Research on Reasoning and Retrieval Methods Based on Mongolian Curriculum Areas of Semantic Web

Hongbo Chen
College of Information Engineering Inner Mongolia University of Technology Hohhot 010051, P.R.China
Email: chb263@163.com

Yila Su and Chuan Fan
College of Information Engineering Inner Mongolia University of Technology Hohhot 010051, P.R.China
Email: suyila@tsinghua.org.cn

***Abstract*—The backwardness of the Mongolian network teaching resources results in its low reuse rates and utilization. For this situation, a retrieval method of semantic web based on Mongolian curriculum areas was set up. Firstly, the method established the Mongolian ontology of course 'Artificial Intelligence ([illegible].) 'in area of teaching, it uses a relationship database MySQL to record ontology information, on the top of that, it uses semantic similarity calculation and reasoning rules established to do the semantic information reasoning, in the end, it builds a query and retrieval prototype system based on the field of Mongolian teaching. Experimental results show that this method can effectively improve the information retrieval recall and precision rates, and it established Mongolian teaching resources which can be shared and interoperable knowledge hierarchy.**

***Index Terms*—Mongolian, semantic Web, ontology, reasoning, curriculum, information retrieval**

I. Introduction

As traditional information retrieval technology is mainly based on keyword matching, syntactical level search, it was lack of processing and presentation in the semantic level, the search results are incomplete and inaccuracy issues. However, solve this problem is from the simple keyword matching converted to semantic matching in the search process. Semantic Web[1][2] is proposed for traditional information retrieval provides a new technique that enables the machine to be understood semantic information. In the information age with rapid development of Internet, online teaching more and more welcomed by the people, which had been accepted and become a new trend at domestic and abroad. This new way of teaching not only make the knowledge and information spread fast, but also provide a platform for the sharing of teaching resources[3]. But in the network exist the large number of teaching and learning resources, mostly using simple semantic knowledge of the hierarchy without sharing and interoperability, so exists lower reuse and utilization rates issues. Applying the ontology technology to the field of education started in the early twenty-first century, it was university of Kalsruhe in Germany that constructed an ontology building tools KAON combined semantic web with E-learning[4][5]. Stanford University analyzes the differences between the traditional network of teaching and teaching based on semantic WEB networks, and it proposed ontology-based metadata that describe the learning resources, it builds a ontology-based courseware resource editor [6]. In China, the ontology research started in early 2002, it was Hui Dong in university of Wuhan, ontology application in digital libraries research [7], and Junfeng Song, Weiming Zhang, ontology-based information retrieval model [8], who is the scholar of engaging in applying the ontology theory to the field of teaching area. Mongolian language [9] [10] is one of the official language in Inner Mongolia, Mongolian computer information processing system was set up by the1980s, which provide good conditions for the promotion and application of computer processing of Inner Mongolian information. Inner Mongolia University, Inner Mongolia Academy of Social Sciences, Inner Mongolia Normal University and other units have developed a Mongolian computer information processing systems. Mongolian international standard code achieved recognition of the National Organization for Standardization and UNICODE Technical Committee. Because of poor conditions, small-scale schools, distribution of concentration issues in Mongolia, led to low levels of Mongolian teaching, greatly influenced the Mongolian teaching progress. Under these conditions, developing Mongolian network teaching can greatly improve.

II. Establishment of "Artificial Intelligence" Curriculum Mongolian Ontology

A. *The Needs Analysis of Mongolian Curriculum Areas Ontology*

The needs analysis[11] is the cornerstone of building knowledge domain ontology[12][13], after understanding the specific objectives and characteristics in knowledge domain ontology, it can be normalized, detailed and

doi:10.4304/jsw.9.7.1701-1706

specific described for ontology. So, in the beginning of building Mongolian curriculum domain ontology, the needs analysis is essential. According the specific purpose and significance of Mongolian curriculum areas ontology mentioned in this paper, specified information is as followed:

First, the purposes of building Mongolian curriculum ontology, ontology used in Mongolian curriculum areas construction, is mainly to achieve a common understanding in a hierarchy of knowledge of Mongolian specific curriculum areas, and meet the knowledge sharing, knowledge representation and knowledge inquiry between different systems, by adding semantic information in curriculum ontology can greatly improve the retrieval recall and precision rates.

