

# An Automatic Answering System using SMS service in Closed-Domain

Navin Gawde M.

Snehal Govilkar M.

Omy Lonare V.

Sneha Waje S.

Prof. Gade D.P

(Computer Department)

Shri Chhatrapati Shivaji College of Engineering

**Abstract**—This paper presents An Automatic Answering System which uses a SMS (Short Message Services) service to get an answer from the system. This paper had shown the working of system with its importance in future and day-to-day life. It tried to cover all the limitation of previous system and shows the increase needs of this system at different levels of organization. The system is able to answer the question of particular domain in which it is build from anywhere and anytime. System is capable to handle both SMS and Natural Language questions. User need to ask related question as an input to the system in any of the two languages and will get an appropriate output in a form of natural language.

**Index Terms**— Question Answering (QA), Short Message Service (SMS), Information Retrieval System (IRS), Template Matching, Frequently Asked Question (FAQ), Natural Language Processing (NLP).

## I. INTRODUCTION

Our system aim is to provide the end user with a solution to answer his queries using mobile phones and SMS with a service of 24x7. It is expected to make the user routine operations easier by offering him a wide range of system which is able to answer the queries as per user required. The systems receives the user's question in form of SMS or natural language via mobile phones, analyze it, parse it and then processed the text in natural language with use of template stored in database [2]. Finally the user will received text on his mobile phones with a concise and accurate answer.

The question answering (QA) system can better satisfy the needs of users as they will provide an accurate, quicker, convenient and effective way of giving answers to user questions [1].

### A. Question Answering System

The significance of the system is based on the fact that it provides a fast retrieval of information and would able to answer number of users simultaneously. This system accept the factual question such as Who, When, Where, What and How (4W1H), question types which helps to easily classify the types of question and found the related answer quickly [7]. This technology of QA is widely noticed as an advanced style

of fusion of Information Retrieval (IR) and Information Extraction (IE). This system gives us not relevant documents but the answers of question [4]. For example, if we asked the question of related domain to system i.e. "Where is SCSCOE college?" the system will first translate SMS language to natural language match with template and then answer the phrase that is stored in our database i.e. "Rahuri Factory, Tal- Rahuri, District- Ahmednagar".

QA system is classified in two domains viz, (1) closed-domain question answering system that deals with question related with specific domain or under specific knowledge which is not available in the public domain; and (2) open-domain question answering system that deals with question nearly everything which is present around this world knowledge. These types of system usually have more data available in the public domain to extract answer [1, 3].

In this paper, we describe the processes of system, the functionality of the software system, and the applications of the system that are built on the top of these technologies. Experimental results of answering system in their applications are also presented

### B. Information retrieval system (IRS)

Information Retrieval system (IRS) is a technique to search for the matches in the document database and retrieves the results to users. The goal of an IR system is to help user to locate the most similar documents that have potential to satisfy the user information needs. The focus of IR is the ability to search for information relevant to a user's needs within a collection of data which is relevant to the user's query [6].

The three main component of IR System is shown in Fig. 1 [10]. It contains Query Subsystem, Documentary Database and Matching Mechanism.

User questions are forward to the query subsystem which contains query language that formulate and collects the rules to handle and generate queries. It contains some components to select the relevant documents.

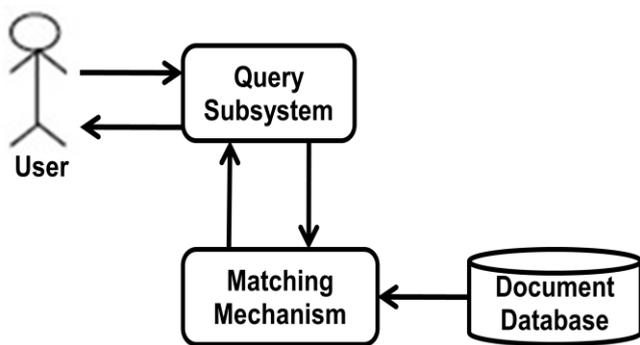


Fig.1. Information Retrieval System framework

Matching mechanism has the name consider matches the documents in the document database with query forwarded by the user to query subsystem. Query match with the document database is termed as relevant documents. It helps to retrieve the results. Document database is used to store documents and their representations [6, 11]. These component also contains an indexer module which helps to generate a representation for each document by extracting the document contents

