

Construction of Politically Sensitive Information Ontology and Semantic Reasoning based on Protege

Yefan Wu^{1,a}, Shujun Li^{1,b}, Ying Jiang^{1,c}, Runbo Zhuang^{1,d}, Zejia Pan^{1,e} and Weichen Li^{1,f}

¹School of Beijing Normal University, Zhuhai 519000, China.

^am13726279603@163.com, ^b13726276332@163.com, ^cjpz6311whu@bnu.edu.cn,

^dzhuangrunbo081@gmail.com, ^epanzejia@iflin.cn, ^fleeweichen11037@163.com

Keywords: Politically sensitive information, ontology, semantics, consistency, reasoning.

Abstract. Ontology construction is a hot topic in semantic research. This paper proposes constructing the domain ontology of political sensitive Information with Political Information as knowledge source data. By the using of ontology itself contains the description logic consistency reasoning verification mechanism, the constructed ontology library can be applied to multiple fields, such as performance evaluation, marking system, semantic network retrieval semantic proofreading and so on. It shows that the research and implementation of the construction of sensitive information ontology based on Protege and semantic reasoning has important practical significance and extensive application prospect.

1. Introduction

1.1 Background and significance of research

Text is an important carrier of human social information, and the information transmission of the Internet is also based on the natural language text. With the rapid development of information technology in the whole society, the importance and urgency of distinguishing the correctness of text information becomes more and more obvious. China has a huge number of traditional media industries, hundreds of newspapers and printing press published a large number of publications which have a non-general influence, the wrong political sensitive information may even cause a country's instability, disharmony. The Internet has become the most influential, the most promising new media in the field of information dissemination. But most of the Internet users do not like politicians or professional journalists who have sufficient political sensitivity of cultivation, So users are more likely to make political errors in the information dissemination, and inadvertently destroyed the security of the Internet environment. But at present, the research on the logical reasoning of Chinese text on the semantic level of china is still relatively weak.

Ontology, as a modeling tool that can describe the conceptual model of information system on the semantic and knowledge level, has become a research hot-spot in the field of information system and artificial intelligence[1]. Ontology provides a way to standardize the description and modeling of information and knowledge, and it also has a good conceptual hierarchy and support for logical reasoning, which is of great significance in building intelligent semantic logical reasoning.

1.2 Research status at home and abroad

1.2.1 The research status at foreign

The systematic analysis and Research on the construction method of domain ontology in foreign countries has been more mature. At present, there are nine main methods of building ontology abroad, according to ontology construction method chanting time to sort are : ① IDEF4 method; ② The skeleton method; ③ the TOVE; ④ METHONTOLOGY method; ⑤ KACTUS engineering

method; ⑥ SSENSUS method; ⑦ seven-step method; ⑧ The loop acquisition method; ⑨ five-step loop method. Through the literature survey can be learned that the first comparative study of several foreign methods is Yang Qiufen. Li Jing added the IDEFS method and the seven-step method to the comparison research of the construction method of the knowledge. Comparing with the standard of IEEE1074-1994, he analyzes the imperfectness of the existing knowledge ontology construction method and the difference with the software development, And summarizes the principles of knowledge ontology construction[2].

1.2.2 Domestic research status

China's ontology research is more late than abroad, the technology is relatively backward, so we learn more ontology construction methods from foreign, and add new content to it, thus forming a new perspective method, especially in constructing domain ontology by ontology Engineering method is the most obvious. Although the method of building ontology of domestic domain has not yet formed a very mature system, some methods have been relatively mature and have certain influence. With the rapid development of network technology, in order to improve the efficiency of information retrieval, Lin Jingliang of Nanjing Agricultural University has researched on the semantic search system based on domain ontology, from the traditional syntax-oriented structure and structure of the key words for the semantic concept reasoning, so that the search of the blog resources up to the semantic level[3]. Thereby which can effectively improve the retrieval efficiency of the blog resources. So far, most scholars have been using the domain ontology to achieve semantic retrieval, semantic web application, knowledge engineering, personalized recommendation, system modeling, digital library and so on. however, the semantic logic reasoning function of natural language has not been realized by constructing the politically sensitive information ontology library.

2 The Significance of Research on Ontology Logic Reasoning

2.1 An overview of politically sensitive information

Political information includes political parties, social groups, social forces, and policies, policies and activities in dealing with national life and international relations, as the most complex and demanding content of the information component. Politically sensitive information refers to news about political issues and controversial issues published by non-authoritative government agencies or media, including controversies on ideological issues, disputes over political and social systems, disputes on terrorism, religion and ethnic separatism National separatism, nationalism, human rights, social morality, national defense, foreign policy and other aspects of major differences and other relevant information[4].

