

A Survey on Types of Question Answering System

A. Chandra Obula Reddy¹, Dr. K. Madhavi²

¹Research Scholar, Department of Computer Science & Engineering, JNTUA,
Ananthapuramu - 515002, A.P., India.

²Associate Professor, Department of Computer Science & Engineering, JNT University, Ananthapuramu -
515002, A.P., India.

Abstract: Search engines like Google and Yahoo! that allow the users to search for documents on the World Wide Web. In Search engines the user has to check each and every document to get useful answer to the question and it is a time consuming process. The Question Answering (QA) system reduces the search time to get exact answer to the question. Question Answering system is an important research area in information retrieval. Research on the area of Question Answering system started in the year 1960 and present lot of Question Answering systems have been developed. Question Answering system combines the research from different domains like Natural Language Processing, Artificial Intelligence, Information Retrieval and Information extraction. The objective of question answering system is to find exact answer to the question asked by user in natural language. Question Answering systems are categorized based on their available resource for answers. The closed domain question answering system gives more exact and correct answers than open domain question answering system. This paper survey a overview on the Question Answering System, types of Question Answering systems and general architecture of Question Answering system.

Keywords: Information Retrieval, Natural Language Processing, Question Answering System.

Date of Submission: 15-12-2017

Date of acceptance: 30-12-2017

I. Introduction

Nowadays, the internet users are increasing in percentage and there are huge growing online sources to get information interrelated to various fields of topics i.e. health, economic, educational, and industry information [1,2]. Search engines like Google and Yahoo, which can be used to find a ranked list of relevant documents on the web in answer to user's formulated questions based on different measures such as keyword matching, frequencies of accessing documents [3]. However, in Search engines the users have to look at every document individually for getting the required information so it is a time taking process. Present Question Answering (QA) systems are used to get most suitable and accurate answer to the questions asked by human in Natural Language. Question Answering is a rapid upward research field, consists of different fields like Computer Science, Information Extraction (IE), Natural Language Processing (NLP) and Information Retrieval(IR) [4]. Lot of advancement is seen in Question Answering associated to the English, Korean, Japanese, Chinese, etc. languages. Question Answering system has a lot of application based on the basis of answers. Like extracting information from document, Online examination system, document management, Language learning, human and computer interaction, classification of document and many more [5]. Question Answering systems are more complex than IR Systems and need general natural language processing techniques to get precise answer to the natural language questions [6]. The main objective of Question Answering system is to get back answers of questions rather than full document. Question Answering provides perfect solution to get valid and accurate answers to user question asked in natural language instead of query.

In the second section, this paper discusses the classification of Question Answering systems. Third section explains the types of questions asked by users. Fourth section describes the general architecture of Question Answering system.

II. A classification of question answering systems

In this section, we discuss details of the classification of Question Answering (QA) systems. We present a description of the classification. The job of producing answers of questions is associated to the type of questions asked. Choice of the domain as a basis of classification of Question Answering systems. Some users need general information on a general topic and other users need specific information from a particular application domain [7].

2.1. Open domain Question Answering system

Open domain Question Answering systems are not restricted to any specific domain and provide a short answer to a question, addressed in natural language. Web is the best source to get information with huge extend usage of internet and most of the Web based Question Answering systems works for open domain[8]. Open domain QA system depends on information such as World Wide Web and universal ontology and can answer with questions almost everything [9]. Open domain QA systems apply a general vocabulary and do not require any domain specific vocabulary. Users do not require knowledge of domain specific for preparing questions. Open domain QA systems consists of large repository of questions and generally search for answers within a large document collection. In Open domain QA systems, Wikipedia can be used as a source of information. In Open domain QA systems, the quality of answers is low which are generated by users. Open domain QA systems are domain independent and depends on world data and basic ontology [10]. In Open domain QA system, there is no restriction of domain which is based on web and questions are related to any subject. Open domain QA systems are more appropriate for huge number of casual users. Restricted domain QA systems are used when Domain experts need specific information in answers for the questions.

