Smart Health: Natural Language Processing based Question and Answering Retrieval System in Healthcare

[1]Warish D. Patel, [2]Dr. Chirag Patel, [3]Bhupendra Ramani, [4]Sourabh Bhaskar

[1]Assistant Professor, [2]Associate Professor, [3]Assistant Professor, [4]Assistant Professor
[1]Department of Computer Science & Engineering
[1]Parul Institute of Engineering and Technology, Parul University, Vadodara, India

Abstract : The healthcare system may be a data-driven trade that consists of large and growing volumes of narrative data obtained from discharge summaries/reports, physicians case notes, pathologists similarly as radiologists reports. This data is typically held on in unstructured and non-standardized formats in electronic healthcare systems that create it tough for the systems to grasp {the data the knowledge the data} contents of the narrative information. Thus, access to valuable and pregnant attention data for deciding may be a challenge. all the same, natural language processing (NLP) techniques are accustomed structure narrative data inattention. Thus, NLP techniques have the potential to capture unstructured healthcare data, analyze its grammatical structure. Question and Answering System is one of the major researches is in Natural Language. Main challenges of Question and Answer system gives an exact answer of question which gives by Patient (User). Question and Answering system can be classified into three categories are an open domain, closed domain, and restricted domain. Using advanced Natural Language Processing tool, we will be developed a framework for question answering system. In this paper, we work on the restricted domain question answering system. Proposed system works on keyword and question matching and returns a precise answer to the question.

Index Terms - **Smart Healthcare Systems, Healthcare, Natural Language processing, information retrieval, semantic similarity, restricted domain, answer extraction, answer ranking.**

I. INTRODUCTION

Although the set of documents that area unit retrieved by the computer program contain heaps data of data of knowledge} concerning the search topic however it should or might not contain precisely that information that the Patient is searching for [1]. Nowadays, Real-time health care observance system attract special attention from researchers and different industries specializing within the medical field. many types of research are exhausted this direction and lots of others current. For this work, we've restricted ourselves to the foremost recent connected works. the fundamental plan behind the question responsive system is that the Patients simply ought to enter the question and therefore the system can retrieve the foremost acceptable and precise declare that question and come it to the Patient. thence in those cases wherever the Patient is searching for a brief and precise answer, question responsive system plays an excellent role instead of Search Engines, that sometimes give an oversized set of links of these web content which could contain the solution 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.

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.

1.1 OVERVIEW OF NLP IN HEALTHCARE

A language could be a mode of communication, either verbal or written, that consists of structured words, sets of symbols (such as digits, letters, and special characters) in addition as sets of rules that govern the composition and manipulation of the words and symbols in a very standard means.
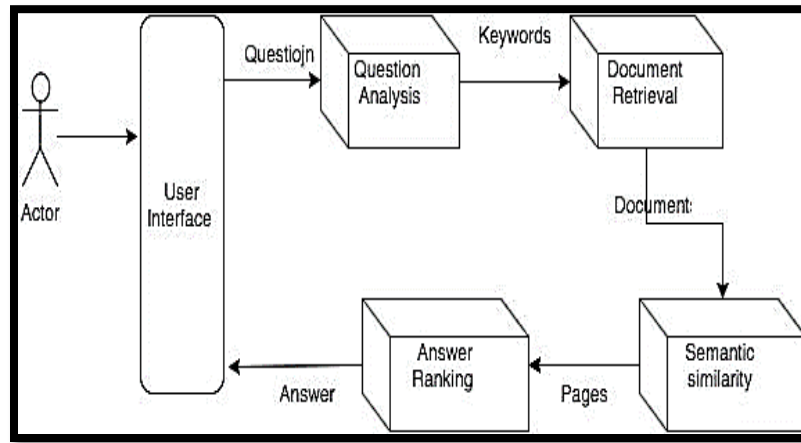
Figure 1. Block Diagram Question Answering System

Natural Languages (NL) are so languages that either spoken or written by human beings for communication. samples of NL embody English, Arabic, Chinese, Japanese, Spanish, and French. There are various definitions for nlp. natural language {processing |NLP |human language technology| information science| informatics| information processing | IP} additionally stated as procedure linguistic is solely the utilization of computers for processing natural languages [18].

II. ARCHITECTURE OF A QUESTION ASNSWERING IN HELTHCARE

In this section, we describe the architecture of our system. The overall architecture of the system can be subdivided into three main modules: (1) pre-processing, (2) question template matching, and (3) answering. Each module is described in detail in the following subsections.