2.2 Research content based on ontology semantic logic reasoning

In this study, the main purpose of constructing politically sensitive information based on protege is to carry out Chinese text semantic logic reasoning. Because the Chinese language is profound and the sentence pattern is ever-changing, we choose to build an ontology library that is similar to the knowledge base. The knowledge engineering of ontology mainly includes knowledge acquisition, knowledge representation and knowledge reasoning. The first step of modeling is to extract the semantic content of politically sensitive information from unstructured Chinese natural language, and then transform the extracted semantic content into structural ontology in protege. Because the ontology reasoning language contains the consistency reasoning verification mechanism based on the description logic, the extracted semantic content is input into the inference engine according to the predetermined order and the domain ontology of the regime. the logical consistency verification reasoning is carried out successively, and the Chinese semantic content of the logical inconsistency in the reasoning result is identified, it is easy to find errors and correct errors in time, so as to achieve the consistency of the Chinese text political logic semantic reasoning validation function.

2.3 Research ideas based on ontology semantic logic reasoning

Step1: Knowledge source data collection. Knowledge source data is the information resources that collected to establish the whole political semantic network. This study collects information

about politically sensitive information, which refers to inappropriate information when it comes to issues such as the title information of national leaders, Hong Kong, Macao, Taiwan, territorial sovereignty, and ethnic groups.

Step2: Information filtering. The further screening of information. Since the amount of information collected is also more confusing, and not all of the information is in the semantic level errors, so the first thing is to delete the error messages not a semantic directly.

Step3: semantic extraction and classification. Processing the specific information content, extracting the semantic content from the unstructured Chinese natural language, and then transforming the extracted semantic content into the structured ontology form. That means extracting key words from a sentence which has semantic error. Generally, three words are selected to form a triad[5], and each keyword is combined and classified. Thus forming the semantic types and semantic relations needed in protege ontology construction.

Step4: Construction of domain ontology. The protege[6] is the main technical support platform to build ontology of politically sensitive information domain. It provides a visual method to support the editing, browsing and updating of ontology class and instance[7]. This paper mainly refers to the most popular ontology construction method: seven-step method.

Step5: Chinese word segmentation tool selection. The author choose Stanford-Parser[6] which is a parsing tool introduced by the Stanford University Natural Language Research Group, It analyzes the structure of the sentence, and marks the different components of the sentence. It is specific to a word segmentation unit, but also tagging them with POS tags, the word function is relatively complete.

Step6: Jena reasoning. Because the semantic relation content that needs to be detected can not be directly reasoned with the ontology library, So we need to use Jena to create the ontology model, use the ontology model to write the extracted triplet into the ontology information by Jena interface, Constructs the new OWL file as the Data to be reasoned, stores it to the hard disk, reads the ontology library file through the ontology model created by Jena and binds it to Schema data, then use the Jena inference rules to match the ontology library with the Data file to achieve the consistency verification reasoning.

Step7: Test the ontology library. Using Jena to mark and print out the reasoning results in the logic of Chinese semantic inconsistencies. Judge the detection result whether the detected logical inconsistency error meets the expectation requirement. If yes, the logical expression written in the ontology library is preserved; otherwise, the logical expression can be modified to be feasible.

2.4 Research innovation point based on ontology semantic logic reasoning

This research is based on the ontology consistency to realize the Chinese text semantic logic reasoning, there are two aspects of innovation and three significant advantages. One of the innovations is the implementation of semantic proofreading of Chinese text, and the other is the innovation of domain ontology construction.

Innovative point one: In the past, most of the methods used in automatic proofreading of text were to build a model to match the text information and calculate the error probability of the sentence according to some parameters. The method used by the author is different from the previous scholars' research methods. By constructing the ontology library, inputting the semantic content with predetermined order and the correct domain ontology to the inference engine in sequence, And the Chinese semantic content of logic inconsistency in the reasoning result is marked to realize the logic reasoning of Chinese text semantics.

Innovation point two: previous scholars to build ontology library is mainly semantic Web application and semantic retrieval for the purpose. Ontology is a hot topic in current semantic research. The former scholars construct the ontology content of the corresponding field according to the research purpose of these two aspects, and construct ontology such as acupuncture, medicine, chemical redox reaction, new media etc. The author constructs an ontology by using political sensitive information as a data source, and verifies the consistency of ontology to achieve many functions, such as Chinese text semantic proofreading of political sensitive information and political response grading system.

