iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** V **Month of publication:** May 2018

DOI: http://doi.org/10.22214/ijraset.2018.5462

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Implementation on Framework for Restricted Domain Question Answering System using Advance NLP Tools & Software in

Shilpa Sharma[1], Prof. Manmohan Singh[2]
[1, 2]*M-Tech Scholar, Professor RKDF School of Engineering, Indore*

Abstract: ***In Computer Science, Question Answering System always helps to user in easier way to find the answers of given question. Information area is new area for human. With help of web crawlers, we can get any data readily available. We are only a tick far from getting to a page at remote corner of the world. We have constantly needed PCs to act savvy . To achieve this undertaking the field of Artificial Intelligence appeared. One of the key obstructions in making PCs clever is comprehension of Natural Language. Normal dialect handling which manages comprehension of dialects is sub division of Artificial Intelligence. Square outline of Question Answering System is appears in beneath figure1.1. Question Answering is a great NLP application. Assignment that an inquiry noting framework acknowledges is given an inquiry and gathering of archives, finds the correct response for the question. It incorporates two integral objectives: first comprehend the different issues in normal dialect comprehension and portrayal and the second to plan regular dialect interface to PCs.***

Keywords:-Natural Language processing, information retrieval, semantic similarity, restricted domain , answer extraction, answer ranking

I. INTRODUCTION

Although the set of documents which are retrieved by the search engine contain a lot of information about the search topic but it may or may not contain exactly that information which the user is looking for .The basic idea behind the question answering system is that the users just have to enter the question and the system will retrieve the most appropriate and precise answer for that question and return it to the user. Hence in those cases where the user is looking for a short and precise answer, question answering System plays a great role rather than Search Engines, which usually provide a large set of links of those web pages which might contain the answer of that question. A typical Question Answering system can be divided into 3 modules namely: Question Processing module, Document Processing or Information Retrieval module and Answer Processing module. Each Processing and Information Retrieval module contains several sub modules and these modules use several Natural Language Processing Techniques in order to extract the proper answer. The usual Question Answering system is designed to answer simple wh-questions like “who”, “what”, “when”, “where”, etc. But the recent QA research focuses on extending the system to answer complex questions, summary questions, opinion questions etc. The paper proposes a Question Answering system that answers simple factoid, wh-questions by using a technique called Semantic Role Labeling.

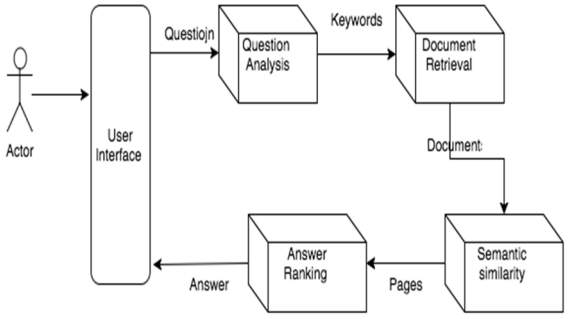


Figure1. Block Diagram Question Answering System

The rest of the paper is organized as follows. The next section describes the general architecture of a Question Answering System. Section 3 discusses some of the related works in this area. The proposed system architecture is described in section 4. The paper concludes with the experimental setup and results.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

II. LITERATURE SURVEY

Jinzhong Xu et al. had worked on Research of Automatic Question Answering System in Network Teaching. This method offer students and teachers to exchange and answer the Question based on Natural Language.

This paper advances a model of automatic question answering system which is based on Natural Language Processing. This paper introduces theory of semantic representation and Ontology. This paper has researched the key technologies of Question answering system based on Natural language Processing. This paper does not gives an accurate answer for complex questions.

Shouning Qu et al. had worked on Research and design of Intelligent Question Answering System. This model proposes a model which supports natural language and finds answer from the intelligent Question answering system.

This system proposes an improved text classification algorithm to classify the database question accurately. The classification algorithm with improved TFIDF method gives more accurate answer than the traditional method. After user inputs text, the system analyses the data using natural processing, positions the target relevant to category, matches the answer and provides efficient answer.

Wael Salloum et al. had worked on A Question Answering System based on Conceptual Graph Formalism . This paper proposes a new text based question answering system which converts knowledge into documents and question into Conceptual Graph formalism. For every question type there is a different conceptual grah formalism, thus for each question many CG's are generated. In this paper projection operator is used to compare questions CG to a sentences CG , and then the exact answer is extracted from it. More reaserch can be done on extracting answers from simple sentences and combining sentences of similar meaning.

Tilani Gunawardena et al. had worked on An Automated Answering System with template Matching for Natural Language Questions. This system uses Closed Domain Question answering system to find the answers. Therefore answers are stored in a database by domain experts.

The final answer extracted in this paper has the ability to answer the question asked in SMS or English language. For this answer extraction process a Template Matching technique is applied in this paper. This paper does not guarantees that the systems gives an accurate answer since the system deals with the i) lack of understanding of problem domain ii) handling SMS abbreviations iii) Handling Spelling mistakes.

