

# A Survey of Text Question Answering Techniques

Poonam Gupta

ME, Computer Science & Engineering  
University Institute of Engineering & Technology,  
Panjab University, Chandigarh

Vishal Gupta

Assistant Professor, Computer Science &  
Engineering Department  
University Institute of Engineering & Technology  
Panjab University, Chandigarh

## ABSTRACT

Question Answering (QA) is a specific type of information retrieval. Given a set of documents, a Question Answering system attempts to find out the correct answer to the question pose in natural language. Question answering is multidisciplinary. It involves information technology, artificial intelligence, natural language processing, knowledge and database management and cognitive science. From the technological perspective, question answering uses natural or statistical language processing, information retrieval, and knowledge representation and reasoning as potential building blocks. It involves text classification, information extraction and summarization technologies. In general, question answering system (QAS) has three components such as question classification, information retrieval, and answer extraction. These components play a essential role in QAS. Question classification play primary role in QA system to categorize the question based upon on the type of its entity. Information retrieval method is get of identify success by extracting out applicable answer post by their intelligent question answering system. Finally, answer extraction module is rising topics in the QAS where these systems are often requiring ranking and validating a candidate's answer.

Most of the Question Answering systems consists of three main modules: question processing, document processing and answer processing. Question processing module plays an important part in QA systems. If this module doesn't work correctly, it will make problems for other sections. Moreover answer processing module is an emerging topic in Question Answering, in which these systems are often required to rank and validate candidate answers. These techniques aiming at discovering the short and precise answers are often based on the semantic classification. QA systems give the ability to answer questions posed in natural language by extracting, from a repository of documents, fragments of documents that contain material relevant to the answer.

## General Terms

Types, Architecture, Applications, Information retrieval.

## Keywords

Natural language processing, Question answering System, Information retrieval.

## 1. INTRODUCTION

NLP focuses on communications between computers and natural languages in terms of theoretical results and practical applications, and on information sharing now that information is exchange as it never has been before and sharing information becoming the leading theme in the domain of

NLP systems. This movement leads to an explosion of activities like information retrieval, natural language understanding, etc. [1][2][3]. Information retrieval is an art and science of searching for information in documents, searching for documents themselves, searching for metadata which describe documents, or searching within databases, whether relational standalone databases or hypertext networked databases such as the Internet, for text, sound, images or data [4].

Question answering is a difficult form of information retrieval characterised by information needs that are at least somewhat expressed as natural language statements or questions, and was used as one of the most natural type of human computer communication. In comparison with classical information retrieval, where complete documents are considered similar to the information request, in question answering, specific pieces of information are returned as an answer. The user of a question answering system is interested in a concise, comprehensible and correct answer, which may refer to a word, sentence, paragraph, image, audio fragment, or an entire document [13]. The main purpose of a QA system is to find out "WHO did WHAT to WHOM, WHERE, WHEN, HOW and WHY?"[11]. QA systems merge information retrieval with information extraction methods to identify a set of likely set of candidates and then to produce the final answers using some ranking scheme [12].

In current years, there has been a marked increase in the amount of information available on the Internet. Users often have specific questions in their mind, for which they expect to discovering out the answers. They would like to find out the answers to be short and precise, and they always prefer to express the questions in their native language without being restricted to a particular query language, query formation rules, or even a particular knowledge domain. The latest approach taken to matching the user needs is to carry out actual investigation of the question from a linguistic point of view and to attempt to understand what the user really means.

A typical pipeline Question Answering System consists of three distinct phases: Question classification, information retrieval or document processing and answer extraction. Question classification is the first phase which classifies user questions, derives expected answer types, extracts keywords, and reformulates a question into semantically equivalent multiple questions. Reformulation of a query into similar meaning queries is also known as query expansion and it boosts up the recall of the information retrieval system. Information retrieval (IR) system recall is very important for question answering. If no correct answers are present in a document, no further processing could be carried out to find an answer. Precision and ranking of candidate passages can also affect question answering performance in the IR phase.

Answer extraction is a final component in question answering system, which is the tag of discrimination[5].

## 2. GENERAL ARCHITECTURE

The user writes a question by means of the user query interface. After that this query is used to extract all the possible answers for the input question. The architecture of Question-Answering system is as shown in Figure 1.

