www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 3 March 2017, Page No. 20453-20457
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i3.09

A Survey on Question –Answering System

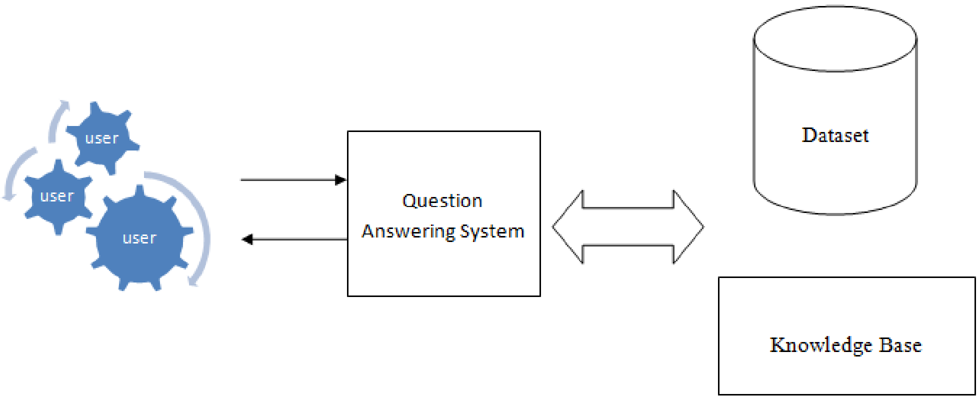
Anjali Saini[1], *P.K.Yadav*[2]
[1] Ph.D Scholar, Computer Science and Engineering, Singhania University, India
[2] Department of Computer Science , Singhania University, Rajasthan, India
[1] angel.anjali43@gmail.com ;[2] pkyadav2018@gmail.com

***Abstract-* Question-Answering(QA) is a new research area/ region in the field of Information science which comes into focus in last few decade. The present study is undertaken to survey about the QA system.The study in this paper will provide the guidelines to the researchers, scholars and practitioners of computer engineering. In this paper, the study is undertaken by planning, conducting, evaluating and reporting the literature review from the past years. A Question –Answering system consists of three core components i.e. question classification ,information retrieval and answer extraction module. This paper aims at giving an overview in this field, evaluating the current and emerging status and visualizing the future scope and trends.**

***Keywords-* Question –Answering system, text mining, Information Retrieval(IR)**

1. INTRODUCTION

A question-answering(QA) system is a simply an IR system in which a query is stated to the system and it transfer the closest or correct results to the specific question asked in natural language.It is a task which is designed to automatically answer a user's question placed in natural language.The answer to a question may come from background data collection of some form and question answering system might require going through the different steps of scientific technology to assemble a probable answer before returning the same to the user.Today with an evanishing increase in information ,the task of developing or making the systems that allow or grant user to quickly search desired information in large text volume is becoming more and more urgent.An example of such system is the question answering one. A QA implementation ,usually,computer system, may construct its answer by quering a structured database of information or knowledge, generally a knowledge base.More commomly, QA can pull answers from an unstructured collection of natural language documents.



Information Retrieval (IR),an open domain question answering system aims at returning an answer in response to the user's question.The answer which we get in return is in the form of short texts rather than a list of relevant documents.The system accepts a Natural language (NL) question as an input rather than set of keywords.The sentence is then direct transform into query through its logical form. Having the input in natural language make the system user friendly,but harder to implement as there are different or many question types and the system will have to find or identify the correct one in order to provide a sensible or relevant answer. Assigning a question's type to the question is a crucial task,the entire answer extraction process depends on finding the correct question type and hence the correct answer type.Key word extraction is the first step for identifying the input question type.Once the question has been identified, an IR system is used to find a set of documents containing the correct keywords.

Text Mining also referred to as text data mining, roughly equal or identical to text analytics, which refers to the process of finding high quality information from text. It involves the process of structuring input text, deriving/finding patterns within the structured data and finally evaluation and interpretation of output. Text analysis involves information retrieval ,lexical analysis to evaluate word frequency distributions, recognition of patterns ,information extraction, data mining. The overall goal is to change text into data for analysis, via application of natural language processing(NLP) and analytical methods or techniques.Web usage mining is the application of data mining techniques and methods to find the interesting usage patterns from web data in order to understand, provide and better serve the needs of web based application.Web content mining is the process of mining extraction and integration of effective or combination of data, information and knowledge from web page content.Web mining itself can be classified further depending on the kind of usage data i.e. web server data, application server data, application level data. Web mining has many merits which makes the technology attractive and enabled e-commerce to do marketing etc.

