Model Design of Semantic Website Construction

Yu Guo, Jie Liu*
College of Information and Engineering, Capital Normal University, Beijing, China
Email: liujxxxy@126.com, guoyugogogo@126.com

***Abstract*—Semantic web is an extension of the current World Wide Web. This paper proposes a new model of constructing semantic website, and takes the construction of semantic website in the field of tourism in Africa as an example, and describes an approach to build the semantic website. It is a common approach and can be applied to the construction of other semantic websites. First, analyze user's requirements, and then build domain ontology based on them requirements, design interactive interface, and analyze the information entered by user. Then, retrieve and reason the ontology by Jena, and provide the required information and links. The proposed method takes full account of the demand of user's interaction, facilitates obtaining the required information on the website. The example shows that the proposed model is effective.**

***Index Terms*—semantic website, domain ontology, semantic retrieval, semantic reasoning**

I. Introduction

In this information age, how to quickly and efficiently retrieve relevant information on the network has become a hot issue in the field of computer research. Current mainstream network search engines, such as Google and Baidu, etc., generally use keyword-based information retrieval method. This method can't directly index and retrieve the semantics of information, and semantic reasoning and retrieval can't complement each other and mutually promote; the retrieval results lack unified form, can't be shared and reused; and associated resources are difficult for users to be retrieved. The main reason for the above problems is that computers can't understand the semantics of the network information. However, the emergence of the semantic web [1] provides a solution to these problems. Information described by semantic web technologies has a clear meaning, so that computers can understand web information, and use it as a basis for knowledge discovery, association and other related services.

Since Tim Berners-Lee proposed the concept of the semantic web in 1998, it has become the focus of people's discussion and study. The study of the semantic web is just in its infancy, however, plenty of scholars have recognized the impact of development of semantic web and related technologies on the future of the Internet, and start and research semantic web and related key technologies and application. Research on the semantic web has achieved certain results, and been hierarchical and deeper from the basic introduction to theoretical research, from theoretical discussion to the test and application.

This paper proposes an interactive information retrieval model based on semantic web technologies to solve the above problems. By extending the existing Internet, add information's meaning to itself and use semantics to re-organize, store and retrieve information, to improve computer's processing power. Meanwhile, in the retrieval, consider the individual needs of different users, treat users differently, and retrieve according to the user's interests and characteristics to get search results that can better meet the demands of users. Compared with the traditional information retrieval, interactive information retrieval based on semantic web technologies has higher accuracy and has very great significance for information retrieval.

II. The Advantage of the Semantic Websites Compared with the Traditional Websites

With the development of the Internet, the network has become one of the important ways to access to information from web pages, and more and more people get the information they need from the Internet. Currently the network resources are very rich, but the information is relatively fragmented, so the status of traditional network retrieval can't meet demands, which means the search mechanism is often based on retrieval of keyword or topic content.

The traditional information retrieval [2] is based on keyword search technology, which requires users with keyword to describe the need more accurately, and finally the information is submitted to the inquiry system in the form of retrieval expressions. Since its literal meaning is not same with its concept extension, it results in the results retrieved traditionally might just related to the literal meanings. However, what people often wander are its concepts and related components, rather than the literal information expressed. In other words, the traditional tourist information retrieval only concerns about the word pattern matching, not semantic concepts and their correlation.

However, changing this situation is the goal of semantic network construction. The goal of the semantic web research is to development of a series of expressing

Manuscript received December 09, 2013; revised January 21, 2014; accepted January 22, 2014.
This paper is supported by the National Nature Science Foundation of China (No.61371194).
Corresponding author is Jie Liu. Email:liujxxxy@126.com

doi:10.4304/jsw.9.8.2142-2147

the semantic information of language and technology that computers can understand and process, to support automatic reasoning in network environment more widely and more effectively. Semantic network founder Tim Berners-Lee on the semantic web is defined as follows: "semantic network is a network, it contains a portion of the document or documents, describes the obvious relationship between things, and contains the semantic information, in order to facilitate automatic processing machine". The understanding and description of the semantic web is different, but can still see some basic features of the semantic web from the description and understanding: (1) the semantic web is different from the world wide web, it is the expansion and extension of the world wide web; (2) the existing world wide web is a document oriented and semantic web for the document the represented data; (3) the semantic web will be more conducive to the computer "understanding and processing", and will have certain judgment and reasoning ability. This paper takes an example for building the model of semantic website in the field of tourism in Africa, to discuss the methods of building this semantic website and others based on the principle.