The architecture which is given in Figure 1 works in 5 stages. The function of each stage is as follows [6]:

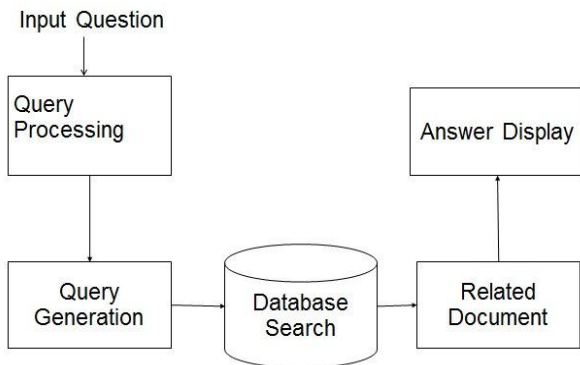


Figure 1. Architecture of Question-Answering System

### 2.1 Query Pre-processing

Given a natural language question as input, the overall function of the question preprocessing module is to process and analyze the input question. This leads to the classification of question as belonging to any of the types supported by the system.

### 2.2 Query Generation

In query generation we will use Query Logic Language (QLL) which is used to express the input question.

### 2.3 Database Search

Here the search of the possible results is done in the stored database, the related results that satisfy the given query with selected keyword and rules are sent to the next stage.

### 2.4 Related Document

The result which was generated by the previous stage is stored as a document.

### 2.5 Answer Display

The result is stored as a document which is in wx format . Then the result is converted into required text which is required by the user and displayed to the user.

## 3. TYPES OF QA SYSTEMS

Different types of QA systems which are divided into two major groups based on the methods used by them. First group of QA system belongs to simple natural language processing and information retrieval methods, while another group of QA systems are dependent upon the reasoning with natural language.

The two QA systems are compared with characteristics of different dimension such as techniques used, question that

deals with and so on. The table 1 provides the detail of the comparisons of these QA systems.

Table 1, Characterization of QA systems

DIMENSIONS	QA system based on NLP and IR	QA systems Reasoning with NLP
Technique	Syntax processing, Named Entity tagging and Information Retrieval	Semantic Analysis or high reasoning
Data Resource	Free text documents	Knowledge Base
Domain	Domain Independent	Domain Oriented
Responses	Extracted Snippets	Synthesized Responses
Questions Deals with	Mostly wh- type of Questions	Beyond of wh- type of questions
Evaluations	Uses existing Information Retrieval	N/ A

### 3.1 Web Based Question Answering Systemiser

With the wide spread usage of internet a tremendous use of data is available, web is one of the best source to obtain the information. Web based question answering systems is using the search engines (Like Google, Yahoo, Alto Vista etc..) to get back webpage's that potentially containing answers to the questions. The majority of these Web based QA systems works for open domain while some of them works for domain oriented also. The wealth of information on the web making it an attractive store for getting quick answers to simple, factual questions[16]. The data that is available on web has the characteristics of semi structure, heterogeneity and distributivity.

The Web Based QA systems mostly handles wh-type of questions such as "who killed Indira Gandhi"?

Or "Which of the following is correct". This QA system provides answers in various forms like text documents, Xml documents or Wikipedia. The common levels that are used by different web based Question Answering systems architectures are as follows [10]:

□ **Question Classification:** This level gives correct answers by classify the user query into one of the question type to which it belongs to. The question classification is made to provide better accuracy in the results.

□ **Answer Extraction :** This level extracts the correct possible answers for different classification of questions.

□ **Answer Selection:** Among the possible answers obtained, ranking approaches are used to find out the best accurate answers based on its weightage factor.

Answer classes generally is of factoid and non - factoid types. The factoid is getting short fact based answers like names, dates, and non-factoid is getting descriptions or definitions[27].

Given a user's natural language question, the system will submit the question to a search engine, then extract all possible answers from the search results according to the question type identified by the question classification module, finally select the most similar answers to return. The

architecture of web based question answering system is shown in figure 2[18].

