

## Design of a Lightweight Automatic Question Answering System in Education Information Service Domain Based On FAQ

Tang MingJing<sup>1,a</sup>, Liang Li<sup>1,b</sup>

<sup>1</sup>Department of Informationization Management, Yunnan Normal University, Kunming China

<sup>a</sup>matthewtmj@ynnu.edu.cn, <sup>b</sup>liangli@ynnu.edu.cn

**Keywords:** Question answering system; Restricted domain; Education information service; FAQ lib retrieval algorithm

**Abstract.** The approach of restricted domains provides a new idea for the development of automatic question answering system which without complicated semantic processing. The paper introduces the design and implementation of a lightweight automatic question answering system in education information service domain based on FAQ (Frequently Asked Question). The workflow and structure of the system were described, and the key techniques were expounded, including Chinese word segmentation based on String-Matching algorithm and the FAQ retrieval algorithm based on keyword lib, etc. Running result shows that the system has high precision on frequently asked questions, and provides an automatic approach to collect question.

### Introduction

With the development of university digital campus construction, a large-scale computer network has been established, and on the basis, some systems has been established which involving the teaching, scientific research, management and service. This changed the traditional mode of teaching and management; campus network has become the necessity of students and teachers. Due to huge number of network access point and campus network users, the maintenance work is under a lot of challenges. Some relevant solution was introduced, such as, telephone service, door to door service, etc. But these solutions have defects, such as, maintenance personnel is not enough, can't immediately respond to user' request and maintenance personnel is engaged in a lot of repetitive work due to the same problem. How to improve the work efficiency of informationization management and maintenance; how to improve campus network user experience has become an urgent problem.

Automatic Question and Answering System is an intelligent automatic customer service system based on a variety of artificial intelligence technology and internet technology, take the Web page or instant messaging tools as form. The current research shows that the type of automatic question answering system include chat robot, question answering system based on knowledgebase, question answering retrieving system and free-text question answering system[1]. Related research results show that the accuracy of the automatic question answering system is still relatively low; especially, the question answering system based on Chinese, accuracy is around 30% [2, 3]. Although automatic question answering system is still at the starting stage, but a growing number of relatively mature applied question answering system appeared gradually, this promote the rapid development of the automatic question answering technology.

According to different application environment, the automatic question answering system can be divided into the restricted domain question answering system, the open domain question answering system and question answering system based on FAQ. The restricted domain question answering system is targeted at specific domain of intelligence Q&A, has such characters as specific domain, simple question classification, good practical values, etc. The open domain question answering system can answer any question in any domain, but large scale knowledge base model of it is difficult, so has low practicability. Question answering system based on FAQ save questions and answers what raised quite often as FAQ library. When the user questions, the system first looks in the FAQ library,

if it finds the same or similar question, it returns the answer of the question directly, without any complicated processes, such as question analysis, information retrieval and answer extracting.

This paper collected and arranged a lot of common problems in the campus network as FAQ set. On this basis, develop a lightweight automatic question answering system based on FAQ. This system offers a quick education information service by integrate of public web service resource. The following introduce the scheme of this system which based on Yunnan Normal University.

## Analysis on System Requirements

**Analysis on Business Requirement.** Before modular design and implementation of this lightweight automatic question answering system, it is necessary to fully understand the business process of the system. The business process of the system is shown in Fig1.

Yunnan Normal University has deployed campus network real-name authentication system. After the integration of automatic question answering system and authentication system, user does not need extra authentication when they access the automatic question answering system, and the system will automatically record and save the user's authentication information. Once the login credentials of a user are verified, user can input question for consultation; the system will return the answer or the candidate question list to the user's input. If the user gets the similar questions or answer, can check the answer and submit the feedback; this feedback is just like "satisfied with the answer?". If the user didn't get the answer, the system automatically records the user's question, and marks this question as unanswered. After the expert user login system, can look at the list of unanswered questions, and answer these questions. When that was done, these questions got into FAQ library.