Second, the Mongolian curriculum domain ontology covered subject areas. In order to avoid gaps in knowledge and meet information interaction between the various cross-disciplinary, in building ontology should take into account all the knowledge points of the curriculum and cross-disciplinary knowledge.

Third, the basic users of Mongolian curriculum areas ontology are teachers, students, and Mongolian researchers.

Fourth, the schedule of constructing the Mongolian curriculum areas ontology is that taking the particularities and knowledge Mongolian comprehensiveness, complexity, and other factors and the rules of practical application and effectiveness into account, we need to allocate time appropriately according to the knowledge level of difficulty.

Fifth, choose a correct description language when building Mongolian curriculum ontology. In order to have a good expression and expansion capabilities in the field of knowledge, we use the OWL ontology language.

B. Establish to Core Ontology Library of Mongolian Teaching Field

First, we should get concepts and relationships between concepts in the field of knowledge in the beginning of building Mongolian curriculum areas ontology, and define the corresponding class. We can through the network, books, authoritative experts and existing ontology to get the organizational structure knowledge areas, according to the organizational structure of the field of knowledge and various properties of concepts to get the top of concepts, one-level concepts and so on. Each of concepts inherits the basic attributes of parent class, and gradually form a hierarchical model. Concept in each layer can be abstracted as the ontology's basic classes, which subclass should inherit all the properties of the parent class, and subclass should have a new property. In this paper, combined with knowledge structure and teaching objectives of "artificial intelligence" curriculum, defines four top classes, include artificial intelligence research focus, applications, search reasoning, knowledge representation, The first layer of subclass more than 20, The second layer of subclass more than 100, coverage all the knowledge points of "Artificial Intelligence ([illegible].)".

Secondly, according to different classes to determine the class attributes. In this paper, according to different relationships between curriculum resources and combination of teaching objectives and tasks, which defined attributes as following table:

TABLE I.
THE MAIN ATTRIBUTES OF CURRICULUM AREAS ONTOLOGY

Data attribute name	Property description
[illegible]	Author
[illegible]	Language
[illegible]	Create date
[illegible]	Using Roles
[illegible]	Difficulty
[illegible]	Keyword
[illegible]	Knowledge synonyms
[illegible]	Key elements (Y / N)
[illegible]	Teaching objectives
[illegible]	The size of knowledge
[illegible]	Importance

TABLE II.
THE OBJECT PROPERTIES OF CURRICULUM AREAS ONTOLOGY

Object attribute name	Attribute description	Attribute explanation
[illegible]	Parent knowledge	A knowledge point corresponds to a parent knowledge
[illegible]	Children knowledge	A knowledge point has one or more sub-knowledge
[illegible]	Refer	Knowledge points refer to other resources
[illegible]	Refered	Knowledge points is refered by other resources
[illegible]	Quote	Articles refer to other resources
[illegible]	Quoted	Articles is refered by other resources
[illegible]	Common reference	Two or more articles refer to the same resource
[illegible]	Common referenceed	Two or more articles are refered by the same resource

Third, combined with teaching objectives of "Artificial Intelligence" curriculum describe the logical relationships between different concepts. This paper defines six basic relationships described curriculum ontology, including the relationship between knowledge and chapters, the relationship between knowledge and resources, the upper relationship between knowledge, the lower relationship between knowledge, the parity relationship between knowledge, the relationship between knowledge, which defined relationship as following table:

TABLE III.
THE BASIC LOGICAL RELATIONSHIPS BETWEEN CLASSES

Relationship	Relationship description
[illegible]	relationship between knowledge and chapters
[illegible]	relationship between knowledge and resources
[illegible]	upper relationship between knowledge
[illegible]	lower relationship between knowledge
[illegible]	parity relationship between knowledge
[illegible]	relationship between knowledge

Fourth, build instance for each specific curriculum class.

