IJRESM
International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792

A Survey on Question Answer System

Kalyani Shinde[1], Anjika Singh[2], Reshma Shitole[3], Pallavi Singh[4], Sumit Harale[5]
[1,2,3,4]B.E. Student, Department of Computer Engineering, ICEM, Pune, India
[5]Professor, Department of Computer Engineering, ICEM, Pune, India

***Abstract*: Question Answering (QA) is the method of automatically answering a question asked by human in natural language utilizing either a pre-structured database or a collection of documents. QA system has numerous applications like online examination, education, health care, sports, geography, etc. In general, question answering system has three modules, for example, question processing, document processing (Information retrieval) and answer processing.**

***Keywords*: Question Answering system, Information Retrieval, Knowledge base**

1. Introduction

Software engineering has dependably helped man in making his/her life less demanding. New period in mankind's history is Information Era. With help of web indexes, we can get any data readily available. We are only a tick far from getting to a page at remote corner of the world. Notwithstanding the splendid research, quicker PC processors and less expensive memory bolstered these extraordinary headways. We have constantly needed PCs to act keen. To achieve this assignment, the field of Artificial Intelligence appeared. Question Answering is an exemplary NLP and data mining application. Task that a question answering system acknowledges is given a question and collection of documents, finds the exact answer for the question.

A question answer (QA) framework is a framework intended to answer Questions presented in natural language. Some QA system draw data from a source, for example, text document or a picture with the end goal to answer a particular query. These "sourced" system can be apportioned sinto two noteworthy subcategories: open domain, in which the inquiries can be practically anything, yet aren't centred around particular area, and closed domain, in which the questions have solid limitations, in that they identify with some predefined source (e.g., a gave context or a particular field, similar to medicine). The closed domain QA system gives more exact answer than the open domain QA system. Along these lines, the thought for building up the Question Answering System for education purpose is proposed. The system will assist school youngsters and students to get data from the System for their homework assignments.

2. Motivation

We are dependably in a mission of Information. Anyway, there is contrast in data furthermore, information. Data Retrieval or web seek is develop and we can get significant data readily available. Question Answering is a particular type of Data Retrieval which looks for learning. We are not just keen on getting the significant pages yet we are occupied with finding particular solution to quires. Question Answering is in itself convergence of Natural Language Processing, Data Retrieval, Machine Learning, Knowledge Representation, Logic and Deduction, Sematic Search. It gives a decent stage to dig into "nearly" all of Artificial intelligence. Question answering system in its being is an art, in the meantime it has science in its substance. Question Answering Systems are required all over the place, be it medicinal science, learning frameworks for understudies, individual partners. It is need in each viewpoint where we require some help from PCs. It's a given that it merits investigating the leaving field of question replying.

3. Related work

Numerous specialists worked on Question Answer System from last numerous years. A portion of some authors are mentioned below:

From the historical backdrop of QA systems the first known QA system is BASEBALL that was created by GREEN et al [1] in 1961, it answer a question about all the baseball matches played in the American association in one season. It is an domain specific QA system with constrained set of data sets available. In 1993 world's first online question answering system was built named START created by Boris Katz [2], it answered the question asked in natural language in all domain.

Avani and Ajay developed a QA system that use a Python factoid question classifier and uses machine learning approach for question classification. It is use to determine the type of question and answer. It only used for Wh-type questions like ‘What’, ‘Which’, ‘Who’, ‘When’, ‘Where’ and ‘Why’. They use the Stanford Dependency Parser to parse the question and retrieve the part of speech tagging. It also retrieves Stanford Universal dependencies between each word in the question. They also used Question Classification(QC) [3] to filter unsuitable candidate answers. QC places constraints on the answer type and provides additional information about question.

Menaha R, Udhaya Surya A, Nandini K & Ishwarya M developed an open domain question answering system. In this system they use web snippets for answer retrieval. Snippet are two line of statements containing the small hint of information. They use google search engine to extract the web pages and snippets from search engine after keyword extraction. After extraction of snippets, snippets contain the relevant answer. They used keyword matching algorithm to filter precise answer by matching exact keywords with snippet content [4].

Web based QA system consists of three modules: question analysis, web information retrieval, answer extraction was developed by Wenpeng Lu, Jinyong Cheng & Qingbo Yang [5]. In question classification, they classify questions based on their interrogative words. In keyword extraction, to extract accurate keywords they use syntax parsing module like Stanford Parser. For keyword expansion they use thesauruses like WordNet and HowNet. To retrieve web information, they used some part of the snippets to compose candidate collection. From the collection of candidate answers they compute similarity among questions and sentences and based on similarity ranks the candidate answer. From ranking sentence with highest similarity is selected as best answer.

Megha and Dr. Sharma developed a user-user centered evaluation model for question answering systems based on the user's perspective and to applying short domain question in Natural language for data query. They generated Domain Dictionary which have two parts, one is main schema and other is the character type data in the database [6]. They only focuses on character type data. They used segmentation tool named as IK Analyzer to segment words because it allows user to define their own dictionary.