To retrieve information on popular traditional tourism site, for example, when the user enters "Egyptian pyramid", the first website shows the group tour information only about traveling to Egypt and pyramids, while another tourism site shows the pyramid location, types and reviews and other relevant information. When users travel in a certain location, what they need not only the content presented above pages, but also more about the attractions information of clothing, food, housing, transportation, weather and other aspects. However, the traditional tourism sites do not provide directly the important messages related to these attractions.

This semantic website model can achieve the needs while above-mentioned website can't. Users enter keywords, word segmentation system will present to the user with all relevant travel information and convenience for users. Through ontology knowledge base, build relationships between classes, achieve the functions traditional tourism website lacks, for example, searching for "Egyptian pyramid", the interface will show all relevant information about "Egypt" and "Pyramid", such as introduction of Egypt, the official language, currency, travel taboos, travel etiquette, visa and other information, attractions introduction, the weather, the surrounding hotels, restaurants, shopping centers and other information, and attractions' recommendations in the same city, which provide the user with a comprehensive search interface.

This model takes full advantage of the many advantages of the semantic web. The ability to locate information based on its meaning. knowing when two statements are equivalent, or knowing that a reference to a person in different web pages are referring to the same individual. Integrating information across different sources by creating mappings across application and terminological boundaries we can identify identical or related concepts. Improving the way in which information is presented to a user. aggregating information from different sources, removing duplicates, and summarizing the data [3].

III. Requirements Analysis and Development Process Ofsemantic Website

A. *Requirements Analysis*

Function: Around the subject that user is willing to learn about, provide user with other interrelated information, including African countries, cities, attractions, restaurants hotels, and other relevant information. Meanwhile, user can click on the relevant links which they are interested in, and learn more details. The following Fig.1 shows the page links graph. Among them, different themes have different links graph.

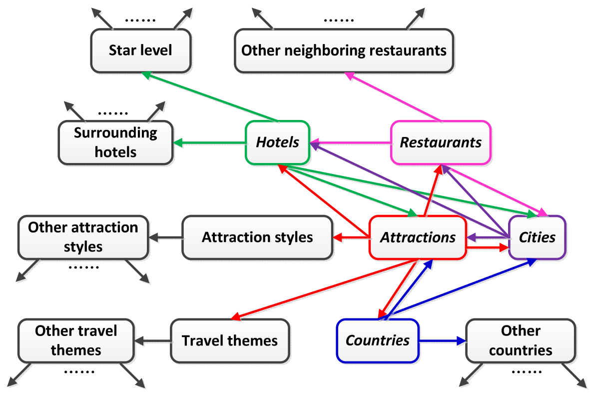

Figure 1. Page links graph

B. *Development Process for Semantic Website*

The development process of semantic website is planning and implementation process to itself, which is universal to lead to the construction of semantic websites. As shown in Fig.2, build the ontology according to the user requirements, then retrieve and reason based on ontology and requirements, and finally present the final result to the user.

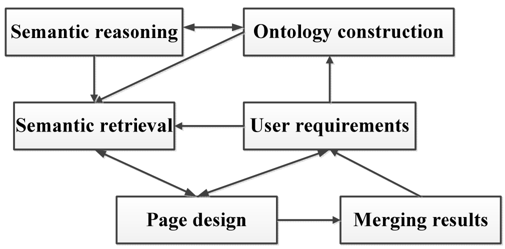

Figure 2. Development flowchart of semantic website

IV. Overall Analysis and Esign

A. *Construction of Domain Ontology Classes, Individuals, Properties*

Ontology [4-5] is a collection of concepts and relations, and abstract description of concepts of domain knowledge. Based on the comparison and induction to a great amount of travel websites and analysis of user's retrieval needs, the system extracts the concepts, individuals, and the attribute relations between them, describes the ontology by protégé as a template tool, and creates a perfect African tourism domain ontology [6] knowledge base.