Question Answering Systems can be classified on the basis of the domains over which it has been constructed.

- Open Domain Question Answering
- Close Domain Question Answering
- Restricted Domain Question Answering

Open domain question answering systems are domain independent. It depends on general metaphysics and world information. sometimes these systems have an outsized assortment of knowledge from wherever the desired answer is to be discovered. Since just in case of Open Domain question answering data content isn't of the specific domain it will answer queries of varied fields but here deep reasoning isn't potential [3].

Close domain question answering systems deal with questions in a specific domain [3]. LUNAR and BASEBALL are examples of close domain QA systems. In this case, the data set contains a very limited amount of focused and structured information. hence in case of close domain question answering systems deep reasoning is possible but the problem with these systems was that due to the very small size of data set they are not more than a 'Toy Systems"[4]. Research in restricted-domain question answering (RDQA) addresses issues associated with the incorporation of domain-specific data into current progressive QA technology with the hope of achieving deep reasoning capabilities and reliable accuracy performance in world applications. In fact, as a not too-long-term vision.

III. LITERATURE SURVEY

In most of the research papers [4, 5, 6] LUNAR [7] and BASEBALL [8] have been discussed as the earlier developed question answering systems. However, there are various question answering systems which have been developed with different concepts since the idea of QA System has been coined.

In a system developed Athira P. M, Et.al [10], presented an architecture of ontology-based domain-specific natural language question answering that applies semantics and domain knowledge to improve both query construction and answer extraction.

Another system developed by Pragisha K. Et.al [11], described the. It receives Malayalam natural language questions from the Patient and extracts the most appropriate response by analysing a collection of Malayalam documents. The system handles four each question.

Research and reviews in question answering system developed by Sanjay K Dwivedi Et.al[12] propose a taxonomy for characterizing Question Answer (QA) systems, a survey of major QA systems described in the literature and provide a qualitative analysis of them. Table [I] presents a comparison of different types of question answering system [22].

In a System developed by Poonam Gupta Et.al [13] A Survey of Text Question Answering Techniques. Question answering is a difficult form of information retrieval characterized by information needs that are at least somewhat expressed as natural language Template Matching Automatic Answering System For natural languages questions proposed by Pachpind Priyanka Et.al [17], Frequently Asked QA System that replies with pre-stored answers to Patient questions asked in regular English, rather than keyword or sentence structure-based retrieval mechanisms.

Table 3.1 Literature Review

S. No	Type of Question and Answering System	Question and Answering System Methods
1	Multilingual Question/Answering	Tokenization and pos tagging., Word Sense disambiguation, Answer type identification, Keywords expansion, Semantic Disambiguation
2	Analysis of the asks Question-Answering System	Query Reformulation, NGram Mining, N-Gram Filtering, N-Gram Tiling.
3	Multilingualism, Spatial-temporal context awareness, Textual entailment	Answering architecture
4	A Question Answering System based on Information Retrieval and Validation	Expected Answer Type, Named Entities Presence
5	A Hybrid Question Answering System based on Information Retrieval and Answer Validation	Module, Hypothesis Generation Module, Document Processing and Indexing
6	A specifiable domain multilingual Question	Answering architecture

IV. PROPOSED SYSTEM

Since each the Open Domain QA System and close Domain QA System have their own pros and cons a brand-new
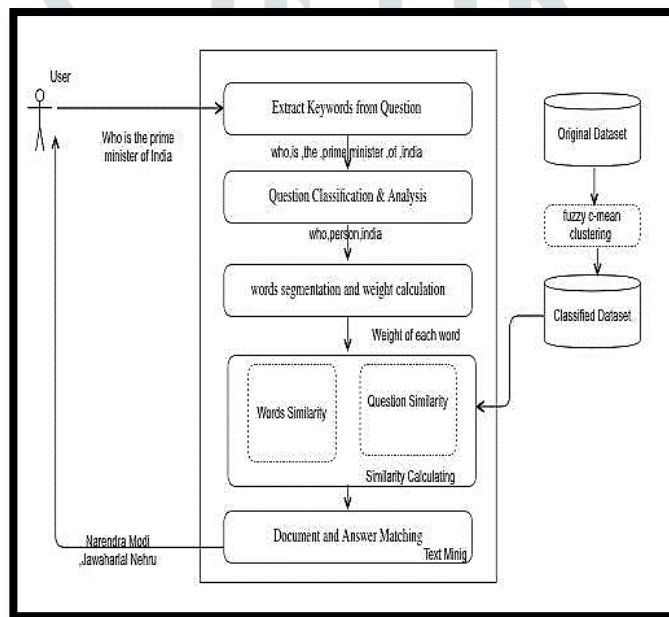

Figure 4.1 Proposed System

conception of Question responsive has been coined by Molla & Vice do [4] known as RESTRICTED DOMAIN QA SYSTEM, that is that the midway of those 2 domains.
We are convinced that research in restricted domains can drive the convergence between structured knowledge-based and free text-based question responsive.