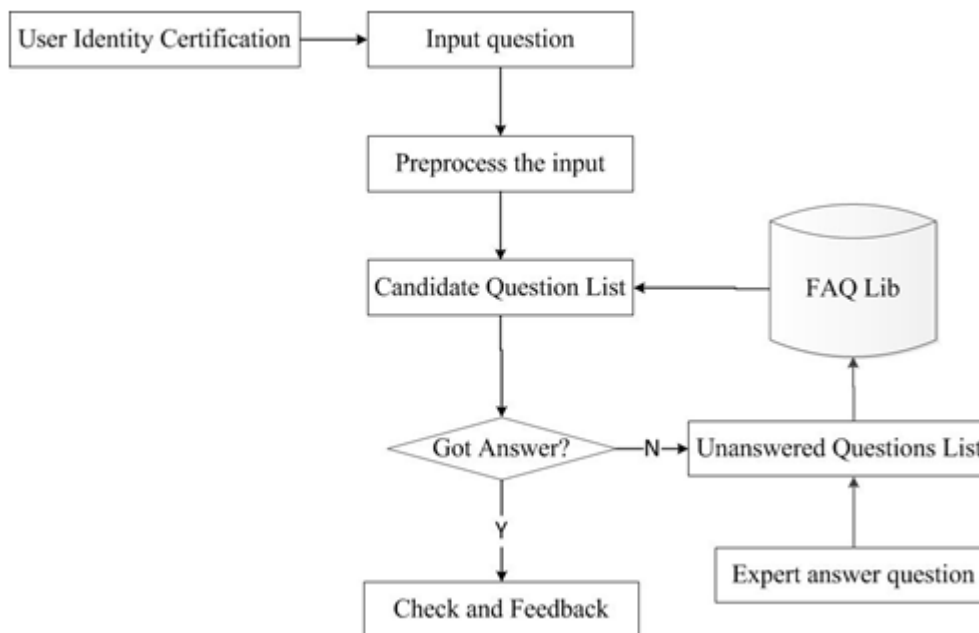


Fig. 1 Flow Chart of Business Process

**Analysis on Non-Functional Requirements.** This lightweight automatic question answering system is a service for university routine administrative management; and will be deployed in the university campus network, as well as provides service for internet users. As the users are very large, the system has to run for 7\*24 hours; and average request response time is less than 5 seconds, maximum response time of less than 15 seconds.

User Interface (UI) is the gateway and the bridge between human and computers, this lightweight automatic question answering system uses real-time interactive way to provide a better user experience by Ajax and XML technologies.

## Design of System

**Design of System Architecture.** The framework of the system adopts three tiers architectures, based on B/S model; and adopts the combination of SSH framework and Java-related technologies to achieve the system implementations [4]. The system architecture design is shown in Fig2. Specifically, the presentation layer and control layer are based on the Struts2 framework technology, the model layer is based on the Spring framework technology, and the database persistence layer is based on Hibernate framework technology. In the page display layer, the Struts2 framework provides the powerful Tablib tag library, implements the code reuse and code abstraction, and improves the efficiency of Web development. With Ajax asynchronous messaging technology, this system is much more interactive. In the business logic layer, the Spring framework provides the Spring IoC mechanism, the business logic is injected into the call module via a configuration file. In the database persistence layer, Hibernate framework achieves optimizing of database performance and common operations of the database interaction.

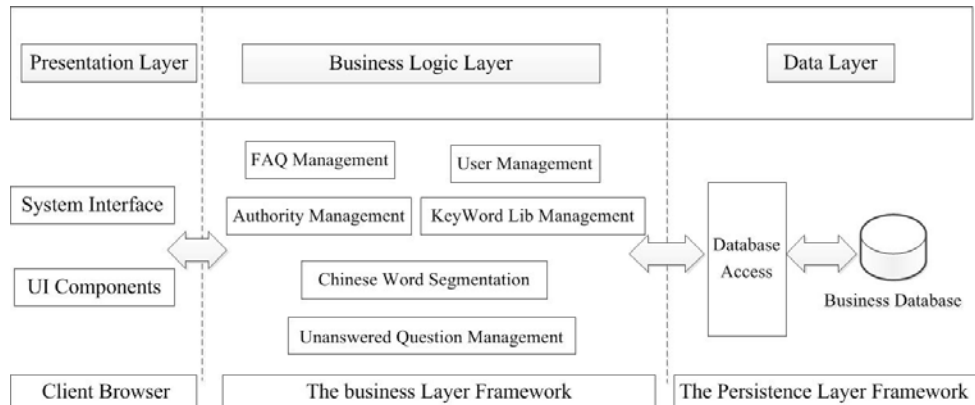


Fig. 2 Chart of System Architecture Design

**Design of System Function.** This system has following functions: consult question, feedback of answer, FAQ lib management, and keyword management, etc.