2. RELATED WORK

Many researchers worked on QA system from last many years in various languages.Some of the work of earlier authors are listed below-

From the history of QA systems the first known QA system is BASEBALL that was developed by GREEN et al [3] in 1961, it answers question about all the baseball games played in the American league in one season. It is a domain specific QA system with limited set of data sets available.

Then later a system named LUNAR which was designed in 1971 a result of Apollo moon mission, LUNAR helps geologists to easily access, compare and evaluate the data of their chemical analysis of lunar rock & soil which was gathered under that mission. In 1993 world’s first online question answering sys- tem was available named START developed by Boris Katz[2], it answered the question asked in natural language in all domain. Further improvements were made in the systems to increase its answering ability and performance.

Nguyen and Le [4] had proposed his research paper and utilize restricted semantic grammars to transform a natural question into an SQL query.

Yen et.al explained a QA framework based on Machine Learning [5], which incorporates the classifier based on questions, manageable documents or passage retrieval. Purely, QA system is a technique of uncovering the exact answers to the questions asked by the user over a massive collection.

Kumar et.al developed system which is domain independent, and identifies a domain by using knowledge base in Hindi Natural Language [6]. The query correctly identifies the domain and translates it into the SQL query.

Many QA systems work on the concept of machine learning such as Support Vector Machine. Machine learning requires similarity functions that can calculate likeness. Similarity Computations require mainly for clustering[7].

A hybrid hierarchy-of-classifiers framework for finding the quality answers in yahoo answers is proposed by Toba,Hapens [8] . Before analyzing the different answers to a given question, the question is analyzed first. In this paper ,the user answers are compared with the expected answers which has been already stored for the different question types. The framework is compared with different questions from yahoo answers and the best answer prediction is good and very accurate .

The authors in [9] proposed a question answering system for Vietnamese named entities, which restricts to only questions of the form "Who?","Whom?" and "Whose?”

3. QUESTION- ANSWERING SYSTEM FRAMEWORK

With the advancement in technology it becomes very easy to fetch the required information on a finger tip by using a single mouse click.Typical QA Comprises of three Phases or modules. These three main modules are-

1) Query Processing module

2) Document Processing module (information retrieval)

3) Answer Processing module

1. Question Processing Module

The goal of question- answering system is to extract a number of pieces of information from the question.The query specifies the keywords that should be used for the IR system to use in searching for documents.For the given natural language query as input, the function of query processing module is to analyze the query and process by creating some representation in some format of the information required. So query processing module must have to do:

- *Analyze* query for representation of main information that is required to answer users query.
- *Classify* the question according to the taxonomy or keyword used in the query, which leads to the expected answer type.
- *Reformulation* of query for enhancing question phrasing and transform query in to semantically equivalent ones which helps in information retrieval process.

These steps will generate set of query terms and pass them to document processing module which performs information retrieval process by using them as shown in the figure 1.

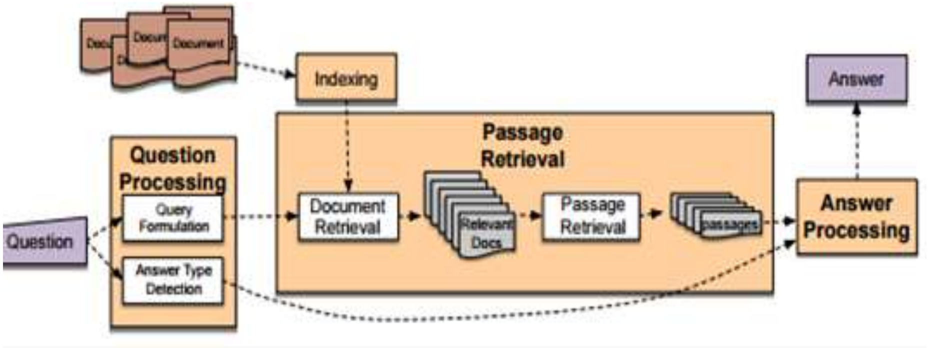


Figure 1 : Question Processing Module

2. Document Processing Module

This module takes reformulated questions an input and submitted that questions to information retrieval systems, which retrieves ranked list of relevant documents. This module mainly depends upon one or more information systems to gather the relevant information and probably uses WWW for information retrieval purpose. Then after that it must filter and order the documents retrieved. Goal of this module is to generate set of candidate paragraphs that contains answers, for that it must do.The task of question classification or answer type recognition is to determine the answer type,the named-entity or similar class categorizing the answer.

