Intelligent Question Answering System for Internet of things data analysis and educational technology

Yanzhuang Chen
Longjiang Vocational and Technical School,Foshan City, Guangdong Province, China
nekoychanjob@126.com

ABSTRACT
Question answering is an indispensable part of the teaching process. The effect of question answering largely depends on the form of the link and whether it can meet the individual needs of students. The traditional teacher learner face-to-face question answering and e-mail question answering can not meet the requirements of learners. Therefore, many educators are committed to the research of intelligent question answering system. This paper analyzes the main obstacles existing in the current domestic question answering system, designs the subject oriented intelligent question answering system, studies the core function modules, and realizes a subject oriented intelligent question answering system with friendly interface, strong pertinence, and adaptability to a variety of inquiry methods.
Keywords: *Subject oriented; Intelligent question answering; Problem handling; Sentence similarity; Answer extraction*

1 INTRODUCTION

With the development of computer and Internet, the information of the current society is growing exponentially. The information on the Internet is open. Disorderness is difficult for a common user to find the information directly from the mass information. Too much useless information affects the retrieval efficiency of users. Users hope to get the answers directly after entering questions that need to be asked. In order to meet the needs of users, the FAQ system came into being. FAQ system is also known as FAQ (frequently asked questions), which is actually the most common question and answer system. With the rapid development of science and technology, more

and more students study online[1]. E-learning refers to learning between teachers and students without the restriction of traditional learning mode, that is, students do not need to learn specific content in a fixed time and space. E-learning has the characteristics of breaking the limitation of time and space, personalized learning, changing learning subjects and interactive learning. Online teaching is a virtual environment provided by the network. Students' learning and teachers' teaching can be carried out synchronously or asynchronously. Through synchronous learning, students can follow the teacher's teaching steps and follow the teacher's teaching contents in real time[2].

The automatic question answering system q/a (automatic question answering) uses natural language processing technology, which can help to understand the questions raised by users and generate correct answers on the other hand. The main difference between the system and FAQ is that the answers of FAQ system come from the established common question answer database, in which the questions and answers have been determined; The answer of the automatic question and answer system comes from the knowledge text base set by the system. The knowledge text in the knowledge text base is very extensive, and can even be the information resource of the whole Internet. With enough knowledge source materials, the automatic question and answer system can answer all the questions raised by users like a knowledgeable expert. The goal of this paper is to establish an intelligent question and answer system (hereinafter referred to as "ETIS") which integrates the FAQ system and the automatic question answering system[3].

2 RELATED WORK

2.1 Research status at home and abroad

At present, network education is very hot. Online learning has become a channel for students' autonomous learning. Online Q & A is a part of network teaching. At present, the intelligent question answering system has not received due attention and corresponding status in the

Copyright © 2022 The Authors. Published by Atlantis Press International B.V.
This is an open access article distributed under the CC BY-NC 4.0 license -http://creativecommons.org/licenses/by-nc/4.0/.

teaching website. The research results of Q & a system are less, and there are still many problems in the research of key technologies of the system. Intelligent question answering system is a computer program. The program can accurately answer students' questions and return the results to users. In this process, it can count the high-frequency questions asked by students, store the new questions asked by students, submit them to relevant teachers or experts for answers, and store the answers to the new questions into the knowledge base to realize the memory and learning of knowledge.

In the 1960s, the intelligent question answering system using natural language came into being. With the development of artificial intelligence and machine learning science, the research of intelligent question answering system has also made a certain breakthrough. Domestic Q & A platforms are generally produced with the development of teaching websites[4]. As an auxiliary module of teaching websites, they help teaching and answer students' questions, so as to consolidate the learning content in class. The foreign question answering system adopts advanced technology in the design process, which can well realize interactive communication and accurately grasp the intention of students' questions. It has a high degree of intelligence and can generally be used as an independent platform. Compared with domestic simple question answering methods, foreign systems can achieve more accurate intelligent answers to knowledge in a specific field.