1. Question Processing: during this module, the given Question is processed to induce some necessary info from it. Steps through that question process Module passes and their descriptions area unit is given below. Steps in Question process Module:

a. finds the kind of given question exploitation Wh-word.
b. finds out the expected variety of answer.
c. Get the Keywords from the Question.
d. concludes the main focus of the question.

he commencement within the QA System is that the Question processing or Question Classification module. varied info, that we are going to get through this module, area unit the kind of Question, Expected Answer sort, Focus or Head Word of the Question and also the Question Keywords.

Table 4.1 Question Type

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

2. 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.

a. Get the question in hand and search relevant documents using a reliable search engine.
b. Take top relevant documents.
c. Extract the content from these documents.
d. Save these contents into a file

3. Answer processing: This module presents algorithms for extracting the potential answer for all the three classes of queries that's Definition variety of Question, a Descriptive variety of Question and Factoid variety of Question.

4. Dataset Clustering: cluster dataset using fuzzy c-mean algorithm then process for the question and answer processing.

V. 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. The graph plotted for computation time, type of
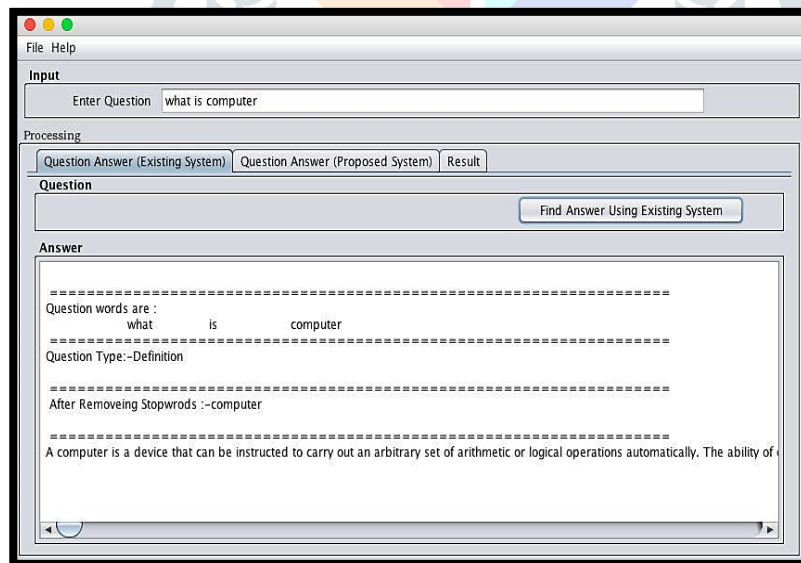

Figure 5.1 Home screen

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 an experiment like Factoid Question, Descriptive and Definition. Wikipedia used as a dataset for search Question answers. Below figure 5.1 shows that home screen of the project.

5.1.1 EVOLUTION PARAMETERS

In Question Answering system focus on the following parameters
- Question type
- Computation Time
- Memory Management

5.1.1.1 QUESTION TYPES

Find the type of Question corresponding to Enter Question for Answer. Using the type of question design template that helps to find a more accurate answer for given entered Question.
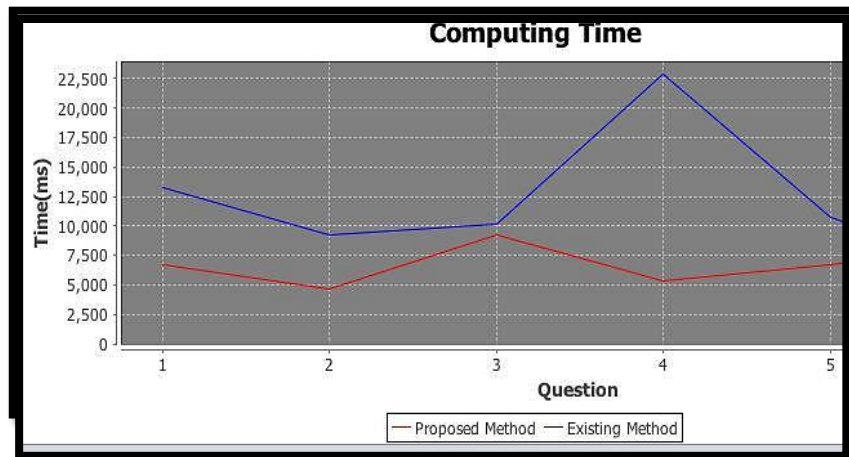