For existing technology, this study has significant advantages such as the completeness of semantic model, the transparency of semantic checking and the intelligence of semantic reasoning.

① semantic objects in text are extracted by ontology technology. The semantic errors and the causes of errors are determined accurately. ②The integrity of semantic model, the author select the OWL1 ontology which contains the Boolean logic, the individual value, the cardinal number, the semantic model of the semantic model, Object definition domain / value domain and other dozens of semantic definitions. (Semantics collocation can use Domain / Range to establish mapping); ③ Intelligence of semantic reasoning. OWL1, its ontology reasoning is based on description logic, which has built-in consistency logic reasoning mechanism, its ontology reasoning algorithm is relatively mature, can compatible with the pellet and other description logic reasoning mechanism for intelligent automatic proofreading semantic errors, without additional exhaustive Sentence or the establishment of fuzzy semantic skeleton.

3. Research on Domain Ontology Modeling of Political Sensitive Information

3.1 Requirements analysis

Text is an important carrier of human social information. The processing of natural language is also very important. By reading a large number of documents and the survey found that political class information is a semantic level is relatively easy to make mistakes and difficult to detect. And semantic logic errors have four major characteristics: territoriality, nature, concealment and severity[8]. The semantic logic reasoning can solve the problem of logical reasoning missing at the semantic level. The author lists several political ontologies can be applied to the following field:

(1) In China, the publication of political sensitive information such as newspapers and publishing houses is very common. By using the semantic logical reasoning mechanism, it can reduce the workload of the staff concerned and avoid such semantic errors. And decrease the threat posed by semantic errors in society.

(2) The construction of politically sensitive information ontology can also be applied to political marking, and the frequency of inconsistency can be scored by logical consistency reasoning. If the frequency is high, it means that there are many political semantic logic errors in the answer, then the score will be relatively low; if the frequency is not high, will get the high score. So you can greatly improve the marking efficiency.

(3) In many party and government organs and enterprises, political and ideological construction is also an important part of the performance appraisal, the use of the logical reasoning mechanism of the political ontology performance evaluation, In saving labor costs, time costs at the same time more fair and equitable to achieve the purpose of performance appraisal.

Constructing ontology of different domains has different characteristics. The purpose of constructing the ontology of political is to make semantic logical reasoning on the Chinese text information, so it is a targeted search to collect the expression forms that are prone to semantic errors, to transform the unstructured language into computer-readable The structured language is incorporated into the ontology .In this regard, the construction of the field of political sensitive information ontology is in line with social needs, user needs, has a certain practical significance.

3.2 The Selecting of the concept of field words

In determining the concept word of the political field, because of the particularity of the contents of this paper, which need to extract the correct and error of two aspects of political information. The author chooses the important terms in the political field from the "Chinese Classified Thesaurus" to ensure the quality of the terminology and the completeness of the ontology. On this basis, the author also refers to the other resources to determine the concept of the political field of words.

After a series of analysis and collation, the author defines the 51 concept words in the political field as follows: "character, Characters collectively, values, meetings, the Communist Party, other, system, country, wrong country formulation, national leaders, national defense, regional, wrong regional formulation, diplomacy, market, ideology, time, institutions, the wrong formulation of institutions, rights, types of cases, ethnic, the wrong formulation of nation, legal judgments,

law-related personnel, Taiwan-related normative terminology, Taiwan-related wrong languages, Hong Kong-related normative terminology, Hong Kong-related wrong languages, Hong Kong, Macao and Taiwan regions, purposes, organizations, the wrong formulation of organizations, economic, economic entity, charges, the wrong formulation of charges, groups, communities, the wrong formulation of communities, professional title, title, the wrong formulation of title, acts, regulations, documents, proceedings, the wrong formulation of proceedings, trend, revolution, territorial dispute, commotion”.

