

A Survey on Recent Trends in Question Answering Systems

B. Srinivas¹, B. Raju²

Assistant Professor, Department of CSE, KITS, Warangal.

Assistant Professor, Department of CSE, KITS, Warangal

Abstract: *Question Answering (QA) Systems is a mechanized way to deal with recuperate cure responses to the questions asked by human in natural language. The principal thinking behind QA framework is to enable man-to machine correspondence. In this paper, we propose scientific categorization for portraying Question Answer (QA) systems, quickly overview major QA systems depicted in literature. At long last, an examination between these approaches based on specific highlights of QA system discovered basic in our investigation has been done, so as to acquire an understanding to explore scope this course. QAS go for fulfilling clients who are hoping to answer a particular question in natural language. In this paper we study different QAS. We give likewise insights and investigation. This can make room and help analysts to pick the suitable answer for their issue. They can see the deficiency, with the goal that they can propose new systems for complex questions. They can likewise adjust or reuse QAS systems for particular research issues.*

Index Terms: *linguistic approach, QA system, statistical approach, templates, pattern based approach, Natural Language Processing(NLP).*

I. Introduction

Modern information retrieval systems enable us to find archives that may have the related information, however the dominant part of them abandon it to the client to separate the valuable information from a requested rundown. For e.g., the question "Who has won the maximum individual medals in Olympics 2012?" ought to get back the reaction "Michael Phelps" yet instead the client is given a rundown of pertinent archives to investigate in the journey of an exact answer. QA is perceived as an ability with awesome potential. QA system empowers clients to get to the information assets in a natural manner (i.e., by asking questions) and to get back a significant and appropriate reaction in compact words.

One of the challenging undertakings for existing QA systems is to understand the natural language questions effectively and find the exact meaning to recover correct reactions. Change in motorized understanding of questions faces issues like question classification, formulation of right queries, ambiguity resolution, semantic symmetry detection, identification of temporal relationship in complex questions. In the comparable route identification of an immaculate answer requires legitimate approval system.

The QA system processing may extensively have three phases, i.e., question analysis: parsing, question classification and inquiry reformulation; report analysis: extricate candidate records, recognize replies; and answer analysis: separate candidate answers and rank the best one. A large portion of the current works integrate a few or these approaches to constructed improved systems that can manage shortage of these approaches. This paper is sorted out as takes after. In Section 3, we talk about classification for characterizing QA approaches. Following in next Section, we talk about upsides and downsides of each approach and present a comparative study of these approaches while final area presents conclusions.

II. Background

Numerous thoughts must be learned before checking chips away at Question Answering Systems.

What is Question Answering System?

Numerous definitions are accessible in the writing:

"For human-PC connection, natural language is the best information get to instrument for people. Henceforth, Question Answering Systems (QAS) have extraordinary importance and points of interest over web crawlers and are thought to be a definitive objective of semantic Web look into for client's information needs" [3].

"Question Answering on the Web is moving past the phase where clients basically type an inquiry and recover a positioned requesting of suitable Web pages. Clients and experts need focused on answers to their questions without incidental information" [4].

In this paper, we center in especially around QAS committed to the Web of records and the Web of information.

QAS for Web of archives and content

In Information Retrieval (IR) and Natural Language Processing (NLP), Question Answering (QA) is the undertaking of consequently giving a response to a question asked by a human in natural language. QA as an assignment can be separated into three fundamental particular subtasks, which are: Question Analysis, Document Retrieval and Answer Extraction [5] (see Fig. 1).

