Research on the Optimizing Method of Question Answering System in Natural Language Processing

Zhang Kunpeng
School of information and software engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054, China
1360434810@qq.com

***Abstract*—Natural language processing technology can not only enrich the functions of computers, but also fundamentally promote the development of artificial intelligence technology. Based on natural language processing technology, many useful systems for people's survival and life have been produced, such as the question-and-answer system described in this thesis. This system mainly uses natural language processing technology and information retrieval technology. Although it is based on text retrieval, it is quite different from traditional search engine. Traditional search engines point out that users need to input a series of keyword combinations, and users can only get a variety of related websites, but also rely on their own discrimination ability to select useful information. However, the question answering system can allow users to input a question in the form of natural language. Finally, according to the search and judgment, the system can get a short and accurate answer to the user, which greatly improves the convenience of people's production and life. This thesis mainly elaborates the content of question answering system of natural language processing and analyses how to optimize it.**

Keywords-Natural language processing; Question answering system; Information retrieval

I. RESEARCH CONTENTS OF NATURAL LANGUAGE PROCESSING

Natural language processing covers a wide range of areas, such as natural language understanding, human-computer dialogue, information retrieval and so on. These aspects can be summarized into four parts: linguistic orientation, data processing orientation, artificial intelligence and scientific cognitive orientation, and language engineering orientation [1]:

A. linguistic orientation

This direction is to analyze and study natural language processing as time-sharing in linguistics, and to study the relationship between language and language processing and computation, but it does not analyze the specific practical process of language on computer. In this direction, the most important research fields mainly involve two theories, one is grammatical formalization theory, the other is mathematical theory. In terms of semantic relations, its main primitives can be referred to in Figure 1.

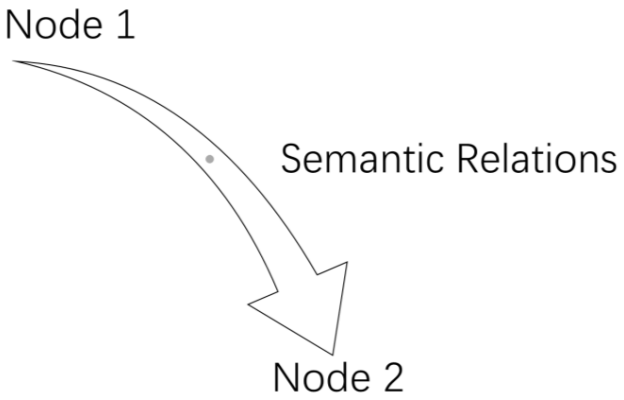

Figure 1. Semantic primitives

B. language engineering orientation

It mainly refers to natural language processing as a kind of language software, which is practice-oriented and has an engineering language. In this direction, people usually call it "human language technology" or "language engineering".

This thesis is trying to make a preliminary exploration and Analysis on the direction of language engineering of question answering system. The system consists of three main components: problem processing, information retrieval and answer extraction. For the questions submitted by users, we should first think about the purpose of users'questions, then classify the questions, and determine that the answers meet the users' questions'purposes and requirements, and then extract the keywords in the user's questions for search and feedback. In the information retrieval stage, the main task is to classify the articles in the system, to construct the information index, and then to search the documents accurately by using the keywords in the user's questions. Because the final feedback of question answering system to users is a relatively short answer, when information retrieval comes out, it is necessary to use the answer extraction link to extract shorter and more accurate answers. Answer extraction refers to the core part of the question answering system. In this link, this thesis uses the advanced technology of constructing the discourse structure tree and calculating the sentence semantic similarity, so as to analyze and discuss in detail the situation that the event in the sentence is the "extracting" verb, so as to truly realize the process of automatic question and answer.

978-1-7281-5050-5/19/$31.00 ©2019 IEEE
DOI 10.1109/ICVRIS.2019.00069

II. DEVELOPMENT TREND OF NATURAL LANGUAGE PROCESSING RESEARCH

Since the beginning of this century, the Internet has been popularized continuously, and the computer processing of natural language has also been popularized in people's production and life. The computer processing technology of natural language is an important means for modern people to acquire knowledge from the Internet. Therefore, at present, all countries in the world have carried out in-depth analysis and Research on natural language, and invested a lot of money. Therefore, although the time of studying natural language is very short, there are also more fruitful results. Its importance and bright application prospects have been known to the world. In some developed countries, such as the United States, the United Kingdom, Japan and so on, natural language processing as the core research topic of artificial intelligence is analyzed and discussed. At the same time, natural language processing has become a popular course in modern efficient computer. From the perspective of knowledge industry, natural language processing software plays an extremely important role. It involves many aspects, such as expert systems, databases, intelligent robots and so on, which need to use natural language. In the long run, natural language understanding system, if it has the ability of text comprehension, can be applied to machine automatic translation, information retrieval and other aspects, and has bright application prospects. [2]

III. RESEARCH STATUS AND EXISTING PROBLEMS

A. Research Status of Question Answering System

Question answering system refers to the user using natural language to ask questions and give concise and accurate answers. In addition, the question answering system has a long history.

