

**INTEGRATING QA INTO GENERAL NOTION OF INFORMATION  
RETRIEVAL**Tejas Rupani<sup>1</sup>, Girish Mulchandani<sup>2</sup>, Nivid Limbasiya<sup>3</sup><sup>1</sup>Computer Engineering, V.V.P Engineering College, rupani.tejas@gmail.com<sup>2</sup>Asst. Prof., Computer Engineering, V.V.P Engineering College, ghmulchandani@gmail.com<sup>3</sup>Asst. Prof., Computer Engineering, V.V.P Engineering College, nlimbasiya24@gmail.com

---

**Abstract**—The Question Answering System is used to retrieve the exact or relevant answer of questions which is written in Natural language by user. This system is specially developed for education system in which user can also provide the data source from which user want to get the relevant or exact answer rather than the list of documents which is given by most of the search engine. As we know most of the search engines are keyword base so the answer of query written in natural language is differ from the query written in keyword. But this system can handle the natural language and also provide only the answer not the entire list of documents as given by search engine. So it's save the time of user and provides accurate and relevant answer. This system normally uses the concept of information retrieval which is the basic part of QA system and also use answer analyzer, question analyzer.

---

**Keyword-** Question Answering System, Information Retrieval, Educational QA

**I. INTRODUCTION**

QA system is natural way to request information that we do not know, check information that we are not sure. Developing of QA system involves Information Retrieval and Natural Language Processing. It is the better way to find the information directly rather than the search that particular information. Create QA system in which user can upload the document from which, user want to extract the information and want to get answer without wasting of time to search large number of documents or data. Also user can search the query from already available source of data. User enters the query or question in natural language which is not handled by most of the search engine. After that question analyzer is used to analyses the question and identified the types of question among available category the question, so that system can judge which kind of answer will be possible for such kind of questions. Question type and its expected answer type are generally identified by looking at question keywords.

In any education system, students want to gain more knowledge by own efforts. And important part in this kind of system searching the information is very useful to solve the doubt. So what happens, generally students search the web and get very large amount of data from large scattered available database. Now it's very tough situation for the student to go thoroughly to all the available data and find any single accurate answer of questions. QA system is such kind of system that only gives the short term answer from very large source of data rather than the list of documents.

Classification of Question Answering System:

1. Content based classification
  2. Information source based classification
  3. Language paradigm based classification
1. Content based classification
    - a. Open domain Question Answering System.  
It provides answer to any type of questions.
    - b. Restricted domain Question Answering System.  
It provides answer to any specific types of questions. For Example gives answer to only medical related question, engineering related question, games related question etc.
  2. Information Source based classification
    - a. Data base oriented Question Answering System.  
In this we already have some fixed kind of data base, means structured data and answer of user's query will be identified from this set of facts and give the result.
    - b. Web-oriented Question Answering System.  
In this answer of user's query is search from web, as done in search engine. So we have large data which will be unstructured and scattered data.
  3. Language paradigm based classification
    - a. Single Language Question Answering System.

In this type of system, Questions, data base and answer all are in single language.

b. Cross Language Question Answering System.

In this type of system, user's query and resource both are in different languages. When users enter queries then it will be translated into resource language and then searching will be done. After that answer is also converted into same language of user.

### Introduction of Information Retrieval

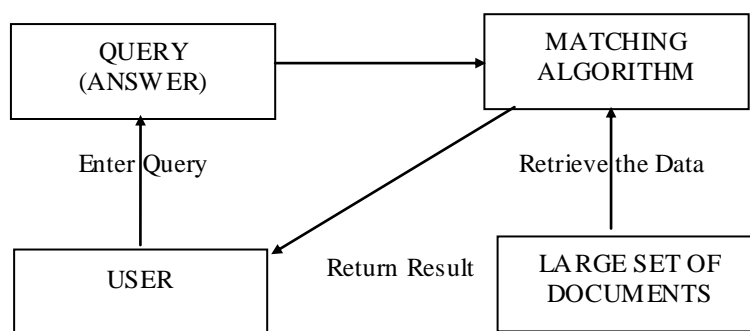
Information retrieval is the process of searching the documents of an unstructured way that satisfied the information requirement of user from large available scattered data. Information retrieval locates relevant documents, on the basis of user input such as query. Web search engines are the most common example of information retrieval.

Simple Example:

There is a library of number of books and students want to formulate the answer of query or question that will be generated into the mind then what students do?

Solution: Read all the related books and find the accurate answer from available resources which takes lots of time and simply not possible!!!!

Another solution: Create a system that read the related books and find the relevant answer.



### INL Question Classification:

1. Factoid questions.  
When, who, how many, where, etc.
2. Yes/No questions.  
Answer of this type of questions is either yes or no.
3. Definition questions.  
Answer of this type of questions is simple definition.
4. List type questions.  
Answer of this type of questions is list of details.
5. Cause & Effect.  
Answer of this type of questions is reason of particular event.
6. Opinion.  
Answer contains different types of view of user or different humans on different things.