## II. RELATED WORK

In late 60s and early 70s, use of natural language to access relational database was practiced in QA system. This system was built in closed-domain, and it was come to practice in 1961 at BASEBALL match, where system was able for answering the question about the baseball games played in American League over one season [11]. This machine was also practiced for LUNAR system where the system answer question about geological analysis of rocks returned by the Apollo missions. Lunar was able to answer 90% of question in its domain when queried by untrained person [2]. Around 90s, research moves further and develop an open-domain system which can be used by public for storing and retrieving data. QA system received considerable attention in 1999 from the research community Text Retrieval Conference [TREC-8] and was addressed first time by natural language processing (NLP) researches [1].

## III. TEMPLATE MATCHING ALGORITHM

In this section, we discuss the templates matching algorithm used and their syntax. Here some manually based templates are stored in a database with coupled of answer. The user questions are match with the question with the best matched template [1]. The quality and accuracy of the system is totally depends on the quality of templates stored in the database. Single question can be asked in multiple ways due to different tenses, various abbreviations, forms, and order of using words and also by using optional or words.

This algorithm is then enhanced by using two techniques i.e. using synonyms list and applying disemvoweling. Disemvoweling is carried out to clarify spelling mistakes and for easy matching of templates [8, 9]. Synonyms list is created since each of users as different terminology than person built the system. Same list is occurs many time so for fast accessing data those file is stored in text file which is mapped into a Hash Map so can be retrieved easily when program loads. Working of algorithm is shown in Fig. 2.

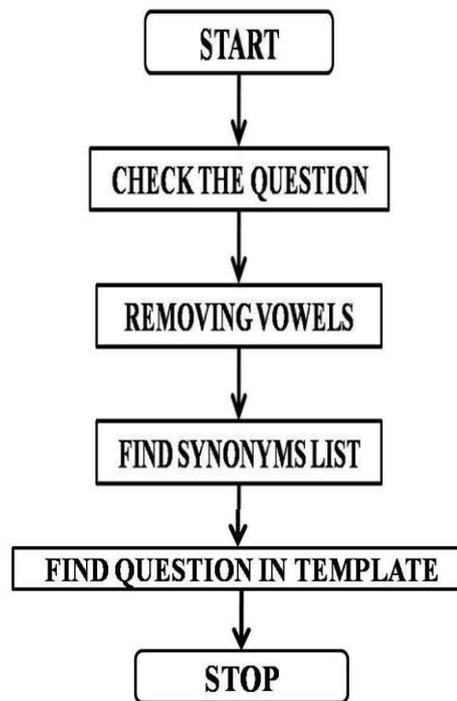


Fig.2. Algorithm Flowchart

Fast retrieval of information can be attained by adding more templates that can save information obtained by Web queries or by saving the data which were used previously by their users.

## IV. SYSTEM ARCHITECTURE

The main goal of question answering system is to make a machine able to answer the question without any error and with appropriate answer in short interval of time [1]. The system need to perform the way it was built and perform suitable operations as shown in Fig.3. QA system architecture contains four main modules viz, Pre-processing module, Question matching module, Answering module and Security module.