Figure 5.2 Question Types

Table 5.1 Number of Question in Types

S.No.	Question Type	No. Of Questions
1	Definition Type	5
2	Description Type	3
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	Description
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

5.1.1.2 COMPUTATION TIME

We calculate computation time for Exiting Question

Figure 5.3 Computation time for Existing and Proposed System.

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.

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	13303	7217
2	Question Number 2	9854	4979
3	Question Number 3	10335	9724
4	Question Number 4	22560	5124
5	Question Number 5	11286	7457

5.1.1.3 COMPUTATION MEMORY
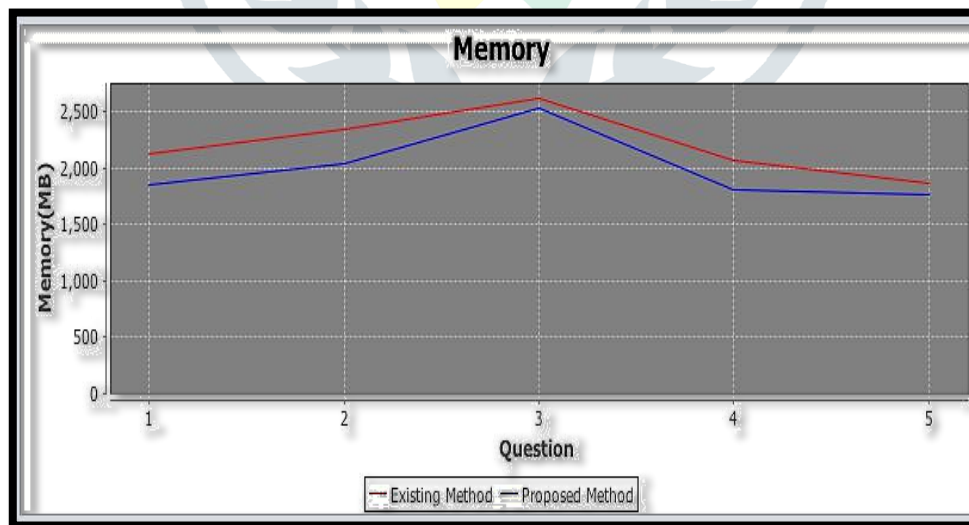
Figure 5.4 Computation Memories for Existing and Proposed System.