At present, some very mature question answering systems have been developed abroad. Moreover, in 1993, the question answering system developed by MIT has been able to answer some simple questions. Moreover, these questions involve a wide range of fields, such as geography, history, culture, etc. In addition, there is a more advanced question answering system, namely AnswerBus, which is mainly a multi-lingual question answering system. It can answer questions raised by users in various languages, such as English, French, Italian and so on. At present, there are also some research institutes in China that have carried out the research and development of question-answering system, such as the Chinese Academy of Sciences, the Hong Kong University of Science and Technology, etc. However, at present, there is still no research and development of a more mature Chinese question-answering system.

B. Existing Question Answering System Problems

At present, there are a series of problems in question answering system, among which the most important one is the lack of reasoning ability in question answering system.

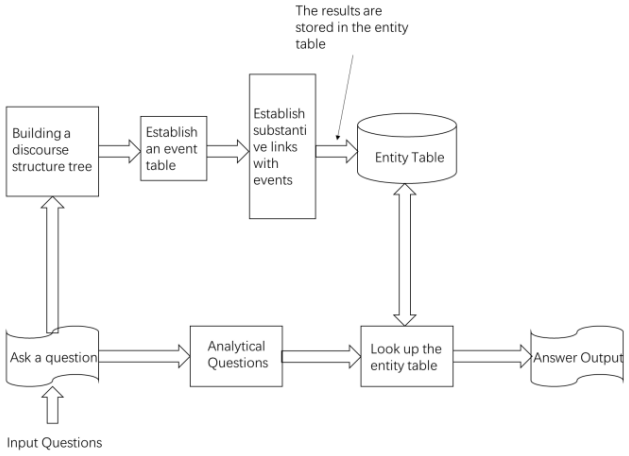

Figure 2. The Implementation and Evaluation of Question Answering System

Montague points out that there is no difference between natural language and logical artificial language. Essentially, natural language and logical artificial language are both symbolic systems. Both of them follow the common law of symbolic structure, which is also a general grammatical thought. The user's problem is parsed into the degree to which Prolog can reasoning, and then Prolog is used as a tool to extract keywords and information from the user's problem. However, there are some limitations in this way, that is, the user's questions are complex and diverse. In the question answering system based on the design and research of this thesis, advanced reasoning technology is also introduced. However, reasoning often belongs to the process of manual reorganization, if only relying on machines, it is basically impossible to achieve. At the same time, language knowledge and personalized rules are added to the question answering system, which are important ways to improve the accuracy of the final answer. At present, scientists have not been able to overcome the formalization of natural language. The main reason is that, at present, neither mathematics nor logic can provide a scientific and efficient tool for use. On one hand, the syntax and semantics of natural language exist in almost every aspect at the same time. In other words, they are interdependent and inseparable. On the other hand, there is still a long way to go to construct a formal language with unambiguous syntax and semantics. [4]

IV. OPTIMIZING METHODS OF QUESTION ANSWERING SYSTEM

In the question answering system, it is very important to deal with user's questions efficiently and scientifically, which plays an important role in whether users can get accurate answers in the later stage. Usually, there are two common ways to deal with problems. One method is based on principle, while the other is based on statistics. Statistical-based problem-solving method refers to the use of a large number of training corpus, so that the computer can automatically count the frequent interrogative phrases in these corpuses, thus taking these interrogative phrases as the basic basis for problem classification. The advantage of the

statistical method is that it can fundamentally reduce the interference caused by human subjectivity on the answers to questions, but this method requires a lot of training corpus. [5]

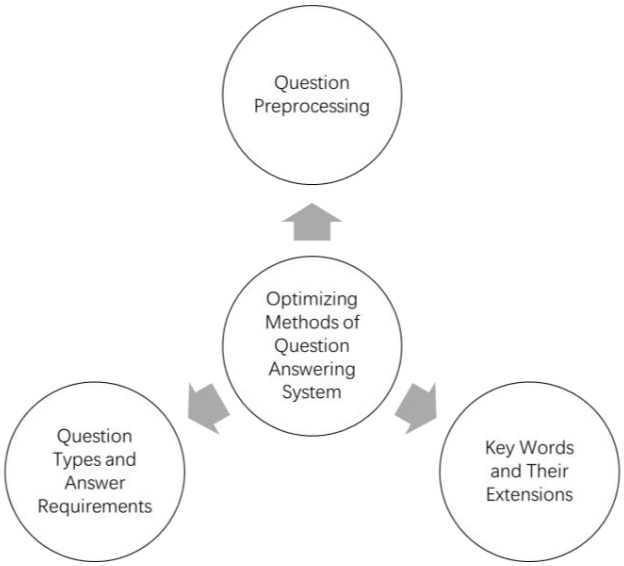

Figure 3. Optimizing Methods

At present, there is a lack of a large number of training corpus, so the second method is used less frequently. The main method adopted in this paper is rule-based method. At present, rule-based methods have been widely used in people's production and life, and have achieved good results. However, this method needs to be based on human experience and knowledge. It has a great subjectivity and a low degree of objectivity. Question processing module needs to complete a lot of work, for example, to classify and extract user natural language forms of questions, so as to determine the type of questions, and determine the final answer to meet the various conditions. In addition, in this link, a series of keywords will be generated and expanded.

