The application of ontology model in knowledge retrieval and recommendation

Liang Zhang[1], Yuanfeng Zhang

Abstract. The traditional way of keywords-based information retrieval mainly uses string matching to obtain search results. This kind of means ignores the semantic relations between keywords and retrieval results so it is difficult to express the real searching requirement of the users. For the above problem, this paper brings forward a novel knowledge information retrieval mechanism based on ontology. By applying the built course knowledge ontology to the Moodle platform, this paper implements a knowledge retrieval engine, which can get concepts information having semantic relations with the input keywords. Thus the search engine is completely different from the traditional keyword-based knowledge retrieval, and it has some features of semantic retrieval. This engine facilitates the establishment of concepts network for learners, with which the learners can easily obtain the semantic relations among concepts.

1 Introduction

Under the massive amount of information in the network, people usually fall back on various retrieval platforms to obtain all the necessary information resources. The existing various retrieval ways, analysing the nature of retrieval thought, we can see that they are essentially based on string matching, namely they all belong to "keywords retrieval". The drawback of keywords retrieval technology is obvious. Search engines take the content of retrieval only as several independent keywords to search results for return, so they often ignore the semantic relationships between retrieval keywords and their results. Although many search technologies have already adopted excellent word segmentation technology to establish index, there will still be a lot of useless information in the returned results. Assuming

[1] L. Zhang (✉)
School of Information Science and Engineering Hangzhou Normal University Hangzhou, China, 310017
e-mail: zhl@hznu.edu.cn

such a situation, a document with the theme of "calculating machine" and without a word of "computer" throughout it, using keyword-based string matching method, when the user input "computer" as keywords to retrieve, the document is clearly not hit, even though we all know that "calculating machine" and "computer" convey the same meaning in most cases. Thus, the effect of the above retrieval pattern is not always satisfactory.

Tim Berners Lee, the father of semantic web, put forward the concept of semantic web in 2002, and pointed out that it would become the next generation Internet. The ideas and technology of semantic web brings an opportunity for solving the above problem, and also points out the research direction for semantic retrieval [1]. Ontology, as a conceptual modelling tools, which describes information in the semantic level and has logical reasoning ability. We can use these characteristics of ontology to extend and reason user retrieval intention, in order to improve the precision and efficiency of the retrieval. Therefore, in the field of information retrieval, especially in knowledge-based retrieval, ontology has been widely used. In the retrieval based on ontology, the first step is to build a comprehensive ontology model, then storing the model to a stable storage medium as the basis of search. Facing the massive information in "information explosion" era, this is a great work of task, and cannot come easily. We firstly stored the curriculum ontology model of computer organization into Mysql Database, and then applied the knowledge model to Moodle platform for learners. For the proposal of knowledge retrieval and recommendation, we develop a knowledge search module, though which learners can get knowledge information having basically semantic relations with the input keywords. It can be seen that this module has some features of semantic retrieval, which is completely different from the traditional keyword-based retrieval.

2 Related Work

Currently, the semantic-based knowledge representation is a hot topic. The research of ontology which reflecting the concept of the Semantic Web is also increasing. Ontology as a conceptual modeling tool, which can describe the information system in semantic and knowledge lever, has caused the attention of researchers at home and abroad. At present, ontology has been widely used in the field of knowledge engineering, intelligent information retrieval, information integration, digital libraries, and natural language process [2].

In the ontology-based information retrieval, in a sense ontology is the knowledge model. The design ideas can be roughly described as: under the participation of computer science experts and ontology engineers, constructing domain ontology based on the concept of the field, then continuing the domain ontology related data in accordance with the prescribed format to the storage medium, and finally with

the help of ontology knowledge model, getting the query results from user search interface.

Facing massive data, the way of ontology storage management directly affects the efficiency of ontology's storage and query. Thus, the study of how to store large-scale ontology efficiently is significant, which is the basis of ontology-related applications. There has been a lot of ontology storage management systems, according to storage media can be divided into: memory storage, text storage, relational database storage.

At present, in the study of knowledge management and knowledge information retrieval based on ontology, famous research projects have: On-To-Knowledge, Ontoweb, Ontobroker, OntoSeek[3]. In addition, there are also some other research results available for reference. [4] proposed a ontology-based semantic retrieval model, verifying the validity of the model by building a Web ontology-based intelligent retrieval prototype system. [5] designed a domain ontology-based information retrieval system, using the domain ontology to understand and expand the keywords for the subsequent information retrieval. [6] realized semantic reasoning and retrieval based on the established curriculum ontology. Throughout above literatures, although the ontology-based retrieval research has made significant progress, many systems is still limited to laboratories. The research on the practical application of ontology is also not too much. This paper explores the use of ontology knowledge model in the Moodle platform, taking the development of a knowledge search forum as an example to illustrate the implementation of ontology-based knowledge retrieval in the Moodle platform.