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 the Proposed Question Answering system less computation memory compare to Existing Question Answering system.
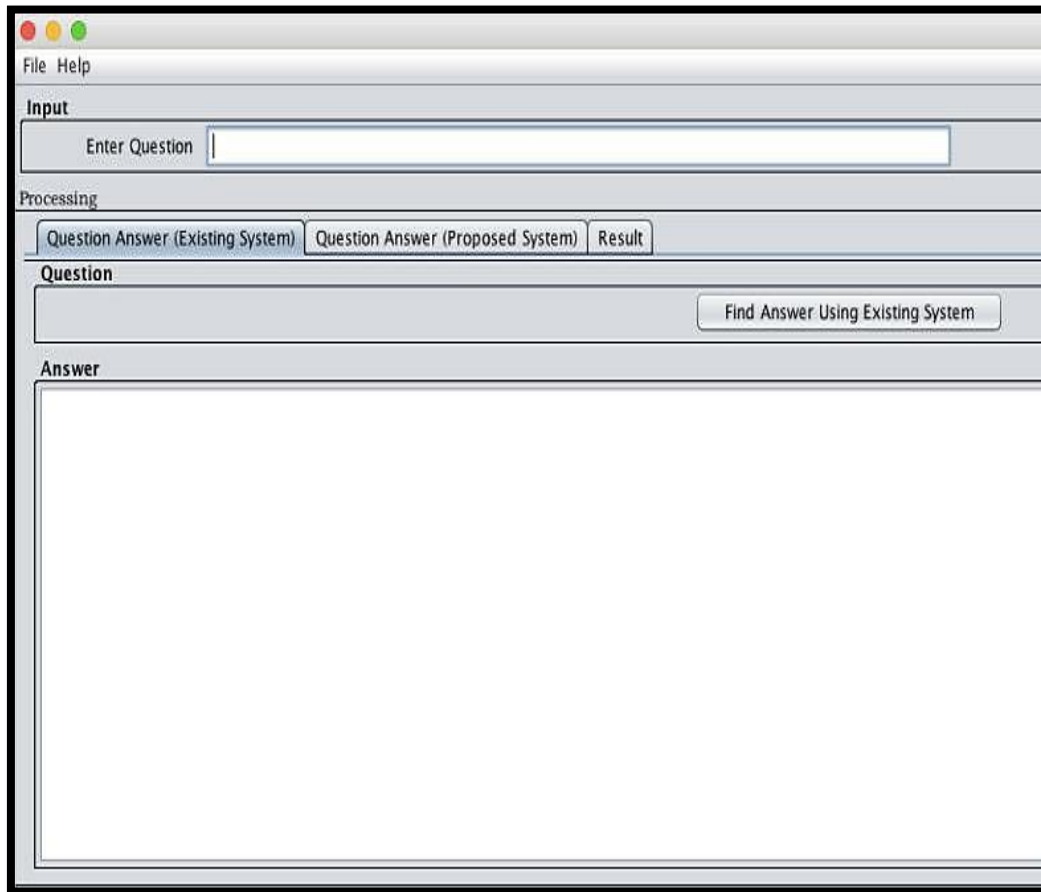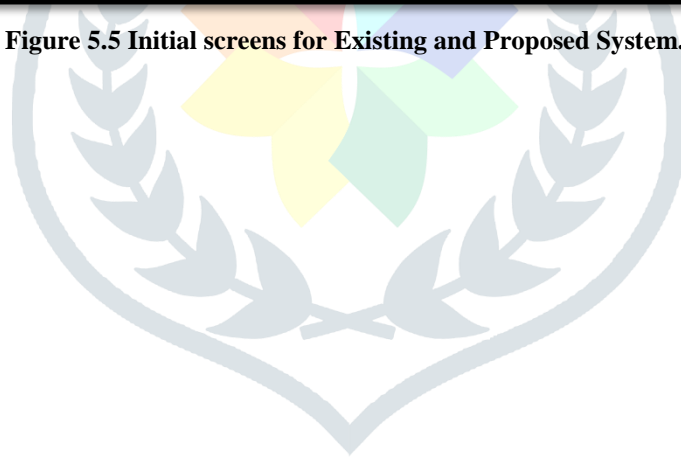
5.2 OUTPUT SCREEN

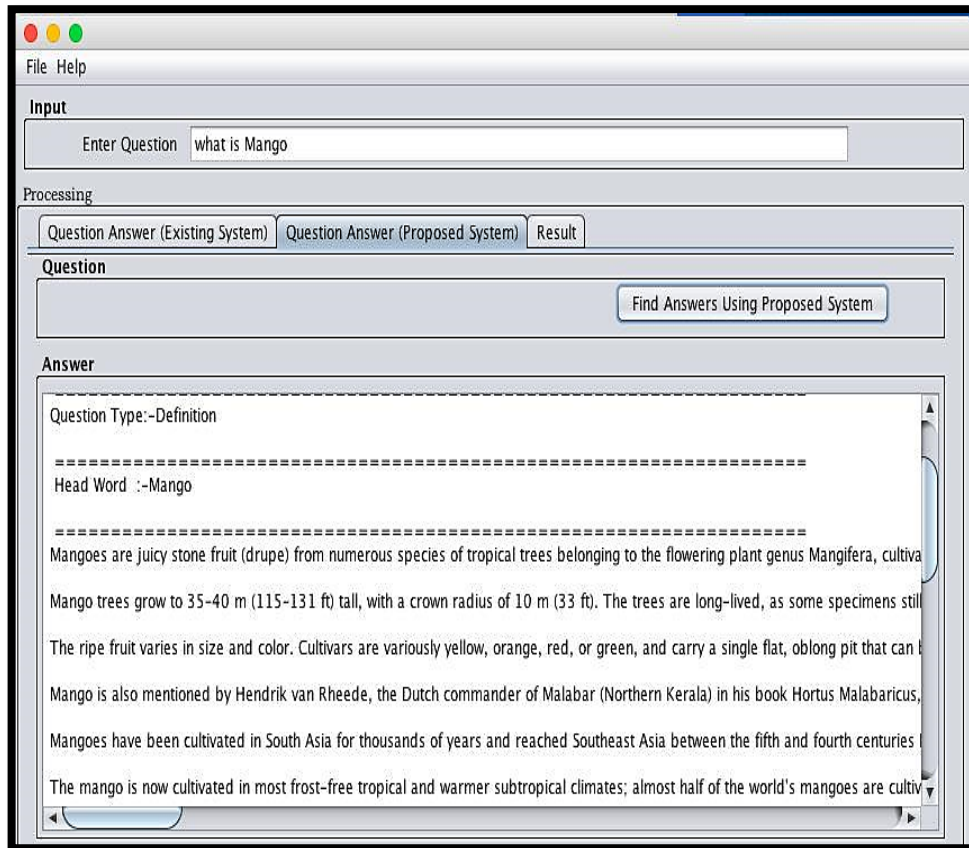
Figure 5.5 Initial screens for Existing and Proposed System.

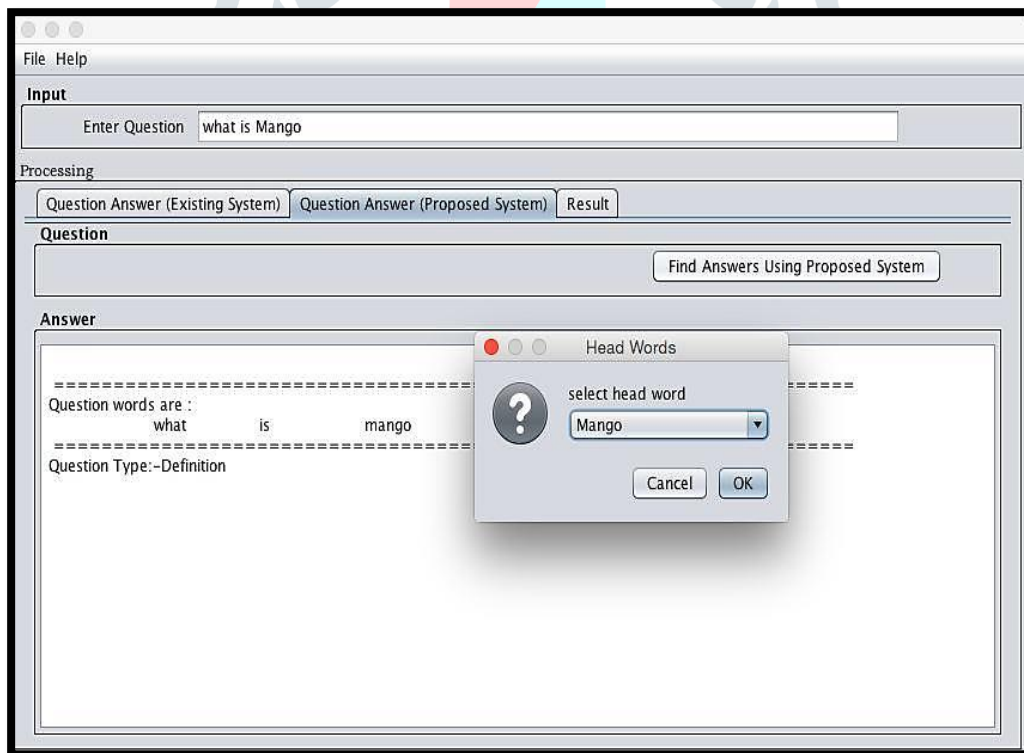**Figure 5.6 Existing Question Answering System**

Figure 5.7 Proposed Question Answering System

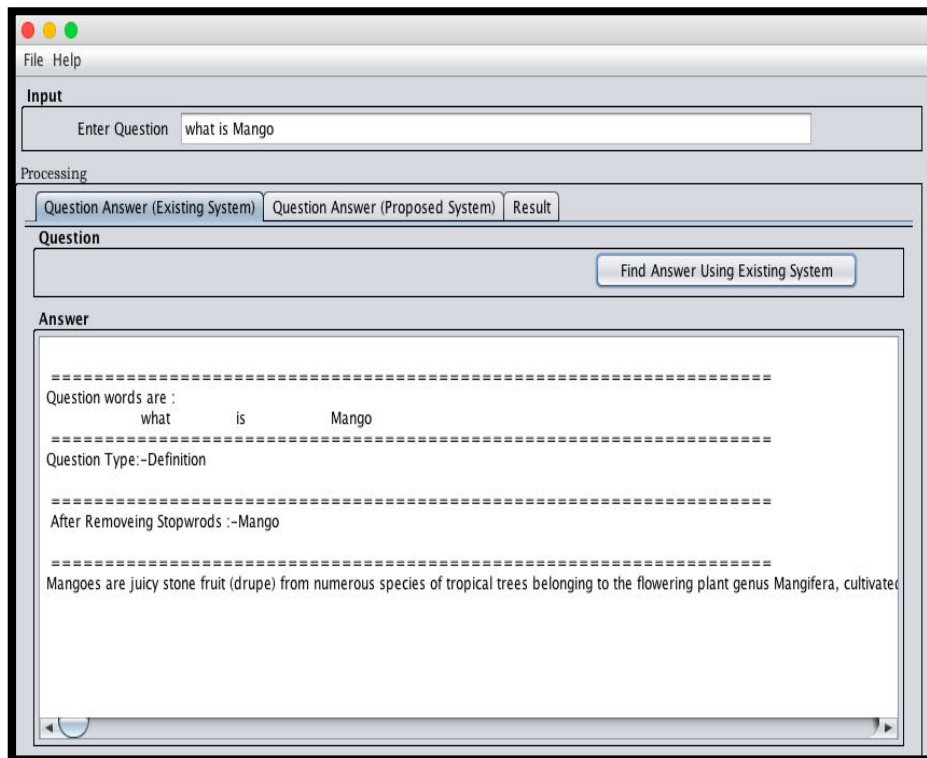**Figure 5.8 Select Head Word of Question**

Figure 5.9 Answer of Proposed Question Answer

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	2173	1935
2	Question Number 2	2443	2017
3	Question Number 3	2517	2471
4	Question Number 4	2154	1940
5	Question Number 5	1787	1815

Comparative Analysis by Chart: -
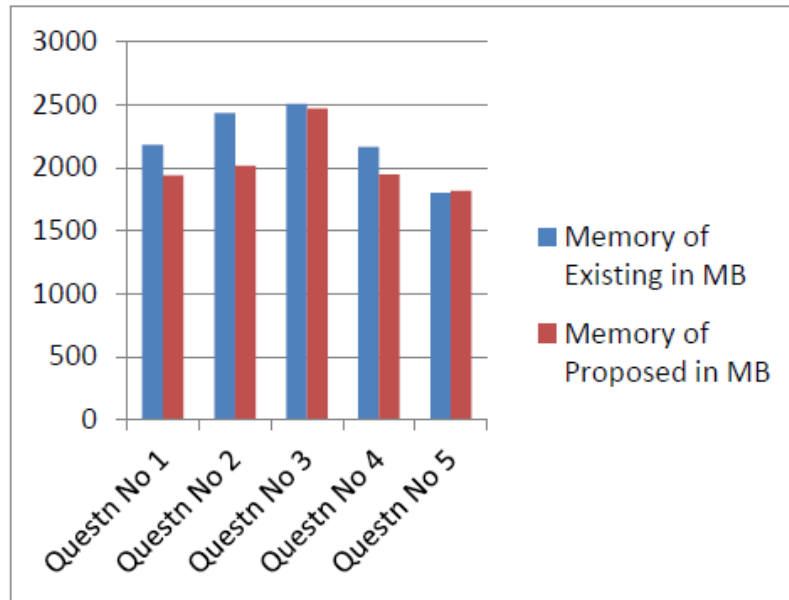
VI. CONCLUTION

In this paper, we've proposed a framework for restricted domain question answering System victimization advanced natural language processing tools and software. The most clinical data area unit typically in kind of narrative texts that are extremely unstructured in nature and therefore not simply understood by the pc. This framework is often accustomed develop a matter responsive System for extracting an actual and precise answer from restricted domain matter information set. The proposed framework not solely provides an easy and implementable framework for developing question responsive System however additionally provides a proper flow of information for answer extraction. Since the proposed model works over keywords and headword and is independent of the question or structure, it's reduced the overhead of question social control. furthermore, since the framework is given for restricted domain, it additionally handles the difficulty of word sense disambiguation. the main drawback that exists with the projected framework is that its performance relies on the performance of the computer program and therefore the used natural language processing tools. natural language processing in aid, the applications of natural language processing in aid, natural language processing systems and resources employed in healthcare and therefore the challenges of natural language processing systems in healthcare are also discussed.

