ISSN: 2321-7782 (Online)
Volume 2, Issue 11, November 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Intelligent Question Answering System

Abhishek Shirode[1]
Department of Information Technology
MIT College of Engineering
Pune, India

Pravin Bibve[2]
Department of Information Technology
MIT College of Engineering
Pune, India

Kunal Tamboli[3]
Department of Information Technology
MIT College of Engineering
Pune, India

Rishikesh Kadupatil[4]
Department of Information Technology
MIT College of Engineering
Pune, India

Rajendra Pawar[5]
Department of Information Technology
MIT College of Engineering
Pune, India

Abstract: ***Everything that we do is digital now. If we want to get information we don't go in libraries. Instead we search it online. Search engines can return ranked documents as a result for any query from which the user struggle to navigate and search the correct answer. This process wastes user's navigation time and due to this the need for automated question answering systems becomes more urgent. We need such a system which is capable of replying the exact and concise answer to the question posed in natural language. The best way to address this problem is use of Question answering systems (QAS). The basic aim of QAS is to provide short and correct answer to the user saving his/her navigation time. The concept of Natural Language's Processing plays an important role in developing any QAS. This project is an implementation approaches for various categories of QAS such as Closed Domain based QAS, Open Domain based QAS, Information Retrieval or Information Extraction (IR/IE) based QAS which will be helpful for new directions of research in this area. The basic idea of this project is to create an interface with natural language processing capabilities that can search any database (e.g. Oracle) and return answer in form of a sentence.***

Keywords: Natural Language Processing (NLP), Question answering systems (QAS), Information Retrieval (IR).

I. INTRODUCTION

Today, when a computer encounters a natural language, it just considers it as a string. Most of the computers cannot understand/interpret what the natural language means. This shortcoming can be rectified by using a NLP based systems. The NLP based systems available today are not perfect when they are to be used for data extraction. Some systems have a good algorithm used for NLP but are ineffective when used for query generations. Thus we are planning to create an intelligent system which will integrate all the good qualities of these systems and will be universally applicable to any database available. What we intend to do is to create an interface which when applied to any database can be used for instantaneous data extraction. This can be attained by using natural language processing. Here the user will give a voice command which will be converted into text using pre-available tools. This text then will be processed using parsing techniques and the sentence will be divided into tokens such that the computer should understand the proper meaning of that sentence. This sentence will then be used to create a logical query. This logical query will act as a simple query stating which database to be searched. By using this logical query a database (SQL) query will be generated if the database is structured. In case of an unstructured database e.g. a book with paragraphs, Natural Language Processing will be used for data extraction. The final output will be represented into a textual form which will be converted to voice using text-to-voice tools.

II. Intelligent Question Answering System (IQAS)

The proposed system, i.e. the intelligent question answering system is a hybrid system which includes the combination of various technologies like speech recognition, natural language processing, and information retrieval. It is supposed to be an intelligent interface which can interact with a user and can provide faster access to database. The basic idea behind the proposed system is that the user will ask a question to the system in voice or text format. The system then will recognise what user needs by using natural language processing, then retrieve the required data from the database, and display the data to the user. Even if this seems as a simple process, it involved a lot of critical processing.

The vital part of this system is the natural language processing unit. This unit's function is to analyse the question asked and deduce what the user wants. This is the trickiest section as one small misinterpretation can lead to wrong answer. When the knowledge of what is required is obtained, it is then the job of system to locate the data in the database. In case of multiple hits, it should be able to understand which the correct data is. Then the obtained answer should be displayed to the user in form of text as well as text-to-speech audio.

The main benefit of the proposed system is that it can provide database access to any person regardless of his technical knowledge. This makes it valuable in the information sector. Since the system to be built is an interface, it's portable. It can thus be implemented on different operating systems. The flexibility of database makes it more useful in real world applications. In case the system fails to deliver, it will have a cognitive learning ability, where the query asked to the system will be forwarded to developer as a bug report, and the developer can remotely solve the problem and increase the processing scope of the system to include vivid scenarios which are related to the query at which the system previously malfunctioned.