2.2. Closed domain Question Answering system

In Closed domain QA system, there is restriction of domain which is based on web and questions are related to a specific domain. Closed domain Question Answering system consists of limited repository of domain specific questions and can answer a limited number of questions. Hence in closed domain QA systems, the quality of answers is high. Closed domain QA systems answer domain specific questions and answers are searched within domain specific document collections. Closed domain QA systems are designed to get answers from structured data (such as databases), unstructured data (free texts) and semi-structured data (such as XML-annotated texts) [11]. Closed domain QA systems make use of domain specific terminology and ontology. There are different Closed domain QA systems developed in the literature such as: medical domain QA system, temporal domain QA system, community based QA system, geospatial domain QA system, patent QA system, etc. By integrating different Closed domain QA systems can get Open domain QA systems. Closed domain QA systems having limited work domain and more useful to domain expert users those who require specialized answers. One of the Closed domain QA system is Green's BASEBALL and that give answers to the questions about one season's baseball data [12].

III. Classification based on types of questions

The task of producing answers to questions which are asked by user's directly associated to type of questions. Thus types of questions asked in Question Answering systems directly have an effect on the answers. We organize types of questions into different categories. The different categories of question types are (1) Factoid type questions, (2) List type questions, (3) Confirmation Questions, (4) Causal Questions, (5) Hypothetical Questions, (6) Complex questions.

We give a brief explanation of each kind of question in the next sub sections.

3.1. Factoid type questions [what, which, when, who, how]

The factoid type questions commonly begin with wh-word. These questions are simple to answer and fact based that need answers in a single sentence or short phrase. For instance, the factoid type question "What is the capital of India?" asks for a city name and it is easy to answer this type of factoid question and reduces the search space for possible answers. The answer types for factoid type questions are generally named entities [13]. Factoid type questions gives a satisfactory performance in answering. Generally factoid type questions are large repository of questions. Factoid type questions do not need complex natural language processing to get answers. Identification of factoid type questions and their sub classification is one of the research issue in Question Answering system. Factoid type questions can be answered by short phrases, such as organizations, persons, dates and locations.

3.2. List type questions

The list type questions need a list of facts or entities as answers e.g. List names of movies in 2017. For the list type questions, the answer types are named entities. Hence, the answers of list questions can give good accuracy. Question Answering systems do not need deep natural language processing to retrieve answers of list type questions. The techniques which are applied in factoid type questions can work well for list type questions [14]. One of the problem asked in list type question is fixing the threshold value for quantity of the entity or the number.

3.3 Confirmation Questions [yes or no]

Confirmation questions need answers in the form of yes or no. For instance, the confirmation type question "Is Jesus Christ God and man?", asks for the answer yes or no. To answer confirmation questions world knowledge, inference mechanism and common sense reasoning necessary. One of the advantages of Confirmation type questions asked in QA systems are some expert users may like to investigate for information which need common sense reasoning and world knowledge for new knowledge [15]. The disadvantages of confirmation questions asked in QA systems are they need a higher level of knowledge gaining and retrieval techniques which are under the progress phase.

Apart from the above confirmation type questions, there can be opinion questions which require subjective information about an event or entity. QA systems use social web and opinion mining techniques to get answers to the opinion type questions. The advantages of opinion type questions are opinionated data sources contain public opinions which can help the users in making judgment about the products. The disadvantages of opinion type questions are detection of spam or fake content in text systems which causes problem in truly opinion mining of the text.

3.4 Causal Questions [why or how]

The answers of causal questions are not named entities as factoid type questions. Causal questions need answers descriptions about an entity. Causal questions are asked by users those who desire answers as reasons, explanations, elaborations etc related to particular objects or events.

e.g., Why professor Y take a speech in auditorium?

Causal questions have descriptive answers which can expand from sentences to paragraphs to a whole document [16].

3.5 Hypothetical Questions

Hypothetical questions request for information associated to any hypothetical event and no specific answers of these questions. Hypothetical questions usually start with 'what would happen if'. The reliability and accuracy of these questions are low and depends upon users and context. The expected answer type is spread for hypothetical type question. For this reason, the accuracy of hypothetical question answering is low [3].