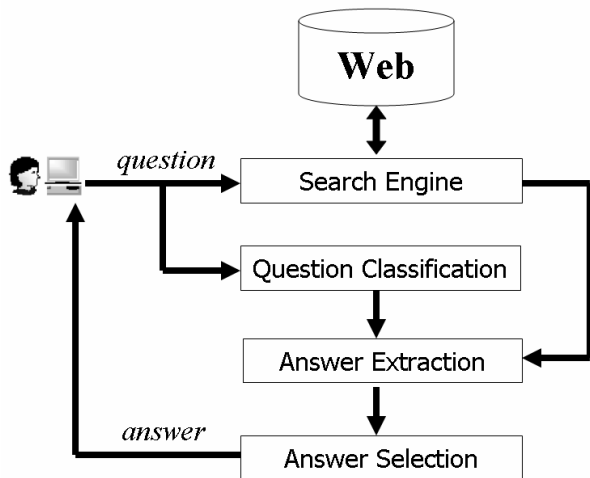


Figure 2. Architecture of Web based question answering system

The design of our question answering system was encouraged by the goal of exploiting the large amounts of text data that is available on the Web and elsewhere as a useful resource [21]. Huge amounts of data provide several sources of redundancy that our system capitalizes on. Answer redundancy such that multiple, differently phrased, answer occurrences enable us to use only simple query rewrites for matching, and facilitates the extraction of candidate answers.

### 3.2 IR / IE Based Question Answering Systems

Most of the IR based QA systems is returning a set of top ranked documents or passages as responses to the query. Information Extraction (IE) system is using the natural language processing (NLP) systems to parse the question or documents returned by IR systems, yielding the “meaning of each word”. IE systems need several resources like Named Entity Tagging (NE), Template Element (TE), Template relation (TR), Correlated Element (CE), and General Element (GE). IE systems architecture is build into distinct levels:

- Level 1 NE tagger is use to handle named entity elements in the text (who, when, where, what etc...).
- Level 2 handles NE tagging + adj like (how far, how long ,how often etc...),
- Level 3 builds the correlated entities by using the most important entity in the question and prepares General Element(GE) which consists of asking point of view. For *Eg*: “Who won the first Nobel Prize in Literature?” The ASKING POINT is clearly defined i.e. Person (Noun) if we by passing this question into the separate levels which was mentioned above.

KEYWORDS such as won, noble, prize etc..., are retrieved. The architecture of IE systems consisting of two common modules, they are

□ **Question processor** which is taking the question as input and generates asking point for the question which in turn helps to match for the answer in the text.

□ **Text Processor** retrieve named entities keywords from the text to produce accurate results. Some of the IR systems like AskJeeves, LaSiE system performs text analysis which use some of the basic modules like Tokenizer, Sentence splitter, Parse process, Name matcher, Discourse Interpreter.

The IR/IE based QA systems depends upon the knowledge base which requires an extension to CE and GE components to handle yes/no types of questions in the text. This systems can answer only wh-type of questions but other than wh-type of questions such as “**How can I assemble a computer?**” are not answered. The Architecture of IR/IE based question answering system is given in figure 3[19].

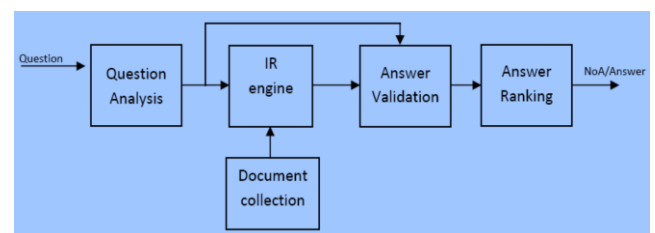


Figure 3. Architecture of IR/IE based question answering

### 3.3 Restricted Domain Question Answering systems

This type of Question answering system requiring a linguistic support to understand the natural language text in order to answer the questions accurately. An efficient approach for improving the accuracy of QA system was done by restricting the domain of questions and the size of knowledge base which resulted in the development of restricted domain question answering system (RDQA). This system have particular characteristics like “System must be Accurate” and “Reducing the level of Redundancy”. RDQA overcomes the difficulties incurred in open domain by achieving better accuracy. Early RDQA systems like LUNAR allows to ask geologist questions about rocks. BASEBALL is another restricted domain QA system, which can only answer about one season’s Baseball data. These early systems has encoded large amount of domain knowledge in data bases.

Question answering on restricted domains requiring the processing of complex questions and offering the opportunity to carry out complex analysis of the text sources and the questions. The main difference between open-domain question answering and restricted-domain question answering is the existence of domain-dependent information that can be used to improve the accuracy of the system[28].

A question is linguistically analysed by the Heart of Gold (HoG) NLP architecture, which flexibly integrates deep and shallow NLP components, for instance, PoS tagger, named entity recognition and HPSG parser. The semantic representations which was generated by the Heart of Gold are then interpreted and a question object is generated that contains a proto query. This proto query can be viewed as an implementation-independent, ‘higher level’ representation of a database or ontology query. From this, an instance of a

specific database or ontology query is constructed. From the result(s) returned by the queried information source, an answer object is generated which forms the basis for subsequent natural language answer generation. This is shown in figure 4 [15].