User can search everything in the field of tourism, including food, housing, transportation, travel, shopping and so on. Based on the user retrieval needs, tourism resource ontology organizes a number of tourism-related resources by use of ontology, and describes the resources.

Through analysis and refining of user needs, the main information in the area of tourism includes attractions, countries, cities, services, weather and other concepts, where in service also includes transportation, catering, hotels and other concepts. Therefore, the African tourism domain ontology is built in a top-down way, with the important information considered in the field of tourism in Africa. The class hierarchy of tourism ontology is: First class: countries, cities, weather, tourists, tourism, attractions, services; Second class: for example the climate and wind are under the weather concept, the cultural customs and travel themes and tips are under the tourism concept, cultural and biological landscape are under the attractions concept, etc.; If further subdivided, the festival class is included in cultural customs, as well as the languages and the currencies are included in travel tips, etc. The individuals come into being if classes are further subdivided, and if no same individuals among classes exist they are called mutex class. Then add the corresponding individuals into ontology. Each class contains a number of individuals. For example, cultural attractions class, as the subclass of attractions class, contains Lesedi Cultural Village, Alexandria National Museum, the Cradle of Humankind site, Museum of Islamic Art, and many other individuals. Furthermore, these individuals constitute the main information of semantic website retrieval.

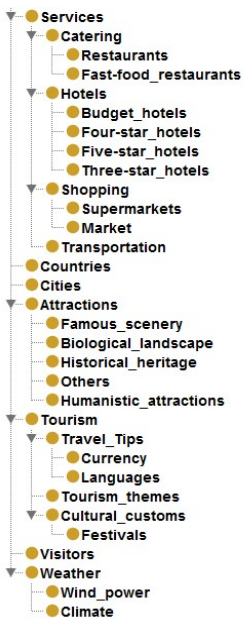

Figure 3. Diagram of classes

However, it's not enough that ontology only contains classes and individuals, because what user needs is not just a single presentation of information, more importantly, but the related information. Thus, if specific information user required needs to be returned, relevant properties should be added, and reasoning if necessary.

Attribute relationship of classes includes object properties and date properties. Object properties describe the relationship between individuals. So it's an important issue needed to be considered when building ontology. With the definition of the object properties, individuals are interconnected, and services are provided for the search and reasoning of ontology. In this paper, object properties of domain ontology mainly are "located in", "has/have" between countries and cities and so on. Data properties describe the relationship between individuals and RDF literals or data types of XML Schema [7] . For example, the nation introduction, in this paper, is defined by data properties whose type is "string".

By analyzing the requirements of user's retrieval, the addition of properties and individuals in ontology knowledge base is analyzed and defined as following:

Take an example retrieving the attraction Lesedi Cultural Village, which is an individual of cultural attractions, subclass of attractions. First, user may need to find out its ticket price, telephone, opening hours and related profiles, which can be achieved by adding data properties; Then, the user may need to know the city where the village is located, which can be achieved by building the relationship between the city and attractions through defining object properties as "locate In". Of course, the user may also retrieve related attractions by looking for travel themes. In the ontology, the folklore individual of travel theme, subclass of tourism, creates object properties "SuitView" with Lesedi Cultural Village.

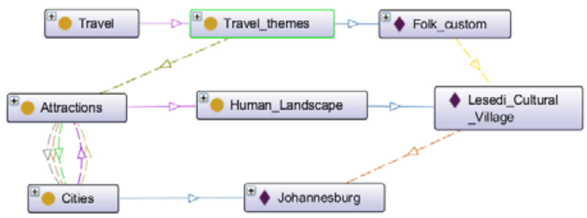

Figure 4. Attractions individual construction in ontology

B. Retrieval and Reasoning Algorithm of Ontology

(1) Semantic retrieval in ontology by Jena [8]

In order to display the information user's required by retrieving the domain ontology, the algorithm uses Jena to query ontology and complete semantic retrieval. Read OWL/RDF [9] file by Jena, output the corresponding information in the following way, call the relevant methods, output corresponding information stored in ontology knowledge base according to user's requirements, and call them by web.