Information Retrieval—

There is lots of information available on web in almost all domains. Therefore, there is need of a system that makes these information available to the user, so information retrieval system is needed. Information retrieval system’s goal is to retrieve possibly accurate results in response to query submitted by user and rank these results according to relevancy. But one thing is to be considered that IR systems uses cosine vector space model to measure similarity between query and document. However QA system wants to retrieve only those documents when all keywords are present. IR systems are evaluated on the basis of their capability of precision and recall as the performance metrics. However QA system has nothing to do with precision, their main focus is on recall. Since QA systems later processes returned documents, the recall of IR system should be prioritize over its precision.

3. Answer Processing Module

The final phase of QA architecture is Answer Processing module which is responsible for identification, extraction and validation of answers from set of generated ordered paragraph received from document processing module. Task to be done by answer processing module is to

1. *Identify Answer Identification –In this* candidate answers from filtered paragraphs. Determining Answer type during question processing module is very beneficial for identification or finding of the answer. Type of answer is totally relies on the questions which have asked.

2. *Answer Extraction* - answer by choosing phrase or words according to question classification. The parser helps in the recognition of answer candidates present in paragraph. So after identifying the answer candidates, we apply set of heuristics for extracting relevant words or phrases that asked question.

DOI: 10.18535/ijecs/v6i3.09

3. *Answer Validation* - providing validation or confidence in correctness of answer. It considers degree of match between passages retrieved corresponding to the question, source reliability, temporal relation- ships, taxonomic classification and semantic relationships.

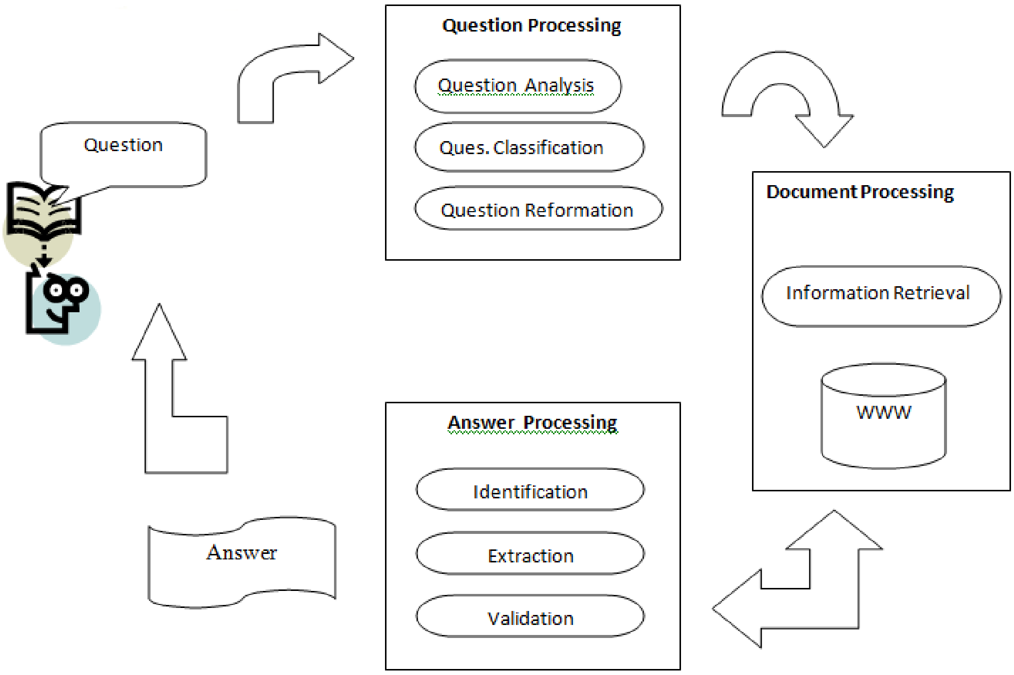


Figure 2. Framework of Question Answering System

QA system enables the user to access or get the information resources naturally by asking query in natural language(NL) and get back the precise,concise and relevant response as result or output. Many techniques and methods from AI(Artificial Intelligence), NLP, Information retrieval, Information extraction and machine learning are combined or integrated together to provide a better and effective QA system. A Question or we can say query and its Answers or Output are defined in terms of natural language statement. Classification is well known approach and is defined as picking out or choosing the correct “Class Label” from the given input (Query). Generally in Classification tasks, set of labels is satisfy or clarify in advance and each input is estimate in isolation from all the other inputs that are given. QA system undertakes queries by the user in Natural language(NL),that searches for the accurate answers from a bundle of catalogues and return with the perfect,concise and accurate one as shown in figure 2.

