International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-3, January 2020

A Comprehensive Retrospection of Literature Reported Works of Community Question Answering Systems

Venkateswara Rao. P, A.P Siva Kumar

Abstract: ***Community question answering CQA) systems are rapidly gaining attention in the society. Several researchers have actively engaged in improving the theories associated with question answering (QA) systems. This paper reviews the literature reported works on question answering QA systems. In this paper, we discuss on the early contributions on QA systems along with their present and future scope. We have categorized the literature reported works into 20 subgroups according to their significance and relevance. The works in each group will be brought out along with their inter-relevance. Finding the question and answer quality is the prime challenge almost addressed by many researchers. Modeling similar questions, identifying experts in prior and understanding seeker satisfaction also considered as potential challenges. Researchers at the most have done experimentations on popular CQAs like Yahoo! Answers, Wiki Answers, Baidu Knows, Brianly, Quora, Pubmed and Stack Overflow respectively. Machine learning, probabilistic modeling, deep learning and hybrid approach of solving show profound significance in addressing various challenges encounter with QA systems. Today the paradigm of CQA systems took the shift by serving as Open Educational Resources to learning community.***

Keywords: ***Community Question Answering Systems (CQAs), Question Answer Retrieval, Question Analysis.***

I. INTRODUCTION

Question answering (**QA**) system is concerned in giving automated answers for the questions posed by humans. Human posed questions predominantly takes the form of natural language statements. Thus the concept of Question answering (QA) takes the support of natural language processing and information retrieval. QA system could be perceived as a software application which tries to structure the unstructured collection of natural language information and attempts in giving appropriate answer for the posed question. Unstructured collection of information could include Wikipedia pages, compiled news articles, collection of reference textual documents, and a subset of voluminous collection of web pages available over Internet. During the attempt of answering, QA deals with wide variety of questions in finding how, why, actual facts and semantics. Sometimes QA systems might encounter with hypothetical and cross-lingual queries too.

QA systems broadly classified into two categories namely closed-domain and open-domain. Closed domain is related to a specific domain (for ex: Sports, medicine, cinema etc.) which takes less effort compared to open-domain for processing the answers because such a system can fetch domain-specific knowledge through frequently formalized ontologies. Also, such systems at the most accept limited questions of descriptive nature of instead of procedural information. On the other hand open-domain QA systems works with world knowledge allows posing questions in any domain. These systems are quite complex because of their wide search through numerous general ontologies and world knowledge.

II. EARLY QA SYSTEMS

Joseph Weizenbaum [1] developed the early natural language processing computer program named ELIZA during mid 1960’s at the MIT Artificial Intelligence Laboratory. Weizenbaum The primary focus of the creation of ELIZA is the demonstration of communication’s superficiality between humans and machines. It was the first chatter bot which attempted the Turing Test The system cleverly created an illusion to the users and simulated conversation with users is achieved by pattern matching and the methodology of substitution. MAD Slip and DOCTOR scripting systems allowed ELIZA in processing user requests. ELIZA has created a strong impact in its user community and thus several academicians believed that the program assists the doctors to work for the treatment of patient [2]. The prime limitation of ELIZA is that it could not converse with true understanding. Even though ELIZA stored data from punched cards. BASEBALL begins its work by searching for words and idioms in the database dictionary then takes further search for phrase structure and syntactic needs to be refined, but still it is proved to be successful for its intelligence and understanding.

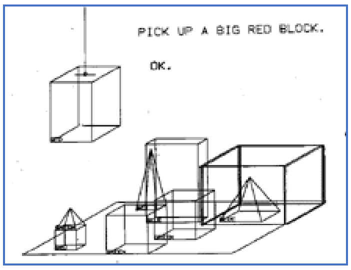
One of the early QA system is BASEBALL [3] conceived by Fredrick, Oliver and Gerald developed at Stanford University during the year 1961. BASEBALL is a computer program tries to answer in simple English with reference to the facts to retrieve the information requested. This QA system belongs to closed domain category and caters the information pertaining to the baseball games.

Revised Manuscript Received on January 30, 2020.
* Correspondence Author
Mr. Venkateswara Rao P*, Assistant Professor, Department of CSE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India.
Dr. A. P. Sivakumar, Assistant Professor, Department of CSE, JNTU Anantapur, Andhra Pradesh, India.

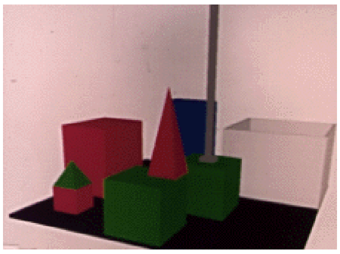
LUNAR [4] is the next significant research prototype for answering the questions related to geological analysis of rocks returned by the Apollo moon missions. The system was first publicly demonstrated at a lunar science convention in 1971 and the system could able to answer 90% of the questions in its domain posed by people untrained on the system. The LUNAR system initially performs syntactic analysis using heuristic information and then tries to determine the meaning of the posed question and finally produces the required output.

