

A Proposed Automatic Answering System for Natural Language Questions

Prof. Dhanshri Patil¹, Abhijeet Chopade², Pankaj Bhambure³, Sanket Deshmukh⁴, Aniket Tetame⁵

¹Nutan Maharashtra Institute of Engineering & Technology, Pune, India
patildhanshri@gmail.com

²Nutan Maharashtra Institute of Engineering & Technology, Pune, India
abhichopade55@gmail.com

³Nutan Maharashtra Institute of Engineering & Technology, Pune, India
pankajbhambure10@gmail.com

⁴Nutan Maharashtra Institute of Engineering & Technology, Pune, India
sanketdeshmukh7206@gmail.com

⁵Nutan Maharashtra Institute of Engineering & Technology, Pune, India
anikettetame203@gmail.com

Abstract: *Though computers are considered as machines, they are expected to give answers to questions in normal English language, just the way human beings can do. To train a computer to answer English language questions is an interesting and challenging problem. The automatic answering systems are classified under two categories: open domain systems and close domain systems. In this paper an answering scheme is proposed that combines close domain and open domain system. The question answering system for closed domain deals with questions under specific domain. Answers to questions from close domain cannot be searched using a search engine. During a close domain, answers to questions are not alive in the public area and so they cannot be search using a search engine. Hence, the answers to such questions are maintained in a database by the domain expert. During retrieval of answers, the best matched answer searched from database is returned to the user. To perform this matching a template matching technique is used. Open domain question answering system deals with all questions covering all domains. These systems can rely search engines to find the required answer. In this paper, we have combined both the approaches and proposed an answering scheme which accepts input in General English language. Thus the main challenge of the proposed scheme is to understand the English language question.*

Keywords: Answering Machine, Closed Domain Problem, Natural Language Processing, Template Matching.

1. Introduction

The interaction between computer and human being is always interesting and challenging. Due to increased use of computers, different mechanisms for computers are developed to answer the questions asked in natural English language. Such mechanisms allow users to ask questions in a natural language and give a concise and accurate answer. Understanding user questions in natural language requires Natural Language Processing (NLP). The goal of the Natural Language Processing is to design and build methods that will analyze, understand, and generate languages that humans use naturally. With the use of NLP, it is possible to address the computer in the same way used to address other person.

Natural language processing is the computerized approach to analyzing text based on both a set of theories and a set of technologies. It will become important to use NLP as the natural English language cannot be understood by the system. NLP includes removal of stop words, removal of vowels,

stemming the words, etc. In the proposed system user is able to ask queries and obtain answers, using natural language (NL) expressions, rather than the keyword based retrieval

mechanisms. The proposed system will better satisfy the needs of users as the system is developed with an accurate, quicker, convenient and effective way of giving answers to user questions.

2. Previous Work Done

In [1] they mention that to answer English language questions using computer is an interesting and challenging problem. Generally such problems are handled under two categories: open Domain problems and close domain problems. In this paper presents a system that attempts to solve both close and open domain problems. Answers to questions from close domain cannot be searched using a search engine. Hence answers have to be stored in a database by a domain expert. Then, the challenge is to understand the English language question so that the solution could be matched to the respective answer in the database. We use a template matching technique to perform this matching. The system is developed such that the questions can be asked using short messages from a mobile phone and therefore the system is designed to understand SMS language in addition to English. One of the main contributions of this paper is the outcome presented of a deployment of this system in a real environment.

In [2] they mention that computers are considered as machines but they can answer the questions in a way just like the human

beings can, which is an interesting but a challenging problem and such problems are handled under two categories: open domain problems and close domain problems. The problems which come under the close domain are the problems which are not available in public domain and therefore they cannot be searched using a search engine. Hence, the answers to such questions are maintained in a database by the domain expert. But to retrieve the answers, the challenge is to understand the questions so that the best matched answer can be searched from the database. There problem comes under this domain and they are using template matching algorithm and SQL mapping to perform this matching. In addition, there target is to access the internet without having an actual internet connection on user's mobile phone. The user will be continuously updated through RSS feeds, can send or check the mails and the user can search whatever he is interested in by just asking a question in natural or SMS language through our android application.