Erfan Najmi et al. had worked on Intelligent Semantic Question Answering System. This paper introduces an approach for Question answering system using Semantic Technologies.

This paper converts the Query into Resource Description Framework (RDF) triples and then searches the answer in the RDF files. The advantage of having this system is that it has less Computation time. The system also has some disadvantages that it does not provide answers to Descriptive or Long answer type Question.

PayaJ Biswas et al. had worked on the A Framework for Restricted Domain Question Answering System. This paper proposes a framework for restricted domain question answering system.

This model makes use of Information Extraction than that of Information Retrieval Process used by search engines. This framework can be used to develop a system which provides exact and precise answer. Also this model provides a proper flow of data for answer extraction. The major issues exited in the proposed model is that the performance depends on the search engine and the NLP tools used.

Varsha Bhoir et al. has proposed Question Answering System : A Heuristic Approach. The proposed model works for specific domain of tourism, which is a restricted domain model.

The main aim of this restricted domain Question answering system is to improve the accuracy of the extracting answer. The system returns precise answers related to the tourism domain. This system uses an integrated answer retrieval technique which combines of web crawler and Keyword oriented procedure.

Sangdo Han et al. has worked on Keyword Question Answering System with Report Generation for Linked Data. This paper introduces a Question answering system that extracts answer from Linked Data and generate report in Natural Language.

This system uses entity disambiguation and distributed word similarity to match each keywords to property in Linked Data. To extract Keyword related Information, this model uses SPARQL query. This system returned the correct answer for 95% of the questions.

Sreelakshmi V et al. has worked on the Open domain Question answering System Using Semantic Role Labeling. In this paper the system finds an answer using online search and Semantic Role Labeling.

The goal of the Semantic Role Labeling is to identify all the constituents that fill a semantic role i.e. to determine the roles like Agent , patient, Location, etc in a sentence. The result of this system was compared to that of the system using pattern matching to

find the answers. This system can be enhanced in future for Complex Questions. The different types of Question answering system described in this paper is as: i) Web based Question answering system ii) Information Extraction based question answering system iii) Restricted Domain Question answering system iv) Rule based question answering system v) Classification of Questioners Level vi) Question answering system based on Information Retrieval.

A. Existing System

The solution of Question Answering system works for a specific domain of tourism, which is a global and routine activity for leisure. The users have to struggle to navigate through these overloaded sites for a short piece of information of their interest. The crawler developed in the system gathers web page information, which is processed using Natural Language Processing and Procedure programming for a specific keyword.

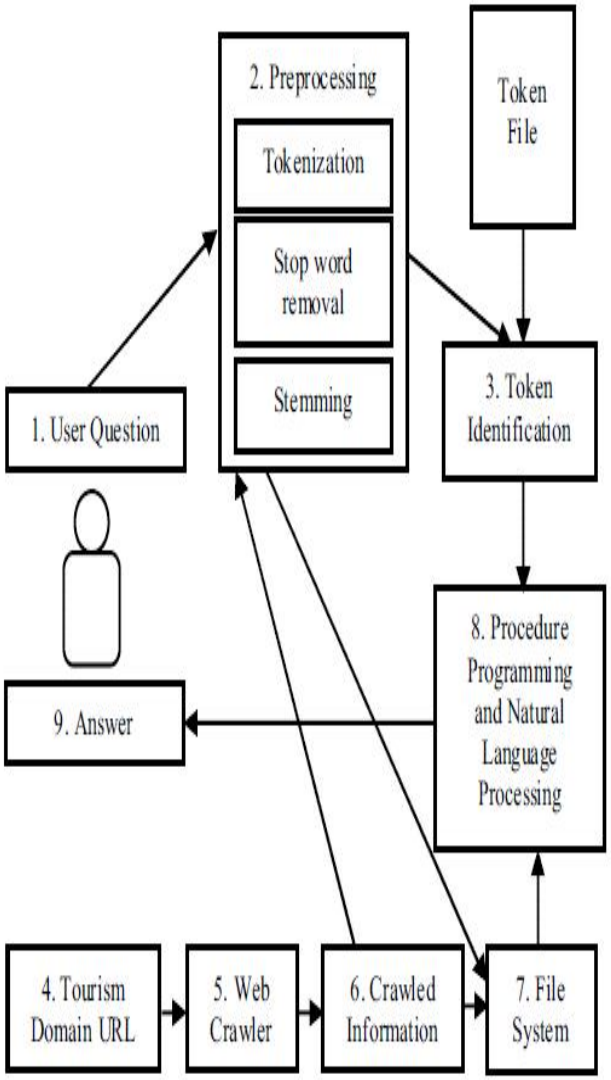


Figure3.2. Work Flow Diagram of Existing Question Answering System

The question expansion emphasize on the comparison of multiple Tokens for a particular keyword. The result shows that the expansions are feasible and efficient to find more similar documents and sentences, as well as retrieving the answers from web resources to the user questions. A straightforward Question Answering structure was implemented using the technique called Semantic Role Labeling. The system consisted of three phases called Query Processing, Document Processing, Answer Processing. In each of these levels, numerous language processing devices were used. The system is web based and therefore uses the search engine to extract information from the web.