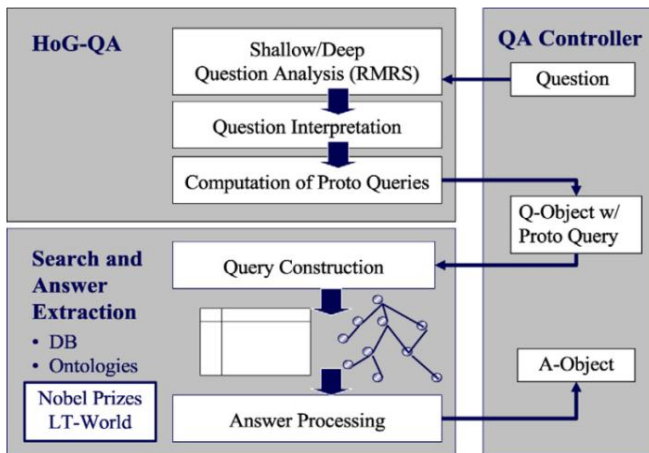


Figure 4. Architecture of Domain Restricted question answering system

### 3.4 Rule Based Question Answering Systems

The rule based QA system is an extended form for IR based QA system. Rule Based QA doesn't use deep language understanding or specific sophisticated approaches. A broad coverage of NLP techniques are used in order to achieve accuracy of the answers retrieved. Some popular rule based QA systems such as Quarc and Noisy channel generates heuristic rules with the help of lexical and semantic features in the questions. For each type of questions it generates rules for the semantic classes like **who**, **when**, **what**, **where** and **Why** type questions. "Who" rules looks for Names that are mostly Nouns of persons or things. "What" rules focuses on generic word matching function shared by all question types it consists of DATE expression or nouns. "When" rules mainly consists of time expressions only. "Where" rules are mostly consisting of matching locations such as "in", "at", "near" and inside. "Why" rules are based upon observations, that are nearly matching to the question. These Rule Based QA systems first establish parse notations and generate training cases and test cases through the semantic model. This system consists of some common modules like IR module and Answer identifier or Ranker Module.

- **IR module:** It gives the set of documents or sentences that includes the answers to the given question and returns the results back to the ranker module.
- **Ranker Module:** Assigning ranks or scores to the sentences which are retrieved from IR module.
- **Answer Identifier:** It identifies the answer substrings from the sentences based upon their score or rank.

### 3.5 Classification of Questioners Levels

In Question Answering system the questions are classified into different levels based upon its context. The questions may be assertive, informative, interrogative or interactive in

normal context. The perspectives of these types of questions may fluctuate but the common goal is to obtain accurate answer from the system. This section presents a classification of different levels of Questioners.

**CASUAL QUESTIONERS:** In this type of questioners normal questions are pose to the system. Majorly it focus in normal "perspective" to handle the questions like Eg: "When he was born?" and "who invented telephone?". All these type of questions are having normal context.

**TEMPLATE QUESTIONERS:** In this type of questioners, templates are generated for the given question, which focuses on the "linguistic" knowledge of the question. For Eg: "How Akshay manage to complete a task?" and "Does any specific reason to invent bulb?"

**CUBE REPORTER:** In this type of questioners the complex questions are broken down into small set of questions. It majorly consists of context and specific relations to answer the questions of this type. The QA system needs to search answers from multiple sources which lies beyond the database search. It can answer the questions like Eg: "Does any specific actions performed by US government after Lincoln's death?". Cube reporter generates small set of questions which are associated to the chief question, that are Eg:

"When did Ram died?" "What was the reason behind his death?" and "What was released by Indian government after Gandhi's death?".

**PROFESSIONAL INFORMATION ANALYST:** These questions are having future perspectives. It is used to identifies different taxonomies and multiple facts which are involved in the questions, but it requires much reasoning techniques for answering, the questions like E.g.: "What are the actions done by Indian government to honour Mahatma Gandhi?"

## 4. Question answering system based on information retrieval

Currently, the accessible information, predominantly obtained through the Internet is gradually increasing. The most significant way to access the information is through information retrieval (IR) systems. IR system takes a user's query as input and returns a set of documents sorted by their relevance to the query. Some standard technologies are used to perform the IR task such as existing web search engine like (Google, Askme, Alta vista etc...).