III. SEMANTIC SIMILARITY CALCULATION

Semantic similarity[14][15][16] definition is two or more different concepts which have the same or similar feature, we then define these concepts are similar, for example, using Sim (C1, C2) indicates similarity between C1 and C2, the similarity calculation should satisfy the following situations:

(1)The similarity of Sim (C1, C2) has a value is any real number in [0, 1].

(2) When the concept of C1 and C2 have same attribute characteristics, namely C1 = C2, then Sim (C1, C2) = 1.

(3) When the concept of C1 and C2 have completely different attribute characteristics, then Sim (C1, C2) = 0.

(4) The relationship is symmetrical between concept of C1 and C2, namely Sim (C1, C2) = Sim (C2, C1).

The impact factors of semantic similarity calculation[17][18] include semantic distance, concept node depth, concept node density and adjusting parameters. Depending on different impact factors have different influence level, summed similarity is calculated as follows:

$$Sim(C1,C2)=(a/Dis(C1,C2)+a)^{\alpha}*[(Depth(C1)+Depth(C2))/(|Depth(C1)-Depth(C2)|+1)]^{\beta}*(1/Density(C1,C2))^{\gamma}$$

In the above formula, a / Dis (C1, C2) + a represents semantic distance can influence the semantic similarity calculation, since semantic distance largely affected the semantic similarity calculation, so α has the larger weights; (Depth (C1) + Depth (C2)) / (| Depth (C1) - Depth (C2) | +1) represents concept node depth can influence the semantic similarity calculation, since concept node depth less affected the semantic similarity calculation, so β has the smaller weights; 1 / Density (C1, C2) represents concept node density can influence the semantic similarity calculation, since concept node density less affected the semantic similarity calculation, so γ has the smaller weights, and α + β + γ = 1.

In this paper, extract part of a collection ontology to calculate semantic similarity. Extract part of a collection of ontology {1、2、3、4、5、6、7、8、9、10、11、12、13、14} denote { (Search and Reasoning), (And or Graph), (Production system), (Heuristic search), (Blind search), (State graph search), (Bidirectional reasoning), (Reverse reasoning), (Global merit search), (Local merit search), (Etc Search), (Exhaustive Search), (Breadth-first search), (Depth-first search)}, as shown below:

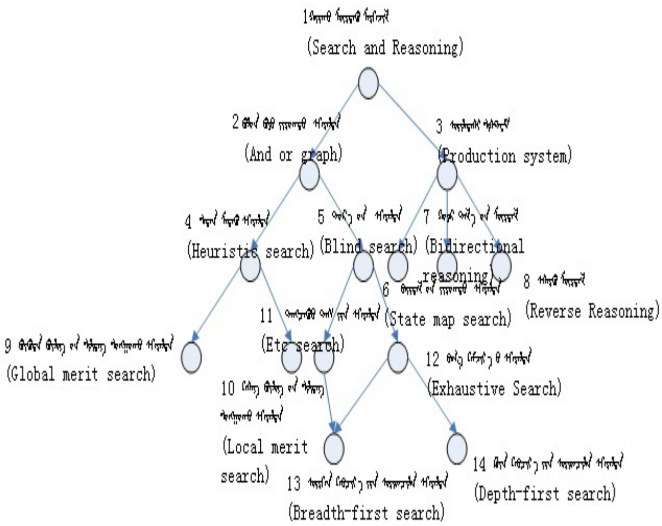

Figure 1. Part of a collection of ontology

This paper take a = 1, α = 0.9, β = 0.05, γ = 0.05, the similarity between each node has been calculated as follows:

TABLE IV.
SEMANTIC SIMILARITY VALUE OF EACH NODE

	1	2	3	4	5	6	7
1	1	0.536	0.536	0.368	0.368	0.368	0.368
2	0.536	1	0.391	0.550	0.550	0.293	0.293
3	0.536	0.391	1	0.293	0.293	0.550	0.550
4	0.368	0.550	0.293	1	0.399	0.246	0.246
5	0.368	0.550	0.293	0.399	1	0.246	0.246
6	0.368	0.293	0.550	0.246	0.246	1	0.399
7	0.368	0.293	0.550	0.246	0.246	0.399	1
8	0.368	0.293	0.550	0.246	0.246	0.399	0.399
9	0.281	0.372	0.235	0.551	0.295	0.203	0.203
10	0.281	0.372	0.235	0.551	0.295	0.203	0.203
11	0.281	0.380	0.238	0.301	0.301	0.205	0.205
12	0.281	0.380	0.238	0.301	0.301	0.205	0.205
13	0.219	0.277	0.190	0.232	0.372	0.169	0.169
14	0.219	0.277	0.190	0.232	0.372	0.169	0.169

	8	9	10	11	12	13	14
1	0.368	0.281	0.281	0.281	0.281	0.219	0.219
2	0.293	0.372	0.372	0.380	0.380	0.277	0.277
3	0.550	0.235	0.235	0.238	0.238	0.190	0.190
4	0.246	0.551	0.551	0.301	0.301	0.232	0.232
5	0.246	0.295	0.295	0.301	0.301	0.372	0.372
6	0.399	0.203	0.203	0.205	0.205	0.169	0.169
7	0.399	0.203	0.203	0.205	0.205	0.169	0.169
8	1	0.203	0.203	0.205	0.205	0.169	0.169
9	0.203	1	0.399	0.250	0.250	0.169	0.171
10	0.203	0.399	1	0.250	0.250	0.171	0171
11	0.205	0.250	0.250	1	0.399	0.171	0.295
12	0.205	0.250	0.250	0.399	1	0.551	0.551
13	0.169	0.171	0.171	0.551	0.551	1	0.399
14	0.169	0.171	0.171	0.295	0.551	0.399	1

In the search " [illegible] (Search and Reasoning)", set the threshold value is 0.5, the result expansion set is { [illegible] (Search and Reasoning), [illegible] (And or Graph), [illegible] (Production system)}, and then use this extended collection conduct information retrieval, greatly improving the information retrieval precision.

IV. Semantic Retrieval System Design Is Based on Mongolian of Curriculum Areas

A. The needs Analysis of Semantic Retrieval System

In this paper, "Artificial Intelligence" curriculum is regarded as research ontology, we finish retrieval system based on Mongolian Semantic Web. In curriculum areas, due to the complex hierarchy of knowledge and teaching objectives, it is difficult to ensure a certain query and retrieval accuracy rate for knowledge. In curriculum areas, the knowledge can be divided into horizontal and vertical knowledge, namely knowledge is at the same level or parallel relationship called horizontal relationship, vertical relationship is parent and child relationship of knowledge points. When learning courses, there have to learn other knowledge before learn certain knowledge, after learning certain knowledge can learn other knowledge, that explain there exist predecessor and successor relationships between knowledge. In addition, there exists interdependent and interaction relationships between knowledge. So can make all the knowledge points more structured and mastery, according to different relationships of knowledge points, forming a hierarchical model between knowledge points.

B. System Model Design

Retrieval system based on Mongolian semantic in curriculum areas can achieve semantic level reasoning and matching, and also can conduct semantic extensions and semantic mapping, after inputting Mongolian keyword, according to inference rules conduct semantic reasoning, retrieved semantic information associated with the keywords. This solves precision rates issue based on keyword query. This paper according to knowledge structure of Mongolian curriculum areas, use "artificial intelligence" as a research ontology and achieve the "Artificial Intelligence" course knowledge query and retrieval, the system consists of four parts, include user query and results display interface, ontology information database, Web information management, semantic extension and reasoning.

User query and results display interface: This function is information exchange between user and the system, namely provide users with entered interface and returns the query results.

Ontology information database: Ontology as the core of semantic retrieval, it defines the concept of knowledge, relationships, and a collection of entities and attributes, so that people can reach for a common understanding of knowledge areas, and achieve information sharing and reuse, so building and managing ontology is very crucial. Ontology information mainly from books, experts and have defined the concept to get, it major complete semantic expansion and semantic annotation.