3 Using Ontology Knowledge Model for Knowledge Retrieval

Protégé[7] was used to construct the ontology knowledge model of computer organization. Protégé is an ontology editor which provides a suite of tools for the construction of domain models and knowledge-based applications of ontology. The process of building ontology mainly contains four steps: (1) the creation of new project, project type such as OWL/RDF files can be selected from the options; (2) the definition of classes and their hierarchy; (3) the definition of attributes and allowed values; (4) the addition of instance. In the initial stage, it's needed to abstract concepts of the curriculum, and then extract relationships among different concepts, finally establish abstracted classes (including concepts' definitions, attributes and relations) according with the syntax of ontology knowledge representation. WorldNet, a quite common ontology, has considerable influence in the international computing linguistics. In order to improve the sharability and reusability of the established curricular ontology, this paper borrows ideas from WorldNet when constructing the ontology knowledge model

This paper adopt the top-down method to divide concepts in the curriculum from the top concept "computer system" downward detailed layer by layer to form the overall conceptual framework of the course. Protégé can define two types of attributes for concepts: Static properties and dynamic properties. Static properties describes the meaning and characteristics of the concept, here mainly define the following static properties: concept identifier (C_Id), concept type (C_Type), concept objective (C_Objective), concept difficulty (C_Objective), concept media Link (C_mediaLink) , father concept set (C_Father), child concept set (C_Children), subsequent concept set (C_Reference). Dynamic properties refer to specific definition, computing, process, events of call, which describe binary relation between concepts. The mainly defined relations include:

1. Generalization relationship (isSuperOf): Concept A is higher than concept B in generalization degree or contains degree.
2. Specialized relationship (isSubOf): Concept A is lower than concept B in generalization degree or contains degree.
3. Appositive relationship (isParallelOf): Concept A and concept B have common father knowledge point.
4. Equivalent relationship (isEquivalentOf): Concept A is exactly the same as concept B.
5. Component relationship (isPartOf): Concept A is part of concept B.
6. Instantiation relationship (isInstanceOf): Concept A is an example of concept B.

Some of relations among different concepts around "Instruction System" in the ontology graph are showed as Figure 1.
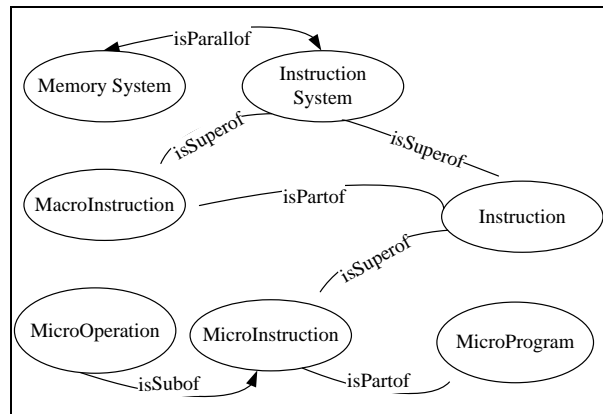

Fig. 1 Some relations among concepts in the ontology

4 Using ontology knowledge model for knowledge retrieval

When users put forward a inquiry in the knowledge-based environment or system, the inquiry which contains what the user want to know or learn can be named as knowledge requirement[8]. Knowledge requirement includes explicit knowledge requirement and implicit knowledge requirement. Explicit knowledge requirement is the surface, intuitive knowledge needs of users, while implicit knowledge requirement is the potential ones. Users' knowledge query can be divided into the input query concept itself (explicit knowledge requirement) and the concepts related to the input query concept (implicit knowledge requirement). Traditional keyword-based query forms can better satisfy the demand for explicit knowledge, but for the potential implicit knowledge requirement, it could do nothing. The query based on ontology knowledge model achieve above two kinds of inquires: The defined ontology classes and properties can be applied for the query directly against the concept itself, and the defined various binary relations for the implicit knowledge query.