There are various ways to express specific information needs. While traditional information retrieval systems rely on structured or unstructured keyword lists as queries, perhaps the most natural way for users to express an information need is through a natural language question. From a processing perspective, various syntactic, semantic and pragmatic aspects of such questions can be exploited by NLQA systems to interpret questions and develop a representation to match against possible answers. The syntactic structure of natural language (NL) questions is defined by the rules that govern question formation in a particular language. NL queries are not restricted by a specific formal structure, such as Boolean logic. Whereas a simple Boolean query might be constructed as "captain" AND "U.S.S. Minnow", the same query in natural language would be "Who is the captain of the U.S.S. Minnow?". For English, various structures, with and without question words, are considered "allowable" NL queries. Examples include "When did the United States enter World War II?" or "Explain U.S. foreign policy in the Middle East." While there is typically more freedom in structuring natural language than is allowed in a restrictive query language, there are rules for question formation that can be exploited in question answering systems.

3. Proposed System

Using computers to answer natural language questions is an interesting and challenging problem. Generally such problems are handled under two categories: open domain problems and close domain problems. This paper presents a system that attempts to solve close domain problems, open domain and if both not work our system will take help of Google API also. Typically, in a close domain, answers to questions are not available in the public domain and therefore they cannot be searched using a search engine. Hence answers have to be stored in a database by a domain expert. Then, the challenge is to understand the natural language question so that the solution could be matched to the respective answer in the database. We use a template matching technique to perform this matching. In addition, given that our target is to use this system with non-native English speakers, we developed a method to overcome the mismatches we might encounter due to spelling mistakes. The system is developed such that the questions can be asked using short messages from a mobile phone and therefore the system is designed to understand SMS language in addition to English. One of the main contributions of this paper is the outcome presented of a deployment of this system in a real

environment. Also some times the closed domains as well as open domain are not able to give output so in our system we are going to search such FAQ or queries in Google with help of Google API and store this query and its answer in database for future used. Following Diagrams show the comparative analysis of previous and our proposed system.

Input and output of system: The input for our system will be the question entered by the user. And the output of our system will be proper answer for user question with image and videos the answer is also provide by SMS to the user.

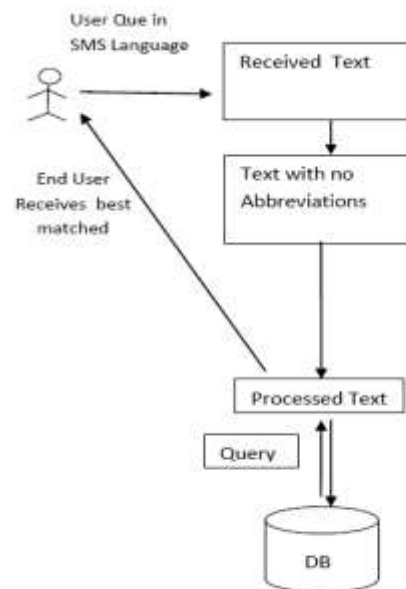


Figure1. Previous System Architecture

Fig 1 shows the previously done system architecture. In this they use closed domain methodology. They are processing the user query with help of data present in the database. If the data related to the user query is not available then this architecture will be give somewhat wrong output and many times no output.