Question answering is an information retrieval task constrained by an expression of all or a part of the information need as a set of natural language questions or statements. IR systems are usually based on the segmentation of documents and queries into index terms, and their relevance is computed according to the index terms they have in common, as well as according to other information such as the characteristics of the documents, for instance number of words, hyperlink between papers.

The number of document returned by the IR system huge means paragraph filtering concept has used to reduce the no of candidate document and to reduce the amount of candidate text from each document[5]. The steps involved for QA system based on information retrieval is given below:

#### **4.1 Filtering candidate document**

The idea of paragraph filtering is based on the principle that the most relevant documents should contain the question keywords in a few neighboring paragraphs, rather than dispersed over all documents. To exploit that idea, the position of the set of question keywords in each document was examined. If the keywords are all found in some set of N successive paragraphs, then that set of paragraphs will be returned, otherwise, the document is rejected from further processing. 'N' is again a configurable number that could be tuned based on an evaluation of system performance under changed tolerances of keyword distance in documents.

#### **4.2 Identifying quality of the document**

To estimate the quality of the selected paragraph quality component has used. If the quality of paragraphs is deemed to be inadequate, then the system returns to the question keyword extraction module, and alters the heuristics for extracting keywords from the question. Then the IR can be performed by using new set of key word retrieved from scratch. The reason of re-determining question keywords stems from including either too many or too few candidate paragraphs after paragraph filtering. In either case, new queries for the information retrieval system are produced by revisiting the question keywords component, and either adding or dropping keywords. This feedback loop offers some form of retrieval context that ensures that only a 'reasonable' number of paragraphs are passed onto the Answer Processing module. Like several other parameters, exactly how many paragraphs constitute a 'reasonable' number should be configured, based on performance testing. Next paragraph ordering is to rank the paragraphs according to a plausibility degree of containing the correct answer.

#### **4.3 Standard radix sort algorithm for paragraph ordering**

This algorithm uses different scores to order the paragraph. The number of words from the question that are recognized in the identical sequence within the recent paragraph window, the number of words that separate the majority of distant keywords in the current paragraph window and the number of unmatched keywords in the recent paragraph window. Paragraph window is defined as the smallest span of text required to capture each maximally inclusive set of question keywords within each paragraph. Radix sorting is performed for each paragraph window among all the paragraphs.

#### **4.4 Lexical and Syntactic Knowledge for IR**

In our suggestion we adopt the format of parsing the query to acquire the set of query terms to calculate the TP information, instead of calculating TP among all possible combinations of query pairs, but we vary from previous approaches in the following three points: first we do not carry out a full parsing of the query but chunking the queries into sets of simple phrases such as noun, prepositional phrases and sequences of verbs. In order to reach a more consistent behavior for different queries, we apply different TP measures depending on the lexical type of each query term. We apply TP measures to phrases as well as terms because phrases represent the concepts expressed in a text more accurately than single words.

#### **4.5 Question Classification**

Question answering is an alternate of information retrieval, which retrieves detailed information rather than documents. A

QA system take a natural language question as input, convert the question into a query and forwards it to an IR module. When a set of appropriate documents is retrieved, the QA system extracts an answer for this question. There are different methods of identifying answers. One of them make use of a predefined set of entity classes. Given a selected question, the QA system classifies it into those classes based on the type of entity it is looking for, identifies entity instances in the documents, and selects the most expected one from all the entities with the same class as the question. There are different types of methods available for classify the question. In the following section we are going to discuss important technique for question classification. Such as identification of question pattern, semantic approach for question classification, sub tree kernel using support vector machine to improve the performance of the question classification.

**Functional Word Questions:** All Non-Wh questions (except how) fall under the category of Functional Word Questions. These questions generally start with non-significant verb phrases.

Example: Name the Ranger who was always after Yogi Bear.

**When Questions:** When Questions starts with "When" keyword and are temporal in nature. The general pattern for When Questions is When (do|does|did|AUX) NP VP X", where AUX, NP, and VP auxiliary verbs, noun phrases, and Verb phrases. '|' indicates Boolean OR operation and 'X' can be any combination of words playing insignificant role in answer type determination.

Example: When did Israel begin turning the Gaza Strip and Jericho over to the PLO?

**Where Questions:** "Where Questions" starts with Where keyword and are related to the location. These may represent natural entities such as mountains, geographical boundaries, manmade locations such as temple, or some virtual location such as Internet or fictional place. The general pattern for Where Questions is Where (do|does|did|AUX) NP VP X?"