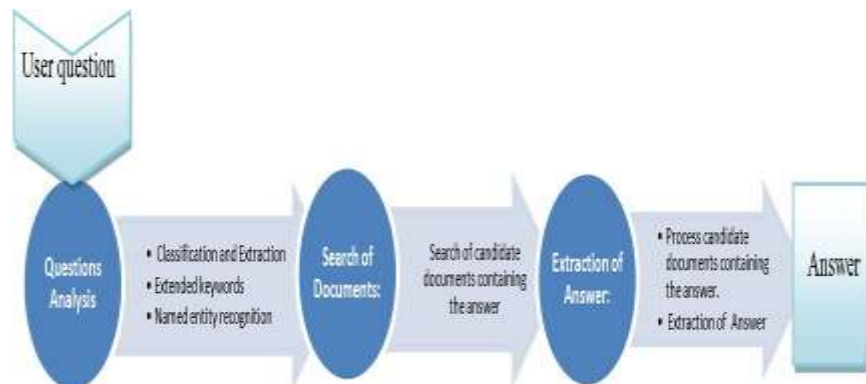


Fig 1: QAS Subtasks

In especially, Factoid questions are those gotten some information about Named Entity (NE), utilizing for instance the words: When, Where, How much/many, Who, and What, which get some information about date/time, place, individual, and association. The Second sort is the questions that get some information about the meaning of term or idea. Questions that utilization the words "Why" or "How" are another write that is difficult to answer and there are next to no if any endeavors done to answer this kind of questions.

QAS for the Web of data

The objective of QA Systems, as characterized by [6], is to enable clients to make inquiries in Natural Language (NL), utilizing their own wording, and get a compact answer. For QAS committed to the Web of data, User makes inquiry in natural language. The procedure begins by etymologically investigating (reliance charts utilizing a linguistic parser with a stage of named elements acknowledgment NER). The following stage is to arrange the question as indicated by one characterized question classification. The SPARQL inquiry is created in two stages (etymologically investigate and question order). An outer metaphysics asset can be utilized for coordinating things produced all the while. At last, when the SPARQL inquiry is produced, the cross examination of the Linked Data is done, and creates the correct answer of the client question.

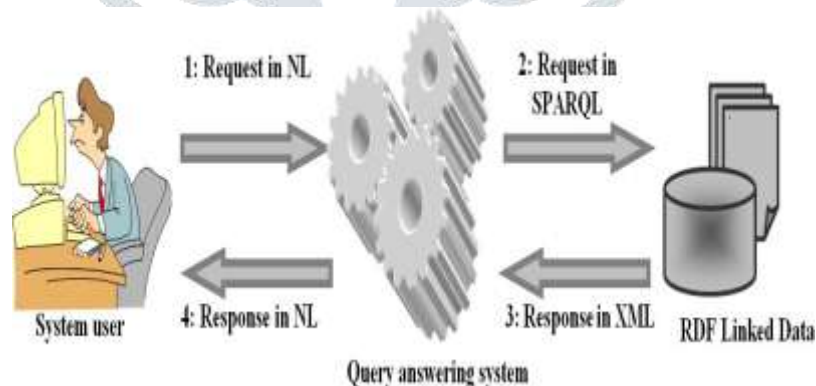


Fig 2: Interacting with QA System towards the Linked Data.

III. Approaches

a. Linguistic approach

QA system requires huge number of past analysts depended on artificial intelligence (AI) based strategies that integrate natural language processing (NLP) procedures and information base or corpus to manufacture QA rationales. The learning information is sorted out in the type of generation rules, rationales, outlines, templates (spoke to with triple relations), ontologies and

semantic systems, which are used during analysis of question-answer match. Linguistic strategies, for example, tokenization, POS tagging and parsing were actualized to client's question for formulating it into an exact inquiry that only concentrates the separate reaction from the organized database. Be that as it may, organization of a particular domain information base stances movability restriction as an alternate application domain requires distinctive punctuation and mapping rules. Moreover, building a suitable learning base is a tedious procedure, so these systems are for the most part connected to issues that have long haul information requirements.

Prior QA systems around 1960s, for example, [1] and [3] were simply natural language front-closes for organized database inquiry systems. The questions displayed to these systems were generally investigated using NLP procedures to deliver a sanctioned shape, which was then used to build a standard database inquiry. Exchange system viz., [2] and [4] likewise utilized organized database as the learning source. The key impediment of these systems is that the learning put away in the organized database was just fit for answering questions asked within the limited domain.

