ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Restricted Domain Question Answering System

Aakriti Gupta[1], Pranay Chauhan[2], Preetesh Purohit[3]

M.Tech Student, Dept of CS, Swami Vivekananda College of Engineering Indore, M.P, India[1*]

Assistant Professor, Dept of IT, Swami Vivekananda College of Engineering Indore, M.P, India[2]

Associate Professor, Dept of CS, Swami Vivekananda College of Engineering Indore, M.P, India[3]

ABSTRACT: Question Answering (QA) is a fast-growing field that focuses on Information Retrieval (IR), Information Extraction (IE) and Natural Language Processing (NLP). Main challenges of question answering system are to give exact answer to the user for the given query. This paper proposes a system which applies three functions for the user question and displays the best matched answer with its similarity calculated.

KEYWORDS: - Answer extraction, answer ranking, semantic similarity.

I.INTRODUCTION

Today, internet is the biggest source of information and the amount of data is increasing at a tremendous rate. The information that is stored is not in the proper format ie it is not in the structure format. To find any information from the internet, we usually do a web search from search engines. It is a faster way to search from the internet than searching in a book. But searching a particular piece of information is not perfect.

A web search engine searches for the information on the internet. Fundamentally, it provides a list of links and URLs of the query given by then user. The user has the extra task to go through that provided links, in order to search the required answer.

QA systems are designed to overcome this problem. Questions answering system is an automated QA machine that is designed to take a query from the user and retrieve the most appropriate and precise answer for the given question and give it back to the user. Question Answering systems are pretty much different from web based search engines. search engines works on the principal of Information Retrieval (IR), in which much focus is given only to retrieve the required document. On the other hand, QA systems work on the concept of Information Retrieval (IR) along with the concept of Information Extraction (IE). That means document is retrieved as well as proper answer is searched from that document also. The given below figure shows the general steps involved in Question Answering System.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

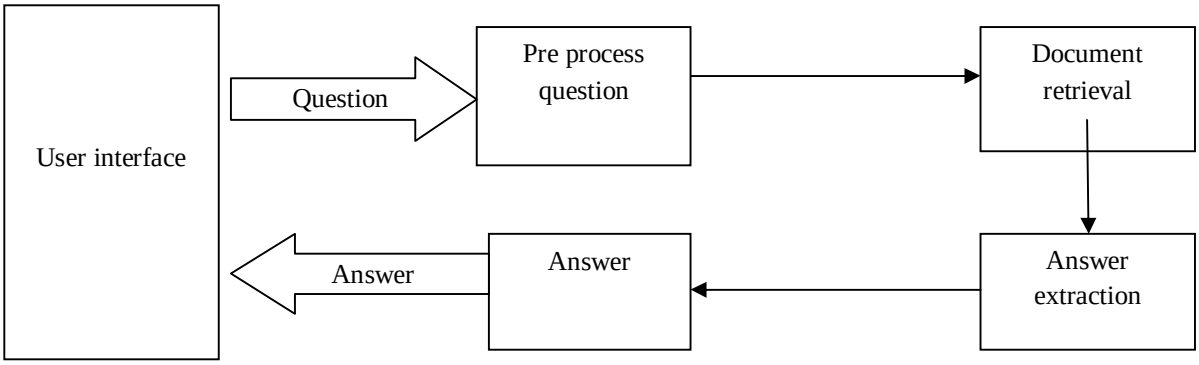

Figure 1 General architecture of QA

II. LITERATURE REVIEW

Since 1960s, various natural language understanding systems were created. Generally, a good number of the question answering systems provide answers to factoid based questions. The early QA systems like BASEBALL [9] and LUNAR [10] were designed to work for the specific domain. BASEBALL QA was designed to answer the queries related to the annual US baseball league. While LUNAR was designed to answer queries about the geological analysis of rocks returned by the Apollo missions.

In system developed by Sreelakshmi and Sangeetha Jamal, the proposed system use language processing components like parser, trigger etc.In this paper, an easy QA process used to be carried out using a process referred to as Semantic role Labeling and the efficiency of the system is in comparison with a QA method that uses sample matching.
In a system developed by Erfan Naumi[2], the proposed approach has some limitations. For instance the work doesn't answer the questions in list form. It doesn't have the deductive power to unite concepts and mostly give long answers to the query. For searching answers, pure RDF repository needs to be used as the source.

Sangdo Han And Shim introduced a question-answering system that responds to a keywords-query by extracting information from linked data and generating reports in natural language (NL). Using entity disambiguation and distributed word similarity, they matched each keyword to a related entity and property in linked data.

Payal biswas proposes a framework for establishing question Answering process for restrained area utilizing developed NLP tools. . The primary objective of the model is to extract the distinct and designated answer for the given query from a colossal dataset. In system developed by Shouning Qu, keyword searching in search engines was not so efficient. And also it was lack in accuracy.
Thus we need to develop an algorithm that overcomes all these issues

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

TABLE 1. II. RELATED WORK

Sr. No.	Year	Author Name	Issue Identified	Proposed Mechanism
1	2013	Erfan Naumi, Khanyyam Hashmi, Fayez Khazalah, Zaki Malik	a. Repository of questions is limited by the no. of QA stored in the DB b. Search engine lacks the power to answer multi level questions.	a. Covert the natural language user query to RDF triples and finds the answer. b. Used Resource Description Framework triples and find the answer without the use of ontology.
2	2014	Payal Biswas, Aditi Sharan, Nidhi Malik	a. Search engines lack the capability to deduce the correct answer. b. A advanced NLP tool is required to understand and implement the QA system.	a. This framework provides a proper flow of data for answer extraction. b. Proposed a restricted domain QA system using advanced NLP tools and software.
3	2014	Varsha Bhoir, M.A. Potey	a. The questions were classified by heuristic rule based on interrogative and its qualifiers. b. A more effective technique of question classification is necessary.	a. Works on restricted domain.. b. The system is an integrated retrieval technique and enhance web crawler and keyword oriented procedure programming.
4	2014	Sreelakshmi V, Sangeetha Jamal	The retrieval of information should be quick for factual data.	a. Retrieval of accurate information from web made quickly for factual questions. b. QA system heavily use language processing components like parser, trigger etc.
5	2015	Sangdo Han, Hyosup Shim, Byungsoo Kim, Sconyeong Park, Seoghan Ryu, Gary Geunbae	a. Mostly queries are fired in keywords. b. The system must interpret every possible user intention and report related answer in natural language form.	a. A concept of keyword question answering system for linked data. b. Used entity disambiguation and distributed word similarity to match each keyword to a related entity and property in linked data.
6	2008	Shouning Qu, Sujan Wang, Yan Zou, Qin Wang	a. Keyword searching in search engines was not so efficient. b. Lack in accuracy.	a.Introduced text classification algorithm. b. Used NLP segment text and calculate weight

III. PROPOSED ALGORITHM

This research work proposes an efficient technique to create a web application that gives exact answer to the user query. We have implemented a web application that answers the query and display the answers that best matches with the given query.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

The given below figure shows the main steps involved in proposed Question Answering System.

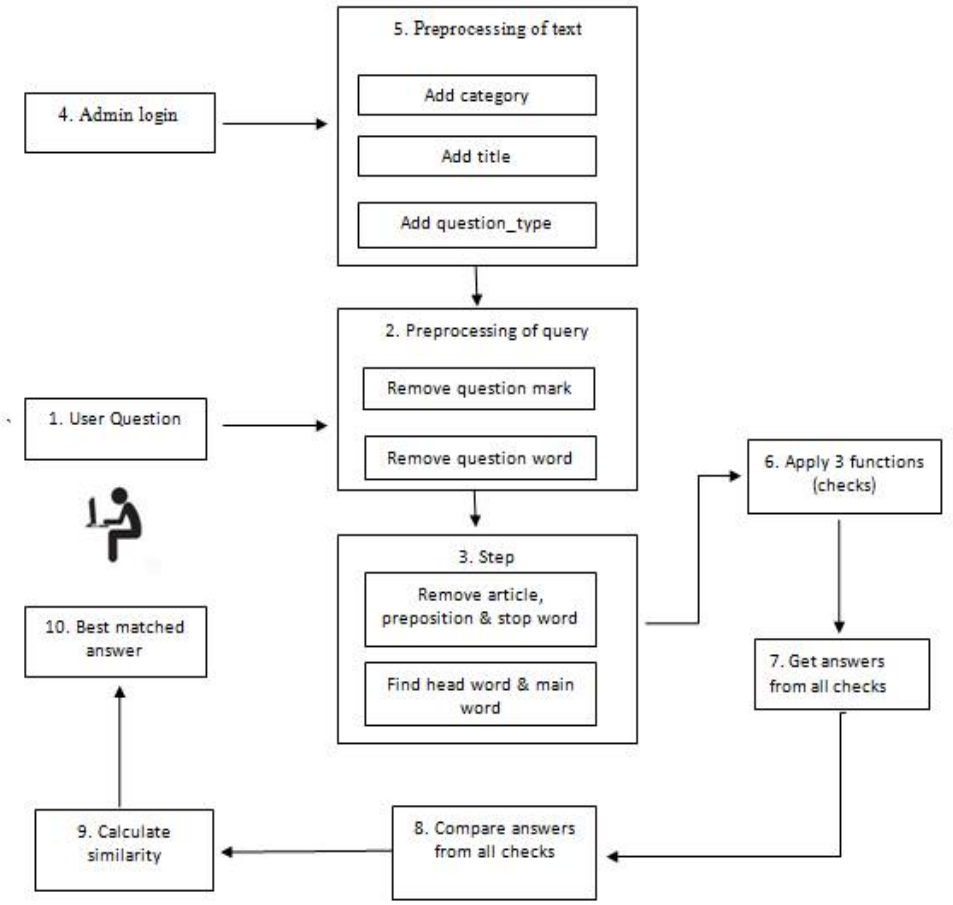

Figure 2.Stepwise Diagram of proposed Question Answering System

The system consists of following major steps as follows:

Step 1: The user enters a query through a user interface.
Step 2: This step involve the preprocessing of user query which have the functioning of removal of question mark, removal of question word and store the question type in an array.
Step 3: This is the main step for answer extraction, where possible keywords and headwords are identified in the search of the best matched answer.
Step 4: This step is the admin step. Admin need to login into the system for the preprocessing of the data stored.
Step 5. Admin converts the raw data into semi structured information. The information is stored in database.
Step 6. This step applies three checks (function). These checks are the different conditions that are applied on the user query to find the best answer.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Step 7. Get all the possible answers from the applied checks. Checks may generate same or different answer according to the function applied.
Step 8. Compare all the answers from each check. If each answer is same then display the same answer. If they are different, then send them for the calculation of similarity.
Step 9. If the answers are different, then calculate the similarity of the answers with the user query.
Step 10. The best matched answer is given to the user with its similarity calculated, and answers of the other checks are also displayed.

In step 6, we are applying three different checks or functions that are applied on the given query.
Algorithm 1: [check 1]
Subject string (wh-word,article,preposition,subject)
Function first_check()
{

1. Remove wh-word
2. Update subject string
3. Remove article
4. Update subject string
5. Remove preposition
6. Update subject string
7. Compare updates subject string with array question_title[]
8. Store the matching strings in another array temp[]
9. Calculate length of each string in array temp[]
10. Store max length string in variable result_1

} End Function first_check

Algorithm 2: [check 2]
Subject string (wh-word,article,preposition,subject)
Function Second_check()
{

1. Check category and question type of subject string.
2. Store it in q_category and quest_type
3. Select title from q_category that has question type as quest_type
4. Store it in array title_result[]
5. Compare title_array with Subject string
6. Store the matching result in array temp[]
7. Calculate length of each string in array temp[]
8. Store max length string in variable result_2

}
End Function Second_check

Algorithm 3: [check 3]
Subject string (wh-word,article,preposition,subject)
Function Third_check()
{

1. Use preg_match() to match strings with regular expressions.
2. Function preg_match(expression,string)
   {
   If
   match
   than
   return true

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

```
                    Else
                            return false
                }
```

3. Store result of true value string in array temp[]
4. Calculate similarity percentage of element of array temp[]
5. Store max similarity percentage string in variable result_3

}
End Function Third check

Now, compare the result string of each check. If two or all the three strings are same , then it is displayed as result. If not, then compare the length of all result strings, and display the string with the maximum length as that of result string.

IV. SIMULATION AND RESULTS

TABLE 2. QUESTION AND ANSWER EVALUATION

Sr. No	Question	Similarity calculated	Answer generated
1	What is computer?	66.66%	good
2	Who is mahatma Gandhi?	44.44%	satisfactory
3	Where was gandhiji born?	100%	excellent
4	When did gandhiji went to south Africa?	Not found	Very bad
5	What is vitamin	63.63%	good
6	Sources of vitamin a	100%	excellent
7	Sources of vitamins	59.25	satisfactory
8	Who is inventor of computer	100%	excellent
9	Who is father of nation?	46.66%	Very bad
10	Who is father of gandhiji?	43.75%	good

In the Table 2, questions are given to the system and the evaluation of the answer generated is given. The system generates the similarity calculated with the documents stored in the database. On the basis of similarity, the rank has been given to the answer. What, where, who type of questions gave a more than good answer based on the similarity calculated. Thus factoid types of questions are answered more satisfactorily. Definition type of questions like what type of questions, the answer generated has good performance.
The performance of the algorithm for extracting other types of questions will be much better in future with the improved performance of tools and softwares used for the testing.

V. CONCLUSION AND FUTURE WORK

The aim of the proposed system is to improve the accuracy and the time to retrieve the answer of the given query. Since similarity is also calculated, so the user has the clear perspective of the answer evaluated.
In the future, we consider reducing the pre-processing of the knowledge base done by the admin. We will also attempt to develop a technique that automatically pre-processes the knowledge base.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

REFERENCES

1. PayaL Biswas, Aditi Sharan, Nidhi Malik, "A framework for restricted domain question answering system".
2. Erfan najmi, Khayyan Hashmi, IEEE 2013 "Intelligent semantic question answering system".
3. M Ramprasath, S Hariharan Improved Question Answering System by semantic reformulation, IEEE- Fourth International Conference on Advanced Computing, 2012.
4. 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
5. Molla D., and Vicedo J., "Question answering in restricted domains: An overview", Computer Linguist, ppAI-6 1, 2007
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 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, 1961.
9. W.A. Woods, 'Progress in natural natural language understanding- An application to lunar geology," in proceeding of the June 4-8, 1973, National Computer Conference and Exposition.
10. 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.
11. 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.
12. Pragisha K. "design and implementation of a QA system in Malayalam".
13. 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.
14. Poonam Gupta, Vishal GuptaAssistant Professor, Computer Science & Engineering Department University Institute of Engineering & Technology Panjab University, Chandigarh.
15. 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.
16. PachpindPriyanka P, BornareHarshita N, KshirsagarRutumbhara B, MalveAshish D BE Comp S.N.D COE & RC, YEOLA," An Automatic Answering System Using Template Matching For Natural Language Questions".