**Consulting Service.** The user input query what is natural-language, the system preprocess this query and return the answer or the candidate question list according to user's query.

**Feedback of Answer.** After user gets the answer, user can give some feedback to evaluate whether or not the result applies to user's question. These feedbacks are useful to system optimization and data statistics.

**FAQ Library Management.** In this module, expert user can add, view, edit, and delete the record of FAQ Library. The content of answer can contain arbitrary text, images, audio and videos.

**Key Word Management.** In order to improve the retrieval efficiency, the system extracts all keywords of question-answer pairs, and records its index relation.

## Key Technologies of System Implementation

**Question Preprocessing.** The question user input is natural-language query, the automatic question answering system need to do some preprocessing, including automatic word segmentation, remove the punctuation marks, and so on.

Automatic word segmentation is that natural language is divided into word sequence according to the phrase library. Because the lightweight automatic question answering system in this paper is an automatic question answering system in restricted domain, the system does not concern about each words of question can be accurate segmented, but focus on the meaningful words of question can be accurate segmented; this paper does not adopt huge Chinese library, but design special word library according to application background. The Chinese word for each sentence, there is no space between words, and structure of Chinese can be tremendous variances; so the Chinese word segmentation has a lot of difficulties. At present, three main methods have been used for Chinese word segmentation, which include character matching method, statistical method and understanding method [5]. These

three Chinese word segmentation algorithms each has its characteristics, but still can't let computer understand Chinese just like human mind. Related research results show that the reverse matching algorithms based on character matching method has higher accuracy and efficiency [6,7]. This paper adopts the method of Chinese word segmentation based on character matching method and reverse max-matching method based on professional lexicon. As the user enters the question, the system removes irrelevant punctuation, and divides the sentence into several strings. Then these strings match with professional lexicon. If the match is successful, the string will be preserved; otherwise the string will be abandoned. For example, the user input question, just like "I am a freshman, how do I apply for Yunnan normal university's mailbox services? Thank you!" After Chinese word segmentation, removing irrelevant punctuation and other preprocess, the system get string array of question just like "freshman, application, Yunnan normal university, and email".

**FAQ Retrieval Algorithm Based On Keyword Library.** In order to improve the efficiency of the system, keyword is created for each question-answer pairs when FAQ library is built, and remove the repetition and the same prefix keyword to form an authoritative keyword set.

The FAQ retrieval algorithm based on keyword library is as follows:

- (1) To build authoritative keyword set of the FAQ library, get an array : keywordLib= $(k_1, k_2, \dots, k_m)$  ;
- (2) Preprocess the user's query, then get question string array: questionArray= $(q_1, q_2, \dots, q_n)$ ;
- (3) Each keyword of keywordLib match each string of questionArray, if the match is successful, then put current keyword  $k_i$  into matchlist keywordList;
- (4) After the match, get question-answer pairs according to keyword of keywordList;
- (5) The result as candidate question set return to the user.

The process of building authoritative keyword set of the FAQ library is initialization of algorithm; in this process, all the keywords are separated based on the specific symbol.

**Ajax Data Exchange Based on XML.** In order to provide users a system interface that is interactive and non-fresh page, the system adopted Ajax, JQuery, and Xml technologies.

Ajax (Asynchronous JavaScript and XML) is a technique for creating better, faster, and more interactive Web applications. Compared with traditional Web applications, Ajax is a technique to retrieve server-side content without requiring the entire Web page to be refreshed, and reduce the amount of communication data between the server and a Web page. JQuery is a lightweight JavaScript library that can be compatible with various browsers, can more easily interact with Ajax. Meanwhile, the system use XML as data exchange format between client and server. For example, data exchange format of candidate question set is shown as Table 1.

Table. 1 Data Interchange Format Based on XML

```
<answer>
<timestamp>1412390476</timestamp>
<time>Sat Oct 04 10:41:16 CST 2014</time>
<key>network</key>
<list count='538'> My Internet account balance is insufficient, how to do?</list>
<list count='414'> How to use the switch </list>
<list count='413'> Why can't I log on the campus website?</list>
<list count='245'> What is network repair call? </list>
</answer>
```

## Evaluation

This paper use LoadRunner [8] as a tool to evaluate the performance of this lightweight automatic question answering system. The evaluation scenario includes user login, question, check answer, and feedback. The Loading plan is to load 50 Vusers every 5 seconds, until 500 Vusers.

Fig3 shows that the Minimum response time is 2.064 seconds, the Maximum response time is 12.24 seconds, and the average response time is 6.642 seconds.

At present, this automatic question answering system has been running for three months, has 3846 users, online number up to around 96 users, and received 20893 visits. Trial run result of the system shows that the system is fit for purposeful, clear questions. There are 1213 feedbacks of FAQ question-answer pairs; the praise rate is greater than 90%.

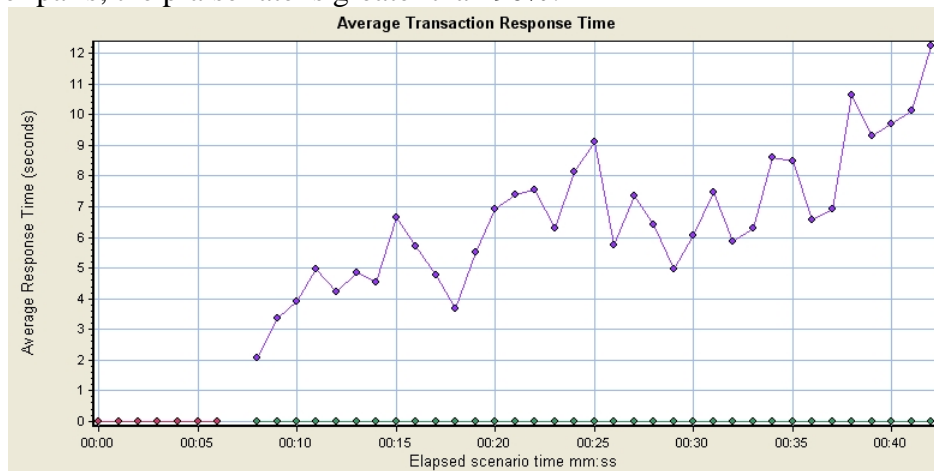


Fig. 3 Average Transaction Response Time

## Conclusion

This paper introduces the design and implementation of an automatic question answering system for education information service domain based on FAQ. After trial running, the system has high precision on frequently asked questions. Meanwhile, the system provides an approach to collect question what the FAQ do not have. Besides the system has good generality and is an effective exploration of university digital campus construction and management.

## Acknowledgment

This work has been supported by the National NSF of China (No. 61262071), the national financial support to local college's development special funds.

## References

- [1] Liu Li, Zeng Qing-Tian. An Overview of Automatic Question and Answering System [J]. Journal of Shandong University of Science and Technology, 2007, 26(4):73-76
- [2] Yin Hong-Zhi, Zhang Fan, Ding Ding, Zhao Bin. AnswerSeeker: Question Answering System Based on Web Mining [J]. Computer Systems & Applications, 2010, 19(1):6-8
- [3] Liu Fang, Yu Fei. Research and Development of Intelligent Q&A System for Medical Industry [J]. Microelectronics & Computer, 2012, 29(11):95-97
- [4] Meng Fan-Qi, Qu Zhao-Yang, Hou Song-Lin. Application of SSH Framework in Developing Integrated Management System for Advisors' Labs [J]. Computer Applications and Software, 2013, 30(8):174-176
- [5] Zhang Hui-Li, Meng Zhao-Peng, Wang Hui-Zhi. Ambiguity Processing in Chinese Word Segmentation [J]. Microcomputer Applications, 2006, 27(6):685-688
- [6] Fan Kang-Xin. Design and Implementation of OTC Question Answering System Based on FAQ [J]. Computer Applications and Software, 2013, 30(7):178-181
- [7] Tang Su-qin, Li Bo, Xu Yong-Min. Intelligent Answer System Based on Sentence Template [J]. Journal of Guangxi Normal University: Natural Science Edition, 2007, 25(2):5-8

[8] Hao Jian-Ying, Yan Hai-Hua. Effective Web Performance Testing Method and Its Application [J].Application Research of Computers, 2007, 14(1):106-109