In any case, in late works, this impediment of the learning base is acknowledged as the capacity to give a circumstance particular answer. This combined approach enables clients to get to reaction of routine questions as well as of those questions that were unexpected at the season of system development. This particular element of QA system is accomplished through inference engine segment.

b. Statistical approach

In the current research scenario, fast development in accessible online text repositories and web information has increased the significance of statistical approaches. These approaches set forward such methods, which can't just manage the vast measure of information yet their heterogeneity too. Furthermore, statistical approaches are additionally independent of organized inquiry languages and can figure queries in natural language frame. These approaches fundamentally require a sufficient measure of information for exact statistical learning however once appropriately learned, deliver preferable outcomes over other competing approaches. Furthermore, the scholarly statistical program or technique can be effectively altered to another domain being independent of any language shape. Be that as it may, one of the significant disadvantages of statistical approaches is that they treat each term independently and neglect to recognize linguistic features for combination of words or expressions.

When all is said in done, statistical methods have been so far effectively connected to the diverse phases of a QA system. Support vector machine (SVM) classifiers, Bayesian classifiers, Maximum entropy models are a few procedures that have been utilized for question classification reason. These statistical measures break down questions for making forecast about clients' normal answer write. These models are trained on a corpus of questions or reports that has been commented on with the particular specified classes in the system.

Berger has investigated the possibilities of applying statistical strategies to answer finding errand in QA and found that these methods performed great depending on the attributes of the underlying information set— vocabulary measure, the cover amongst question and replies, and between different answers, and so forth. Statistical procedures, for example, N-gram mining, sentence likeness models and Okapi closeness estimation are connected to answer finding assignments in a QA system. These strategies examine question and record based on different comparability features so as to determine the closeness of candidate archives or replies as for question. The thought of answer approval could likewise be actualized incorporating statistical approaches through pertinence input system. Answer choice phase of this system depended on different heuristic separation measurements to look for an answer. Moschitti actualized comparability estimation display for calculating the closeness score amongst inquiry and reports or sentences from corresponding accumulations. The closeness demonstrate exhibited depended on a sentence comparability model to figure the similitude amongst question and reply. This model accounted on various features, for example, watchword likeness, length closeness, arrange similitude and separation comparability of the catchphrases utilized as a part of question and reply. A system created by Soricut et al.

c. Pattern matching approach

This approach utilizes the expressive energy of text patterns to supplant the refined processing involved in other competing approaches. For instance, the question "Where was Cricket World Cup 2012 held?" takes after the pattern "Where was <Event Name> held?" and its answer pattern will be indistinguishable "<Event Name> was held at

<Location>". Straightforwardness of such systems makes it very great for little and medium-measure sites, which can't bear the cost of complex arrangements that require much time and uncommon human aptitudes to install and maintain the system. The majority of the patterns matching QA systems utilize the surface text patterns while some of them likewise depend on templates for response generation.

d. Surface Pattern based

This approach extricates answers from the surface structure of the recovered reports by relying on a broad rundown of patterns. Reply to a question is recognized based on comparability between their reflecting patterns having certain semantics. These patterns resemble normal articulations. Despite the fact that designing such arrangement of patterns requires a considerable measure of human expertise and time yet the approach has demonstrated high exactness as well.

Initially, the surface pattern based strategy go for finding answers to verifiable questions, as their answer is constrained to maybe a couple sentences. With a specific end goal to plan an ideal arrangement of pattern, the majority of the current surface pattern based system utilized strategy portrayed by [13]. They executed an automatic learning technique which utilized bootstrapping to assembled a vast arrangement of patterns starting just with a couple of cases of QA combine from the web. Inspiration driving their work is surprising quality of such patterns proposed by [11] in the TREC-10 Question answering assessment track.