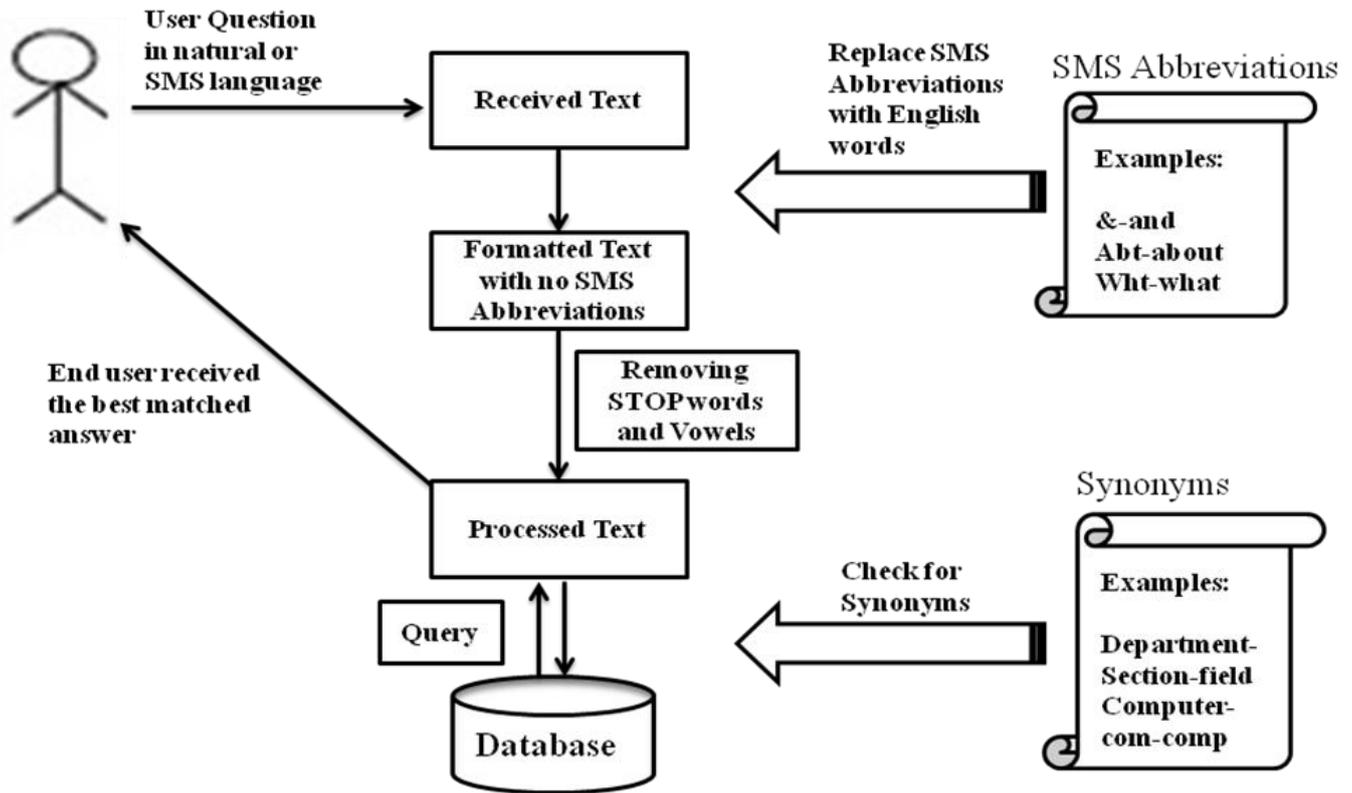


Fig.3. Architecture of Question Answering System

### A. Pre-processing Module

Pre-processing module is the first module of the system which handles the input text send by user through SMS. It mainly consists of three important operations: (1) Matching SMS abbreviations (2) Removing stop words, and (3) Removing vowels.

#### 1) Matching SMS abbreviation:

The system must be able to answer the both SMS and natural language system. But computer are able to understand only natural language [1]. So this module is used to convert SMS language to English words so that question can be processed further. This is done by referring the previous pre-stored abbreviations.

#### 2) Removing Stop Words:

In our sentence few words are present which doesn't have any effect on the meaning of sentence [12]. Even if we removed those words the meaning of sentence will remain same. This module removes those words to increase effectiveness of system by saving time and disk space.

#### 3) Removing Vowels:

Most of the mistake occurs due to addition or out of order vowels. Removing vowels would reduce the amount of spelling mistakes encountered in sentence. The process of removing vowels in text is known as Disemvoweling.