III. Natural Language Processing

Natural language processing (NLP) is a method to translate between computer and human languages. It is a method of getting a computer to understandably read a line of text without the computer being fed some sort of clue or calculation. In other words, NLP automates the translation process between computers and humans.

NLP includes the following steps:

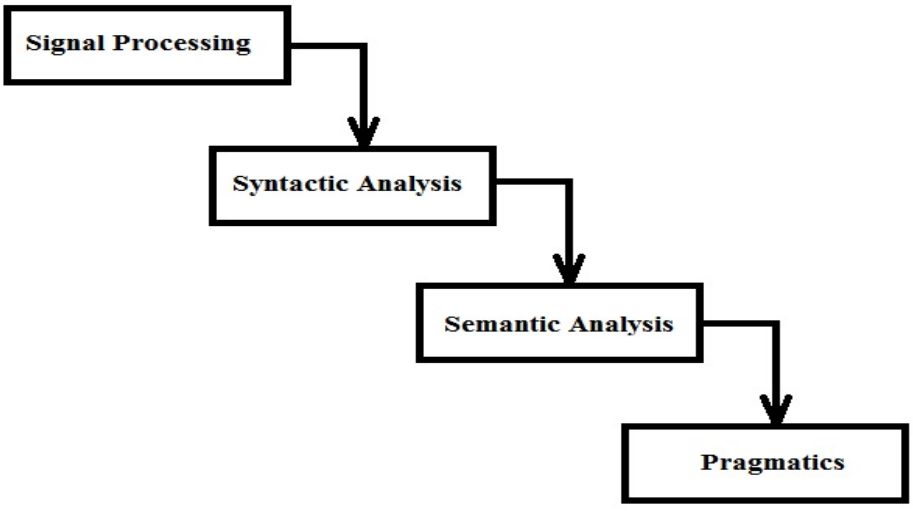

Fig. 1 Natural Language Processing (NLP)

1. Signal Processing

Signal Processing takes spoken words as input and turns it into text. Tokenization is a part of signal processing. Tokenization is the conversion of an input signal into parts (tokens) so that the computer can process it.

E.g.

Input: RAM is a volatile memory.

Output: RAM, is, a, volatile, memory

2. *Syntactic Analysis*

Syntactic Analysis takes care of the structure or grammar of the sentences. Computer system will read the text word by word and in the end produce a structural description, but problems can arise such as a word may function as different parts of speech in different context, there may be several possible interpretations of the structure of a sentence. Processing a sentence syntactically involves determining the subject and predicate, the place of noun, verbs, pronouns, etc. This is usually done by creating a parse tree.

E.g. Input: He ate the pizza

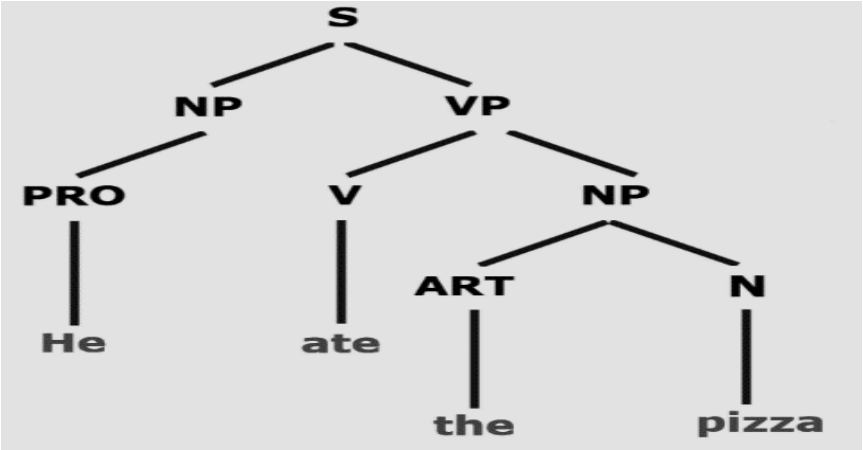

Fig. 2 Parse Tree created in Syntactic Analysis

3. *Semantic Analysis*

Semantic Analysis deals with the meaning of the words and sentences, the ways that words and sentences refer to elements in the world. Semantics analysis is used to represent:

i. Entities – individuals such as a particular person, location or product, e.g. John F. Kennedy, Washington, D.C., Cocoa Puffs

ii. Concepts – the general category of individuals such as person, city, breakfast cereal.

iii. Relations between entities and concepts

4. *Pragmatics*

Pragmatics concerns how the meaning of a sentence depends on its function in everyday life. In processing the Natural Language, some types of ambiguity arise. This can be eliminated by involving some kind of General Knowledge as well as specific knowledge about the situation.