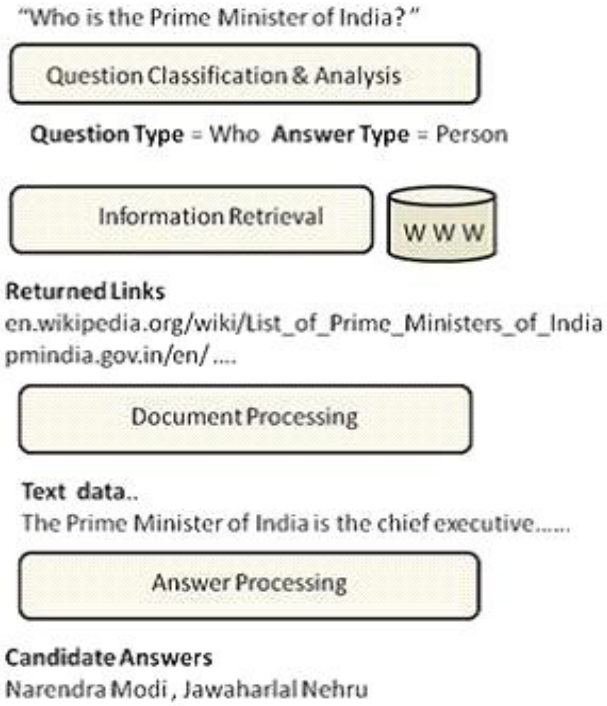


Figure3.3. Existing System Work Flow Diagram

A basic Question Answering framework was actualized utilizing the method called Semantic Role Labeling. The system consisted of three phases called Query Processing, Document Processing, and Answer Processing. In each of these phases, several language-processing components were used. The system is web based and hence uses the search engine to extract information from the web.

III. PROPOSED SYSTEM

A. Proposed System Show in below Figure 4.1.Proposed System Performs Following Operations

1) Question Processing: In this module the given Question is processed to get some important information from it. Steps through which question Processing Module passes and their descriptions are given below. Steps in Question Processing Module:

2) Find the Type of given question using Wh word.

3) Find out the expected type of answer.

4) Get the Keywords from the Question.

5) Find out the Focus of the question.

The first step in the QA System is the Question Processing or Question Classification module. Various information, which we will get through this module, are the Type of Question, Expected Answer Type, Focus or Head Word of the Question and the Question Keywords.

Table 4.1 Identification of Question

WH word	Question Type		
	Factoid Type	Definition Type	Descriptive Type
Question	Who When What Where Which	How What Why	What

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

Document Processing: Once the question has been processed we will move towards the document processing module. In this module the documents which are relevant to the given question are retrieved and processed. Following steps used in document processing.

6) Get the question in hand and search relevant documents using a reliable search engine.
7) Take top relevant documents.
8) Extract the content from these documents
9) Save these contents in to file
10) Answer processing**:** This module presents algorithms for extracting the potential answer for all the three categories of questions that is Definition Type of Question, Descriptive Type of Question and Factoid Type of Question.
11) Dataset Clustering: Cluster dataset using fuzzy c-mean algorithm then process for question and answer processing.

The architecture of the proposed system is shown below. The question asked by the user goes through different stages from preprocessing to Question identification, and then from clustering the data sets and finally template matching is performed to get the required answer.