Typical question answering systems are:

(1) Start Q & a system. The system is designed and developed by MIT. It is an intelligent question answering and retrieval system based on natural language technology. It is also one of the earliest intelligent question answering systems in the world facing the Internet. The core of the system is the construction of knowledge base, and the basis of question and answer is also the information search of knowledge base. The range of knowledge that the system can answer includes geography, politics, art and other fields. The Q & a content is rich (including video, audio, etc.) and the accuracy is high, but the system can not recognize and understand the questions in Chinese or other language forms, and can only realize English Q & A, which is also a huge defect of the system.

(2) Askjeeves system. The system can ask questions in natural language, which is developed and designed by askjeeves company. Systematically analyze the questioning sentences and further communicate with the questioner, so as to truly and comprehensively obtain the questioner's ideas. However, the system can not directly provide users with concise answers, which makes users need to identify and filter information according to the returned results, which is also one of the great defects of the system.

(3) An intelligent question answering system developed and designed by Peking University. In this Q & a system, forums and online discussions are the mainstream ways of Q & A. Among them, there are several knowledge areas in the forum area. Students enter the discussion areas of different topics according to their own needs.

2.2 Chinese automatic word segmentation technology

There are many research methods of Chinese word segmentation. This paper mainly introduces three mainstream word segmentation algorithms.

(1) Mechanical word segmentation algorithm

The mechanical word segmentation method is also called character matching method. Its implementation principle is to match the existing string with a sufficiently large dictionary. If the corresponding string is found in the dictionary, the matching is successful. There are many matching methods. According to the scanning method, it can be divided into forward matching and reverse matching, and according to the length first method, it can be divided into maximum matching and minimum matching[5]. Therefore, the common mechanical word segmentation methods are: reverse maximum matching method, forward maximum matching method, two-way maximum matching method and minimum segmentation method.

Advantages: high performance. Because the matching rule is simple, as long as new words are added to the dictionary, it can well support the recognition of unlisted words, so it has a high utilization rate in engineering applications.

Disadvantages: disambiguation ability is weak. The rules are too simple to recognize new words that are not in the dictionary.

(2) Understanding based word segmentation

Word segmentation based on understanding means that the computer recognizes words and understands sentences by simulating human behavior. In the process of word segmentation, ambiguity is handled through syntactic and semantic analysis. The characteristics of Chinese language determine the complexity of Chinese language knowledge. At present, the processing of Chinese language by computer is not perfect. Computers can not fully understand Chinese language information and convert it into machine recognizable language. The implementation of this method is based on a large amount of language information, so this method is rarely used at present[6].

(3) Word segmentation method based on statistics
This method judges whether a string constitutes a word according to the frequency of its occurrence in the corpus. The typical algorithms of this method include CRF, HM, maximum entropy model and so on. Among them, CRF has better effect, and it has weaker context independent hypothesis than HMM.
Advantages: this method is based on statistical algorithm for segmentation. In theory, as long as there is enough training corpus, it can deal with the problems of "unlisted words" and "ambiguity".
Disadvantages: because this method needs enough corpus for training, the training time is long. In industrial applications, there are often new professional field corpora. At this time, it needs to be retrained, which is more troublesome. Therefore, this method is not used in some industrial applications with high performance requirements[7].
The maximum matching algorithm mainly includes forward maximum matching algorithm, reverse maximum matching algorithm, two-way matching algorithm and so on. The main principle is to cut out a single word string, and then compare it with the thesaurus. If it is a word, record it. Otherwise, continue the comparison by adding or reducing a word until there is still one word left. If the word string cannot be cut, it will be treated as unregistered.
The algorithm flow is shown in Figure 1:

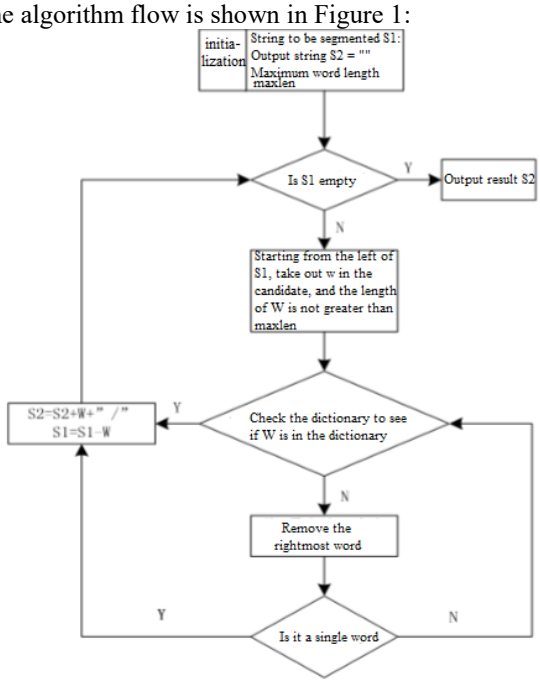

Fig. 1. Forward maximum matching method

In the Chinese language system, the smallest unit is word. But in our natural language, the individual words can not express our meaning accurately to a large extent. So, in fact, the most basic unit of composition is words. Therefore, in the automatic question and answer system, the primary task is to divide the continuous Chinese character strings into the word sequence, that is, the automatic word segmentation. Only after the sentence is converted into a word can the computer understand the sentence[8]. According to the different word segmentation principles, Chinese automatic word segmentation methods are divided into four categories: dictionary based segmentation (or word list method), statistical segmentation method, knowledge segmentation, neural network word segmentation method.

2.3 sentence similarity calculation

In the natural language processing, sentence similarity calculation is a basic and core link. In the automatic question and answer system, many parts need to calculate the similarity of sentences. Sentence similarity refers to the semantic matching degree of two sentences, which is represented by the real number of [0,1]. The larger the value, the more similar the two sentences are, and when the value is 1, the two sentences are identical in semantics; The smaller the value, the lower the similarity between the two sentences, and the semantic difference between the two sentences when the value is 0. The calculation of sentence similarity is very important in all fields of natural language processing, but compared with the English sentence similarity calculation, the calculation of Chinese sentence similarity is more difficult[9].
The existing sentence similarity calculation methods are TF IDF algorithm and sentence meaning based similarity algorithm. Objective analysis of the advantages and disadvantages of these methods can better select the development technology of the system, and is conducive to the realization of more efficient system. The formula of f-idf algorithm is as follows.

$$Sim(T', T^{-}) = \frac{\sum_{i=1}^{n} (T', T^{-})}{\sqrt{\sum_{i=1}^{n} (T') \bullet \sum_{i=1}^{n} (T^{-})}} \quad (1)$$

Full text retrieval is to establish an index for each word, record the frequency and position of the word in the article, and then retrieve the index through the search engine, extract the position of the word according to the index, and the frequency of the word in a certain range. This process is similar to the process of looking up a word through a search list in a dictionary.

Generally speaking, the full-text retrieval system needs to have the basic function of establishing index and providing query. Knowledge text index is actually a module to manage a large number of texts. This module is different from the traditional database, it is essentially a search engine. Its function is to provide massive database management and query, so as to quickly, accurately and comprehensively search the information needed by users [10].

3 DATA ANALYSIS

3.1 *subject oriented intelligent question answering system architecture*

The workflow of ETIS is as follows: the question and answer system first obtains the questions put forward by users, and then preprocesses the questions. After the processing of word segmentation module, the keyword combination is formed, and the formed keywords are analyzed to determine the focus of the problem. Search the common question database, compare the similarity between the user's questions and the same kind of questions in the common question database, when the similarity is greater than the set threshold, directly return to the user's corresponding question answer. Otherwise, search in the knowledge text base, when the similarity is greater than the set threshold, directly return to the user the corresponding question answer; If it is less than, the user will be prompted that there is no exact answer and the corresponding reference answer will be returned. The above workflow is shown by the frame diagram[11]. The overall architecture of the system is shown in Figure 2 below. The whole system is divided into three layers: interface, user operation layer and data layer.