E.g. Input: Bill was told to print the file by him.

In this sentence, it is not possible to know what “him” refers to unless we scan the previous and the next input. This task is taken care of in pragmatics stage to understand the sentence.

IV. WORKING OF SYSTEM

To use IQAS, user will have to first login to the system (for authorization).If the user is authorized, he will be directed to the next page which will provide an interface to the user to ask the question. The following steps will be performed in order to deliver the output to the question asked by the user.

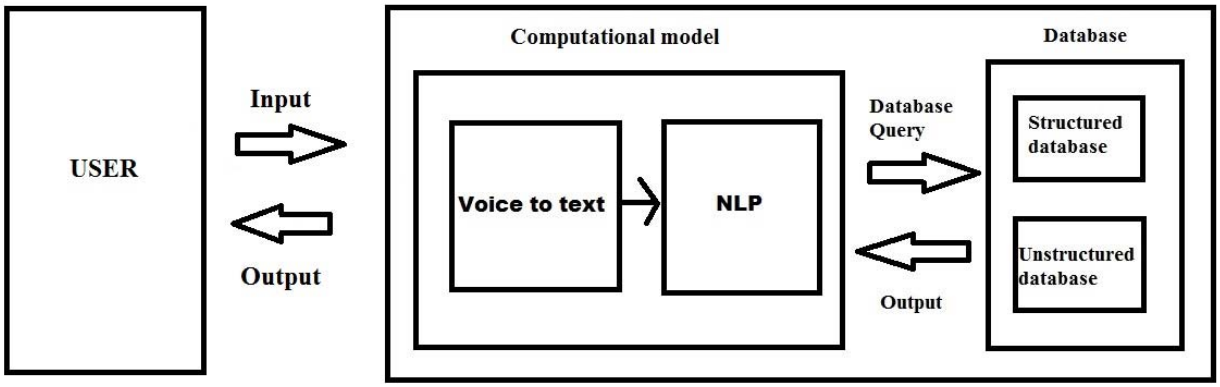

Fig. 3 Architecture of IQAS

1. Ask a question

After logging in to the system, IQAS will provide an interface to the user to ask the question in either voice or text format.

2. Speech to text conversion

If the user asks a question to the system in voice format, it needs to be converted to text format for processing the text and to extract the information from the database. So, for converting voice format to text format, the speech to text conversion tool will be used which will recognize the voice using microphone and will convert it to text format. In case, if microphone is not working, user can type the question using the interface provided to him.

3. Query generation using NLP

The question in the text format will be provided as input string to the natural language processing unit (NLP). The NLP unit will then process this text using different phases like signal processing, syntax analysis, semantic analysis, pragmatics etc. (as mentioned above), and will generate the database query.

4. Information extraction

The database query generated by the NLP unit will be fired to the database. The database will execute the query and will give solution again to NLP unit. The database used for storing the data from different domains in IQAS can be structured or unstructured. The information from structural database can be retrieved by simple execution of queries on the database. For unstructured database, we need to search data using a keyword and sort through the probable solutions to determine which is the correct information requested by user.

5. Deliver output

The result after executing the query on the database will be given as input to NLP unit. The NLP unit will convert this output to natural language. The purpose of converting database output to natural language is to make the system more user friendly. The output provided to NLP unit will be converted to voice format using text to speech conversion tool. And, finally the output will be delivered to the user in both text as well as voice format.