Web information management: This paper mainly achieved query and retrieval the resources on the Web, in order to ensure precision rates we must have sufficient resources of Web information are available for retrieval. In this function module, ontology repository add Web information which has been collected, can also add some Word, PDF documents, which will help the integration of information resources.

Semantic extension and reasoning[19][20]: This function module implements when user enter keyword retrieving information, according to ontology library information which has been built, we should map original query keyword and ontology library information to conduct semantic similarity calculate and semantic extensions, at the same time according to inference rules conduct semantic information inference, which get the ontology information similar to original query words, added new query words to conduct retrieval, so improve the retrieval accuracy. In this paper, semantic extension mainly include synonyms extensions, contain extensions, hypernym extensions, parity word extensions, hyponym extensions and other extensions. Ontology reasoning according to inference rules apply relevant inference engine analysis ontology OWL document, this paper mainly involves some relevant curriculum knowledge reasoning.

C. Add Inference Rules

Jena has ontology query and reasoning mechanisms, if achieving Jena reasoning, we have to formulate inference rules, the inference engine according to inference rules conduct reasoning, and get a collection of concept which associated with given the concept. In this paper, in order to achieve the "Artificial Intelligence" curriculum knowledge areas ontology query and retrieval, practical application and effectiveness must be given as followed:

(1) As in object-oriented learning process has given class inheritance, so the class inheritance concept is also applicable to ontology, the class inheritance inference rules are as follows:

Rule 1: (? C1, rdfs: subClassOf, ? C2), (? P, rdfs: type, ? C1) -> (? P, rdfs: type, ? C2)

Note: if c1 is a subclass of c2 and p is an entity or attribute of c1, then p also belongs to an entity or attribute of c2.

(2) In the ontology concept, if two classes are the same class, the two classes have the same attribute value, inference rules are as follows:

Rule 2: (? C1, owl: equivalentClass, ? C2), (? P, rdfs: type, ? C1) -> (? P, rdfs: type, ? C2)

Rule 3: (? C1, owl: equivalentClass, ? C2), (? P, rdfs: type, ? C2) -> (? P, rdfs: type, ? C1)

Note: in the Rule 2 and Rule 3, c1 and c2 are the same class, then c1 and c2 have the same attribute values.

(3) Classes' contain relationships , inference rules are as follows:

Rule 4: (? C1, rdfs: subClassOf, ? P), (? P, rdfs: subClassOf, ? C2) -> (? C1, rdfs: subClassOf, ? C2)

Note: if c1 is a subclass of p and p is a subclass of c2, then c1 is c2 subclass.

(4) Synonymous class and the current class has the same characteristics, the inference rules are as follows:

Rule 5: (? C1, owl: SameAs, ? C2), (? C1, p# ConnectWith, ? C3) -> (? C2, p # ConnectWith, ? C3)

Note: c1 and c2 are synonymous class, c1 and current class have the same characteristics p, then c2 and c3 also have the same characteristics p.

(5) Citing and cited, the inference rules are as follows:

Rule 6: (? C1, # citing, ? P), (? C2, # citing, ? P), notEqual (? C1, ? C2) -> (? C1, # bothciting, ? C2)

Note: c1ass c2 cite resource p, c1 and c2 are different, then c1 and c2 simultaneously cite p.

Rule 7: (? P, # citing, ? C1), (? P, # citing, ? C2), notEqual (? C1, ? C2) -> (? C1, # bothcited, ? C2)

Note: p cite resources c1 and c2, c1 and c2 are different, then c1 and c2 are simultaneously cited by p.

D. Storage Curriculum Ontology Information

Jena provide some interfaces which RDF data can storage relational database, the interfaces used to access and maintain RDF data, and the interface includes Model, Resource and Query and so on. Jena also provide some program interfaces which can support relational database, including MySQL, HSQLDB, SQL Server, etc. This paper will use a relational database MySQL storage ontology data, the main steps include:

(1) Load the database JDBC driver;

(2) Create a database connection;

(3) Create a ModelMaker database model for the database;

(4) Create a model for ontology information.