3.3 Define facets of property and property of classes

Ontology has two main properties: Object Property and Data Property. The definition and definition of property is an important part of ontology construction, from the internal factors of things to illustrate the concepts and categories of the meaning and internal and external relations, but in the process of constructing political domain ontology, only the Object Property is used. The author draws from the sentence that produces the political category semantic error, selected many Object Property of the political domain class, mainly as follows: “ not allowed, not belong to, not, not belong, to all, not accept, can not object, can not contract ,can not directly intervene, with the border of Shenzhen, with organizations, among, the subject is, represent, occupy,take advantage of, publish, feed, sentence, others, yours, join, assist, contain, contain all, participate, oppose, to learn, to inform, and, and countries, and regions, and our country, and organizations, groups, united union, around, flag, Ministry of Foreign Affairs,university, school, Official documents, realize, the essence is,to Hong Kong, lead, belong to, belong to all, belong to the state, belong to all of them, construct, build, develop, extradite, the president is,served as president, arrest, contract, admit, Investment, in the report, report, to serve as, arrest, detention, have,have a passport, put forward, nominations, political organizations, have no right to, promise, have no right to organize, have no right to, transfer, is, is a part of, to protect, to develop, is legal, is unique, is the goal, to exercise, is the territory, the highest position, the highest authority, apply, apply for a passport, the, the goal of the establishment, of the document,of the organization,of the,nation,of the law,of the organization,of the title,Of the primary goal,direct intervene,show good to, call, sign, performance, by, was opposed, was arrested, was directly intervened, discuss, access, transfer, change, conduct, withdraw, elect, pass, arrest, change with the development, promulgate, in China, institution stationed in China.”

4. Build the political domain ontology based on protege

4.1 The construction of the political domain ontology class

The construction of ontology can be completed in a short time, often need to build a lot of time and energy, through continuous research and practice to step by step to modify and improve it. The domain ontology is constructed by using the tool protege 3.5. The author here only to a level Four category graph constructed as an example, as shown in fig. 1.

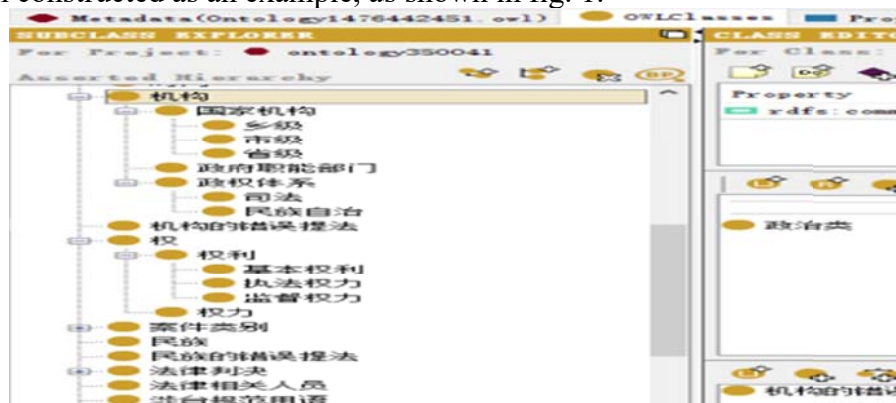


Fig. 1 level Four category graph

4.2 Application of inference mechanism

(1) Disjoint with constraints. This is the most reasoning is the most used one. Countries and regions are semantically disjoint classes, and can be semantically constrained by using Disjoint with.

Figure Domain as semantic constraints, which is the only federal state has served as president in the "担任总统"(served as president) Range semantic constraints of semantic relations, so as not to have the president of this area is the title, if there is "马英九担任台湾总统"(Ma Ying-Jeou served as president of Taiwan) that will appear inconsistency semantic error.

(2)Cardinality constraints. A region can belong to only a country and can not belong to two countries at the same time. Such as the territorial dispute Diaoyu Islands region, Diaoyu Islands is regarded as Domain, the property is belongs to,the Domain semantic constraints is the country. So it is necessary to carry out the Range part of cardinal constraints, the use of MaxCardinality=1 or Cardinality=1 cardinal constraints, making the Diaoyu Islands must belong to a country, can not belong to two countries. If the Diaoyu Islands belong to China and Japan, then it will reasoning China and Japan is same, there is no logical mistake itself. So you need to add Chinese is different from Japan, China and Japan are different from each other, then the consistency of reasoning will produce a logical error: the cardinality reasoning that they are the same, but the two are not the same as specified in the different from.