4. Types of question answer system

Distinctive sorts of QA frameworks which are dependent on systems techniques and the questions it handles [7].

A. QA system depend on web source

Web is the best source supplier to get data with the wide spread utilization of web, where client get a tremendous information. Web based question answering frameworks is utilizing the web crawlers Like Google, Yahoo, Alto Vista and so on., to find ask for site page's that containing solutions to the inquiries. The dominant part of these Web based QA system works for open domain and some of them works for closed domain. The information that is accessible on web has the qualities of semi structure, heterogeneity and distributivity. The Web Based QA frameworks generally handles why-sort of questions, for example, "who is the first prime minister of india"? Or then again "Which of the following is not right". This QA system gives replies in different structures like content archives, Xml reports or Wikipedia.

B. QA system depend on IR/IE

IR based QA frameworks are giving a set of top positioned documents as response to the client question. Information Extraction(IE) system is utilizing the characteristic of natural language processing(NLP) to parse the query or documents returned by IR system, yielding the "importance of each word". Open domain QA systems give a brief response to a query, tended in natural language that isn't confined to a particular field. The knowledge base of a QA system is typically a substantial gathering of documents in natural language. Contingent upon the measure of the information included, numerous QA frameworks use IR modules in their design, in view of their strategies to process and store the data in a way that empowers a query over a lot of information to be recovered in a sensibly brief time. IR systems process and store huge amounts of unstructured information, so it can rapidly return the data that is applicable to a given demand. Information is contribution to the IR system through the document idea.

A document will be returned, by the IR system, as a match to a question to the system. The returned documents are ordered by a scoring method that endeavors to decide the significance of the document to the query. In a QA framework, the IR segment is commonly used to filter through the documents that have nothing to do with the query, holding just the documents that are related, for further preparing [7].

C. QA system depend on restricted domain

This sort of Question answering system required a semantic back up to comprehend the natural language text so as to answer the queries precisely. An effective technique for enhancing the exactness of QA system was achieved by confining the domain of questions and the size of knowledge base which brought about the advancement in restricted domain question answering system (RDQA). These systems have unique characteristic like "System must be Accurate" and "Reducing the level of Redundancy". RDQA beats the issues find in open domain by accomplishing better accuracy.

D. QA system depend on rule based

The rule based QA system is an extension for IR based QA system. Rule Based QA doesn't utilize profound language understanding or explicit advanced methods. A wide inclusion of NLP procedures are utilized so as to accomplish precision of the answers received.

Rule a based QA system produces heuristic rules with the assistance of lexical and semantic highlights in the questions. For each kind of questions it produces rules for the questions, for example, who, when, what, where and Why type questions. Example, "Who" questions produces rules which contains Names that are for the most part Nouns of people or things, for example, Who is the president of India?. These Rule Based QA systems initially build up parse documentations and produce preparing cases and experiments through the semantic model. This system comprises of some basic modules like IR module and Answer identifier or Ranker Module.

5. Classification of questions

The different kinds of questions are asked in Question Answering systems which straightforwardly affect the answers. We compose kinds of questions into various classifications. The classification of question types are (1) Factoid type questions, (2) List type questions, (3) Causal Questions, (4) Confirmation Questions, (5) Hypothetical Questions, (6) Complex questions [7].

A. *Factoid type questions*

The factoid type questions normally start with wh-word. These questions are easy to answer and reality based that require answers in a solitary sentence or short expression. For example, the factoid type question "What is the capital of Maharashtra?" requests a city name and it is simple to answer this sort of factoid question and reduce the search space for conceivable answers. The answer types for factoid type questions are for the most part named elements. Factoid type questions gives an acceptable execution in answering. For the most part factoid type questions are vast store of questions. Factoid type questions needn't bother with complex natural language processing to find solutions. Factoid type questions can be answered by short expressions, for example, associations, people, dates and areas.

B. *List type questions*

The list type questions require a list of actualities or elements as answers for example List names of movies released in 2018. For the list type questions, the answer types are named entities. Henceforth, the answers of list questions can give great precision. Question Answering systems don't need vast natural language processing to receive answers of list type questions. The strategies which are used in factoid type questions can function admirably for list type questions. One of the issue asked in list type question is settling the value for the entity or the number.

C. *Causal Questions*

The answers of causal questions are not named entities as factoid type questions. Causal questions require answers descriptions around an entity. Causal questions are asked by clients the individuals who want answers as reasons, clarifications, elaborations and so on identified with specific items or occasions. Example, Why don't atoms weigh anything?. Causal questions have illustrative answers which can extend from sentences to paragraphs to an entire report.

D. *Confirmation Questions*

Confirmation questions require answers as yes or no. For example, the confirmation type question "Is Earth revolves around a sun?", requests the answer yes or no. To answer confirmation questions world information, derivation system and sound judgment thinking is important. The hindrances of confirmation questions asked in QA systems are they require a larger amount of information picking up and retrieval procedures which are under the advancement stage. Aside from the above confirmation type questions, there can be conclusion questions which require subjective data around an occasion or substance. QA systems utilize social web and opinion mining strategies to find solutions to the opinion type questions. The benefits of opinion type questions are opinionated information sources contain general opinions which can help the clients in making judgment about the items. The disadvantage of opinion type questions are identification of spam or phony information in systems which causes issue in genuinely opinion mining of the content.