Another related idea based on surface patterns is proposed who increased surface patterns with 'support' and 'certainty' measures from information mining group. This system indicated high exactness however low review. [16] integrated surface patterns with named element tagger to sum up these patterns induced from free text. Some other QA systems have likewise taken after this approach to enhance their question answering instrument.

e. *Template based approach*

This approach makes utilization of preformatted patterns for questions. The focal point of this approach is more on illustration rather than interpretation of questions and replies. The set for templates is worked keeping in mind the end goal to contain the ideal number of templates ensuring that it enough cover the space of issue, and every one of its individuals speaks to an extensive variety of questions of their own kind. Templates have element spaces, which are missing components bound to the concept of the question that has to be filled to produce the inquiry template to recover the corresponding response from the database. The response returned by question would be raw data, which is come back to the client. System created by [15] also uses answer templates to represent the appropriate response in a designed way.

The fundamental principle took after by template based question answering system is much like the robotized FAQ (Frequently Asked Questions) answering system that responds with pre-put away responses to client question but not at all like static FAQs, the question templates are filled progressively with parameters. This system utilized pre-prepared text to distinguish best coordinated template-answer combine put away in database. Every one of such templates is defined to coordinate a wide range of variations of a similar question but not the question of same kind making it excessively constrained. While for Question Assistant [15], Sneider's planned templates with the goal that single template could cover an extensive variety of data instances important to the element opening and all questions of its compose. Substance spaces within a question template speak to concepts or elements contained in database and relationship between these concepts is spoken to by templates themselves. In any case, if new relationship has to be included, another template is required.

IV. Performance of Question Answering

In this, 2 types of QA systems are proposed : Ontology based QAS and Text based QAS:

Ontology based QA Systems

To demonstrate the execution of the ontology based QA Systems we took a gander at the assessment comes about did in the writing, outstandingly those compressed in the review paper [5]. At that point we set up the histogram of the accompanying figure.

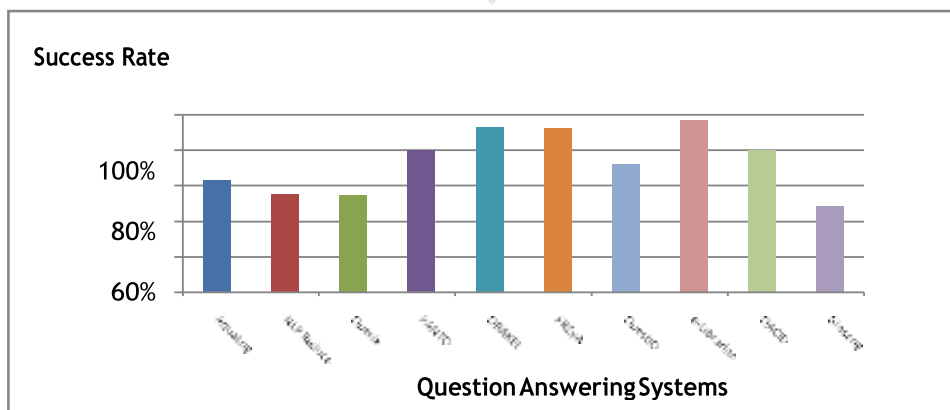


Fig 3: Performance results of the ontology-based Question Answering Systems.

Execution of the Ontology based QAS is spoken to by the achievement rate (adjust answers to questions) in the chart above. We found that the achievement rate of these QA systems fluctuates in the vicinity of 49% and 89%. These outcomes rely upon two criteria: (1) the algorithms and methods of natural language processing utilized by the system, and (2) the predetermined area to be questioned by the system.

Text based QA Systems

The QA4MRE centers around the perusing of single records and the recognizable proof of the right and NoA answers to an arrangement of questions, over the two years 2012 and 2013. NoA implies that the system chose not to answer the question.