The query for explicit knowledge requirement

When the user explicitly express their query, namely users know the request entity in certain degree, then the system only need to return the entity directly related to the input. In the ontology model, the query for concepts and attributes belongs to this type of query. For example, when the user input "macroinstruction" as request to retrieve, for the query of concepts, the retrieval engine will return entities which introduce or describe it in details; For the query of attributes, the retrieval engine will return entities which have an attribute whose value is equal to it.

The query of entity concept

This type of query only needs to return the concepts directly matched with the input concept. The basic idea of ontology-based query for this type is as the follow:

The query concept is denoted as Q_c, ontology concepts set denoted by $E(e_1,e_2,e_3,…,e_n)$. Traversing set E, judging whether e_i matched with Q_c, if matching, then the concept corresponded with e_i will be included in set $O(o_1,o_2,o_3,…,o_n)$ as output.

The query of entity property

The properties are specific description from various aspects for entity. The query for entity property returns concepts which contain a property whose value is the same with the input one. The basic idea of ontology-based query for this type is as the follow:

The query of concept's property is denoted as Q_p, ontology properties set for each e_i denoted by $P(p_1,p_2,p_3,\ldots,p_n)$. Traversing set P of each e_i in set E, judging whether p_i matched with Q_p, if matching, then the concept corresponded with p_i will be included in set $O(o_1,o_2,o_3,\ldots,o_n)$ as output.

The query for implicit knowledge requirement

Sometimes the users may not clearly express their demand for the actual query. In other words, they may only enter with an entity related to the concept of the actual request. In this occasion, the mining for users' potential requirement is needed. Ontology-based queries can take advantage of the defined relations between concepts to deduce the user potential query.

The basic idea of ontology-based query for this type is: After explicitly matching, the output of concept set $O(o_1,o_2,o_3,\ldots,o_n)$ has been obtained. In the next step, traversing set R for o_i in set O, finding the concepts which have strong binary relations with o_i, putting them into $I(i_1,i_2,i_3,\ldots,i_n)$ as output.

To ensure the completeness of semantic association between concepts, mutual information can be used for studying the potential association between knowledge of concepts. Mutual information is applied to measure the degree of association between two events which can also be used to measure the association between two random variables. Mutual information between the two events can be expressed as the probability function of two events. The mutual information of random variables x and y is denoted by MI(x,y), then MI(x,y)=log(p(x,y)/p(x)p(y)), P(x) and P(y) respectively denote the independent probability of user reading concept x or y, P (x, y) represents the probability when user reading both x and y. The different size of MI(x,y) mean [9]:

- if MI(x,y)>>0, showing that the correlation between x and y is strong;
- if MI(x,y)≈0, showing that the correlation between x and y is poor, just occasionally appears at the same time;
- if MI(x,y)<<0, showing that the distribution of x and y are complementary, there is no correlation.

In case, concept x is take for granted an explicit requirement of user. For deduce implicit concepts, we can calculate the mutual information between concepts by MI (x, y), if MI (x, y) is bigger than threshold value ε then select concept y as implicit one.

The realization of knowledge retrieval based on ontology

This paper discusses the application of curriculum knowledge ontology in Moodle platform, taking the realization of a knowledge search model to illustrate the specific implementation process. Moodle, a course management system based on con-

structivist theory, is developed by an Australia teacher-Martin Dougiamas. As an opening source software, recently Moodle has been widely used in various countries. The development and perfection of forum suitable for practical application is still the focus of attention [10]. In the application of Moodle platform, this paper abandoned the traditional forms of knowledge organization and presentation, achieved a knowledge search engine based on established curriculum ontology.

Building and configuring Moodle platform

Moodle platform is constructed on three-layer B/S structure: (1) the logical layer uses Apache as server to parse the PHP script language; (2) the database layer uses Mysql database; (3) the client uses Ajax in order to provide faster interactive service for users. HTML, CSS and DHTML are applied to show the layout and content of the page, also XMLDB and ADOBD are responsible for defining database objects and masking differences in the underlying database.

The development community of Moodle has a lot of third-part plug-ins developed by a variety of modules that can cope with some situations, but still cannot fully meet the specific needs, so that the blocks customized for specific needs is very necessary. There are three forms of secondary development for Moodle: (1) language pack, providing language packs for the Moodle interface, such as translating en_utf8 to zh_cn_utf8; (2) style theme, providing a different look for the Moodle, such as Theme or Skin plug-like; (3) modules and plug-ins, extending for specific needs, including activity pages, function block, resource types and so on.

For secondary development of Moodle, the first step is to establish development environment, downloading the latest versions for windows in Moodle office website, then running the installation package and configuring the server. When the configuration is successful, users can locally run the Moodle platform, and use the super user to login.