3.6 Complex Questions

Complex questions are more difficult to answer and whose answers generally consists of list of "nuggets". Complex question such as "What are the reasons of Air Pollution?" often require inferring and synthesizing information from multiple documents to get multiple nuggets as answers. Complex procedures are needed to answer complex questions. Complex question require multiple, different types of information and making answer is difficult. According to [17], the complex question consists of multiple independent questions and each question seek for answer from multiple documents.

IV. General Architecture of Question Answering System

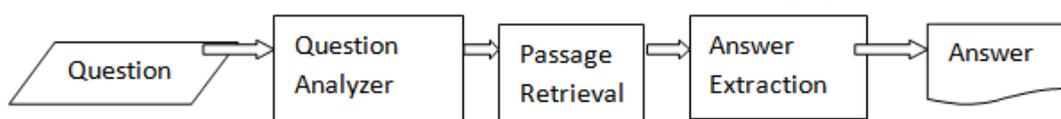


Fig. 1: General architecture of Question Answering System

A Question Answering System consists of three modules: Question Analysis, Passage Retrieval, and Answer extraction. Figure 1 represents the General architecture of Question Answering System [18].

4.1. Question Analysis:

The need of question analysis module in QA system is which helps to recognize the type of question asked by the user. Identification of question type is most important to get relevant information to user question. Question analysis is the most important step in QA system. The aim of Question Analysis module is to identify the type of question. In question Analysis part , first when the user post a question on the web, the question is sent to the server side, where the tokenizer breaks the question into separate words [19]. Next, the answer type detector recognizes the type of the answer. The expected answer can be determined based on question classification as illustrated in Table1 [20].

Table1: Example of expected answer type for question classification

Question Classification	Expected Answer Type
When	DATE
Which	PERSON/DATE/LOCATION
Why	REASON
What	PERSON/DATE/LOCATION
Who	PERSON

4.2. Passage Retrieval:

This module searches a list of related documents and extracts a set of paragraphs based on the focus of the question. The answer is in terms of paragraphs. Most appropriate passage is selected according to passage score for an answer. The passage retrieval module selects most related documents from the database and extract the passage that contains the candidate answer text [21].The distance between question and documents is calculated based on the Term-Frequency(tf) and Inverse Document Frequency(idf). The document that is relevant to the answer can be retrieved based on ranking. Most Question Answering Systems are based on Information Retrieval(IR) methods that have been adapted to work on passages instead of the whole document.

4.3. Answer Extraction:

Answer Extraction module selects the scored passages and the answer can be extracted from the top ranked sentences [22]. In answer extraction module most related answer is considered as the desired answer and the rest of the answers are candidate answers. This module is responsible for selecting most relevant answer from the answers list for the question. The most relevant answer for the question can be selected by answer-similarity checker and answer ranker. Similarity measured by counting the number of matching keywords between question and each retrieved answer. Answer ranker, sorts the answers in descending order of their relevance.

V. Conclusion

In this paper, an survey of Question Answering is presented and architecture of Question Answering system is introduced. This survey paper describes different types of Question Answering systems. Closed domain question answering system is restricted to a specific domain and the quality of answers is high. There is Open domain question answering system not restricted to any domain and the quality of answers is low. This paper also describes different category of questions. These questions are asked by users in human natural language. In this paper we observed that various researchers work on Question Answering with different languages. This paper explains Question Answering system works on structured, unstructured and non-question form queries. We also observed that most researchers work with closed domain question answering system rather than open domain question answering system.

