Research of Online Intelligent Question Answering System Based on Curriculum Domain Ontology[*]

Wei Sun Shuhan Cheng Hongmei Zhu Weiyan Li

College of Information Science and Engineering,Shandong Agriculture University
Taian, 271000, China
E-mail: sw@sdau.edu.cn

Abstract

At present question answering system mostly take the keyword matching as the foundation, it has ignored the keywords itself contains semantic information, and the keyword matching accuracy is low. But the intelligent question answering system is on this basis, it added the understanding to the semantic relations. In the article we proposed the intelligent question answering system based on curriculum domain ontology, it can significantly enhance the recall and precision of the information retrieval by the semantic inference.

Keywords: Ontology, Intelligent question answering system, The semantic web, Relationships

1. Introduction

Online intelligent question answering system integrates high and new technology such as network communication, artificial intelligence, etc. This makes the behavior to break through the limitation of time and space, answering questions and widely extended the problem solving form and resource acquisition way to present content, resources, provides the high efficiency, high quality of the students' interaction and education resources, not only to support the network teaching, and become the beneficial supplement of the traditional classroom. Intelligent question answering system can answer the students' problems in time, eliminating barriers to students' answering questions, is an important part of the network education platform, it can strengthen the communication between teachers and students, help students to clear problem and really need to get their information has an irreplaceable role.

Intelligent question answering system based on the traditional question answering system to provide a more friendly interface of answering questions, students can ask questions in the form of natural language, system segmentation, similarity computation and semantic understanding fuzzy matching method to give an answer. System involves the natural language processing, information retrieval, such as fuzzy matching technology.

Online intelligent question answering system based on ontology is mainly in view of the special knowledge domain (takes "Database Principles" course for example). The construction of domain ontology is a repeated process of superposition needs to constantly improve.

2. The domain ontology construction process

Process of construction of domain ontology closely involved in related field. This paper takes "Database Principles" course for example to put forward the main processes for professional domain knowledge base construction.

2.1. Establish ontology framework

On the basis of the content of the subject, knowledge framework of this subject is formed. Figure 1 is Part of "Database Principles" course's knowledge framework.
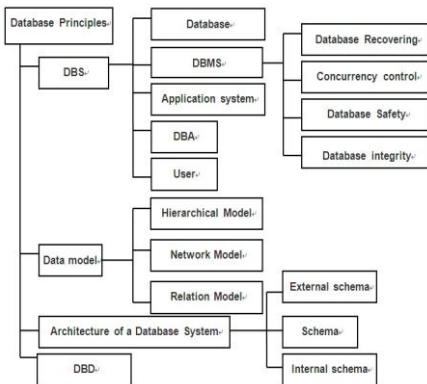

Fig. 1: Database principles course's knowledge framework (part).

2.2. List concepts of professional domain ontology

In general, sememe is the most basic and the smallest unit of meaning which is not easy to divide. For instance, Database Management System is a concept that cannot be split up, so think of it as a sememe.

For professional domain knowledge base, the collection of sememe can extract through the analysis of the keywords which is obtained on the basis of professional keywords of this domain.

First of all, professional keywords are extracted. So-called professional keyword refers to the professional field of words to represent a particular concept, such as "database", "hierarchical databases", "relational algebra", "normal form", "concurrency control", "data independence ", "distributed database system", "deadlock prevention", etc. Then, we can get the exact sememe through investigation and analysis to the extracted keywords. Keywords can be further divided break down into multiple sememe 。 For example, "data independence" can be further broken down into two sememe "data" and "independence"; "deadlock prevention" can be further divided into two sememe "deadlock" and "prevention".

2.3. Determine the relationships between concepts

In the domain ontology, concepts interact with each other through inter-concept relations. HowNet unveils inter-concept relations and inter-attribute relations of the concepts as connoting in lexicons of the Chinese and their English equivalents. In Hownet [1], it describes a total of eight kinds of relationships between concepts: hyponymy, synonymy, antonymy, correlativity, properties-host relationship, meronymy, material-finished product relationship, event-role relationship.

According to the characteristics of question answering system in specific areas, we built the domain ontology using four kinds of relationship: hyponymy, synonymy, correlativity, meronymy [1].