4. BENEFITS AND FEATURES OF QA SYSTEM

1. It searches for information which is in Natural Language.The basic motive behind the development of any QA system is to provide or give the precise or concise responses to the questions written in natural language in order to save time and effort.
2. It provides greater relevance of found questions.
3. We can found the answers to the questions in world wide web(WWW) or in corporate documents.
4. With the help of Question Answering system document and knowledge management will greatly increase.
5. It uses semantic technology to find the information and give the answer to such type of questions like who? where? how? Etc.
6. It also searches for information at E-libraries/catalogues etc and can be used at document and knowledge management systems.

5. CONCLUSION AND FUTURE SCOPE

In this paper, we presented a research perspective on Question –answering (QA) System.By the survey it is seen that, the interest in seeking ,finding and sharing questions and answers through the QA system has been increased for finding the best answer in such systems is one of the thee main task that we are facing today.The complete task of overall accuracy and similarities are used which gives a good platform for user to ask query in Natural Language Processing (NLP) and get back result/ answer in return.

One of our goal for future is to extract more features in this system. Firstly, question references in answers which it means how many answers which of other question has been referenced to this question. Secondly, self-answering which it means whether a user answer to himself/herself and so on.Thirdly,working on comments and unlabelled data are and her future work.

DOI: 10.18535/ijecs/v6i3.09

REFERENCES

[1] Ferucci D. et al. "*Building Watson: An overview of deep QA project*", AAAI magazine fall 2010.

[2] Katz.B, Gary Borchardt, "*Natural language annotations for question answering*", AAAI magazine fall 2007.

[3] Green B., et al. "*BASEBALL: An automatic question answerer*", In proceedings of western joint IRE-AIEE-ACM Computing Conference, Los Angles, p.219-224, 1961.

[4] AnhKim Nguyen and HuongThanh Le. Natural language interface construction using semantic grammars. In *PRICAI 2008: Trends in Artificial Intelligence*, volume 5351 of *Lecture Notes in Computer Science*, pages 728–739. Springer Berlin Heidelberg.

[5] Show-Jane Yen, Yu-ChiehWu, Jie-Chi Yang, Yue-Shi Lee, Chung-Jung Lee, Jui-Jung Liu," *A support vector machine-based context-ranking model for question answering,"* Information Sciences, ScienceDirect, pp. 77-87, 2013.

[6] Mohit Dua, Sandeep Kumar, Zorawar Singh Virk, " *Hindi Language Graphical User Interface to Database Management System,*" International Conference on Machine Learning and Applications, IEEE, 2013.

[7] Sneha Bagde, Mohit Dua, and Zorawar Singh Virk, *" Comparison of Different Similarity Functions on Hindi QA System,*" ICT4SD, Springer, in press.

[8] Toba, Hapnes, et al. "*Discovering high quality answers in community question answering archives using a hierarchy of classifiers*." Information Sciences 261 (2014): 101-115.

[9] Mai-Vu Tran, Duc-Trong Le, Xuan-Tu Tran, and Tien-Tung Nguyen. ""*A model of vietnamese person named entity question answering system*. *26th Pacific Asia Conference on Language, Information and Computation (PACLIC 26),*page 325, 2012.

[10] Mai-Vu Tran, Duc-Trong Le, Xuan-Tu Tran, and Tien-Tung Nguyen. A model of vietnamese person named entity question answering system. *26th Pacific Asia Conference on Language, Information and Computation (PACLIC 26),*page 325, 2012.

[11] Vu Mai Tran, Vinh Duc Nguyen, Oanh Thi Tran, Uyen Thu Thi Pham, and Thuy Quang Ha." *An experimental study of vietnamese question answering system*". In *International Conference on Asian Language Processing, 2009. IALP'09.*, pages 152–155.

[12] Peng F., et al "*Combining deep linguistic analysis and surface pattern learning: A hybrid approach to Chinese definitional question answer- ing*", In proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), p.307-314,2005

[13] Zhang D., & Lee W. "A *Web-based Question Answering system",* Massachusetts Institute of Technology (DSpace@MIT), 2003.

[14] Zhang D., & Lee W. "*Question Classification using Support Vector Machines*", In proceedings of the 26th annual international ACM SIGIR Conference, 2003.

[15] E. M. Voorhees and H.T. Dang, *'Overview of the TREC 2005 Question Answering Track'*, 2005.