References

- [1]. Waheeb Ahmed, Babu Anto P, Arabic Question Answering System Based On Data Mining, International Journal Of Scientific & Technology, 6, 2, February 2017,237-239.
- [2]. Nabeel Neamah, Saidah Saad, Question Answering System Using Similarity And Classification Techniques, International Journal of Computer Systems, 03,07, July, 2016,542-548
- [3]. Amit Mishra, Sanjay Kumar Jain, A survey on question answering systems with classification, Journal of King Saud University – Computer and Information Sciences , 28, 2016 , 345–361.
- [4]. Gursharan Singh Dhanjal, Sukhwinder Sharma, Advancements In Question Answering Systems Towards Indic Languages, International Journal of Research in Computer Science, 5, 1,2015,15-26
- [5]. Ajith Kumar, Khillare, Namrata Mahender, Question Answering System, Approaches and Techniques: A Review, International Journal of Computer Applications, 141, 3, May 2016, 34-39.
- [6]. Suresh Kumar, Zayaraz, Concept relation extraction using Naive Bayes classifier for ontology-based question answering systems, Journal of King Saud University – Computer and Information Sciences, 27, 2015,13–24
- [7]. Anjali Saini,, P.K.Yadav, A Survey on Question–Answering System, International Journal Of Engineering And Computer Science, 6,3,March 2017, 20453-20457
- [8]. Deepa, Manjunath, Ravindra, A Survey of Intelligent Question Answering System Using NLP and Information Retrieval Techniques International Journal of Advanced Research in Computer and Communication Engineering, 5, 5, May 2016,536-540
- [9]. Sreelakshmi, Sangeetha, Open Domain Question Answering System Using Semantic Role Labeling, IJRET: International Journal of Research in Engineering and Technology, 03, 15, Dec-2014, 104-107
- [10]. Payal Biswas, Aditi Sharan, Nidhi Malik, A Framework For Restricted Domain Question Answering System,IEEE, 978-1-4799-2900-9/14, 2014,613-620
- [11]. Harpreet, Rimpi, A Review on Novel Scoring System for Identify Accurate Answers for Factoid Questions, International Journal of Science and Research (IJSR), 2, 9, September 2013, 154-157.
- [12]. aspreet Kaur, Vishal Gupta, Effective Question Answering Techniques and their Evaluation Metrics, International Journal of Computer Applications, 65, 12, March 2013,30-37
- [13]. Comparative Study on Question Answering Systems and Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, 4, 11, November 2014, 775-778
- [14]. Youzheng, Hori, Hisashi, Leveraging social Q&A collections for improving complex question answering, Computer Speech and Language, 29, 2015, 1–19.

- [15]. Tanwar, Prasad, Datta, An Effective Reasoning Algorithm for Question Answering System, (IJACSA) International Journal of Advanced Computer Science and Applications, 2014,52-57
- [16]. Pechsiri, priyaku, Developing a Why-How Question Answering system on community web boards with a causality graph including procedural knowledge, Information processing in Agriculture,3, 2016, 36-53
- [17]. Setio Basuki, Ayu Purwarianti, Statistical-based Approach for Indonesian Complex Factoid Question Decomposition, International Journal on Electrical Engineering and Informatics ,8, 2, June 2016,356-373
- [18]. Bodor A.B. Sati, Mohammed A.S Ali, Sherif M. Abdou, Arabic Text Question Answering from an Answer Retrieval Point of View:a survey,(IJACSA) International Journal of Advanced Computer Science and Applications, 7, 7, 2016, 478-484
- [19]. Heba Kurdi, Sara Alkhaider, Nada Alfaifi, Development And Evaluation Of A Web Based Question Answering System For Arabic Language, International Journal on Natural Language Computing (IJNLC), 3, 2, April 2014,11-32
- [20]. Rosy Madaa, A.K.Sharma, Dixit, Deepti, Renu, A comparative study of web based and IR based question answering systems, An International online open access peer reviewed journal, 2277-9140, April 2012,188-194
- [21]. Rohini, Avinash, Keywords based closed domain question answering system for Indian penal code sessions and indian Amendment laws, I.J. Intelligent systems and applications, 12, 2015,57-67
- [22]. Pooja,Vinodh Krishnan, Geetha Manjunath, An Efficient Passage Ranking Technique for A QA System, International Journal of Computer Science & Information Technology (IJCSIT) 4, 6, December 2012,65-76

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

A. Chandra Obula Reddy "A Survey on Types of Question Answering System." IOSR Journal of Computer Engineering (IOSR-JCE) 19.6 (2017): 19-23.