With the circular iterative algorithm, the Ontology subsystem defines the ways to access to classes, properties and individuals, which are based on "OntModel". By using the read method of "OntModel", we can import an ontology file. By the "listClasses" method of "OntModel", we can get the lists of all the classes in the ontology whose type is "OntClass". With the method of "listobjectproperties" and "listDatatypeproperties", we can get the collection of properties, and an iterator of object attribute collection is returned by "hasobjectProperties" method. In Jena, the object properties correspond with "Objectproperty" class, the data properties correspond with "Datatypeproperty" class, and the attribute domain and range, property restrictions and other information can be obtained by calling them. Get domain by "getDomain" method, and obtain range by "getRange" method. Get the individuals by "hasIndivuduals" method, which returns an iterator of collection of individuals. Of course, we can also use RDF triples to retrieve domain ontology, and easily provide information for web pages by distinguishing subject, predicate and object.

With the method, we can get all kinds of classes in domain ontology; get all the subclass and parent class both direct and indirect; get all the properties of a class, like: property name, property type, property value; get all individuals and their corresponding attributes, if it is numeric property, then directly output its value, otherwise output object properties.

(2) Using Jena to implement OWL-based semantics reasoning

Although the specific contents can be retrieved by calling the front end and returned to the user, the user can find the information established in the ontology in the search process. Some fixed relationships are built in ontology knowledge base, however, the user's retrieval needs can't be meet. For example: The user may also need to learn about what are the attractions, hotels and restaurants surrounded by Lesedi Cultural Village, which can't be achieved by constructing simply ontology object properties, but by ontology reasoning. (As shown in Fig.5: Green: data properties, Blue: object properties, Purple: individuals, Red: reasoning)

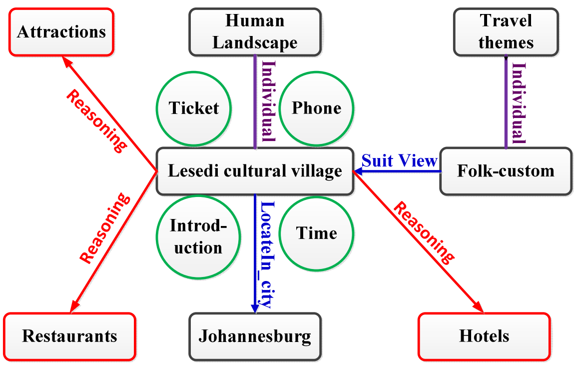

Figure 5. Relationship between attractions and individuals

In the semantic website model, Jena is used for reasoning owl ontology file. If only the relationships between classes exist, and no corresponding individuals exist, user will not be able to obtain the required information. However, with reasoning rules, the relationships between classes can reach relationships between corresponding individuals, thus desired information can successfully be returned to the user.

SPARQL [10-14] is a query language for RDF data. It uses SQL statements, and can be embedded into Java code, so that the ontology data query is realized.

Using SPARQL for ontology reasoning:

- "InfModel" is an extension of the conventional Model, supports any relevant reasoning ability;
- "QueryExecution" implements the query of SPARQL;
- "ResultSetFormatter" obtains query result sets;
- Reasoning utilizes ontology file named Ontology_Africa.owl.

Using the SPARQL to achieve reasoning, according to relationships between the classes established such as countries and attractions, can reason out all the relation between attractions' and national individuals, and is convenient for the user to view the current list of national scenic spots, and travel query.

Construction rules are as follows: (In rule1 if 'x' sits in city 'y' and 'y' is located in the country 'z', then 'x' is located in the country 'z'.)