V. APPLICATIONS

This system is flexible and can be applied to any data sources. These systems can thus be used in

1. Business Domain

IQAS can be used in various industries by Business Analysts, Decision Scientists in order to analyze the data and to take decisions accordingly. Every company have their own database. Even though product based companies focus on product development, in order to solve their problems related to the market or financial conditions they need to approach Consultancy Firms. Each consultancy firm has team of Business Analysts and Decision Scientists.

Business Analysts and Decision Scientists will analyze the data provided by the Company. And by analyzing various trends they can take decisions accordingly. So, basically this system will be useful in decision making in order to increase the sales and revenue as well as to solve the business related problems in the industries.

2. Medical field

In healthcare, IQAS's natural language, hypothesis generation, and evidence-based learning capabilities allow it to function as a clinical decision support system for use by medical professionals. To aid physicians in the treatment of their patients, once a doctor has posed a query to the system describing symptoms and other related factors, IQAS first parses the input to identify the most important pieces of information, then extracts patient data to find facts relevant to the patient's medical and hereditary history, then examines available data sources to form and test hypotheses and finally provides a list of recommendations based on its likelihood. The sources of data that IQAS uses for analysis can include treatment guidelines, electronic medical record data, notes from doctors and nurses, research materials, clinical studies, journal articles, and patient information.

3. Education

Consider if the database selected for IQAS is a reference book of a particular subject. Now using natural language processing the system should be able to understand the whole content of the book, thus making it an expert system. If a student is to ask any question to the system, it should be able to retrieve it without any difficulty. This basic idea creates an environment where data availability is insured.

Any person who doesn't have proper skill set for handling a database can also use this system without extra knowledge. Since it is supposed to be a basic interfacing structure, any database can be used.

VI. CONCLUSION

In this paper, we proposed a question answering system which can be used in versatile environments. When fully developed, the system can be considered as benchmark for information retrieval systems. Due to flexible operating system and database support, the proposed system thus can be implemented in real life scenarios. It eliminates the need of trained and technical personals for accessing any database. If developed efficiently, the proposed system can be further crated into an expert system.

ACKNOWLEDGEMENT

We express our profound gratitude and deep regard to our guides Prof. Anil Hiwale (HOD, Department of Information technology, MIT College of Engineering Pune) and Prof. Rajendra Pawar (Assistant Professor, MIT College of Engineering Pune) for their exemplary guidance, constant supervision and encouragement throughout the course of this thesis.

References

1. "Data-intensive question answering", in D. Harman and E. Voorhees (eds): Proc of TREC 2001, NIST, Gaithersburg, USA, 2001, E. Brill, J et al.

2. Implementation Approaches for Various Categories of Question Answering System, ICT, 2013, Ms. Pooja P. Walke, Mr. Shivkumar Karale,

3. "Natural language interfaces: a new philosophy", Sun Expert Magazine, 1991, J.L. Binot et al.

4. "Natural Language Interfaces to Databases: An Introduction", Journal of Natural Language Engineering, 1 (1), 1995, I. Androutsopoulos, G. D. Ritchie, P. Thanisch.

5. "Automated Question Answering: Review of the Main Approaches", Andrea Andrenucci, Eriks Sneiders Stockholm University/ Royal Institute of Technology, Sweden {andrea, eriks}@dsv.su.se

AUTHOR(S) PROFILE

Abhishek Shirode, currently pursuing B.E from Department of Information Technology in Maharashtra Institute of Technology College of Engineering, Pune. (MIT-COE) (2011-2015 Batch).

Pravin Bibve, currently pursuing B.E from Department of Information Technology in Maharashtra Institute of Technology College of Engineering, Pune. (MIT-COE) (2011-2015 Batch).

Kunal Tamboli, currently pursuing B.E from Department of Information Technology in Maharashtra Institute of Technology College of Engineering, Pune. (MIT-COE) (2011-2015 Batch).

Rishikesh Kadupatil, currently pursuing B.E from Department of Information Technology in Maharashtra Institute of Technology College of Engineering, Pune. (MIT-COE) (2011-2015 Batch).

Rajendra Pawar, received the BE in Information Technology in 2005 and M.Tech degree in 2013. He is currently an Assistant Professor with in Maharashtra Institute of Technology College of Engineering, Pune. (MIT-COE)