The language parser of LUNAR followed the general parsing algorithm which actually been developed at Harvard University. LUNAR QA system is developed with the grammatical rules for a large subset of English. It also consists of a set of semantic interpretation rules for interpreting requests to perform chemical analyses and ratios. A dictionary of 3500 words and with user helpful functions are integrated with LUNAR for interrogating the database. LUNAR system is quite fluent in a few specific constructions and could able to recognize a large number of variations on the request. However, the prototype LUNAR needs to be greatly enhanced for its linguistic processing, semantic representations and inference mechanisms.

In late 60s and early 70s, Terry Wingard at MIT developed SHRDLU [5], an early natural processing understanding program which gained prominent recognition as a question-answering system. Micro Planner and LISP programming are used to write programs. The program simulates the operation of a robot to attend the Blocks World problem. During the simulation the program make a conversation with the humans trying to describe the state of the world. Figure 2.1a depicts the original screen display of Blocks World with SHRDLU. Its corresponding color rendering is shown in figure 2.2b which is colored by the University of UTAH.



a) Original Screen Display



b) color rendering

Fig 2.1: SHRDLU Blocks World Geometric Representation

The biggest strength of this program is that the rules of physics are quite positively encoded in a computer program. A famous demonstration of SHRDLU is shown in the following figures (as follows:

Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM Maintaining
Computer: OK.
Person: What does the box contain?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
Person: What is the pyramid supported by?
Computer: THE BOX.
Person: How many blocks are not in the box?
Computer: FOUR OF THEM.
Person: Is at least one of them narrower than the one which i told you to pick up?
Computer: YES, THE RED CUBE.
Person: Is it supported?
Computer: YES, BY THE TABLE.
Person: Can the table pick up blocks?
Computer: NO.
Person: Can a pyramid be supported by a block?
Computer: YES.
Person: Can a pyramid support a pyramid?
Computer: I DON'T KNOW.
Person: Stack up two pyramids.
Computer: I CAN'T.
Person: Will you please stack up both of the red blocks and either a green cube or a pyramid?
Computer: OK.
Person: Which cube is sitting on the table?
Computer: THE LARGE GREEN ONE WHICH SUPPORTS THE RED PYRAMID.
Person: Is there a large block behind a pyramid?
Computer: YES, THREE OF THEM: A LARGE RED ONE, A LARGE GREEN CUBE, AND THE BLUE ONE.
Person: Put a small one onto the green cube which supports a pyramid.
Computer: OK.
Person: Put the littlest pyramid on top of it.
Computer: OK.
Person: Does the shortest thing the tallest pyramid's support supports support anything g reen?
Computer: YES, THE GREEN PYRAMID.
Person: What color is it?

Computer: BY "IT", I ASSUME YOU MEAN THE SHORTEST THING THE TALLEST WILL SUPPORT THE PYRAMID SUPPORT.

Computer: RED

Person: How many things are on top of green cubes?

Computer: I'M NOT SURE WHAT YOU MEAN BY "ON TOP OF "IN THE PHRASE "ON TOP OF GREEN BLOCKS".

DO YOU MEAN:

1 - DIRECTLY ON THE SURFACE

2 - ANWHERE ON TOP OF?

Person: 2

Computer: THREE OF THEM.

Person: Had you touched any pyramid before you put the green one on the little cube?

Computer: YES, THE GREEN ONE.

Person: When did you pick it up?

Computer: WHILE I WAS STACKING UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.

Person: Why?

Computer: TO EXTINGUISH IT.

Person: Why did you do that?

Computer: TO CLEAN OFF THE RED CUBE.

Person: Why did you clear off that cube?

Computer: TO PUT IT ON A LARGE GREEN CUBE.

Person: Why did you do that?

Computer: TO STACK UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.

Person: Why did you do that?

Computer: BECAUSE YOU ASKED ME TO.

Person: How did you do it?

Computer: BY PUTTING A LARGE RED BLOCK ON THE TABLE; THEN LETTING GO OF IT; THEN PLACING A LARGE GREEN BLOCK ON IT BY LETTING IT MOVE THE BLOCK AWAY PLACING IT RED CUBE ON THAT CUBE BY LETTING IT MOVE THE BLOCK AWAY

Person: How many objects did you touch while you were doing it?

Computer: FOUR OF THEM.

Person: What did the red cube support before you started to clean it off?

Computer: THE GREEN PYRAMID.

Person: There were five blocks to the left of the box then.