[rule1:(?x fa:locateIn ?y)(?y fa:locateIn_country ?z)→(?x fa:LocateIn_country ?z)]
[rule2:(?x fa:locateIn ?y)(?z fa:commonIn ?y)→(?x fa:HasCommonHotel ?z)]
[rule3:(?x fa:locateIn ?y)(?z fa:locateIn_City ?y)→(?x fa:HasRestaurant ?z)]

Figure 6. Reasoning rules

The above reasoning by SPARQL also includes: attractions& nations, attractions & currency, attractions & hotels, attractions & restaurants, attractions & languages, cities & climate, cities & currency, cities &languages, hotels & restaurants, etc., in total 25 inference rules. These rules construction has been basically able to meet the general users' requirement, of course, with the continuous improvement of domain ontology, more rules are built to improve the user's retrieval demands.

C. *The Human-machine Interface and the Retrieval Model Design*

(1) Analysis the information that user requests

The segmentation extracts keywords for retrieval and reasoning, then how to classify keywords to filter out the main keyword priority to searches? First, simulate a user testing, explore the composition types of keywords input, and the results are as following:

1).Individual & class name

For instance, what are attractions in South Africa, which are hotels around the Sphinx, etc. All of those are combination of individuals and class names.

2).Individual & property

For example, how to apply for a visa to South Africa, and how to get the synopsis of Sphinx and so on, are all combination of individuals and properties.

Another way to search is to retrieve individuals, such as South Africa, the Sphinx and so on. By analyzing the test results, the retrieval model designed of system deals with two keywords in maximum, which can meet the users' requirements.

(2) Filter two or more keywords

Processing of information user input is also together with the status after segmentation to analyze. In the lexicon, there exist two words like hotel and a hotel's name, for example the President Hotel. If the user enters the President Hotel, the segmentation result are Hotel and President Hotel, and are still two before retrieval, then how to deal with the problem that no matter the input is President Hotel or Cairo hotels, users can get the information they need?

The solution is that the results should be classified after segmentation and before search. Put the keywords into an array like hotels, attractions, profiles, tickets and etc., and then the rest will be retrieved. For example, the user input the attractions of Cairo, the word in the retrieval is Cairo, and after retrieval, all relevant information about Cairo will shows, including attractions in Cairo, how to access to the attractions information users need to query preferentially. The attractions, synopsis and other words are input into an array, so that the extracted attractions in array can be matched with the information retrieved, and priority to be provided for user.

(3) Screening and filtering process of reasoned information

If the user searches for something like "What are attractions in South Africa", then the information returned to the user will be related to reasoned results. Organization of information after processing is as shown:

Tunisia = [Archaeological Site of Carthage, amphitheatre of Thysdrus, Great Mosque of Ez-Zitouna];
Libya = [Leptis Magna].

(4) Implementing process of the semantic website

The whole process as it's shown in Fig.7:

1. First, segment the information user has input , and store them in an array A;
2. Provide user with information that should be around individuals, so the country and attraction classes and their property names should be stored in an array;
3. Retrieve the keyword of individuals, according to the type of returned term, call different reasoning function, and set different returned page parameters;
4. According to the word in the array, set different parameters and returned values, so that different redirected pages can be configured.

Finally, return the results of integration, retrieval and reasoning to the user.