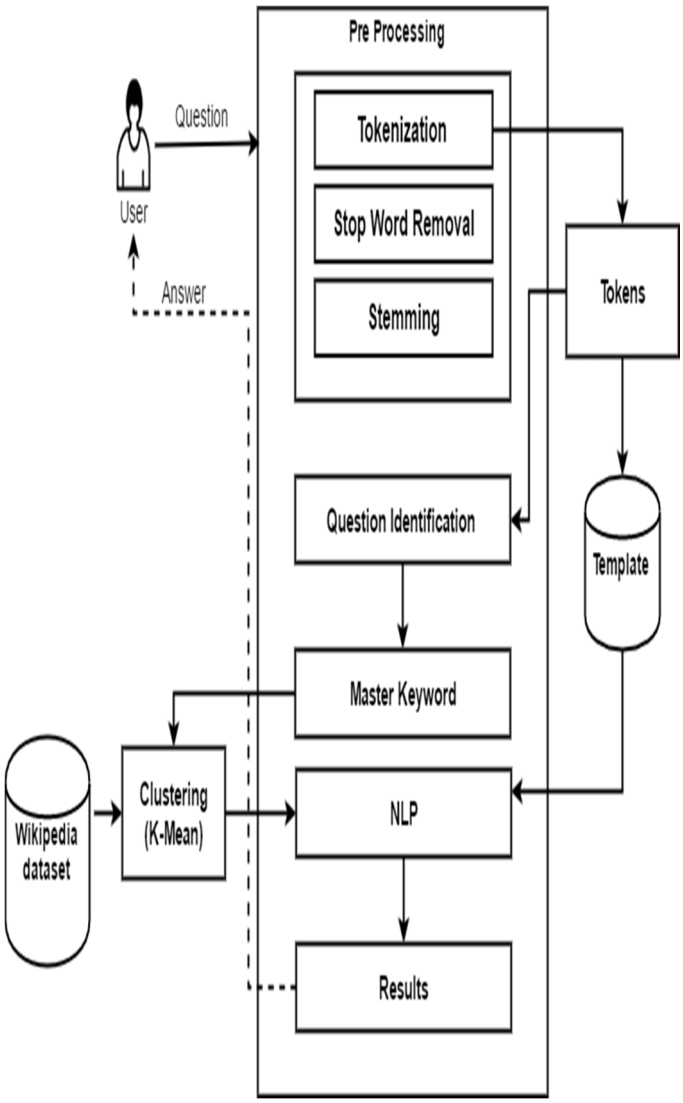


Figure 4.1: Proposed System Architecture

Below shown is the flow chart of the proposed system. The flow chart describes the way in which user gets an answer for its Query. If the question entered by the user is not valid, then the system returns an error. If the question is valid it performs the steps shown to get a exact answer.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue V, May 2018- Available at www.ijraset.com

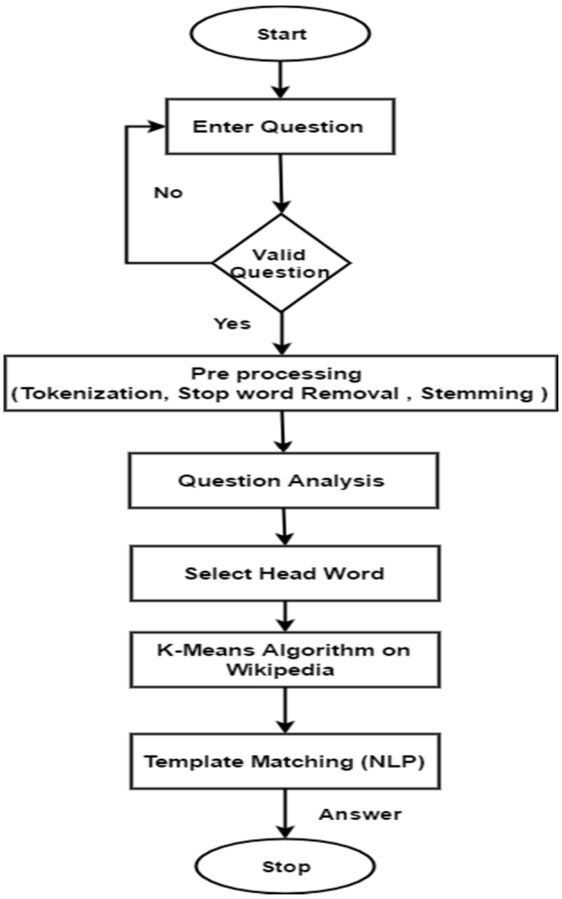


Figure 4.2: Flow Chart of Proposed System

B. Recommendation Algorithm

```
Algorithm Question_Answering (question){
stopword[] //string type array
//delimiter ="";
tokens[] = split_string(question,del);
for i=1 to tokens.length {
        for j=i to stopword.length
                if(tokens[i]==stopword[j])
                        remove tokens[i] }
for i=1 to tokens.length {
        stemming (tokens[i])
}
type_of_question=questionIdentification();
//generate template
 print question type
//select HeadWord from tokens;
K-mean(wikipedia_database)
ans[]= search wikipedia (HeadWord);
        for i=1 to ans.length
                matchTemplate(template, ans[i])
                add to answer
return answer;}
```

IV. RESULT ANALYSIS

Question and Answering System is developed in this research with help Java (JDK1.8) and Net Beans IDE8.02 on window operating sytem7. All forms of Question Answering System design in Swing. Graph plotted for computation time, type of question and memory management using JFree Chart Library. In Result Analysis compare Proposed Question Answering system with existing Question Answering system in term of computation time and memory. In Question Answering System took each type of questions for experiment like Factoid Question, Descriptive and Definition. Wikipedia used as dataset for search Question answers. Below figure 5.1 shows that home screen of project.

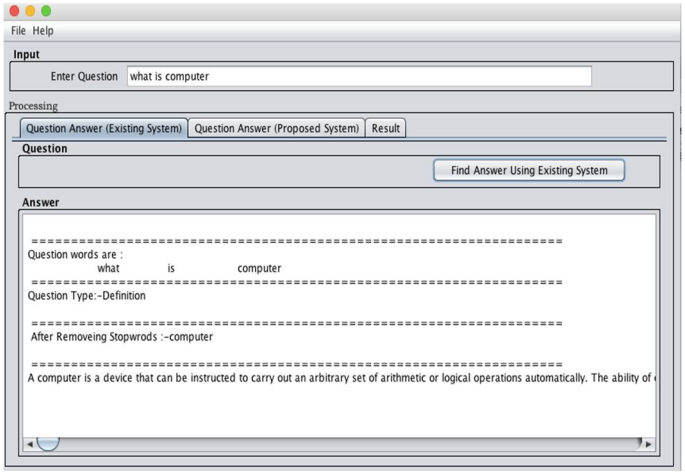


Figure 5.1 Home screen of project.