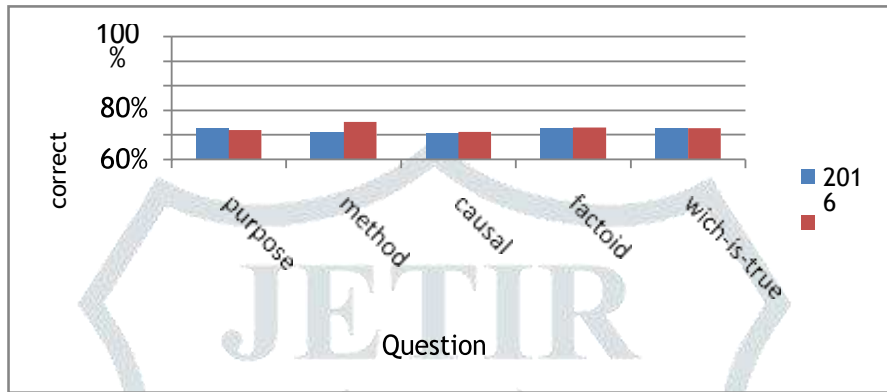


Fig 4: Percentage of different question types answered correctly in the versions of QA4MRE challenge.

V. Natural Language Processing

Computer science has dependably helped man in making his/her life less demanding. New time cruel history is Information Era. With help of web search engines, we can get any information readily available. We are only a tick far from getting to a page at remote corner of the world. Notwithstanding the splendid research, quicker computer processors and less expensive memory bolstered these incredible progressions. We have constantly needed computers to act canny. To achieve this undertaking the field of Artificial Intelligence appeared. One of the key hindrances in making computers shrewd is understanding of Natural Language. Natural language processing which manages understanding of languages is sub division of Artificial Intelligence. Question Answering is a great NLP application. Errand that a question answering framework acknowledges is given a question and gathering of reports, finds the correct response for the question. It has two complementary goals: first understand the different issues in natural language understanding and portrayal and the second to create natural language interface to computers.

We are dependably in a mission of Information. However there is contrast in information and knowledge. Information Retrieval or web search is develop and we can get applicable information readily available. Question Answering is a specific type of Information Retrieval which looks for knowledge. We are occupied with getting the pertinent pages as well as we are keen on finding particular solution to questions. Question Answering is in itself convergence of Natural Language Processing, Information Retrieval, Machine Learning, Knowledge Representation, Logic and Inference, Sematic Search. It gives a decent stage to dig into "nearly" all of AI. On the off chance that an announcement is made that "Question Answering is a definitive AI", the announcement will be univocally acknowledged. Question answering framework in its being is a craftsmanship, in the meantime it has science in its quintessence. Question Answering Systems are required all over, be it restorative science, learning frameworks for understudies, individual collaborators. It is need in each perspective where we require some help from computers. It's a given that it merits investigating the leaving field of question answering.

VI. Discussion

Distinctive approaches that have been so far examined in above section perform genuinely well for their domain of extension. QA systems relying on linguistic approach were essentially based upon a knowledge base for particular domain, which gives a productive and solid response for short answers. Answer extraction component from the knowledge base is supported by profound linguistic analysis to recognize the significant answer. In addition, building of a fitting knowledge base with handmade tenets requires a ton of human ability and time and stances compactness limitation. Some of these prior systems also depended on heuristic tenets to distinguish question class and connected shallow NLP methods, but accomplishment of this administer based component is so far confined to the systems having only text archives as their knowledge resource. In addition to it, construction of legitimate tenets required adequate measure of training data and time along with handy human exertion.

Statistical approach is destined to be helpful for huge amount of data having enough word for statistical comparisons to be considered noteworthy. The conspicuous decision of extensive data set for this approach is made to give the adequate measure of learning data while training statistical models. Be that as it may, once statistical models have been appropriately trained, these systems could effectively give the response of even complex questions.

All the elementary approaches talked about so far in this paper, perform genuinely well for their separate application zones but experience the ill effects of certain limitations when executed beyond that. This reality prompts the improvement of QA system with half and half the approach that would conquer the constraints as well as endeavor the potential, resulting from an individual approach. As of late, therefore, numerous systems were produced combining the capacities of individual elementary approaches. The vast majority of QA systems relying on half and half approach demonstrate diverse level of proficiency for various sort (acronym, definition, list, and so on.) of questions.

