Feature set3	Feature set4
Average accuracy (%)	65.1	72.1	75.7	87.6
Average recall rate (%)	61.2	66.7	71.3	78.5
F-Score	63.1	69.3	73.4	82.8

It can be seen clearly from the results of extraction that different feature sets lead to different performances. When considering word feature alone, F-Score is only 63.1%. After adding POS, chunk, named entity and synonyms of words, F-Score is improved to 73.4%. Particularly after adding semantic features of ontology elements, the effects of instance extraction and classification are enhanced and F-Score is up to 82.8%.

To further investigate the effect of instance extraction and classification, we chose randomly 20%, 30%, 40%, 50%, 60%, 70% and 80% of instances in every ontology class respectively as training data in training resource. The learning resource is not changed. The F-Score of 4 feature sets are shown in Fig.4.

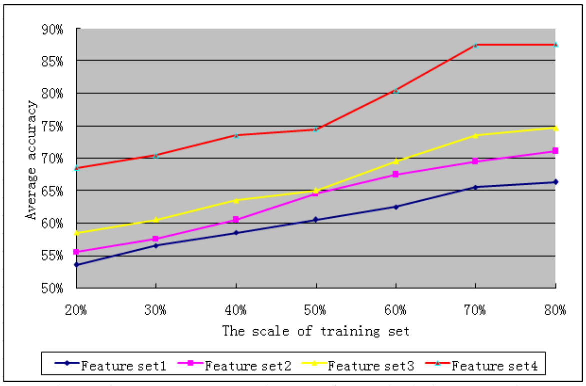

Figure 4 Instances extraction results analysis in proportion

From Fig.4 we can find that the scale of training set has great effect on instance extraction and classification. When the proportion of training set is smaller, the effect is not ideal, for every feature set F-Score is lower. The maximum of F-Score is 63.1% when the proportion of training set is 20% and the feature set 4 is used. The reason is that too little training data results in the classification not obtaining the full feature information. With the proportion of training set increasing, the effect is improved obviously. In figure 4, when the proportion of training set is gradually increased according to 30%, 40%, 50%, 60% and 70%, the slope of classification line shows obvious increasing trend. But when the training set data is increased to certain degree, the performance of classifier gradually tends to be stable. It displays that: in figure 4, when the proportion of training set is 70%-80%, corresponding curves of 4 feature sets tend to be horizontal, and the effect is better, and the performance of classifier at 70% is near that at 80%.

After analyzing the experimental intermediate data, some reasons effecting experimental results are found such as wrong annotation of non-listed words, chunk and NE, part of class words used as candidates, and so on.

After extracting and classifying new instances, and reusing the existing elements such as C, HC, R, rel, P, v, r, lp and lvr, a new initial ontology is generated.

VI CONCLUSIONS

This paper formally defined ontology structure according to the needs of answers extraction and semantic analysis of questions. A method of ontology reuse was presented that can reduced the process of ontology building in RDQA. Further study will be carried out in the following aspects:

- Take account of contextual relation to choose more effective classification features;
- Optimize classification features, so as to obtain the most effective feature;
- Optimize and reduce dimension of feature;
- Weight value calculation of classification features.

ACKNOWLEDGMENT

This work is supported by The Beijing Natural Science Foundation (No.4123094), National Nature Science Foundation of china (No.60863011), The Science and Technology Project of Beijing Municipal commission of Education (No.KM201110028020, No.KM20101028019). This work is supported by Beijing Key Construction Discipline “Computer Application Technology”.