VII. ACKNOWLEDGMENT

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, pp-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", Colin 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 natural language 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. The 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 the 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, Bornare Harshita N, Kshirsagar Rutumbhara B, Malve Ashish D BE Comp S.N.D COE & RC, YEOLA," An Automatic Answering System Using Template Matching for Natural Language Questions".

[18].   C. K. Bretonnel and Lawrence Hunter, "Natural language processing and systems biology," Artificial Intelligence and Tools for Systems Biology, vol.5, pp. 147-173, 2004.

[19].   Sahu, Archana, Kaushik Baruah, Sumit Negi, and Om D. Deshmukh. "Method and system for content processing to query multiple healthcare-related knowledge graphs." U.S. Patent Application 15/355,085, filed May 24, 2018.

[20].   Beller, Charles E., Sean L. Bethard, William G. Dubyak, Alexander C. Tonetti, Sean T. Thatcher, and T. Yu Julie. "System and method for generating improved search queries from natural language questions." U.S. Patent Application 15/359,010, filed May 24, 2018.

[21].   Erpenbach, Eric L., Andrew J. Lavery, Richard J. Stevens, and Fernando J. Suarez Saiz. "Natural language processing review and override based on confidence analysis." U.S. Patent Application 15/333,793, filed April 26, 2018.

[22].   Ren, Hao, Hongwei Li, Yuanshun Dai, Kan Yang, and Xiaodong Lin. "Querying in the Internet of Things with Privacy Preserving: Challenges, Solutions, and Opportunities." IEEE Network 99 (2018): 1-8.

[23].   Kumar, Krishna, Sushil Kumar, Omprakash Kaiwartya, Yue Cao, Jaime Lloret, and Nauman Aslam. "Cross-layer energy optimization for IoT environments: technical advances and opportunities." Energies 10, no. 12 (2017).

[24].   Kothari, Aansi A., and Warish D. Patel. "A novel approach towards context based recommendations using support vector machine methodology." Procedia Computer Science 57 (2015): 1171-1178.

[25].   Kothari, Aansi A., and Warish D. Patel. "A Novel Approach Towards Context Sensitive Recommendations Based on Machine Learning Methodology." 2015 Fifth International Conference on Communication Systems and Network Technologies. IEEE, 2015.

[26].   Patel, Warish, Sharnil Pandya, and Viral Mistry. "i-MsRTRM: Developing an IoT Based Intelligent Medicare System for Real-Time Remote Health Monitoring." 2016 8th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2016.

[27].   Patel, Warish, and Dinesh Vaghela. "A Review on Adaptive Join Algorithms for Efficient Query Processing On Heterogeneous Data Sets." International Journal 2.1 (2014).

[28].   Patel, W. D., Patel, C., & Valderrama, C. IoMT based Efficient Vital Signs Monitoring System for Elderly Healthcare Using Neural Network.

[29].   Patel TP, Patel WD. An Enhanced MEIT Approach for Itemset Mining using Levelwise Pruning.

[30].   Kothari, Aansi A., and Warish D. Patel. "A Contemporary Overview on Feature Selection and Classification Techniques in Opinion Mining." International Journal of Computer Applications 110.10 (2015): 10-14.

[31].   S. P. Mudnur, S. Raj Goyal, K. N. Jariwala, W. D. Patel and B. Ramani, "Extraction of Maximum Secret Information Hidden in Approximate Band of Haar Wavelet Transform of an Image," 2018 Conference on Information and Communication Technology (CICT), Jabalpur, India, 2018, pp. 1-5.

[32].   S. P. Mudnur, S. Raj Goyal, K. N. Jariwala, W. D. Patel and B. Ramani, "Extraction of Maximum Secret Information Hidden in Approximate Band of Haar Wavelet Transform of an Image," 2018 Conference on Information and Communication Technology (CICT), Jabalpur, India, 2018, pp. 1-5.

[33].   W. D. Patel, S. Pandya, B. Koyuncu, B. Ramani, S. Bhaskar and H. Ghayvat, "NXTGeUH: LoRaWAN based NEXT Generation Ubiquitous Healthcare System for Vital Signs Monitoring & Falls Detection," 2018 IEEE Punecon, Pune, India, 2018, pp. 1-8.