Semantic Web was constructed mainly by professionals who are familiar with the domain. It describes concepts and inter-concept relations. Usually between

concepts of the following four relationships: synonymy (mainly alias and a bilingual); hyponymy; correlativity; meronymy.

According to the four relationships, we built a sememe relationships diagram based on sememes that have been extracted. Figure 2 is part of the sememe relationships diagram. After the construction of the sememe relationships diagram, semantic similarity between sememes can be decided by their position in the semantic web and the kinds of relationships between them [2].
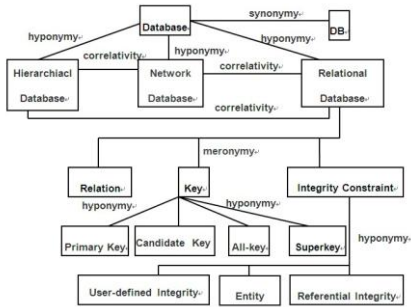

Fig. 2: Manual established sememe relationships diagram.

2.4. Defining the concept of class properties

We use the class to the interpretation of the concept of domain ontology, concept, object properties, and so on was described. Only through the class is not enough to accurately describe a domain, as a result, once defined class will describe the internal structure of this class. Choice in creating the concept of class, most of the rest of the concepts may be these attributes of a class. In addition to define classes must describe concept between the inner structures. If the object is structured, it can be part of the concrete or abstract elements, as well as to describe the relationship between individual members of the class and other classes. In addition to initially determine some of the

properties, also need to add some other properties. Any class of all subclass inherits the attributes of the class [2]. An attribute should be attached on the biggest class which has the properties.

2.5. Definitions of attribute values

Attribute values can be either a numeric value or it can be a class. We will attribute value as a class, is called the attribute class, attribute class also have attributes, through these attributes to that value type, the number of values and the values of the other features [3].

2.6. Create an instance

We create individual instances of concept classes. First, we should choose the class and then create the concepts class instances. As to the concept "database management system", you can add "Microsoft SQL Server" and "Oracle Server", etc. They can be an instance of the concept class.

2.7. The coding and formalization of domain ontology

Choose appropriate ontology description language to encode and formalize the above building of domain ontology [3]. Ontology formalization model can provide stricter than natural language format, can improve the readability of the machine, automatic translation as well as the exchange, and is advantageous for the ontology model, automatically logical reasoning and inspection.

This phase, we use OWL to formalize representation of concepts and relations; the most important thing is to define the classes, subclasses, and properties and their respective features. All ontology is completely constructed by Protégé4.2.0, after the completion of the ontology file suffix for the OWL file format. Figure 3 is Protege4.2.0 automatically generated diagram. From left to right in figure display "Database Principles " classes in the

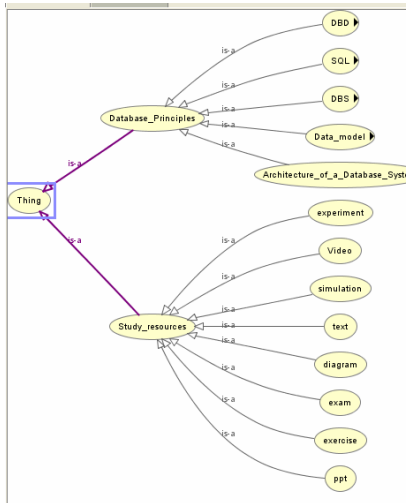ontology level, the class on the right is a subclass of the class on the left .

Fig. 3："Database Principles" class hierarchy ontology diagram.

2.8. Inspection and evaluation of the ontology

After ontology formalization, whether meet the requirements and the established criteria of ontology, whether the ontology of the terms is clearly defined, the concepts and their relationships are complete in the ontology are needed to test and evaluation after the ontology building. Ontology test evaluations require the participation of experts. And ontology library also need to be in the process of the use of the system constantly to improve and maintenance.