## II. COMPONENT OF QA SYSTEM

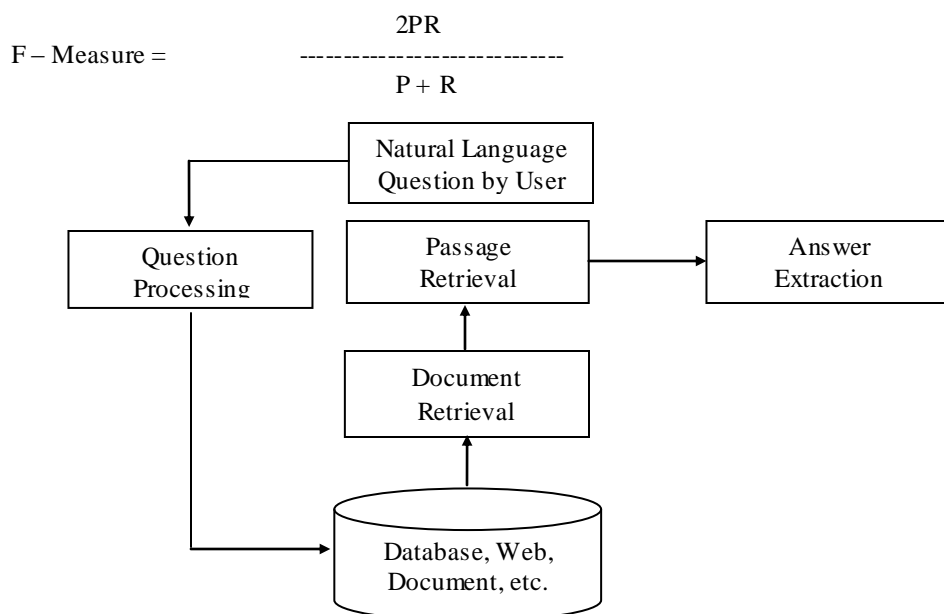
First when users enter their query using natural language then it will be processed. Question processing system identified the type of question and based on that identify the type of answer of that particular question. Based on different keywords first of all document is retrieved from large data base or from internet or large data source that are available. Now from that passage is retrieved as possible way to find answer of particular question. Now from the available passages, answer extraction is used to extract the answer with best possible way. Answer preparation is used to identify the higher rank of answer from short listed answer retrieved by answer extraction.

There are so many different systems are available for evaluating the performance and correctness of QA system as well as Information Retrieval System. The measure required the collection of document and a query. Every document is either relevant or non relevant based on the query. Consider following parameter for performance and correctness.

Precision – It is the probability of retrieved document which is correct or relevant.

Recall – It is the probability of relevant document is retrieved in a search.

F – Measure – It the ratio of multiplication of precision and recall with addition of them.



**Figure1.** Simple View of QA System

### III. IMPORTANT ASPECT OF QA

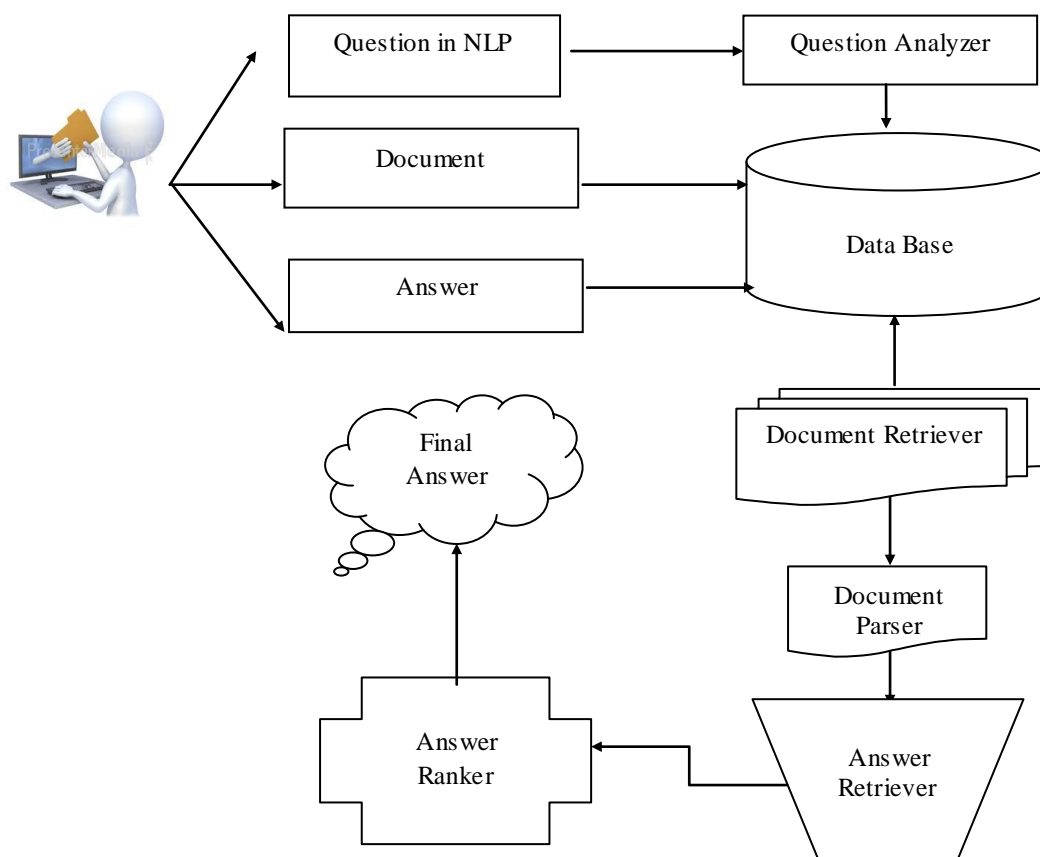
There are two main points which are very important for QA system.