Example: Where is Italy?

**Which Questions:** The general pattern for Which Questions is Which NP X"? The expected answer type of such questions is decided by the entity type of the NP.

Example: Which company manufactures sports kit?

**Who/Whose/Whom Questions:** Questions falling under this category have general pattern(Who|Whose|Whom) [do|does|did|AUX] [VP] [NP] X? Here [word] indicates the optional presence of the term word in the pattern. These questions usually ask about an individual or an organization.

Example: Who wrote 'Hamlet'?

**Why Questions:** Why Questions always ask for certain reasons or explanations. The general pattern for Why Questions" is "Why [do|does|did|AUX] NP [VP] [NP]" X".

Example: Why do heavier objects travel downhill rapidly?

**How Question:** "How Questions" have two types patterns of syntax: "How [do|does|did|AUX] NP VP X?" or "How [big|fast|long|many|much|far] X?" For the first pattern, the answer type is the explanation of some process while second pattern return some number as a result.

Example: How did the jack gets its name?



**What Questions:** What Questions have several types of patterns? The most general regular expression for What Questions can be written as “What [NP] [do/does/did/AUX] [functional-words] [NP] [VP] X? What Questions can ask for virtually anything.

Example: What is considered the costliest disaster for insurance industry? Many What Questions are disguised in the form of “Functional Word Questions”.

## 5. MULTI-STREAM QUESTIONANSWERING

The selection of the final answer is complicated by the fact that the final answer has to be selected from various pools of ranked candidates found by different streams[25]. In other words, the selection of the correct answer from a given set of replies corresponding to different QA systems. In particular, it propose a supervised multi-stream approach that decides about the correctness of answers based upon a set of features that describe: (i) the compatibility between question and answer types, (ii) the redundancy of answers across streams, as well as (iii) the overlap and non-overlap information between the question–answer pair and the support text[14].

The general scheme of the proposed multi-stream QA approach. It consists of two chief stages. In the first stage, called QA stage, several QA systems extract—in parallel—a candidate answer and its corresponding support text for a given question. Then, in the second stage, called selection stage, a classifier evaluates all candidate answers and assigns to each of them a category (correct or incorrect) as well as a confidence value (ranging from 0 to 1). At the end, the correct answer having the highest confidence value is selected as the final response. In the case that all answers were classified as an incorrect result, the system returns a nil response.

## 6. QUESTION ANSWERING SYSTEM FOR INDIAN LANGUAGES

**HINDI LANGUAGE** Hindi QA system research attempts to deal with a wide range of question types like when, where, what time, how many. The developed Question-Answering system in Hindi is using Hindi Shallow Parser. The shallow parser gives the analysis of the sentence in terms of the morphological analysis, POS tagging, Chunking etc. Apart from the final output, intermediate output of all the individual modules is also available. All outputs are available in Shakti Standard Format (SSF).

**TELUGU LANGUAGE** Telugu is an important language in India belonging to the Dravidian family. The important component of our QA system is the Dialogue Manager (DM), to handle the dialogues between user and system. It is necessary in generating dialogue for clarifying partially understood questions, resolving Anaphora and Co-reference problems[20].

**BENGALI LANGUAGE** The language Bengali is one of the Indo-Aryan languages of South Asia with over 200 million native speakers. Bangla was written in the Brahmi-derived Bangla script. Bangla underwent a period of vigorous Sanskritization that was started in the 12th century and continued throughout the middle ages. The Bangla lexicon consists of tatsama (Sanskrit words that have changed pronunciation, but have retained the original spelling), tadbhava (Sanskrit words that have changed at least twice in

the process of becoming Bangla), and a fairly large number of “loan-words” from Persian, Arabic, Portuguese, English and other languages. Also a large number of words are considered to be of unknown etymology. A translation based on transliteration and a table look-up method is proposed as an interface to the actual QA task. The implementation part thus involves transliterating a Bangla question as an equivalent Latin alphabet (English) version that could be used in an actual QA task. The Bangla lexicon consists of a good number of “loan-words” from Arabic, Persian, English and other languages. And most of them are pronounced almost the same way as would be pronounced in the original language. Entire work can be divided into two components, the translation based on transliteration with table look-up and the question answering part[16].