VII. Conclusion

Our exertion has been to take a comprehensive overview of the question answering research to address the difficulties because of information explosion in this information and communication technology era. We watched that the decision of a strategy is exceptionally issue particular. Often a hybrid approach, prudently blending clearly extraordinary methods, gives enhanced outcomes in the type of faster speed, increased relevancy, and higher precision and recall measures. Note that a standout amongst the most imperative highlights of QA Systems is their capacity to give correct answers, in light of the fact that distinctive sources are the objective of these systems. At that point, the client makes an inquiry utilizing a natural language without knowing the structure of the sources to be questioned. A few languages are better off than others, because of the development of research in the nations talking this language. So the exploration in natural language processing is primordial for creating Question Answering Systems for unstructured and organized data. The Arabic semantic Web is exceptionally a long way from the advancement of Arabic Question Answering System over semantic Web, which is our definitive objective to accomplish.

References

- [1] Green BF, Laughery K. Baseball, Wolf AK, Chomsky C : An automatic question answerer. In Proceedings of Western Computing Conference, Vol. 19, 1961, pp. 219–224.
- [2] Weizenbaum - a computer program for the study of natural language communication between man and machine. In Communications of the ACM, Vol. 9(1), 1966, pp. 36-45.
- [3] Woods. Progress in Understanding of Natural Language - An Application to Lunar Geology. In Proceedings of AFIPS Conference, Vol. 42, 1973, pp. 441–450.
- [4] Bobrow DG, Kaplan RM, Kay M, Norman DA, Thompson H, and Winograd T. Gus, a frame-driven dialog system. Artificial Intelligence, Vol. 8(2), 1977, pp. 155-173.
- [5] Clark P, Thompson J, and Porter B. A knowledge-based approach to question answering. In Proceedings of AAAI'99 Fall Symposium on Question-Answering Systems, 1999, pp. 43-51.
- [6] Soubbotin MM and Soubbotin SM. Patterns of Potential Answer Expressions as Clues to the Right Answer. In Proceeding of the TREC- 10, NIST, 2001, pp. 175-182.
- [7] Susu Jiang and Zhu Yonghua, "The Design for Semantic Based Intelligent Q&A System", First International Workshop on Education Technology and Computer Science, IEEE,2009, pp.90-92.
- [8] Min-kyoung kim and Han-joon Kim, "Design of question Answering System a with automated Question Generation", IEEE, 2008, pp.365-369.
- [9] Boris Katz and Jimmy Lin "REXTOR: A System for Generating Relations from Natural Language",Proceeding of ACL workshop on NLP & IR 2000.
- [10] Rich and Kelvin Knight, "Artificial Intelligence", Mcgraw hill companies, chapter 15, pp.377-426.
- [11] Yasunori Ishihara, Hiroyuki Seki and Tadao Kasami "A Translation Method from Natural Language Specifications into Formal Specifications Using Contextual Dependencies",
- [12] J. A. Goguen, J. W. Thatcher, and E. G. Wagner, "An initial algebra approach to the specification, correctness and implementation of abstract data types," IBM Research Report, 1976
- [13] Paeseler, H. Ney: "Continuous Speech Recognition Using a Stochastic Language Model" IEEE. Int. Conf. on Acoustics, Speech and Signal Processing, Glasgow, pp. 719-722, May 1989

- [14] H. Seki, et al., “A Processing System for Program Specifications in a Natural Language,” Proc. 21th HICSS, pp. 754–763, Jan. 1988.
- [15] Hang Cui, Renxu Sun, Keya Li, Min-Yen Kan, and Tat-Seng Chua. “Question Answering passage retrieval using dependency relations”. In Proceedings of the 28th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, pages 400–407, Salvador, Brazil, 2005.
- [16] E. Sneider, “Automated Question Answering: Template-Based Approach”, PhD thesis, Stockholm University / KTH press, Sweden, 2002.