(3)A variety of expression forms. Chinese language and culture expression is ever-changing, broad and profound. For example, "钓鱼岛属于日本"(Diaoyu Islands belong to Japan) the formulation in the text may exist in a variety of forms, there are "钓鱼岛坐落于日本"(Diaoyu Island is located in Japan), "钓鱼岛位于日本"(Diaoyu Island is located in Japan) and other descriptions, obviously these are conveyed a sense of the same, There exists the same political semantic error, it is the correct expression is "钓鱼岛属于中国"(Diaoyu Islands belong to China). The same semantic error types have different forms of expression. Through the analysis of sentence structure, it is found that there are differences in the definition of property constraint, namely "属于"(belong), "坐落于" (located in), "位于"(located in) , "是...的"(is)and so on. Apparently they express the same semantic meaning. Therefore, the author use the function label that called add-label in property, No matter any kind of property semantic relation of that appearing, it can conclude the error contents successfully.

5. Test the ontology library

5.1 Test procedure

The author adding triples to the Jena test directly, and print out the results of the test results. There are two types of results. One is that there is no logical conflict, it indicates that the detected triplet content is consistent with the constructed correct ontology library semantic constraint. One is the result of logic inconsistency, this situation requires manual judgment, two cases: ①The contents of the sentence is wrong to detect is expected to want the wrong content; ②The content of the sentence error is not expected to be detected; If the result of ②, the corresponding content in the ontology library is modified and perfected. The detection of logical consistency reasoning of ontology base by adding three triple content, according to the results of continuous modification and improvement of the construction of semantic content in the field of politically sensitive information ontology library. The test procedure is shown in Fig. 4.

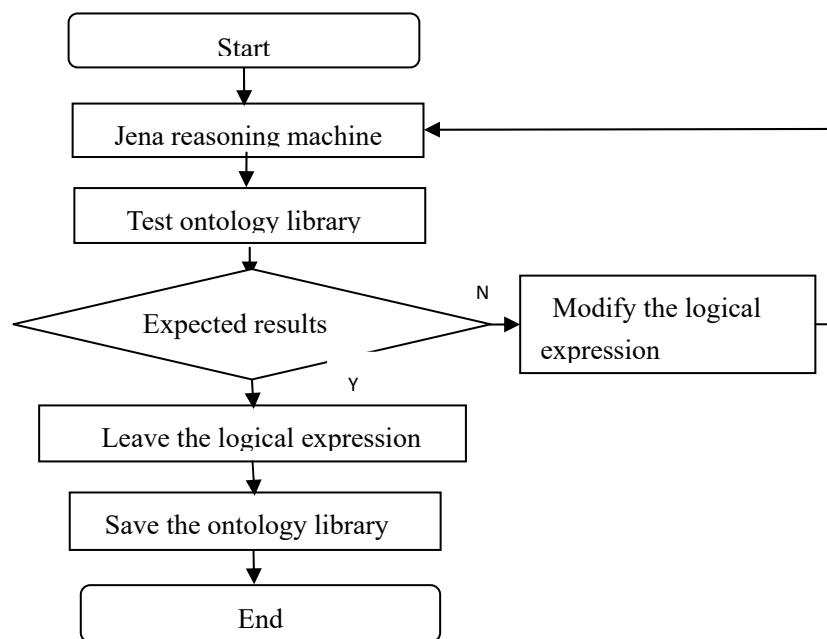


Fig. 4 Test procedure

5.2 Examples of test data

Table 1 Examples of test data

Logical error type	Domain	Property	Range	Disjointwith	Number of instances
Range semantic constraint	李克强 (Keqing Li)	担任 (Served)	总经理 (General manager)	职场职称 (Professional title)	8
	村里 (The village)	选举 (election)	村长 (village head)	职称错误提法 (flase Professional title)	8
	日月潭 (Sun Moon Lake)	是领土 (Is the territory)	中华民国 (zhonghuamin guo)	涉台错误用语 (Taiwan - related Wrong language)	40
Domain semantic constraint	中国 (China)	助力于 (Powered by)	香港 (Hong Kong)	国家 (country)	1
	欧共体 (European Community)	召开 (Held)	会议 (conference)	组织错误提法 (Misrepresentation of organization)	2
	台湾 (Taiwan)	总统是 (The president is)	马英九 (Yingjiu Ma)	地区 (area)	67
Value constraint conflict	南海 (the Nanhai Sea)	属于 (belong)	菲律宾 (Philippines)	中国 (China)	13
	南沙群岛 (Nansha Islands)	不是 (not)	中国 (China)	越南 (Vietnam)	13
	法律 (legal)	是保护 (Is protected)	弱者 (Weak person)	守法者 (Law-abiding person)	8