A. *Evolution Parameters*

In Question Answering system focus on following parameters

1) *Question Types:* Find type of Question corresponding to Enter Question for Answer. Using type of question design template that helps to find more accurate answer for given entered Question.

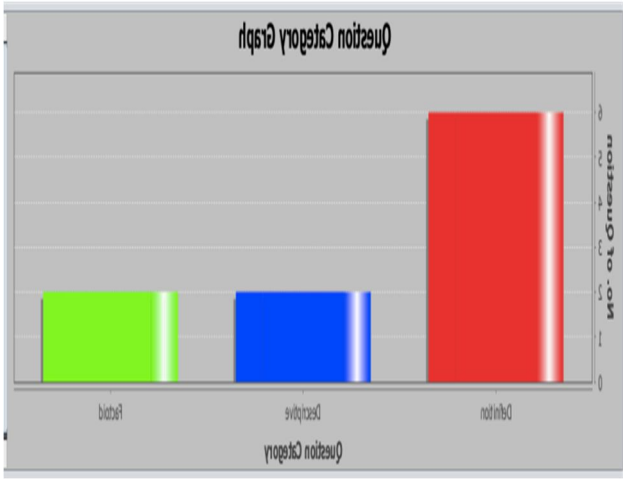


Figure 5.2 Question Types

Table 5.1 Number of Question in Types

S.No.	Question Type	No. Of Questions
1	Definition Type	6
2	Description Type	2
3	Factoid Type	2

Table 5.2 Question Type of each Question

S.No.	Question Number	Question Type
1	Question Number 1	Description
2	Question Number 2	Definition
3	Question Number 3	Definition
4	Question Number 4	Definition
5	Question Number 5	Description
6	Question Number 6	Definition
7	Question Number 7	Factoid
8	Question Number 8	Description
9	Question Number 9	Factoid
10	Question Number 10	Definition

2) *Computation Time :* Calculate computation time for Exiting Question Answering system and Proposed Question Answering system. And results shown with help of graph. From experiments found that Proposed Question Answering system less computation time compare to Existing Question Answering system.

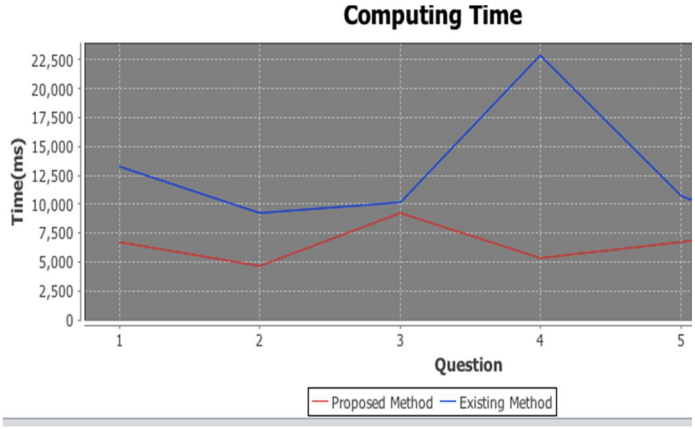


Figure 5.3 Computation time for Existing and Proposed System.

Table 5.3 Computation time for Existing and Proposed System.

S.No.	Question Number	Computation Time of Existing Question Answering System (MS)	Computation Time of Proposed Question Answering System (MS)
1	Question Number 1	13203	7215
2	Question Number 2	9853	4978
3	Question Number 3	10340	9734
4	Question Number 4	22565	5123
5	Question Number 5	11287	7460

3) Computation Memory : We calculate computation memory for Exiting Question Answering system and Proposed Question Answering system. And results shown with help of graph. From experiments found that Proposed Question Answering system less computation memory compare to Existing Question Answering system.

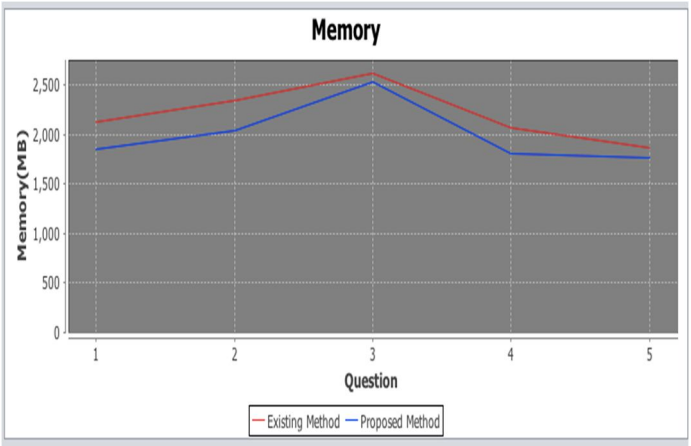


Figure 5.3 Computation Memories for Existing and Proposed System.