REFERENCES

[1] Shifu Zheng, Ting Liu, et al, “Overview of Question-Answering”, Journal of Chinese Information Processing, 2002, 16(6):46-52. (Chinese)
[2] Shuxi Wang, Qun Liu, Shuo Bai, “An Expert System About Relationship”, Journal of Guangxi Normal University(Natural Science. 2003, 21(l):31-36. (Chinese)
[3] Yinfei Huang, Fang Zheng, Pengju Yan, et al, “The Design and Implementation of Campus Navigation System: EasyNav”, Journal of Chinese Information Processing, 2001, 15(4):35-40. (Chinese)
[4] Bing Qin, Ting Liu, et al, “Question answering system based on frequently asked questions”, Journal of Harbin University of Civil Engineering and Architecture, 2003, 35(10):1179-1182. (Chinese)
[5] Xiaozhong Fan, Hongqiao Li, et al. A Study on a Bank-Domain Automatic Chinese Question- Answering System BAQS. Journal of Beijing Institute of Technology, 2004, 24(6):528-532. (Chinese)
[6] Peng Wu, SiKun Li. Social network analysis layout algorithm under ontology model. Journal of Software, 2011, 7(6): 1321-1328.
[7] Ohshima Hiroaki, Tanaka, Katsumi. High-speed detection of ontological knowledge and bi-directional lexico-syntactic patterns from the web. Journal of Software, 2010, 5(2): 195-205.
[8] Wei Huang, Liyi Zhang, Jidong Zhang, etc. Semantic focused crawling for retrieving E-commerce information. Journal of Software, 2009, 4(5): 436-443.
[9] PDiego Molla, Jose Luis Vicedo. Question answering in restricted domains: an overview. Computational Linguistics, 2007, 33(1): 41-61.
[10] Qinglin Guo, Ming Zhang. Question answering based on pervasive agent ontology and Semantic Web. Knowledge-Based Systems 22 (2009) 443–448.
[11] Jie Liu, Xiaozhong Fan, Tao Wang. Chinese Restricted Domain Question Answering System Based on Ontology. Journal of Guangxi Normal University(Natural Science Edition). 2009, 27(1), 169–172.
[12] Jie Liu, Li-zhen Liu, Yan Peng, etc.Answer Extraction of Chinese Restricted Domain Question Answering System Based on Ontology. Journal of Computational Information Systems. 2010, 6(1): 155-165.
[13] Paul Doran, Valentina Tamma, Luigi Iannone. Ontology Module Extraction for Ontology Reuse: An Ontology Engineering Perspective. Proceedings of the 16th ACM International Conference on Information and Knowledge Management, Lisboa, Portugal.2007, 61-69
[14] Elena Paslaru Bontas, Malgorzata Mochol, Robert Tolksdorf. Case Studies on Ontology Reuse. In Proceedings of the 5th International Conference on Knowledge Management, Graz, Austria. 2005, 345-353.
[15] Bernardo Cuenca Grau, Ian Horrocks and Yevgeny Kazakov. Modular Reuse of Ontologies: Theory and Practice. Journal of Artificial Intelligence Research. 2008, Vol. 31, 273-318.
[16] V. N. Vapnik. Statistical learning theory [M]. New York: Wiley, 1998.
[17] Hongqiao Li, Changning Huang, Jianfeng Gao, et al. Chinese chunking with another type of Spec. 42nd

Association for Computational Linguistics SIGHAN workshop, Barcelona Spain. 2004.
[18] Hong-kui Yu, Hua-ping Zhang, Qun Liu. Chinese named entity identification using cascaded hidden Markov model. Journal on Communications. 2006, 27(2):87-94. (Chinese)
[19] TongYiCiCiLin (extended). http://www.ir-lab.org/

Jie Liu is an associate professor at College of Information and Engineering, Capital Normal University, Beijing, P. R. China. He received PH.D degree in Computer Application Technology at Beijing Institute of Technology. His main research interests are the restricted domain Question Answering System and ontology application. *The corresponding author. Email: liujxxxy@126.com

Yun Ma is a Master degree candidate whose specialty is Application of Computer at College of Information and Engineering, Capital Normal University, Beijing, P. R. China. She received Bachelor degree in Information Management and System (E-commerce) at Capital Normal University. Her main research interests are ontology application and natural language processing.

Liming Luo is an associate professor at College of Information and Engineering, Capital Normal University, Beijing, P. R. China. He received master's degree in Computer Application Technology at Capital Normal University. His main research interests are ontology application and natural language processing.

Zhengtao Yu, received his Ph.D. degree in computer application technology from Beijing Institute of Technology, Beijing, China, in 2005. He is currently a professor in the School of Information Engineering and Automation, Kunming University of Science and Technology, China. His main research interests include natural language process, Chinese question answering system and machine learning.