Moodle, as a famous open source curriculum management platform, has its own criterion for developers. The developer must abide by its secondary development specification in order for the stability and consistency of Moodle platform. Below are some of the criteria that developers should observe [11]:

- All the code files should make use of .php as the expanded-name.
- All the template files should make use of .html as the expanded-name.
- All the PHP files sign should be a "full" one, not a "short" one, such as using <?php ?> not <? ?>.
- All the files should include a config.php file.
- No creating and using global variables besides $CFG,$SESSION、$THEME and $USER.

Applying curriculum ontology on Moodle platform

The technology map of this paper is: (1) using protégé to build curriculum ontology of computer organization, (2) storing the built ontology to relational database under the help of Jena, (3) realizing parsed function and releasing them to webservice in order for the call of Moodle development environment, (4) finally achieving ontology analytical content and obtaining the application results.

The implementation process involves the interaction between two languages, for the parsing of ontology is in Java development environment, while the achievement of blocks in Moodle is carried out with PHP scripting language. This paper will release the corresponding Java code to webservice for the call of PHP, in order to solve the communication between the above two languages. After users have input the concept, which will be sent as parameters to the corresponding function of webservice by Moodle, the ontology model will be parsed by means of the interface on the Jena ontology model.

Here we mainly present the query output reflecting implicit knowledge requirement, for the explicit knowledge requirement is also can be achieved by traditional keyword-based query which is not our focus. The realized knowledge search engine has some features of semantic retrieval which is completely different from traditional keywords-based retrieval.

When the user input "instruction system", the search engine will return concepts which have close semantic relations to the input one. These return concepts may reflect users' implicit knowledge requirement to a certain extent. Under the help of this engine, learners can quickly obtain the semantic relations among concepts in order to build concepts network in their mind. Here we just list the retrieval results, later some improvements will be made to sort the results by importance and show them in a better way.

5 Conclusion

Along with the rising and developing of semantic Web, ontology knowledge representation and processing technology gradually becomes the focus of attention in related field, which embody thoughts of semantic Web ideas. The developed knowledge search engine in Moodle platform, it's depended on ontology knowledge model, so it possesses features which are different from traditional keywords-based retrieval. Concretely embodying in its results returned, which to a certain extent reflects the inner semantic association between concepts. But there are still some places needed to perfect. In the future, some exploration will be taken to improve the current work, such as sorting the return by importance and linking concepts with their corresponding resources.

6 References

1. T. Berners-Lee, J. Hendler. And O. Lassila. The semantic web. In Scientific American, May 2001.
2. DU Xiao-Yong, LI Man, WANG Shan. A Survey on Ontology Learning Research. Journal of Software, Vol.17, No.9, Sep. 2006, pp.1837−1847.
3. Ming-Che Lee, Kun Hua Tsai1,Tzone I. Wang. A practical ontology query expansion algorithm for semantic-aware learning objects retrieval. Computers & Education,Volume 50, Issue 4, May 2008, Pages 1240-1257.
4. YIN Huan-liang,SUN Si-ming,ZHANG Feng.Research of Web Intelligent Retrieval Based on Ontology.Computer Engineering, Vol.35, No.23, Dec. 2009, pp.44−49.
5. SONG Er-wei, LIU Zong Tian, XU Li-Bin,CHEN Guang.The Realization Mechanism of the Web Information Retrieval Based on the Domain Ontology. Computer Science, Vol.34, No.5, 2007, pp.104-110.
6. Pan Ying,Wang Tianjiang, Jiang Xueling.Knowledge modeling based on ontology.Computer Applications and Software, Vol.25, No.1, Jan. 2008, pp.267−269.
7. John H. Gennari, Mark A. Musen et al. The evolution of Protégé: an environment for knowledge- based systems development. International Journal of Human-Computer Studies, Vol.58, Jan.200, pp.89-123.
8. Xiaoli Yue, Cungen Cao, jinxi Si, Qiangze Feng. A Meta-Level Elicitation of Knowledge Requirements. KEST'04. pp.273-279. 2004.
9. Wei Qu, Shibing Zhu.the Foundation and Application of Information Theory, Tsinghua University Press, 2005.
10. Eladio Gutiérrez, María A. Trenas,Julián Ramos,Francisco Corbera,Sergio Romero.A new Moodle module supporting automatic verification of VHDL-based assignments[J].Computers and Education,2010,54(2):562-577.
11. http://docs.moodle.org/20/en/Development