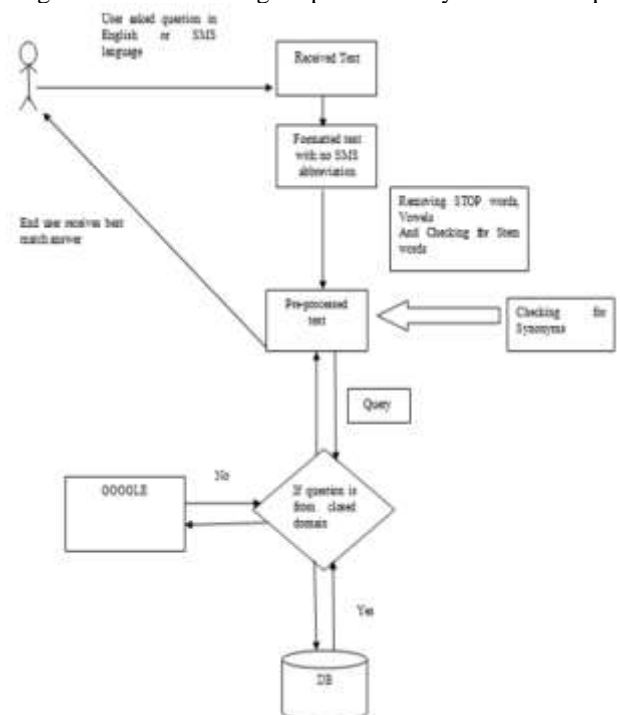


Figure2. Proposed System Architecture

Fig 2 shows the architecture of our proposed system in which we are going to implement closed and opened domain based system. Also if these two methods failed then we are using Google API to search engine. In our system if user insert some query then our system first check in our database i.e. closed domain system. And if the system will not got the output in database then it will search in Google search engine and also store this answer of query in database for future use. We will search output of query in Google with the help of Google API.

4. Conclusion

Thus a smart, user friendly automatic answering system with the ability of detecting and answering questions asked in English language is developed. There are many answering systems which use the concept of natural language processing to answer the questions but they are not so accurate in retrieving the appropriate answer. They also require specific predefined queries with specific format. For such systems users must ask the questions in the given format only. The answering system explained in this paper has multiple improvements as compared to the systems mentioned above. In the proposed system, users can enter the queries in the way they want. No specific format is required. If in some cases, the query does not match with any predefined query, the best matched query will be suggested to the user. Thus the system is more efficient and accurate from all the perspectives.

References

[1] Automatic Answering System for English Language Questions (IJCSIT) International Journal of computer Science and Information Technologies, Vol. 5, 2014.

[2] An Interactive Answering System using Template Matching and SQL Mapping for Natural Language Processing. International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 4, April 2014).

[3] N. Kerdprasop, N. Pannurat and K. Kerdprasop, "Intelligent Query Answering with Virtual, Mining and

Materialized Views". World Academy of Science, Engineering and Technology, 2008.

[4] J. Palme (2008, August 22). "Web 4 Health" [Online]. Available: <http://web4health.info/en/answers/project-search.htm>

[5] J. Palme and E. Sneider, "Natural Language Question Answering System Classification Manual", REPORT D2.2-part B Revision 1.0, 2008.

[6] Maxwell, Kerry (2007, August 13). "Disemvowelling or disemvoweling" [Online]. Word of the Week Archive. Macmillan Available:<http://www.macmillandictionaries.com/wordofthewek/archive/070813-disemvowelling.htm>

[7] Fellbaum, Ch. (ed). 1998. WordNet: An Electronic Lexical Database. Cambridge: MIT Press.

[8] A. Andrenucci and E. Sneider, "Automated Question Answering": Review of the Main Approaches, in International Conference on Information Technology and Applications, 2005.

[9] S.R. Balasundaram and B. Ramadoss. "SMS for Question-Answering in the m-Learning Scenario", Journal of Computer Science 3 (2): pp. 119-121, 2007.

[10] M. Suktarachan, P. Rattnamanee and A. Kawtrakul, "The Development of a Question-Answering Services System for the Farmer through SMS: Query Analysis", in Proceedings of the 2009 Workshop on Knowledge and Reasoning for Answering Questions, pp. 3-10, Suntec, Singapore, 2009.

[11] Afrati, F. et al, 2001. Generating efficient plans for queries using views. Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 319-330.

[12] Aragao, M. and Fernandes, A., 2004. Logic-based integration of query answering and knowledge discovery. Proceedings of 6th International Conference on Flexible Query Answering Systems