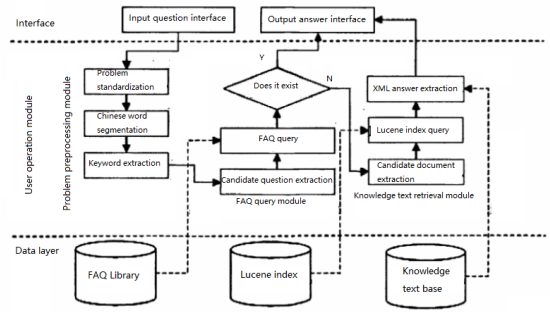

Fig. 2. Overall architecture of subject oriented intelligent question answering system

3.2 *function design of automatic question answering system*

According to the overall architecture of the subject oriented intelligent question answering system, this paper divides the question answering system into the following functional modules: question processing module, FAQ query module, knowledge text retrieval module and index module. The specific functions of each module and the relationship between them are shown in Figure 3 below

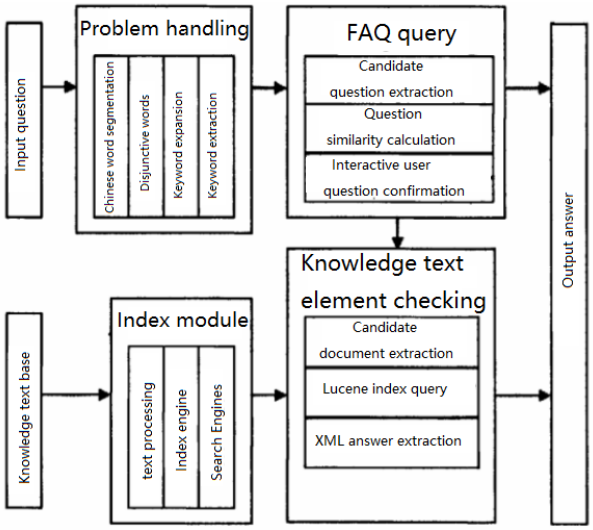

Fig. 3. Function module diagram of intelligent question answering system

3.3 *database design*

The FAQ of this system is stored in the form of database. The database uses Access2003. The entity set of database includes FAQ database and knowledge text database. FAQ database is the collation of common questions in the discipline, including: questions with high frequency, typical questions in the discipline. Through the collection and collation of various information on the Internet, we finally summed up 319 common questions as common questions in educational technology, and put them into the FAQ database.

$$\Delta u_{k+1}(t) = u_d(t) - u_{k+1}(t) = u_d(t) - \Big(u_k(t) + L(t)\big(\dot{e}_{k+1}(t) + e_{k+1}(t)\big)\Big) = u_d(t) - u_k(t) - L(t)\big(\dot{e}_{k+1}(t) + e_{k+1}(t)\big) = \Delta u_k(t) - L(t)(\dot{C}(t)\Delta x_{k+1}(t) + C(t)\Delta\dot{x}_{k+1}(t) + C(t)\Delta x_{k+1}(t)) \quad (2)$$

The knowledge text inventory stores the chapter number, title and representative keywords of each section of educational technology. Its main attributes are: chapter number (ID), chapter number (titleid), title (title), chapter

level (type),. Parent (parent), path (URL) of the text file corresponding to the chapter, repeatability (value) of keywords and user question keywords, and record the word segmentation result (extendkey) of the title field.