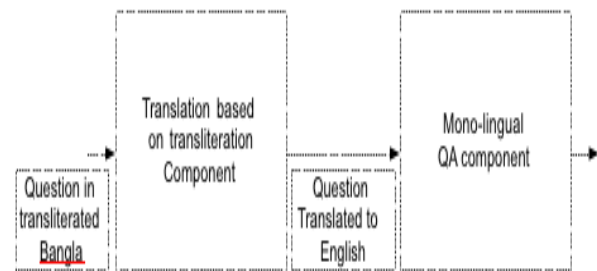


Figure 5. Components of the System

An approach to transform the Bangla question could be

- tokenizing the transliterate version of the Bangla question,
- using translation based upon the transliteration to translate the named entities (medical terms)
- translating the remaining question by a simple table look-up method

## 7. APPLICATIONS OF QUESTION ANSWERING SYSTEM

Question answering has many applications. We can subdividing these applications based upon the source of the answers: structured data (databases), semi-structured data (for example, comment fields in databases) or free text. We can further distinguish among search over a fixed set of collections, as used in TREC (particularly useful for evaluation); search over the Web, search over a collection or book, e.g. an encyclopedia or search over a single text, as done for reading comprehension evaluations. Most of the companies can use Question Answering techniques internally for the employees who are searching out the answers for the similar questions. Another application is in education and medical fields can also find uses for Question Answering in fields where there are frequently asked questions that people wants to search [23].

We can also distinguish between domain-independent question answering systems and domain specific systems, such as help systems. We can even imagine applying question answering techniques to material in other modalities, such as annotated images or speech data. Overall, we would expect that as collections become larger and more heterogeneous, finding answers for questions in such collections will become harder indicates that having multiple answer sources (answer redundancy) increases the likelihood of finding the correct answer for a given question.

## 8. CONCLUSION

The goal of a question answering system is to retrieving answers to questions rather than full documents or best-matching passages, as most information retrieval systems. In this paper we discussed some of the approaches used in the existing QA system and proposed a new architecture for QA system retrieve the exact answer. Answering system has become an important component of the online education platform.

A survey of different QA techniques have been elaborated. Question answering system for Indian languages like hindi telugu, bengali is discussed. No Punjabi QAS is discovered.

The focus of the system has been mainly on four kind of questions of type What, Where, How many, and what time. On analysis of the system the overall efficiency of the system was found to be significant.

The next generation of question answering systems will have to take into consideration presently available multimedia data. There exists a mixture of natural language text, images, video, audio, user added tags, and metadata. On the question side, users may express their queries using a variety of modalities.

## 9. ACKNOWLEDGEMENT

Many thanks to Mr . Vishal Gupta Assistant Professor in UIET, Panjab University Chandigarh, for doing this literature review.

## REFERENCES

- [1] Li, DU. Jia. and Fang, YU .Ping. 2010. Towards natural language processing: A well-formed substring table approach to understanding garden path sentence. 978-1-4244-6977-2/10, IEEE.
- [2] Suarez, O. S., Riudavets, F. J. C., Figueroa, Z. H., and Cabrera, A. C. G. "Integration of an XML electronic dictionary with linguistic tools for natural language processing" Journal of Information Processing & Management, vol. 43, 2007, 946-957.
- [3] Metais, E. "Enhancing information systems management with natural language processing techniques," Journal of Data & Knowledge Engineering, vol. 41, 2002, 247-272.
- [4] Zhang, Wen., Yoshida,Taketoshi., and Tang, Xijin. 2008. TFIDF, LSI and Multi-word in Information Retrieval and Text Categorization. International Conference on Systems, Man and Cybernetics. 1-4244-2384-2/08, IEEE.
- [5] Ramprasath, Muthukrishan. And Hariharan, Shanmugasundram. "A Survey on Question Answering System", International Journal of Research and Reviews in Information Sciences (IJRRIS) Vol. 2, No. 1, 2012, 171-178.
- [6] Sahu, Shriya., Vasnik, Nandkishor., and Roy,Devshri. "Proshanttor : "A Hindi Question Answering System"International Journal of Computer Science & Information Technology (IJCSIT) Vol 4, No 2, 2012, 149-158.
- [7] Kangavari, Mohammad. Reza., Ghandchi, Samira. and Golpou, Manak. "A New Model For Question Answering System", Journal of World Academy of Science, Engineering and Technology 42, 2008. 506-513.
- [8] Hammo, Bassam., Abu-Salem, Hani. and Lytinen, Steven. A Question Answering System to Support the Arabic Language.
- [9] Hirachman, L. and Gaizauskas, R. "Natural Language Question Answering: The View From Here". Journal of Natural Language Engineering 7 (4). 275{300. c 2001. Cambridge University Press DOI: 10.1017/S1351324901002807. 275-299.
- [10] Guda, Vanitha., Sanampudi, Suresh. Kumar. and Manikyamba, ILalkshmi , "Approaches For Question Answering Systems" , Vanitha Guda et al. / International Journal of Engineering Science and Technology (IJEST) ISSN : 0975-5462 Vol. 3 No. 2011. 990-995.
- [11] Moreda, Paloma., Llorens Hector., Saquete, Estela. and Palomar, Manuel. "Combining semantic information in question answering systems" Journal of Information Processing and Management 47, 2011. 870- 885. DOI: 10.1016/j.ipm.2010.03.008. Elsevier.
- [12] Ko, Jeongwoo., Si, Luo., and Nyberg Eric. "Combining evidence with a probabilistic framework for answer ranking and answer merging in question answering" Journal : Information Processing and Management 46, 2010 541-554. DOI: 10.1016/j.ipm.2009.11.004. Elsevier.
- [13] Kolomiyets, Oleksander. and Moens, Marie-Francine. "A survey on question answering technology from an information retrieval perspective". Journal of Information Sciences 181 , 2011.5412-5434. DOI: 10.1016/j.ins.2011.07.047. Elsevier.
- [14] Tellez-Valero, Alberto., Montes-y-Gomez, Manuel., Villasenor-Pineda, Luis. and Padilla Anselmo Penas. "Learning to select the correct answer in multi-stream question answering". Journal of Information Processing and Management,2010. 856 – 869. DOI: 10.1016/j.ipm. Elsevier.
- [15] Frank, Anette., Krieger, Hans-Ulrich., Xu, Feiyu., Uszkoreit, Hans., Crysmann, Berthold., Jörg, Brigitte. and Ulrich Schäfer. "Question answering from structured knowledge sources". Journal of Applied Logic 5 , 2007. 20 – 48. DOI: 10.1016/j.jal.2005.12.006. Elsevier.
- [16] Haque, Nafid. and Rosner, Mike. A prototype framework for a Bangla question answering system using translation based on transliteration and table look-up as an interface for the medical domain. University of Malta Gertjan Van Noord, University of Groningen.
- [17] Zhang Dell. and Lee Sun Wee. A Web-based Question Answering System.
- [18] Rodrigo, Alvaro., Perez-Iglesias, joaqum., Penas, Anselmo., Garrido, Guillermo. and Araujo,Lourdes. A Question Answering System based on Information Retrieval and Validation.
- [19] Reddy, Rami., Reddy, Nandi. and Bandyopadhyay, Sivaji. Dialogue based Question Answering System in Telugu.
- [20] Susan Dumais, Michele Banko, Eric Brill, Jimmy Lin, Andrew Ng "Web Question Answering: Is More Always Better?"

- [21] Zhenqiu, Liang. “Design of Automatic Question Answering System Base on CBR”. *Journal of Procedia Engineering* 29, 2011. 981-985. DOI :10.1016/j.proeng.2012.01.075. Elsevier.
- [22] Badia, Antonio. “Question answering and database querying: Bridging the gap with generalized quantification”. *Journal of Applied Logic* 5,2007. 3-19. DOI:10.1016/j.jal.2005.12.007. Elsevier.
- [23] Gupta, Vishal. and Lehal, Gurpreet S. “A Survey of Text Mining Techniques and Applications”. *Journal of Emerging Technologies in web Intelligence*, VOL. 1, No. 1.
- [24] “Introduction to the special issue on question answering”. *Editorial of Information Processing and Management* 47,2011. 805-807. DOI: 10.1016/j.ipm.2011.04.004. Elsevier.
- [25] Jijkoun, Valentin. and Rijke, Maarten de. “Answer Selection in a Multi-Stream Open Domain Question Answering System”.
- [26] Kwok, Cody.,Etzioni, Oren. and S. Weld, Daniel. “Scaling Question Answering to the Web”. *ACM Transactions on Information Systems*, Vol. 19, No. 3, 2001, 242–262.
- [27] Quarteroni, S. and Manandhar S. “Designing an Interactive Open-Domain Question Answering System”. *Journal of Natural Language Engineering* 1. 1-23.
- [28] Molla ,Diego. and Vicedo, Jose Luis. “Question Answering in Restricted Domains: An Overview”. *Association for Computer Linguistics*. 41-61.