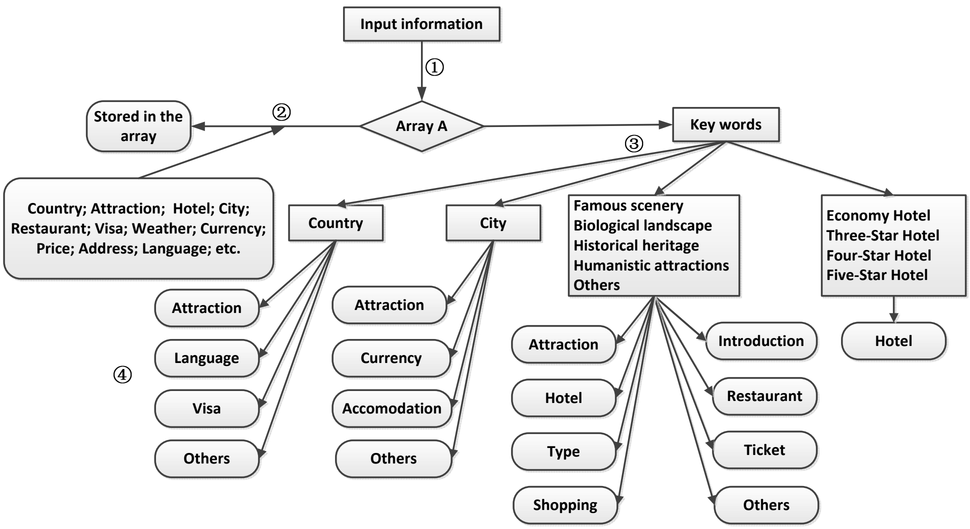

Figure 7. Overall processing

V. Experiment and Result

Tourism is related to clothes, food, shelter and travel, shopping and others, and the information is much and scattered. Thus, in consideration of the users' demands of information retrieval, the form of interface is particularly important, which should not only fully reflect the organization structure of ontology knowledge base, but also provide users with a friendly and interactive platform.

The ultimate goal of website design is to better serve users and meet the needs of users. Thereby, interaction is the main objective of website design [15-16]. The method proposed in this paper fully explains the good interaction between users and machine. The sample page of semantic website is shown as Fig.8.

Taking the users' search of *Townlodge Roodepoort Hotel* as an example, users can get the information of data attributes in ontology such as the city, the price etc. At the same time, the website also offers users attractions around the hotel, nearby restaurants and other links for users to query and access to. The link information results from the user needs, organization way of ontology and the reasoning rules, and fully reflects the advanced nature of the semantic web technology, with better interactivity and usability compared with the traditional site.

VI. Conclusions

This paper proposes a new model of constructing semantic website, and provides users with a new, advanced, and interactive web browsing way, involving many techniques, such as domain ontology construction method, retrieval and reasoning etc. The test showed that the sample website could meet users' retrieval demand for African hot tourism, and its effect was good.

Construction of the semantic website is a field covering much knowledge and there are many issues needed to be further researched. The proposed method has been proved in the semantic website of African hot tourism, which can also be applied to the construction of other semantic websites. Furthermore, this method needs further verification and improvement in the process of the construction of other semantic websites.

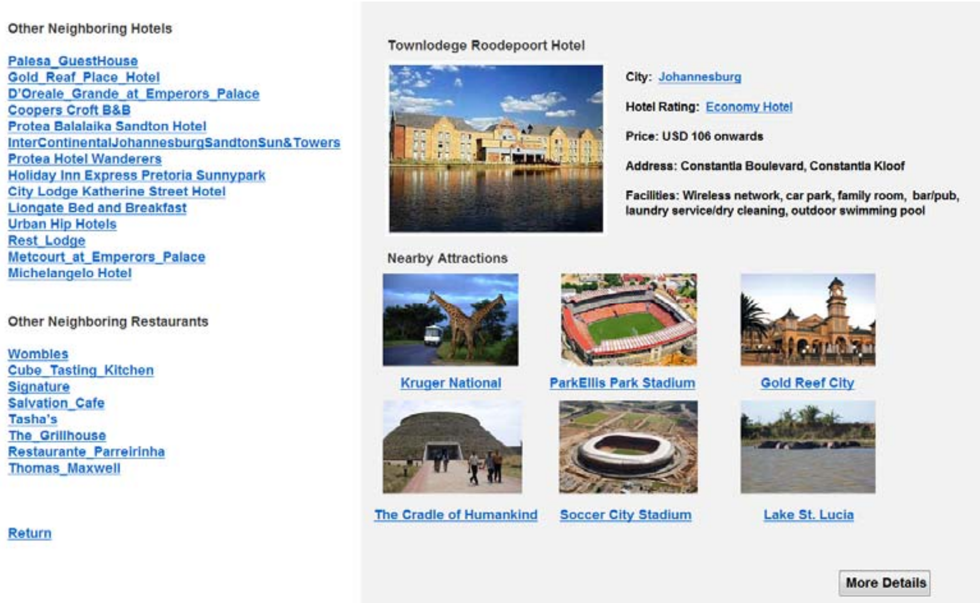

Figure 8. Model run example

ACKNOWLEDGMENT

This paper is supported by the National Nature Science Foundation of China (No.61371194)