- 1) Questions 2) Answers

**Questions:** We can divide questions depends on answer types. We can distinguish different kinds of question like yes/no question, “wh” questions, list questions, opinion questions. Some of the questions are harder than others.

**Answer:** Answer can be long of short, they may be list or narrative (story) types. Also there are different methodology are used for answer creation. Extraction and generation are two most popular methods for that.

### IV. PROPOSED SYSTEM



## **VI. COMPONENT OF PROPOSED SYSTEM:**

- 1) Question Analyzer
- 2) Document Retriever
- 3) Document Parser
- 4) Answer Retriever
- 5) Answer Ranker

➤ **Question Analyzer:**

This component gets the value from user. User enters the query or question in natural language form. Question analyzer receives the question and based on the different classification of question as we know it will prepare the type of expected answer. After that this component also finds the different keyword from the question and sends it to next component document retriever.

➤ **Document Retriever:**

This component used the keyword which will be provided by the Question analyzer and based on that keyword start to search the documents from the database. So, this component finds the relevant documents from the available large amount of document. Then these retrieved documents are passing to the next component document parser.

➤ **Document Parser:**

It is used to find the relevant passage from the relevant document which is the input from document retriever. Then multiple passages are transfer to next component answer retriever.

➤ **Answer Retriever:**

It takes the input from document parser and analyzes it and finds the accurate and relevant answer based on different expert knowledge, FAQ and types of possible answer. These possible answers are transferred to next component answer ranker.

➤ **Answer Ranker:**

Answer ranker takes input from answer retriever and give the rank to particular passage or particular answer which is very near relevant or accurate.

As per the figure user first enter the query in Nature language. User can also enter the document from which user want to extract information. User also can enter answer of particular question so that can be used in frequently asked question answer. Now question analyzer analyzes the questions and from that classified the question into category. Also it retrieve the keyword from query enter by user, and match it with available data base. Document retriever retrieve the document from database which is relevant to keyword of query enter by user. Document parser parse the document and retrieve passage which is related to question based on its types and keywords. Answer retriever retrieves the different types of answer from the different passage and give the rank. In this user can also give their feedback related to answer so that can be used to create knowledge database. Base on rank, answer will be given to user.

## **VII. CONCLUSION**

The given proposed system can be utilize to complete its objectives to get the answer in accurate or exact information form rather than the list of information that need to be viewed by the user of question written in natural language. It is very useful for the student who want to get right information from given source of data without referring much of data available in web. As we know there are so many question answering systems, but QA system for education purpose are not available. All the system use web content to get the answer, so web crawling is must and also for that keyword is very important so problem occurs with natural language questions. In this system user can give the document from which user want to get the answer but not want to read entire document.

## **VIII. ACKNOWLEDGEMENT**

I wish to express my gratitude to my guide Prof. Girish H. Mulchandani, Computer Engineering Department, for introducing me to the problem and providing valuable advice. He not only provided me help whenever needed, but also gives motivation for the same.

#### **REFERENCES**

- [1] Intelligent question answering system-Review and Research, Tomljanovic in IEEE conference 2014.
- [2] A Question Answering System Supported by IR, Rohini Srihari and Wei Li In IEEE conference 2010.
- [3] A Simple Question Answering System, Richard J Cooper and Stefan M Ruger in Imperial College of Science 2010.
- [4] A Hybrid Question Answering System based on IR and A V, Sivaji Bandyopadhyay in Jadavpur University 2011.
- [5] Designing An interactive open domain Question Answering System, S.Quarteroni, S.Manandhar in university of your 2007.
- [6] Question classification and answering from procedural text in english, Somnath banerjee in Jadavpur University 2012.
- [7] A NL Question Answering System as a Participant in Human Q&A Portals, Tiansi Dong, Alrich furbach in Germany.
- [8] Designing an interactive open domain question answering system, S. Manandhar, the University of York 2007.
- [9] Question Answering System: State of the art and future directions, Farah Benamara (IRIT), Bangkok.
- [10] Question Answering, Chin-Yew lin, senior researcher, knowledge mining group, Microsoft research Asia.
- [11] Creating missing classes automatically to improve question classification in question answering system, Bhaktyar M, Doudpota D.M, In IEEE conference 2012.