E. Design and Achievement of Semantic Retrieval Based on Mongolian Curriculum Areas

1. This design based on a curriculum areas of Mongolian semantic retrieval system is based on B / S structure, achieve Mongolian semantic retrieval priority is display properly Mongolian characters on the browser, due to most browsers do not support the Mongolian characters, therefore, in this paper, the design is required before download and install IE Tab Multi plug to solve Mongolian not display correctly on the browser issue.

2. The curriculum areas Mongolian semantic retrieval main purpose is that when we input knowledge point need to retrieve, click submit button, the system will conduct semantic similarity calculation and semantic expansion according to keyword, we will get the ontology information similar to original query words, added new query words to conduct retrieval, so improve the retrieval accuracy.

3. In search module, it has not only the basic information and meanings of search words, more importantly, it also has the all the semantic information corresponding to the search words by the system of semantic query expansion. For instance, given the point of search word “[illegible] (and_or_graph)”, we achieve semantic relationships through semantic query expansion: the relationship of knowledge and chapters, the relationship between knowledge and resources, knowledge, relationships between the upper, lower, relationship between knowledge, knowledge of the relationship between the parity, the relationship between knowledge points. We use this information to call the corresponding class in order to find the specific semantic relations under the relevant knowledge, including the author, creation date, language, size, using roles, synonyms, parity words, the upper word, knowledge point range, whether it is key elements, relevant content, degree of difficulty, teaching objectives, citation, cited the same arguments as well as with other. Specific semantic query flow is as follows:

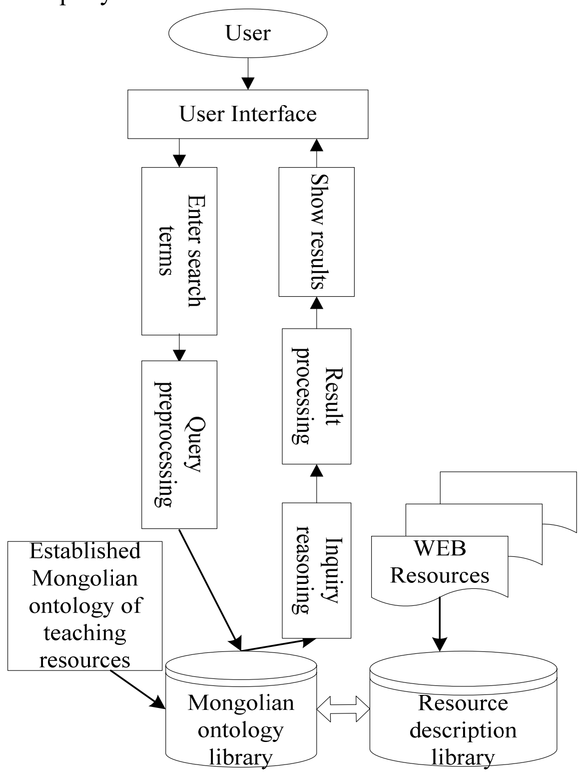

Figure 2. The chart of semantic query flow

V. CONCLUSIONS

This paper according to Mongolian network teaching resources exist the information can not be shared and non-interoperable issue, we study and design information retrieval system based on Mongolian curriculum areas semantic WEB. The main research work is as follows:

(1) This paper carried out a detailed analysis for "artificial intelligence" curriculum, we established curriculum ontology based on Mongolian and

defined class attributes, and relationships.

(2) This paper analysis the impact factors of semantic similarity calculation, extracted part of a collection of ontology to calculate semantic similarity, we get semantic similarity of each knowledge points.

(3) By using the MySQL relational database storage "Artificial Intelligence" curriculum ontology information.

(4) Designed and Achieved retrieval system based on Mongolian curriculum areas Semantic Web.

ACKNOWLEDGMENT

This work is partially supported by National Natural Science Foundation of China (61363052), Natural Science Foundation of Inner Mongolia Autonomous Region (2012MS0904), College and University Scientific Research Project of Inner Mongolia Autonomous Region (NJZZ12046) and Inner Mongolia Science and Technology Plan Project (20130362), Foundation of Inner Mongolia University of Technology (ZD201001).