B. Output Screen

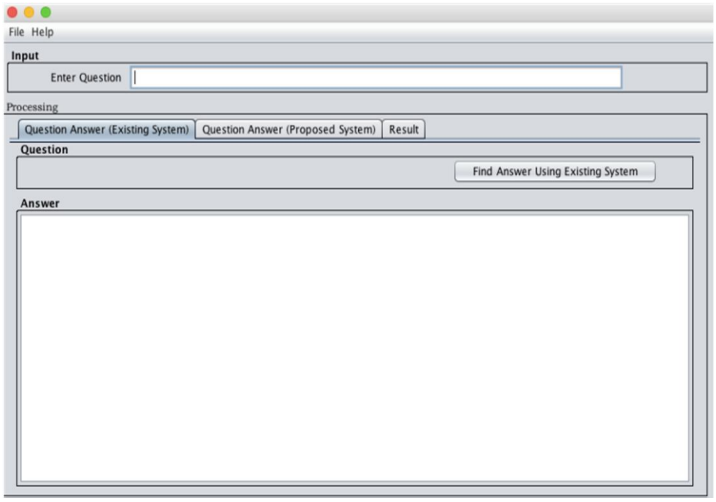


Figure 5.5 Initial screens for Existing and Proposed System.

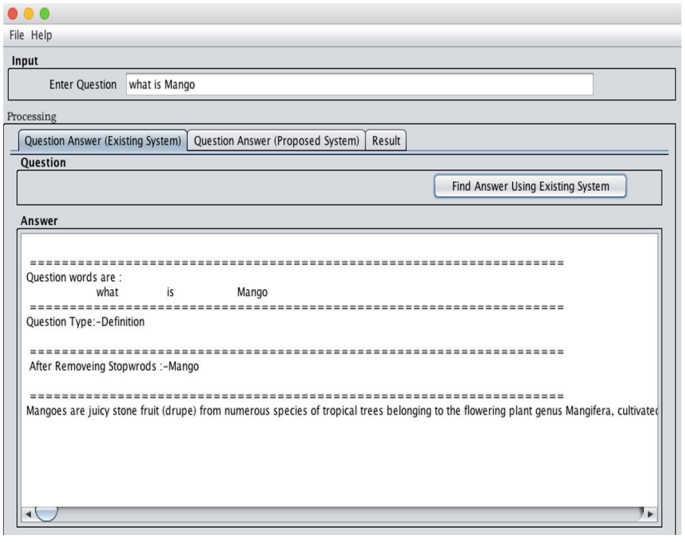


Figure 5.6 Existing Question Answering System

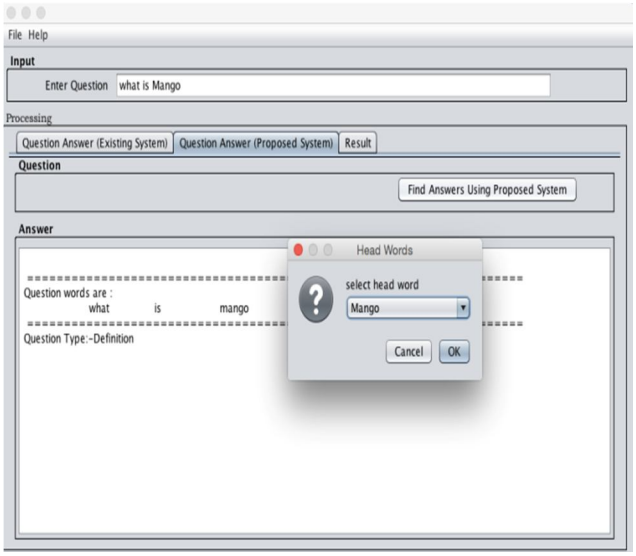


Figure 5.7 Proposed Question Answering System

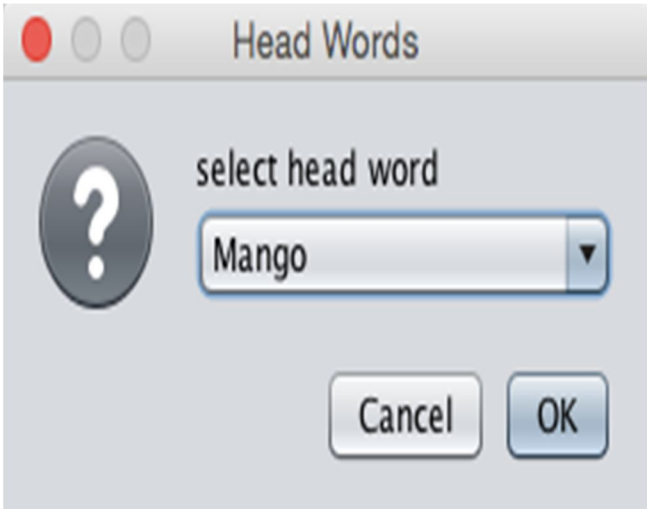


Figure 5.8 Select Head Word of Question

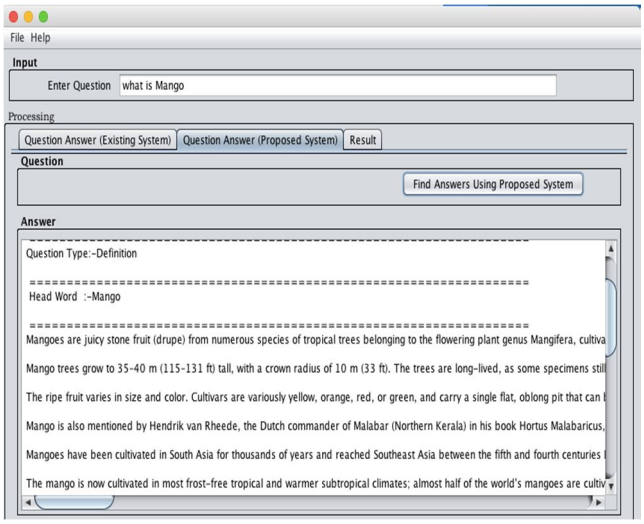
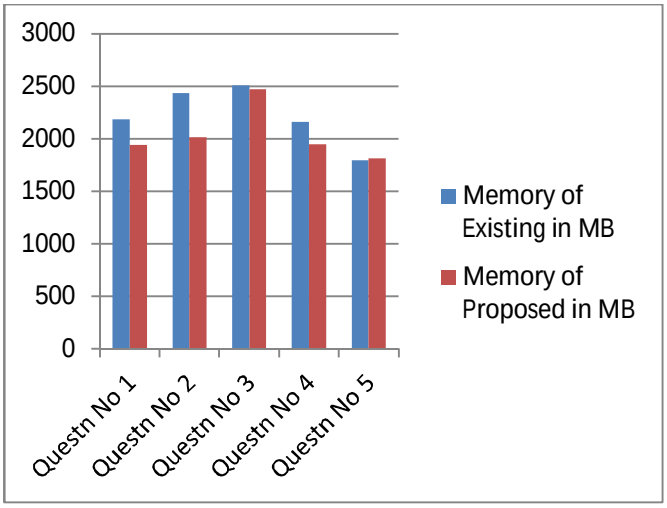


Figure 5.6 Answer of Proposed Question Ans

Table 5.4 Computation memory for Existing and Proposed System

S.No.	Question Number	Computation Memory of Existing Question Answering System (MB)	Computation Memory of Proposed Question Answering System (MB)
1	Question Number 1	2183	1940
2	Question Number 2	2434	2015
3	Question Number 3	2507	2473
4	Question Number 4	2164	1946
5	Question Number 5	1798	1816

Comparative Analysis by Chart



V. CONCLUSION

In this paper we have proposed a framework for restricted domain question Answering System using advanced NLP tools and software. This framework can be used to develop a Question Answering System for extracting exact and precise answer from restricted domain textual data set. The proposed framework not only provides a simple and implementable framework for developing question Answering System but also provides a proper flow of data for answer extraction.