A. Question Preprocessing

By using the method of dividing paragraphs, the vocabulary in user's questions can be extracted more accurately, and these vocabularies can be analyzed, then these vocabularies can be filtered in natural language form. At the same time, in order to improve the accuracy of information retrieval module, it is necessary to expand the extracted key words scientifically and efficiently.

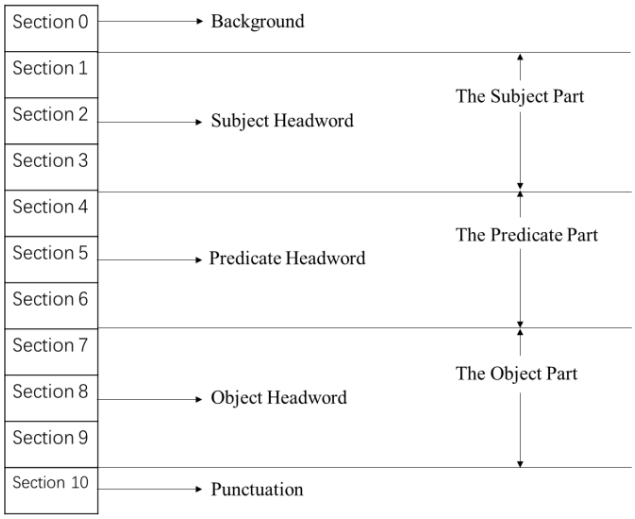

Figure 4. 11 Paragraphs of Sentences

B. Question Types and Answer Requirements

In order to answer users' questions correctly, the first step is to clearly understand the types of questions they ask, and to determine what requirements the final answers to users need to meet. Question type analysis plays a very important role in question answering system. Therefore, at present, many question answering systems begin to focus on the analysis of question type. When the type of questions and answers can meet the needs of users, it is very easy to determine the most suitable answers in the candidate answers in the answer extraction process. [6]

C. Key Words and Their Extensions

The extraction of keywords in question sentences has a great probability that will affect the effect of information retrieval. The search keywords are mainly nouns, verbs, adjectives and so on, and the keywords do not contain the interrogative words in question sentences. Keywords can be divided into general keywords and "must contain" keywords. General keywords can quickly get the correct answers through massive information retrieval, while "must contain" keywords basically do not have the correct answers. Because the keywords "must contain" have strong limitations, such as "most" and "first", it is difficult to find the right answer that meets the requirements.

V. CONCLUSION

At present, the research of natural language processing in China has achieved fruitful results, but compared with western developed countries, it still has a long way to go. In addition, there is no significant innovation or breakthrough in the research and development of natural language processing in China, either in theory or in the application system. Basically, our research is still a follow-up study, but a lack of creative research and important research with original ideas. Therefore, we should not be complacent with the current research results. China must study foreign

advanced research results, actively carry out independent innovation, improve innovation research ability, strive to catch up with the international advanced level as soon as possible, so as to enter the international advanced ranks and master the current international first-class technology and research results, so as to improve our scientific research ability.

At present, the question answering system can not answer as well as human beings. At present, China's question-and-answer system has many shortcomings in many aspects, especially in the ability of thinking and reasoning, and there are many problems. Although the question answering system has a long way to go from our ideal goal, in recent years, the rapid development of the question answering system has been fundamentally promoted by the rapid development of network technology and information technology and the urgent need of the people for the question answering system. The experimental results show that the research idea of this thesis is feasible and scientific. In the near future, the question answering system will probably replace the traditional search engine and get huge pictures, which will be widely used in people's production and life, and help people accurately find the information they need in the mass of information.

REFERENCES

[1] An Yupu. The design and implementation of natural language question answering system [D]. Harbin: Harbin University of Technology, 2003.

[2] Shen Yanna, Design and Implementation of Natural Language English Question and Answer System [D]. Shenyang: Northeast University, 2003.

[3] Ye Dexue. Algorithms for Question Understanding in Chinese Question Answering System [J]. Digital Technology and Applications, 2012.

[4] Qin Bing, Liu Ting, Wang Yang, etc. Research on Chinese Question Answering System Based on Frequent Question Set [J]. Journal of Harbin University of Technology, 2003.

[5] Huang Yong. Research on Formalization of Sentence Meaning and Its Application in Chinese Question Answering System [D]. Beijing: Beijing Normal University, 2005.

[6] Chen Yongping, Yang Sichun, Mao Wansheng, et al. Question Understanding Based on Theme and Focus in Chinese Question Answering System [J]. Computer System Applications, 2011.