

Survey: Restricted Domain Question Answering System Using Semantic

Aakriti Gupta¹, Mr. Pranay Chauhan²

Student M.Tech, Swami Vivekananda College of Engineering, Indore, M.P, India¹

Assistant Professor, Swami Vivekananda College of Engineering, Indore, M.P, India²

Abstract: The information present on the World Wide Web is continuously increasing at a tremendous rate. The rate at which the information is growing, requires new techniques to handle and organize this data. Question and Answering System is used to answer the query given by the user in the natural language. Main challenges of Question and Answer system is to give exact answer of question given by the user. This paper gives a thorough overview on the QA system, Types of question, its methods and the related work done on QA systems.

Index Terms: Answer extraction, answer ranking, semantic similarity.

I. INTRODUCTION

Automated question answering systems are used to facilitate to find the correct answer of the question given by the user. Current QA systems uses "search engine method" where a user is provided with the various links which are arranged in descending order of their ranks. User need to visit each pages and filter them to find the answer she/he needs. If the QA systems are supported through semantic technologies, it can use machine intelligence to return the accurate answer and thus the efficiency is improved. In recent years, there have been various attempts to add this semantic functionality to search engines.

Google have implemented a QA system for simple question that searches for answer by parsing the search records and then rank each site. [1]. This approach lacks the deductive power to answer multi-level questions.

This approach is computationally expensive. The time and space complexity is more in this approach.

Search engine like ask.com keeps a storage database of question and answers shaped from user queries and experts answers. This approach has limitation that the storage repository of question has limited number of questions that are stored in it.

Thus it is observed that a solution is needed that addresses two major concerns: First, it should be computationally inexpensive and feasible. Second, it should be able to expand its question repository and be able to answer the query with the ever changing nature of the information.

The goal of this paper to have a broad look at the approaches used in the QA systems and have a deep study regarding the methods, types, techniques used in current QA systems.

The rest of the paper is organized as follows. Section 2 discusses some of the related works in this area. The next section describes the general architecture of a Question Answering System. The question classification and its types are described in section 4.

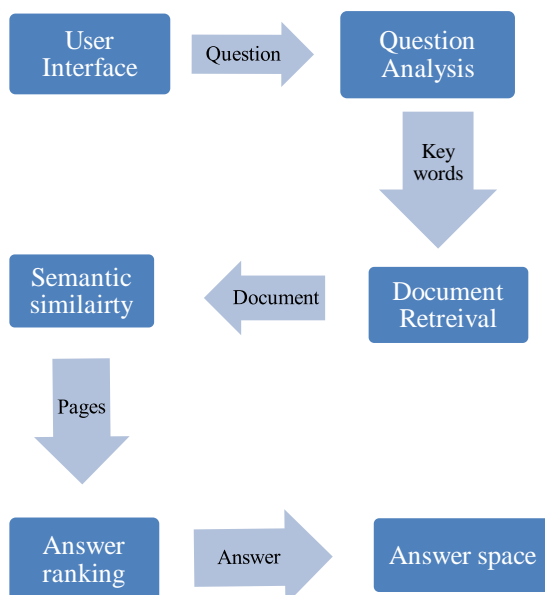


Fig 1 Stepwise Diagram of Question Answering System

II. RELATED WORK

In most of the research papers [1, , , ,].described in Table I, there are various question answering systems which have been developed with different concepts.

In a system developed by Erfan Naumi[1], 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 connect concepts and mainly provide long answers. For searching answers, pure RDF repository needs to be used as the source.

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

III. ARCHITECTURE OF A QUESTION ANSWERING

In this section we describe the architecture of a question answering system.[16]The overall architecture of the system can be subdivided into three main phases: (1) pre-processing of question, (2) question template matching, and (3) answering the query In most of the research papers [1, , ,],described in Table I, there are various question answering systems which have been developed with different concepts. In a system developed by Erfan Naumi[1], 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 connect concepts and mainly provide long answers. For searching answers, pure RDF repository needs to be used as the source.

Each step involved in general architecture of QA system is described in detail below

1. Process question: Processing of question is done to get some important information from it. Steps involved in this module are given below.

- a. Find the Type of given question using Wh word.
- b. Find out the expected type of answer.
- c. Get the Keywords from the Question.
- d. Find out the Focus of the question.

2. Process document: After the question has been processed, document processing module is used. In this module the documents which are relevant to the given question are retrieved and processed. Following steps used in document processing.

- a. Get the question and search relevant documents with the help of a search engine.
- b. Take top relevant documents.
- c. Extract the content from these documents.
- d. Save these contents.

3. Process Answer: This module describes algorithms for extracting the relevant answer for all categories of questions described by the system.

4. Clustering of data set: cluster dataset using fuzzy c-mean algorithm process question and answer processing.

IV. QUESTION CLASSIFICATION

Question Classification is the first step in a QA system. The goal is to classify questions and find the type of the answer. There are multiple ways to categorize questions. Many works try to be as specific as possible about getting the type of the answer, but since our main focus to find the answer is not a specific answer type, we can look at the problem from a more general perspective. There are many kinds of questions we can potentially consider, consisting of factoids, lists, definitions, hypothetical, causal, procedural and confirmation queries.

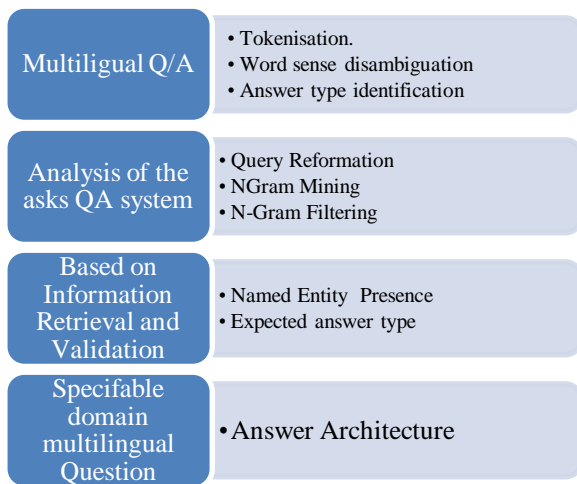


Fig.2 Various types of QA system and its methods.

TABLE. II QUESTION TYPES AND SUBTYPES

WH word	Question Type		
	Factoid Type	Definition Type	Descriptive Type
Question	Who When What Where Which	How What Why	What

Question answering systems can be classified on the basis of the domains over which it has been constructed. [16] For example:

1. Closed domain QA system
 2. Open domain QA system
 3. Restricted domain QA system
- A. Closed domain QA system works on the topics within a specific domain for example: cloud computing.
- B. Open domain QA systems are not domain dependent. The system has large collection of data from which answer is retrieved.
- C. Restricted domain QA system a convergence between open domain and closed domain. It means the system is not as domain independent as open domain and is also not as limited as closed domain QA system. During the

study of question answering systems, it is observed that restricted domain QA systems are more reliable and addresses problems with more accuracy.

V. CONCLUSION

In this paper we have studied the related work done in this field. Types of questions are studied carefully. An architecture of general question answering system is also presented. The survey work not only provides a simple and implementable framework for developing question Answering System but also provides a proper flow of data for answer extraction.

REFERENCES

- [1]. Erfan najmi, Khayyan Hashmi, IEEE 2013"Intelligent semantic question answering system".
- [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 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, 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]. Payal Biswas, Aditi Sharan, Nidhi Malik, "A framework for restricted domain question answering system".
- [17]. 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".