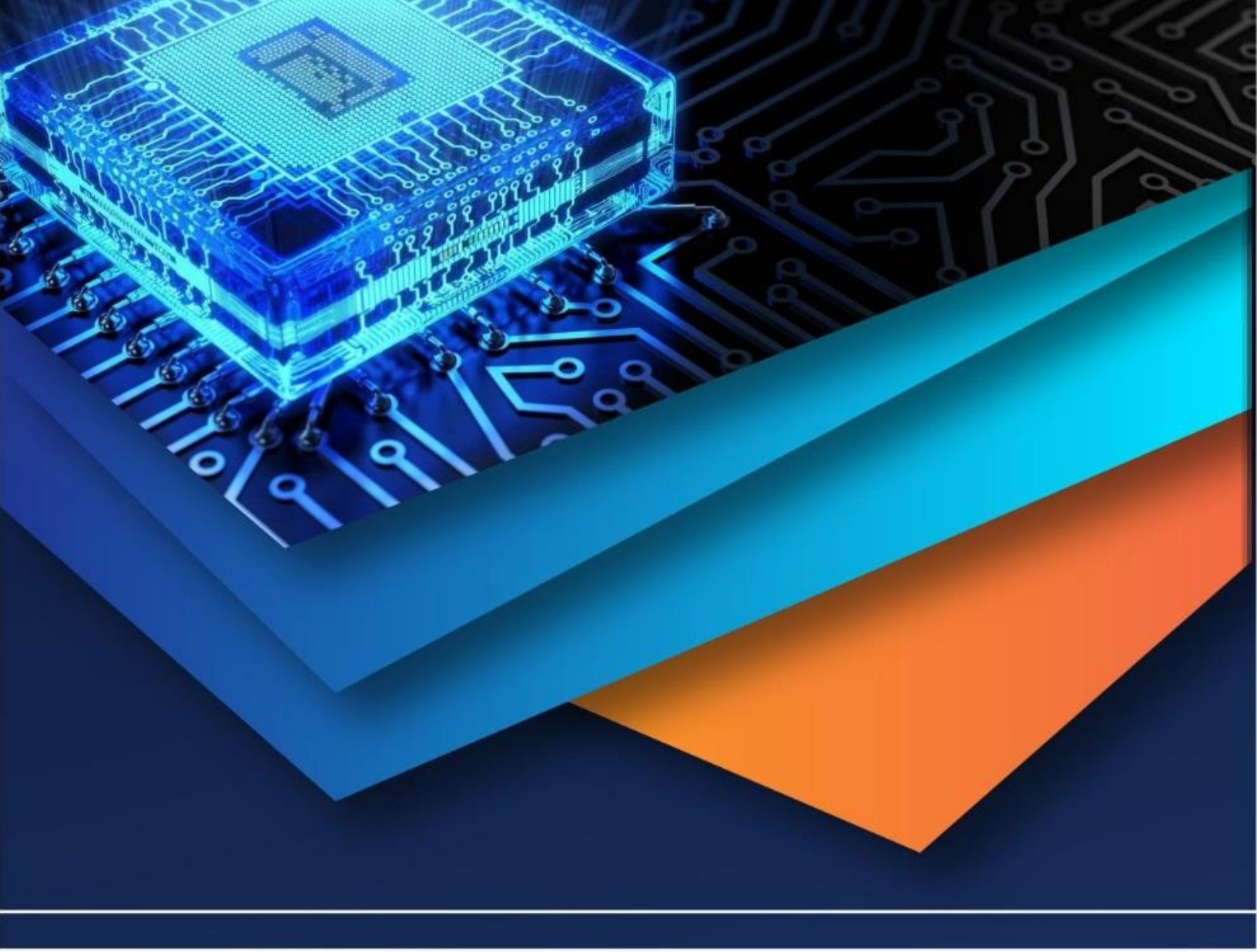
Since the proposed model works over keywords and headword and is independent of the question or sentence structure, it has reduced the overhead of question normalization. Moreover since the framework is given for restricted domain, it also handles the issue of word sense disambiguation. The major problem which exists with the proposed framework is that it's performance is dependent on the performance of the search engine and the used NLP tools.

REFERENCES

[1] Sreelakshmi V, Sangeetha Jamal, Survey Paper : Question Answering Systems, in National Conference on Computing and Communication - (NCCC), March 2014, GEC Idukki

[2] M Ramprasath, S Hariharan Improved Question Answering System by semantic reformulation, IEEE- Fourth International Conference on Advanced Computing, 2012

[3] Ali Mohamed Nabil Allam, and Mohamed Hassan Haggag, The Question Answering Systems: A Survey, International Journal of Research and Reviews in Information Sciences (IJRRIS), September 2012 Science Academy Publisher, United Kingdom
[4] Molla D., and Vicedo J., "Question answering in restricted domains: An overview", Computer Linguist, ppAI-6 1, 2007
[5] Moreda P., Llorens H., Saquete E., & Palomar M., "Combining semantic information in question answering systems", Information Processing & Management, pp.870-885, 20 1 1.
[6] Svetlana Stoyanchev, and Young Chol Song, and William Lahti, "Exact Phrases in Information Retrieval for Question Answering", Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering (IR4QA), pp. 9- 16 Manchester,UK. August 2008".
[7] Woods W.A, Kaplan R.A, Nash-Webber.B, "The lunar sciences naturallanguage information system" , Final report: BBN Report #2378. Technical report, Bolt Beranek and Newman Inc.,Cambridge, MA., June 1972.
[8] Green RF, Wolf A.K., Chomsky, K. Laughery, "BASEBALL: An automatic question answerer", in: Proceedings of Western Computing Conference, vol.19, pp. 2 19-224, 196 1.
[9] Ittycheriah A, Franz M, Zhu WJ, Ratnaparkhi A and Mammone RJ. IBM˝s statistical question answering system. In Proceedings of the Text Retrieval Conference TREC-9, 2000.
[10] Athira P. M., Sreeja M. and P. C. Reghuraj Department of Computer Science and Engineering, Government Engineering College, Sreekrishnapuram, Palakkad, Kerala, India, 678633. Architecture of an Ontology-Based Domain-Specific Natural Language Question Answering System.
[11] Pragisha K. “design and implementation of a QA system in Malayalam”.
[12] Sanjay K Dwivedi, Vaishali Singh. Research and reviews in question answering system Department of Computer Science, B. B. A. University (A Central University) Luck now, Uttar Pradesh, 226025, India.
[13] Poonam Gupta, Vishal Gupta Assistant Professor, Computer Science & Engineering Department University Institute of Engineering & Technology Panjab University, Chandigarh.
[14] 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.
[15] 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.
[16] 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.
[17] Pachpind Priyanka P, BornareHarshita N, KshirsagarRutumbhara B, Malve Ashish D BE Comp S.N.D COE & RC, YEOLA,” An Automatic Answering System Using Template Matching For Natural Language Questions”.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY