

CLOSED DOMAIN QA SYSTEM WITH EXTENSIVE CONVICTION OF IR

TEJAS RUPANI & GIRISH MULCHANDANI

Computer Engineering, V.V.P Engineering College, Gujarat, India

ABSTRACT

Search engines are keyword based, So when user enter query in user language (Natural Language) then its result is differ from the result given by keyword. Search engine also give list of document from which user need to identify the answer. So it is very time consuming process.

Closed Domain QA system is used to identify the precise and accurate answer from given documents, So there is no need to view all the documents and identify the correct answer as we done in search engine.

Thus Closed Domain QA system obviate the time of user and provide precise and correct answer what user actually want. Closed domain QA system uses the concept of information retrieval It is main part of QA system and also use answer analyzer, question analyzer. This is closed domain, Repository oriented and Single language Question Answering System. This system is domain specific so it may be used in education system in which student can get the answer of query which will be available in various books but rather to view the entire book in details, system give direct answer which is more relevant.

KEYWORDS: Closed Domain QA, Question Answering System, Information Retrieval

INTRODUCTION

IR (Information Retrieval) is the process of finding material or documents of an unstructured nature that satisfies an information need from large collection of scattered data. A commonly observed that user of search engine (Information retrieval) issue a query in natural language, and want accurate answer of that particular query.

But instead of such accurate answer what happens, user gets the large set or list of documents as in link of web pages. Closed Domain QA (Question answering) System can be part of information retrieval system that provides the way to get the answer of query that will be asking in natural language. The answer must be aphoristic answer instead of informative documents.

Closed Domain Question answering system can act as “programmed system having the capability of answering natural language question in human like manner with quick and perfect result”.

In current scenario, there has been marked increase in the amount of information available on the Internet. Users often have specific questions in their mind, for which they expect to discover out the answers. Users want to find the answer to be short and precise. Also users want the answer in their own language.

PROPOSED THEORY

Closed Domain QA system is natural way to request information that we do not know, check information that we are not sure. Developing of Closed Domain QA system involves Information Retrieval and Natural Language Processing.

It is the better way to find the information directly rather than the search that particular information. Create Closed Domain QA system in which user can upload the document from which, user want to extract the information and want to get answer without wasting of time to search large number of documents or data. This document must be verified or checked by the expert because of irrelevant document may create invalid answer and decrease the accuracy of the system. Also user can search the query from already available source of data which is already verified by the expert. This can be used in FAQs type of question in which there are so many question and answers are already available but need to search it. User enters the query or question in natural language which is not handled by most of the search engine. After that question analyzer is used to analyses the question and identified the types of question among available category the question, so that system can judge which kind of answer will be possible for such kind of questions. Question type and its expected answer type are generally identified by looking at question keywords. Consider following table.

Table 1: Classification of Question and Its Expected Answer

Question Type	Expected Answer Type
Who	Person
When	Date/Time
Where	Location
What	Object
How	Measure

Now this identify the different keywords from the questions and based on that keyword fist of all identify the different paragraph that consists this keywords. Once system get the number of paragraph from available documents now system need to identify the relevant paragraph that is relevant to the questions. There is various answer extraction method like Heuristic, Pattern Based, Relation Based and logical based. From that method retrieve the answer from the paragraph. Answer extraction takes the input as expected type of answer and set of paragraph retrieved from the available source of data. In this process similarity between the keywords of the question and keywords founds in passage is computed in order to get best passage in a ranked list.

PROPOSED SYSTEM

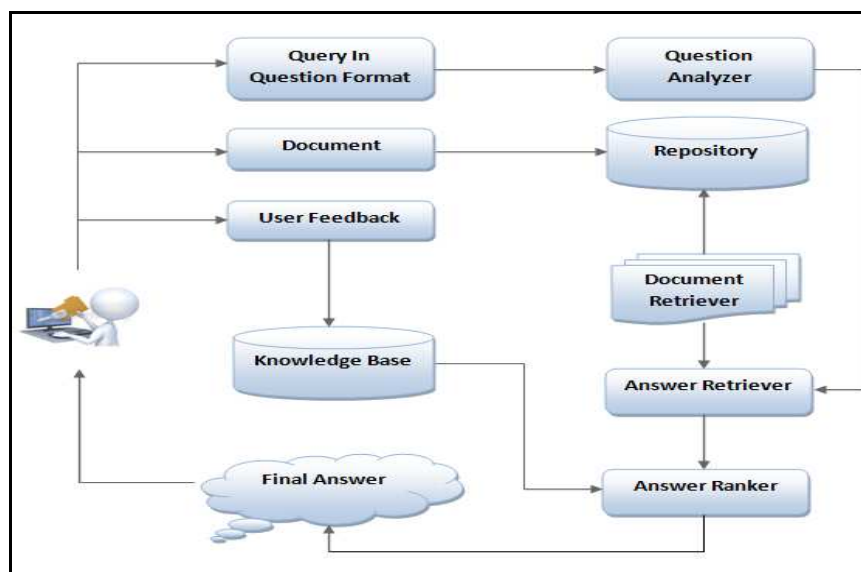


Figure 1: Proposed QA System

Consider following Component of Proposed System

- Question Analyzer
- Document Retriever
- Answer Retriever
- Answer Ranker
- Knowledge Base

Question Analyzer

This component gets the value from user. User enters the query or question in natural language form. Question analyzer receives the question and based on the different classification of question as we know it will prepare the type of expected answer.

Example: Who is prime minister of India?

So, when users enter this question as query then first based on the question “WHO”, it will be identified that the possible answer of this question is some person. After that this component also finds the different keyword from the question and sends it to next component document retriever.

After that it will be parsed and identify the keywords from the question like Prime Minister, India etc. So, after completion of this stage we have two things.

- Expected type of answer
- Keywords which we have to compare with no of documents

Document Retriever

This component used the keyword which will be provided by the Question analyzer and based on that keyword start to search the documents from the database. So, this component finds the relevant documents from the available large amount of document. Based on the rank of document relevant document will be identified and then from that document we have to find the relevant passage. After that number of retrieved document now we need to find the relevant passage from the relevant document. Then multiple passages are transfer to next component answer retriever. For scoring the document simple model is used which is known as TF-IDF model.

$$Score(Q, D) = \sum_{i=1}^n tf(q_i, D) * idf(q_i)$$

$tf(q_i, D)$ = frequency of term q_i in document D

$idf(q_i)$ = $\log(N/\text{Number of documents containing } q_i)$

For Example

Document 1 has Conference is in Ahmadabad

Document 2 has Conference is in ADI

Document 3 has Summit is in Rajkot

Query Q is Conference ADI

So, now ranking will be calculated as per the formulas

$$\text{Score (Q, D1)} = (1 + 0) * (0.64) = 0.64 \quad \text{Rank 2}$$

$$\text{Score (Q, D2)} = (1 + 1) * (0.64) = 1.28 \quad \text{Rank 1}$$

$$\text{Score (Q, D3)} = (0 + 0) * (0.64) = 0.0 \quad \text{Rank 3}$$

Answer Retriever

It takes the input from document retriever and analyzes it and finds the accurate and relevant answer based on different expert knowledge, FAQ and types of possible answer.

These possible answers are transferred to next component answer ranker.

Answer Ranker

Answer ranker takes input from answer retriever and give the rank to particular passage or particular answer which is very near relevant or accurate.

Knowledge Base

It will be build with the user's feedback and expert's comment on uploaded documents. It may contain the FAQ's of relevant document from which user need to find the correct or precise answer.

ANALYSIS

The main purpose of the Question Answering System is to find the precise and correct answer from predefined repository. There is no need to view all the documents from which user want to find answer. So in proposed system, first of all user need to select the repository from which user need to find answer. This repository must be validated by expert because if invalid repository is used then user will always get incorrect answer. After that index will be build on that repository and then user need to enter query in question format which is in natural language form. User will get the answer from highest rank document which is precise and may be correct. Proposed system works effectively for factoid type of questions. Some kind of limitations which are as below:

- Repository must be validated
- Complex questions can't handle
- Answer in image format or video format can't be extracted by system
- Range can't be calculated automatically

CONCLUSIONS

Closed Domain Question Answering System work effectively for factoid types of questions. It gives precise and accurate answer of the question which will be asked in natural language. The overall effectiveness of system is good and it

full fill its objective. This system can be used with any specific domain and can give the precise answer rather than the list of documents which will be given by search engine. If repository of system is valid then user can get answer of the question which must not be complex and not give answer from image or in video format.

ACKNOWLEDGEMENTS

I take this opportunity to express gratitude to Prof G. H. Mulchandani, Computer Engineering Department, for their help and support. I am extremely thankful and indebted to him for sharing expertise, and sincere and valuable guidance and encouragement extended to me. I am also take in mind, my sense of gratitude to one and all, who directly or indirectly help me.

REFERENCES

1. Olga Lebedeva and Lariss zaitseva, "Question Answering System in Education and their Classifications." Riga Technical University, 2014.
2. S. Quarteroni and S. Manandhar, "Designing an Interactive Open Domain Question Answering System." The University of York, United Kingdom, 2008.
3. L. Hirschman and R. Gaizauskas, "Natural Language question answering the view from here." Natural Language Engineering, Cambridge University, 2009.
4. Prof. Kohei Ara and Anik Nur Handayani, "Question Answering System for Effective Collaborative Learning." IJACSA, 3, 1, 2012.
5. María-Dolores Olvera-Lobo and Juncal Gutiérrez-Artacho, "Evaluation of Open – vs. Restricted Domain Question Answering System in the Biomedical Field." journal of Information Scienc, 2009.
6. Sucunuta, M.E and Riofrio, G.E., "Architecture of a Question Answering System for a specific repository of document." ICSTE, by IEEE, 2010.
7. Najmi, E. , Hashmi, K., Khazalah, F. and Malik, Z."Intelligent Semantic question answering system." CYBCONF, by IEEE, 2013
8. Biswas, P., Sharan, A. and Malik, N." A framework for restricted domain Question Answering System." ICICT, by IEEE, 2014.
9. Somnath banerjee, "Question classification and answering from procedural text in English." Jadavapur University, 2012.
10. Sivaji Bandyopadhyay, "A Hybrid Question Answering System based on IR." Jadavapur University, 2011.
11. Richard J Cooper and Stephen M Rigger, "A Simple Question Answering System." TIPTSTER, 2009.
12. Anietie Andy and Mohamed Choukha, "Exploiting Synonyms to Improve Question and Answering System." International Journal of Computer Science, 2014.
13. Rupani Tejas, Girish Mulchandani, ' Integrating QA into General Notion of Information Retrieval', IJAERD, December 2014.