E. *Hypothetical questions*

Hypothetical questions ask for data related to any hypothetical occasion and no particular answers of these questions. Hypothetical questions mostly begin with 'what might happen if'. The unwavering quality and precision of these questions are low and relies on clients and context. The normal answer type is spread for hypothetical type question. Therefore, the exactness of hypothetical question answering is low.

F. *Complex Questions*

Complex questions are progressively hard to answer and whose answers are comprises of list of "pieces". Complex question, for example, "What are the reasons of Global Warming?" regularly require inducing and blending data from various documents to find numerous chunks as solutions. Complex methods are expected to answer complex questions. Complex question requires various, distinctive types of data and making answer is troublesome. The complex question comprises of various questions and each question look for answer from different documents.

6. Framework of question answer system

A question answer system consists of three Phases. The three main phases are [8],

A. *Query processing module*

It accepts single questions as an input to the QA system. The objective of this module is to distinguish the question keywords and supposed set of answer documents.

B. *Document processing module*

After distinguishing the question keywords, the appropriate answer candidates are retrieved from the gathered document for answer matching. The collected document which has total keyword match with the question keyword are chosen for answer extraction.

C. *Answer processing module*

Answer matching has two sub-modules as, Scoring and Ranking and Answer extraction. In Scoring and Ranking of the candidates answer selection of the suitable answer is done by coordinating window sizes candidate answer which has the highest score is chosen as the best answer and the result of best suitable answer is processed for answer extraction.

7. Approaches in QA systems

A. Linguistic approach

Linguistic approaches, for example, tokenization, POS tagging and parsing were applied to form questions into a right query that absolutely extracts the matching answers from the structured database. The type of Questions Handled in this methodology are Factoid questions and there is a profound Semantic comprehension. It is troublesome as knowledge base are generally designed just to deal with their pre-stored questions. It is most dependable as answers are extracted from self-maintained knowledge base. It is scalable and complex as new guidelines must be presented in the knowledge base for each new idea.

B. Statistical approach

These methodologies propose strategies that not just deal with the immense measure of information but also with their heterogeneity as well. The significant limitation is that they consider each word independent and fail to locate the linguistic properties of a combination of terms. They Support vector machine (SVM) classifiers, Bayesian classifiers, Maximum entropy models are a few methods that are utilized for question categorization reason. This methodology is very suitable in dealing with huge volume.

C. Pattern matching approach

This methodology utilizes the communicative power of text patterns to substitute the advanced processing concerned in other competing approaches at present, a significant number of the QA systems automatically study such text patterns from text passages instead of making utilization of complicated linguistic knowledge or tools viz., parser, named-entity recognizer, ontology, WordNet, etc. It requires much time and uncommon human abilities to install and keep up the system. This methodology best suits to small and medium size websites, Semantic web. Different kinds of questions which are handled are Factoids, definition, abbreviation, birth date and is having a semantic understanding less than other competing approaches. It relies upon the validity of knowledge asset and scalability is less as new patterns must be scholarly for each new idea.

8. Conclusion

In this paper we depicted about the survey of a QA system for an English language. It gets natural language questions from the user and chooses most proper answer. This survey paper also portrays the different types of question, different question answering approaches and different types of question answering system. QA Systems can be developed for resources like web, semi-structured and structured knowledgebase domain. The Closed domain QA Systems give more exact answer than that of open domain QA system.

References

[1] Green B., et al. “Baseball: An automatic question answerer”, in proceedings of western joint IRE-AIEE-ACM Computing Conference, Vol. 19, pp. 219-224, 1961.
[2] Katz. B, Gary Borchardt, “Natural language annotations for question answering”, AAAI magazine fall 2007.
[3] A. Chandurkar and A. Bansal, "Information Retrieval from a Structured Knowledge Base," *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, San Diego, CA, 2017, pp. 407-412.
[4] Menaha R, Udhaya Surya A, Nandhni K, Ishwarya M, “Question Answering System Using Web Snippets”, International conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud), (I-SMAC 2017).
[5] Wenpeng Lu, Jinyong Cheng, Qingbo Yang, “Question Answering System based on Web”, 2012 5th International Conference on Intelligent Computation Technology and Automation.
[6] M. Mishra, V. K. Mishra and H. R. Sharma, "Leveraging knowledge based question answer technology to address user-interactive short domain question in natural language," *2012 2nd National Conference on Computational Intelligence and Signal Processing (CISP)*, Guwahati, Assam, 2012, pp. 86-90.
[7] Deepa Yogish, Manjunath T. N, P. Ravindra S. Hegadi, “A Survey of Intelligent Question Answering System Using NLP and Information Retrieval Techniques”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2016.
[8] Anjali Saini, P. K. Yadav, “A Survey on Question – Answering System”, International Journal of Engineering and Computer Science, Volume 6 Issue 3, March 2017.