### B. Question Matching module

The pre-processed text is matched against each and every pre stored template stored in database. In this module, words that are considered to have synonyms are referred in a synonym file. It is worth noting that the templates here are for questions and not for answers. The main target of this system is to identify the closest template that matches the question we have received from the user [1, 8].

### C. Answering module

This module just matches the question with the answer which is present in the database. The best matched template is found and the respective answer will be returned to the end user.

D. Security

It's important to maintain the database appropriately without any mistake and empty answers [5]. So to maintain database and its security, only an authorize person must be able to update or add new data in database when the system needed. So this is also an important module for our system maintenance and working.

V. SYSTEM WORKING

A. Working of QA system:

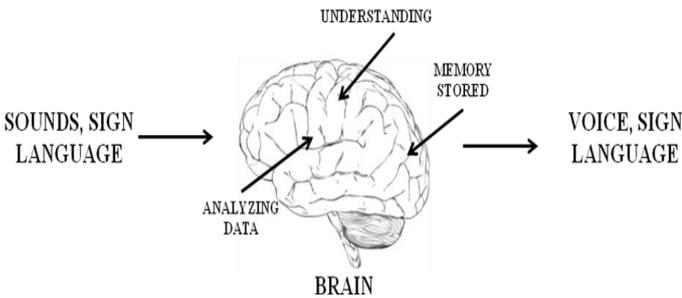


Fig.4. Human question answering system

Here working of QA System is compare with the human brain as shown in Fig.4. Since our brain is the complex structure phenomenon it is able to handle all the operations simultaneously. Our sensory organs such as eyes, ears and mouth behave as a modem which collects input in form of sounds and sign language and pass this message to brain where it will analyze the input. Simultaneously it will understand the input and gave response in fraction of seconds in form of voice or sign languages. Huge part of our brain cerebellum behaves as a database where the memory are stored and retrieved.

B. Working of QA system:

Our system needs three components such as GSM modem, QA system and Database to handle all the operation which brain is capable to do as shown in Fig.5.

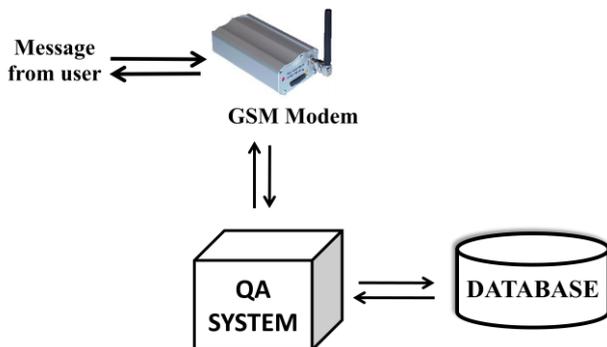


Fig.5. Computer question answering system

GSM modem is used to send/receive message from user to system and vice versa. GSM Modem is just a communication medium between two nodes. This message is then pass to QA system which will performs all the operations such as Matching SMS abbreviation, removal vowels, etc and then with use of template matching algorithm it searches the related answers from the database and then passes the message to the users following same path [3]. All this operations are completed with short interval of time and system is ready to handle next queries.

VI. ADVANTAGES

1. With use of security module we can protect our data from any unauthorized person to access or modify the data. Only an authorized person can change the data as per system required.
2. In these system user can ask the question using short message service (SMS) and get answer at any time and any place within a short interval of time. Service will be available for 24x7.
3. System can update or notify the administrator if any new word is found.

VII. APPLICATIONS

1. The system mainly used in the management organization likewise bus service, weather forecasting, etc.
2. Teaching: This system will make study more interesting and enjoyable by getting answer of question 24x7. Question for similar answer asks by different students can be answered by the system easily. The Shy students may feel more comfortable talking to a computer rather than a teacher to ask any queries.
3. Customer Service: Busy telephone service requires large staff to handle the problem differently. If we split those large documents into various documents and then store in our QA system then we can answer those query with no hours, queues and also single person can maintain the whole system.