REFERENCES

[1] Berners-Lee T. Semantic Web. W3C Recommendation [EB/OL].http://www.w3.org/2000/Talks/1206-xml2k-thl/, 2000

[2] Wei Song, Ming Zhang. A First Step Towards the Semantic Web [M].Beijing: Higher education press,2004

[3] Perez A G, Benjamins V R. Overview of Knowledge Sharing and ReuseComponents Ontologies and Problem Solving Methods[C]. In Stockholm V R, Benjamins B, Chandrasekaran A, eds. Proceeding of the IICAI99 workshop on Ontologies and Problem Solving Methods(KRR5) 1999, 1-15.

[4] Ashok Patel, Kinshuk, David Russell. Designing Life-long Learning Systems for Numeric Domains, Educational Technology, September-October, 1999, pp.8-11.

[5] Vossen G, jaeschke P, Oberweis A. Flexible Workflow Management as a Central E-learning Support Paradigm. In: Proc. 1st European Conference on E-Learning(ECEL), Uxbridge, UK, 2002, pp. 253-267.

[6] Stanford medical Informatics. Welcome to the Protégé Project[Z].Germany,http://www.ontoprise.de/products/ontoedit_en, 2008.

[7] Wenhua Du, Hui Dong. Ontology and its application in digital libraries research [D]. Wuhan University doctoral dissertation, 2005, 6:12-15

[8] Junfeng Song, Weiming Zhang, Weidong Xiao, Jiuyang Tang. The research of information retrieval model based on ontology [J]. Journal of Nanjing University, 2005,41 (2) :189-197.

[9] Inner Mongolia University Mongolia College of Mongolia Language Institute.Modern Mongolian [C]. first edition. Hohhot: Inner Mongolia people's publishing house, 1964

[10] Garidi. The status of Mongolian natural language information processing. Several important issues about Chinese information processing. Science Press, 2003, 11:7-9.

[11] Ruling Lu, Zhi Jin, Gang Chen. The requirements analysis oriented ontology [J]. Journal of Software, 2000, 11 (8): 1009-1017.

[12] Lingyan Jia, Yiping Lu. Talking about ontology method and its development [A]. National advanced manufacturing technology forum and manufacturing automation, information technology symposium [C]. 2005.

[13] Xifeng Wang, Xiaoluan Zhang. Ontology technology and its application [J]. Journal of Baoji University (Natural Science), 2006, 2:10-12.

[14] Ling Song. Semantic similarity calculation and its application [D]. Jinan: Shandong University dissertation .2009, 5:32-35.

[15] Hongzhe Liu, De Xu. Semantic similarity and correlation computation Research based on ontology [J]. 2012, 2:15-17.

[16] Qiuyu Zhang, Peng Wang, Huijuan Yang. Applications of Text Clustering Based on Semantic Body for Chinese Spam Filtering [J]. Journal of Computers. 2012, 7 (11) :12-15.

[17] Bin Shi, Jianzhuo Yan, Pu Wang, Liying Fang. Semantic similarity measure based on ontology [J]. Computer Engineering. 2009,35 (19) :23-25.

[18] Zhongping Zhang, Hailiang Zhao, Zhihui Zhang. Semantic similarity calculation based on ontology [J]. Computer Engineering 2009, 35 (7) :19-23.

[19] Yingjie Song, Rong Chen, Yaqing Liu. A Non-Standard Approach for the OWL Ontologies Checking and Reasoning [J]. Journal of Computers. 2012, 7 (10) :54-60.

[20] Chouyin Hsu. Development of Semantic-CBR Framework for Virtual Enterprises in Project Management [J]. Journal of Computers. 2011, 6 (3) :434-439.

Hongbo Chen (1988-), male, graduate, major research areas: Web intelligence, semantic web, etc.

Yila Su (1964-), male, doctor, professor, mainly engaged in artificial intelligence, machine learning, data mining, computer networks and other aspects.