4 EXAMPLE ANALYSIS

4.1 system development environment

The system adopts B / S mode, rich interaction mode and better user experience. Users can input and submit questions after opening the system page
Java is a simple, object-oriented dynamic language. It has the following characteristics:
(1) Simple. Java was originally designed as a language for integrated control of household appliances, so it must be simple and clear. Its simplicity is mainly reflected in the rich class libraries provided by Java.
(2) Object oriented. Object oriented is the most important feature of Java. The design of Java language is completely object-oriented, it does not support the process oriented programming technology like C language. Java supports static and dynamic style code inheritance and reuse.
(3) Distributed. Java includes a sub library that supports TCP / IP protocols such as HTTP and FTP. Therefore, Java applications can open and access objects on the network by means of URL, and its access mode is almost the same as that of local file system [4].

4.2 realization of main function modules

(1) User interface. It is the interface for the system to realize human-computer interaction, and its main functions include: receiving user input questions and outputting answers to users. This system uses the borderless transparent window, can drag the portrait and pop-up balloon, rich pop-up menu as the interface, in order to enhance the user experience.
(2) Problem handling. The problem processing module includes Chinese word segmentation, stop words removal, keyword expansion and keyword extraction. In this module, using the Chinese Academy of Sciences ictlas word segmentation, can achieve word segmentation, on this basis, in order to make the system more intelligent, this paper continues to complete to stop words, keyword expansion, keyword extraction and other matters.
(3) FAQ query. FAQ query module includes: candidate question extraction, question similarity calculation, interactive user question confirmation.

4.3 realization of chat function

Since Alice is an open source software, many scholars have studied its principle. At the same time, Alice is also maturing in the continuous improvement of many people. Different researchers have different preferences for development languages, so there are many versions of Alice. This system uses the Java version of Alice. In order to make Alice more suitable for this system, some improvements have been made in the process of introduction
Chinese word segmentation is added. Alice was originally designed for English, using spaces to identify English words. However, some Chinese Alice systems only support splitting Chinese into single Chinese characters, which greatly reduces the semantic function. After analyzing the operation mechanism of Alice, the author intercepts the original call and adds Chinese word segmentation to make the system support Chinese word matching.

5 CONCLUSION

This paper only puts forward a prototype system, there are many deficiencies, need to be further improved and perfected. At the same time, the system may not be able to answer some basic questions correctly. Therefore, for the whole system, we need to expand the capacity of the question base, improve the semantic similarity algorithm, and combine the weight table of educational technology professional vocabulary with Chinese word segmentation and HowNet, so as to understand the user's questions more accurately and retrieve the answers efficiently.

REFERENCES

[1] Jia Zongfu, Wang Zhifei. Research on Chinese sentence similarity calculation [J]. Sci tech information, 2009.11:402-403.
[2] Zhang Yufang, Peng Shiming, Lu Jia. Improvement and application of TFIDF method based on text classification [J]. Computer Engineering, 2005, 32 (19): 76-78.
[3] Zhang Liang. Research on problem processing technology of open domain oriented Chinese question answering system [D]. Nanjing: Nanjing University of technology. 2005.
[4] Zhang Huili. Research on Chinese automatic question answering system in computer field [D]. Tianjin: Tianjin University. 2006.
[5] Wang Qiong. Analysis on the current situation of online teaching support system. 1999

[6] Shen Ruimin, Wang Jiajun, Tang Yiyang, automatic question answering system based on Web
[7] Liu quanbo, Huang Ronghuai, he Kekang, design and implementation of intelligent question answering system,
[8] Zhou Hongtao, adaptive automatic courseware generation tool, master's thesis opening report
[9] Chen Dewei, Li Kedong, research on web-based adaptive learning guidance system and creative tools, gccce2002 proceedings
[10] Yu Shengquan, adaptive learning_ Development trend of distance education. Educational technology communication
[11] Tu Fei, Zhang Xiaozhen, research on adaptive teaching strategy adjustment based on multi agent, gccce2002 proceedings