In the ontology database, the semantic logic conflicts appear in the form of three kinds of logic errors, which are shown in Table 1, the author extracts some examples from the three kinds of logic

error types as examples. The italics are representation the individual which occur the logic conflict in table1, such as “李克强担任总经理”(Li Keqiang take office manager), Prime minister belong to “title”. There is the inconsistency logical conflicts because of “Title” disjoint with “workplace title”. The principle of Domain semantic constraint conflict is similar to the Range semantic constraints. In the conflict of value constraint, such as detection “南沙群岛不是中国的”(Spratly Islands does not belong to China), can’t make the difference between them use of the disjoint with between the class and class because of China and Vietnam both belong to the republic country. Which achieved defined the different from the two individual in the same class by use “differentfrom” and MaxCardinality=1, and can be inferred logic errors through the consistency reasoning.

A Property can associate two classes in a specific way. The definition of Domain and Range semantic constraints are implementation based on the definition of class, there are multiple individuals in a class, so the author in the test data examples only enumerates an individual information in Table 1. The logical inconsistency conflict can be deduced by reasoning on the logic consistency of the ontology itself in the Jena engine, which can identify and print the sentences exists logically conflicting.

6. Summary

6.1 Research Conclusion

In the course of constructing the politically sensitive information ontology database, the author also takes the sharing and reuse into account. The author collects all kinds of information which may cause the logic error of politically sensitive semantics, to make it covers the political information as much as possible. Political topic is sensitive and common in the country. People are easy to make semantic logic errors in the process of expression, now logical consistency reasoning can be used to verify that if a logical conflict has occurred. The construction of politically sensitive information ontology library can be applied to many fields. For example, it can effectively help political teachers to review papers and so on. The political domain ontology also needs to be constantly updated and maintained. such as the list of national leaders. In short, the construction of politically sensitive information ontology and the realization of semantic reasoning are the essential step in the development of social progress.

6.2 Future prospects of the project

The author constructs the politically sensitive information ontology database, which can be developed to develop the Chinese semantic proofreading software package in the form of independent software; It can be achieved through the expansion of XML rules or Java rules to customize the perspective of integration, and ultimately in the form of plug-ins to integrate into the syntax checking tool. In the later period, unsupervised method can be implemented to extract the semantic content from the unstructured Chinese natural language, and then the extracted semantic content is converted into the structure of the form, so as to realize using the politically sensitive information ontology database to detect the corpus. Then with the plug-in form and syntax checking tool LanguageTool, Microsoft Office, Kingsoft WPS or Office Open to achieve semantic proof of Chinese text, not only can the content of semantic logic errors be identified by logical consistency reasoning, also can print the corresponding suggestions for the user. Therefore, it has broad application prospects and great commercial value for the research and implementation of the protege political sensitive information ontology construction and semantic reasoning.

Acknowledgments

This work is supported by a project from 2016 Guangdong College Students Science and Technology Innovation Cultivation Special Funds (Project No. pdjh2016b0898). It’s also supported by a grant from Science and Technology Plan Project of Guangdong Province (Project No. 2014A080804001) and a grant from National Statistics Science Research Project (Project No.

2015LZ43). The corresponding author is Ying Jiang (jpz6311whu@bnuz.edu.cn)

References

- [1] Li Qingsai. “Research on Construction of Tourism Domain Ontology”[D]. [Master's degree thesis] Zhengzhou: Zhengzhou University .2015.
- [2] LI Jing, MENG Li-sheng. “Comparative Study of Constructing Knowledge Ontology Method System” [J] .New Technology of Library and Information Service, 2004 (7): 17-22.
- [3] Lin Jing-liang. “Research on Semantic Search System of Library Information Based on Domain Ontology” [D]. Nanjing: Nanjing Agricultural University .2010.
- [4] Tan Guisheng. How to Avoid Political Errors [J] .Chinese editors, 2006 (06): 53-56.
- [5] <http://zh.transwiki.org/cn/rdfprimer.htm>
- [6] Protege. <http://protege.stanford.edu>
- [7] Matthew Horridge, Holger Knublauch, Alan Rector, Robert Setvens, Chris Wroe, A Practical Guide to Building OWL Ontology Using the Protege-OWL Plugin and CO-ODE Tools[EB/OL].
- [8] Yefan Wu,Runbo Zhuang,Ying Jiang,Fan Li. “Research and realization of Chinese text semantic correction Based on Rule” [J].Proceedings of 2015 3rd International Conference on Education,Management,Arts,Economics and Social Science (ICEMAESS 2015).