3. Model of intelligent question answering system based on ontology

System knowledge base adopts ontology for knowledge representation, problem analysis was carried out on the user's natural language questions, determined by the ontology concepts expressed by the keyword semantic hierarchy, from the semantic understanding, process, and retrieve the user's query request, for guar-

antees accurate solutions of the problem. System framework is shown in figure 4.
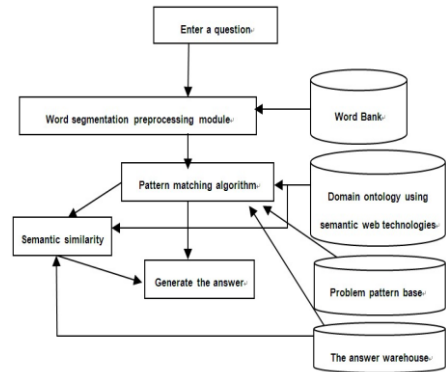

Fig. 4: The intelligent question answering system model.

Students ask questions in natural language, word segmentation pretreatment module to word segmentation problem, corresponding solution to the problem of a group of original sememes, and then organized into a sememe vector. The process used to domain thesaurus and common words. In the process of pattern matching problem, because just relations between the original and the original, and must fully consider the semantic relationships between them. When a student to ask questions, and through the word segmentation module come just after the sememe vector, and from the same material for word segmentation sememe answer material the sememe vector. Established by the first step of the semantic web, which can be concluded that each group of original sememe, the semantic similarity between this problem and the matching of material comes down to the matching problem between questions sememe vector and answers to the questions for material sememe vector. Through a vector matching method can calculate the semantic similarity between two vectors [7]. The highest similarity of a set of meaning of the sememe material is returned to the student.

For example: students' problem: rules of all kinds of integrity constraints.

A. Under the traditional way of answering questions, we might find is about integrity constraints; it cannot find such as entity integrity and referential integrity rules.

B. Under the intelligent question answering system, we can remedy the shortcomings of without the semantic web, because the entity integrity, referential integrity and other likewise concept has a high correlation between "integrity constraints", can find out the answer to the related system.

At the same time, the semantic web can enhance the tolerance of the search, and can also play a heuristic teaching effect. The words because the student questions may not be very standard, in the meaning of original extract process can have limit these irregular words are included in the corresponding original righteousness, so that in the case of a student's questions is not very standard can find accurate answers. System is given and the problems related to a set of answers, do both answered students' questions, and make students can refer to the questions related with himself in advance and did not expect the relevance of the results, has the good effect of heuristic teaching [8].

4. Conclusion

This paper builds an intelligent question answering system based on ontology, due to considering the semantic web, system of intelligent is better. Through the actual test on the system, get better test results. To get better results, of course, have to build a tree in the face of the whole subject of the semantic web, this task is very difficult, it will take a long time for professionals to be familiar with knowledge in the domain, and it is difficult to consider comprehensive, but arti-ficial building ontology is accurate, able to objectively reflect the relationship between the concepts, in the future to consider using machine automatic build and artificial building, make up for the inadequacy of artificial building.

5. References

[1] Mc Guinness, D.: Question Answering on the Semantic Web. IEEE Intelligent Systems 19(1),pp. 82–85, 2004

[2] Dzbor, M., Domingue, J., Motta, E.: Magpie – Towards a Semantic Web Browser. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 690–705, 2003.

[3] Lopez, V., Motta, E.: Ontology Driven Question Answering in AquaLog. In: Meziane, F., Métais, E. (eds.) NLDB 2004. LNCS, vol. 3136,pp. 89–102, 2004.

[4] W3C, OWL Web Ontology Language Guide

[5] Burger, J., Cardie, C., Chaudhri, V., et al.: Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A). NIST Technical Report ,2001.

[6] De Boni, M.: TREC 9 QA track overview

[7] Litkowski, K.C.: Syntactic Clues and Lexical Resources in Question-Answering. In: Voorhees, E. M., Harman, D. K (eds) Information Technology: The Ninth Text REtrieval Conference (TREC-9), NIST Special Publication 500-249. Gaithersburg, MD: National Institute of Standards and Technologypp. 157–166, 2001,.

[8] Basili, R., Hansen, D.H., Paggio, P., Pazienza, M.T., Zanzotto, F.M.: Ontological resources and question answering Workshop on Pragmatics of Question Answering, held jointly with NAACL 2004 Boston, Massachusetts ,May 2004.