VIII. EXPECTED RESULTS

In this section, we describe the experimental results of proposed QA system. We get quick answer within short interval of time. If users ask the question to the system then the system will find the answer in the database and provide to the user within the short time. If person need information

about the particular domain then they can send message through their mobile phones whenever they need it and they want it.

## IX. CONCLUSION

The final result is QA system is smart, user friendly system with ability of its detecting and answering the question in both languages viz, SMS and Natural languages. We introduced the structure of system and simultaneously shown the working of human brain during this operations. The accuracy of system was among top challenges but with use on new algorithms and its needs, system will able to handle all type of query with accurate answer needed by user and would capable to answer all question in any languages.

## X. REFERENCES

- [1] Tilani Gunawardena, Medhavi Lokuhetti, Nishara Pathirana, Roshan Ragel and Sampath Deegalla "An Automatic Answering System with Template Matching for Natural Language Questions" in IEEE 2012 International Workshop on Information and Electronics Engineering (IWIEE) 978-1- 4244-8551-2/10 2010.
- [2] Vanessa Lpez, Supervisors: Enrico Motta, Victoria Uren, and Marta Sabou "State of the art on Semantic Question Answering A Literature Review" in Knowledge Media Institute, Technical Report kmi-07-03, May 2007, MK7 6AA, United Kingdom.
- [3] Radu Soricut, Eric Brill "Automatic Question Answering: Beyond the Factoid "in Information Sciences Institute University of Southern California 4676 Admiralty Way Marina del Rey, 2004, CA 90292.
- [4] Tatsunori Mori, Tomohiro Ohta, Katsuyuki Fujihata, Ryutarō Kumon "A \* Search Algorithm for Question Answering" in The Third NTCIR Workshop, 2003 Graduate School of Environment and Information Sciences, Yokohama National University 79-7 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan.
- [5] Lucy Vanderwende "Answering and Questioning for Machine Reading" in Microsoft Research Redmond, 2006, WA 98052.
- [6] Philomina Simon, S. Siva Sathya. "Genetic Algorithm for Information Retrieval" in IAMA 2009, No. 978-1-4244-4711-4/09/\$25.00 ©2009 IEEE
- [7] K. Fujihata, M. Shiga, and T. Mori. "Extraction of numerical expressions by constraints and default rules of dependency structure" SIG Notes 2001-NL-145, Information Processing Society of Japan, Sept. 2001. (In Japanese).
- [8] Fellbaum, Ch. (ed). 1998. WordNet: "An Electronic Lexical Database. Cambridge" MIT Press.
- [9] Maxwell, Kerry (2007, August 13). "Disemvowelling ordismvoweling" [Online]. Word of the Week Archive
- [10] Michael Gordon, "Applying probabilistic and genetic algorithms for document retrieval", Computer Practices, 1208 -1218 , 1988
- [11] Androutsopoulos, I., Ritchie, G.D., Thanisch P.: "Natural Language Interfaces to Databases - An Introduction. Natural Language Engineering", 1(1) (1995) 29-81.
- [12] 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.

## AUTHORS

Navin Gawde M.  
Computer Department,  
ShriChattrapati Shivaji College of Engineering  
Rahuri Factory, Ahmednagar, India  
navingawde912@yahoo.com

Snehal Govilkar M.  
Computer Department,  
ShriChattrapati Shivaji College of Engineering  
Rahuri Factory, Ahmednagar, India  
1991@gmail.com

Omy Lonare V.  
Computer Department,  
ShriChattrapati Shivaji College of Engineering  
Rahuri Factory, Ahmednagar, India  
omy.lonare@gmail.com

Sneha Waje S.  
Computer Department,  
ShriChattrapati Shivaji College of Engineering  
Rahuri Factory, Ahmednagar, India  
wajesneha@yahoo.com

Prof. Gade D.P (Guide)  
H.O.D of Computer Department,  
ShriChattrapati Shivaji College of Engineering  
Rahuri Factory, Ahmednagar, India