REFERENCES

[1] Yingjie Song, Rong Chen, Yaqing Liu, A Non-Standard Approach for the OWL Ontologies Checking and Reasoning, Journal of Computers, Vol 7, No 10 (2012), 2454-2461.
[2] M.Greaves, P.Mika, Semantic Web and Web 2.0, Web Semantics Sci Serv Agents World Wide Web, 2008.
[3] Davies J, Fensel D, van Harmelen F (eds). Towards the Semantic Web: Ontology-Driven Knowledge Management. John Wiley & Sons, Ltd, 2003.
[4] D.Roman, U.Keller, H.Lausen, J.deBruijn, R.Lara, M.Stollberg, A.Polleres, C.Feier,C.Bussler, D.Fensel, Web services modeling ontology, Journal of Applied Ontology 39 (1) (2005) 77–106.
[5] Jie Liu, Yun Ma, Liming Luo, ZhengTao Yu, Reuse of Chinese Domain Ontology for the restricted domain Question Answering System, Journal of Computers, Vol 7, No 11 (2012), 2684-2690.
[6] WANG Tengfei, ZHANG Guojin, LI Shijin. Semantic web services discovery based on domain ontology. World Automation Congress (WAC), 2012, pp. 24-28.
[7] "Appendix B. OWL Example in XML Syntax", [Online] Available: http://www.w3.org/TR/owl-xmlsyntax/apd-example.html, January, 2003.
[8] McBride, Brian, "An Introduction to RDF and the Jena RDF API", 2007. http://jena.sourceforge.net/tutorial/RDF_API/index.html.
[9] Goyal, Sunil, Rupert Westenthaler, RDF Gravity (RDF Graph Visualization Tool), 2009. http://semweb.salzburgresearch.at/apps/rdf-gravity/.
[10] José María García, David Ruiz, Antonio Ruiz-Cortés. Improving semantic web services discovery using SPARQL-based repository filtering. Journal of Web Semantics , 2012.
[11] Olivier Corby, Leila Kefi-Khelif, HacèneCherfi, Fabien Gandon and Khaled Khelif, "Querying the Semantic Web of Data using SPARQL, RDF and XML",inria-version 1,February, 2009.
[12] Jun Zhai, Jianfeng Li, Yan Lin, Semantic Retrieval Based on SPARQL and Fuzzy Ontology for Electronic Commerce, Journal of Computers, Vol 6, No 10 (2011), 2127-2134.
[13] Tengfei Wang, Guojin Zhang,Shijin Li. Semantic web services discovery based on domain ontology. World Automation Congress (WAC), 2012, pp.24-28.
[14] Wu, J.: Research and its Implementation of Semantic Retrieval System Based on Domain Ontology. Taiyuan University of Technology, May 2010.
[15] Jake Rocheleau, Guide to Building Semantic Web Forms, 2012. http://dzineblog.com/2012/01/guide-to-building-semantic-web-forms.html
[16] Ankit Mathur, Web 3.0: Building the Semantic Web, 2012. http://www.linuxforu.com/2012/05/web-3-0-building-the-semantic-web/

Yu Guo is a Master degree candidate whose specialty is Software Engineering at College of Information and Engineering, Capital Normal University, Beijing, P. R. China. She received Bachelor degree in Intelligence Science and Technology at Capital Normal University. Her main research interests are ontology application and semantic web technology.

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 Semantic Web and ontology application. *The corresponding author. Email: liujxxxy@126.com.