Implementation of Question and Answering Retrieval System over Watch Word & Head Word

Author: Dr. Vivek Uprit

Associate Professor, Sanghvi Institute of Management & Science, Rau Pithampur Bypass, Behind IIM, Rau, Indore-453351, (M.P.), India

E-Mail-vivekuprit@gmail.com, vivek.uprit@sims-indore.com

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 [1]. 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 [2]

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

1. 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 [1].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.

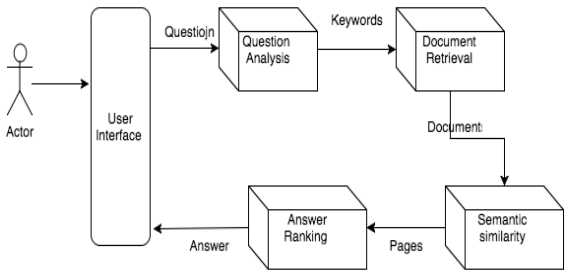


Figure 1. Block Diagram Question Answering

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.

2. ARCHITECTURE OF A QUESTION ANSWERING

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 relies on general ontology and world knowledge. Usually these systems have a large collection of data from where the required answer is to be found out. Since in case of Open Domain question answering information content is not of particular domain it can answer questions of various fields however here deep reasoning is not possible [3].

Close domain question answering systems deal with questions in a specific domain [3]. LUNAR and BASEBALL are the example 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 problems related to the incorporation of domain- specific information into current state-of-the-art QA technology with the hope of achieving deep reasoning capabilities and reliable accuracy performance in real world applications. In fact, as a not too-long-term vision,

3. 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 about the. It receives Malayalam natural language questions from the user and extracts most appropriate response by analyzing 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 taxonomy for characterizing Question Answer (QA) systems, survey of major QA systems described in literature and provide a qualitative analysis of them.

Table [I] presents comparison about different types of question answering system [22].

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	Multilingnality, 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

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 user questions asked in regular English, rather than keyword or sentence structure based retrieval mechanisms.

4.Problem Identification

As early as 2002 a group of researchers3 wrote a roadmap of the research in the field of question answering. They also recognized the issues associated to question answering. The subsequent conversation is based on the issues they acknowledged during their research.

1. Question classes
2. Question processing
3. Context and QA
4. Data sources for QA
5. Answer extraction
6. Answer formulation
7. Real time question answering
8. Multilingual (or cross-lingual) question answering
9. Interactive QA 10. Advanced reasoning for QA
10. Information clustering for QA 12. User profiling for QA

5. Proposed System:

The architecture of the proposed system is explained below. The Architecture consists of various steps.

1. Query Section:- The User enter a Question in the section.
2. Preprocessing:- In the step the question entered by the user undergoes three methods
   i. Tokenization- Here the question entered is converted into tokens or single words
   ii. Stop word removal- All the stop words such as is, am are, etc., are removed in this process.
   iii. Stemming-Stemming refers to the reducing of the word to its root by filtering out prefix and suffix of the word.
3. Token Identification: - The Next step after preprocessing is Token Identification. This is the important step for answer extraction where tokens present in the question are identified for an efficient answer extraction process.
4. Question Analysis :- This phase is broadly divided into three categories :
   i. Definition Type :- Definition Type of question requires one or two sentence as an answer
   ii. Descriptive Type :-Descriptive Type of question requires few set of sentences or a paragraph as an answer
   iii. Factoid Type: - Factoid Type of questions require one or two word answer. For ex. Why, How and Explain Question are asked for descriptive type of answers. Who, When, Where, What, Which are generally asked for Factoid type of answers here who signifies the name of a person, When signifies Time/Date, Where signifies Place/Location.
5. Head word Selection: - After the Question Analysis and Token generation the next phase is Head word Generation. Here the Tokens which are generated in third phase (Token Identification) are chosen as a Head Word. This Head word will be useful in the next phase of Clustering.
6. Clustering: - Here Clustering technique is applied on Wikipedia data set for answer

extraction process. Clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other group. The Clustering Technique used in this paper is K-Means. K-means Clustering is a method of Vector Quantization. It aims to partition n Observations into k clusters in which each observation belongs to the cluster with the nearest mean.

The Head Word Selected in the above phase is used in this phase to form clusters. To apply K-Means on the data set TF-IDF has been used.Three clusters are formed using K-means algorithm, from where and the answers will be retrieved.

7. Templates Matching :- . Templates are the predefined format of the answer which will be presentedin front of the user. Templates are formed by the Head word , selected in the 5th phase and by the Question Format decided in the Question analysis Phase.

   The Answer Extraction process will match the following Templates in its Database to give an exact answer.

   Templates such as:

   i. "Head Word" is . . . . . .
   ii. "Head Word" means . . . . . .
   iii. "Head Word" is known as . . . . . .
   iv. "Head Word" is called. . . . . .
   v. "Head Word" can be defined as . . . . . .

8. Answer Extraction: - In this Phase the templates generated in the above phase will be matched to that in the Clusters formed in the clustering phase.

6.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.

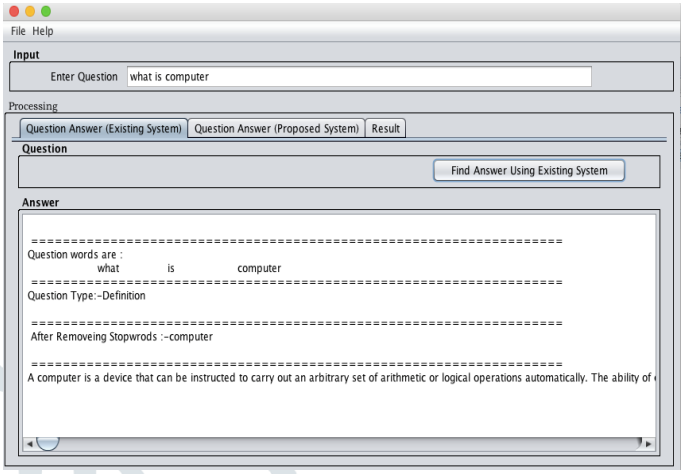


Figure 5.1 Home screen of project.

6.1.1 Evolution Parameters

In Question Answering system focus on following parameters

- Question type
- Computation Time
- Memory Management

6.1.1.1 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.

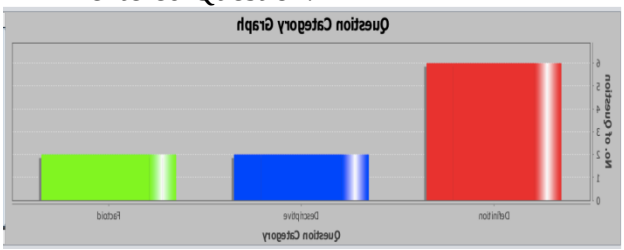


Figure 5.2 Question Types

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

6.1.1.2Computation Time

We 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.

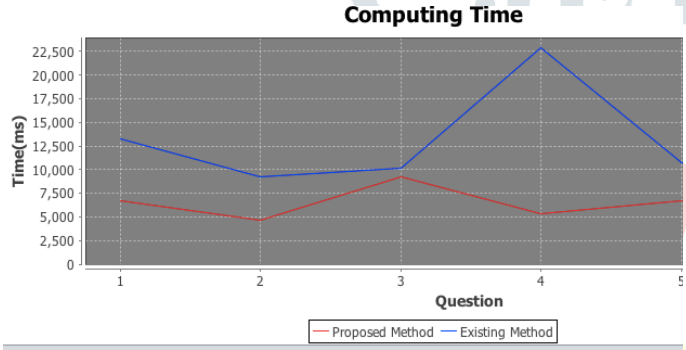


Figure 5.3 Computation time for Existing and Proposed System.

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

6.1.1.3Computation 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.

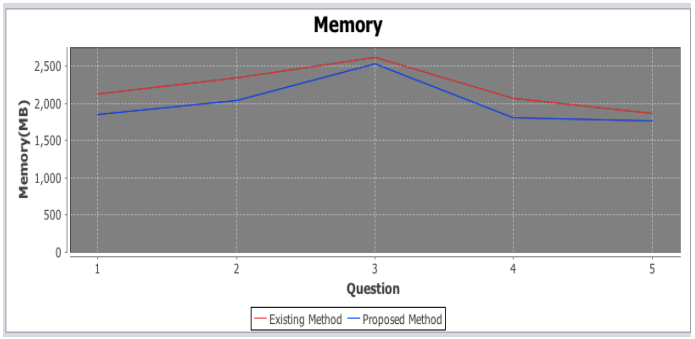


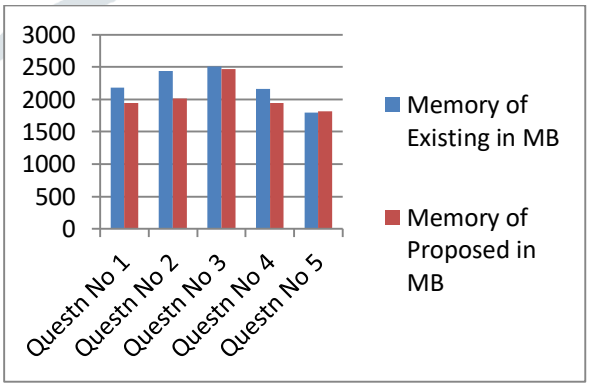
Figure 5.3 Computation Memories for Existing and Proposed System.

5.2 Output Screen

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:-



5. 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.

6. 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”.