Computer: NO, ONLY FOUR OF THEM: THE RED CUBE, TWO LARGE GREEN CUBES AND A LARGE RED CUBE.

Person: Put the blue pyramid on the block in the box.

Computer: OK.

Person: Is there anything which is bigger than every pyramid but is not as wide as the thing that supports it?

Computer: YES, THE BLUE BLOCK.

Person: Does a steeple

Computer (interrupting): SORRY, I DON'T KNOW THE WORD "STEEPLE".

Person: A "steeple" is a stack which contains two green cubes and a pyramid.

Computer: I UNDERSTAND.

Person: Are there any steeples now?

Computer: NO.

Person: Build one.

Computer: OK.

Person: Call the biggest block "superblock".

Computer: OK.

Person: Have you picked superblock up since we began?

Computer: YES.

Person: Why did you drop it?

Computer: BY "IT", I ASSUME YOU MEAN THE SUPERBLOCK.

I CAN'T EXPLAIN A NON-EXISTENT EVENT.

Person: Is there anything to the right of the red pyramid?

Computer: YES, FOUR THINGS: THE BOX, THE RED CUBE, THE BLUE PYRAMID AND THE BLOCK NAMED SUPER BLOCK

Person: Thank You.

Computer: YOU ARE WELCOME!!

Fig 2.2 A demonstration of SHRUDLU algorithm

Terry Winograd atmost has succeeded in attempting Turing test by his SHRDLU. There are several tricky questions posed by humans, still, SHRDLU didn't turn off.

During 1970's the development of knowledge bases resulted in building expert systems .To produce quotable valid responses to the questions confined in the area of knowledge. These expert systems emulate the ability of a human while making decision based on its knowledge. Edward Feigenbaum is the early researcher who developed expert system DENDRAL [6] at Stanford University. Bruce G. Buchanan, Joshua Lederberg, and Carl Djerassi, are the supporting key contributors to the DENDRAL project. Expert systems closely resemble modern QA system. However, the internal architecture of expert systems and modern QA Systems has a significant variation. Expert system relies on its organized knowledge base whereas modern QA system works on unstructured natural language.

Comprehensive theoretical research in natural language processing started rise during 1970s and 1980s. Unix Consultant (UC) [7] is a pioneer project developed by Rober Wilensky during this period. Unix Consultant is a QA system gives valid responses to the user who wish to know about Unix Operating System. UC has associated with comprehensive hand-crafted knowledge base pertaining to its domain. Another significant QA system developed during this period is LILOG which is aimed to serve for German Tourism. However, these two systems doesn't reach the society, still, they laid significant path in the computational linguistics and reasoning.

During the early years of millennium, QA systems are improved to associate a question classifier module that finds the type of question and the type of answer. The multi-agent question-answering architecture is developed during this period, In which every domain represents an agent. Such an agent attempts to answer the posted questions with the support of its domain knowledge. In parallel, a meta–agent for every agent monitors the cooperation among the question answering agents fetches the most relevant answers. During the rapid growth of QA systems one of the significant observation found is that the burden of a QA system could be minimized if the right information appears in many forms. This practice also minimizes false positives to most extent. Numerous QA systems at the most rely on automated reasoning and are predominantly been developed in Prolog and Lisp programming.

III. HOW QA SYSTEM WORKS?

QA system aims in giving an appropriate answer to the user posed question. Most of the cases, the retorted answer is provided in short texts rather than the relevant documents. In conventional scenario, QA system receives the input as natural language question in place of set of keywords. To illustrate, "When is the republic day of India?" Input natural language input is transformed into query through its logical form. QA system is made user friendly with the help of the idea of natural language question which results in complicated processing to get sensible answer. In an open domain, tagging a posed question to its appropriate category is a big deal; since the process of answer extraction depends upon determining the correct question type which deduces the correct answer type.

As a part of finding the question type, keyword extraction is the initial step to be performed. In certain cases these keywords are more specific and could indicate the type of the question directly. For instance, words like "Who", "Where" "How many", indicates QA system to return the answer of "Person", "Location", "Number" type respectively. For instance, referring to the question in previous paragraph, the QA system should respond with a date. Tagging based on parts of speech (POS) and applying syntactic parsing analysis will determines the answer type. For instance "Indian Republic day" is a subject ,"is" the predicate, "when" the adverbial modifier and "Date" the answer type. Haplessly, interrogative words like "Which", "What" "How" show ambiguity in giving clear answers because of their representation. In the posed questions the other words should be taken into account to deal with these situations. The focus should be to identify the word or words which exactly indicate the meaning of the question.

Numerous lexical dictionaries are available in the literature and the processor can take the support of any of these dictionaries for better understanding of discourse.

Once the identification of the question type is done, helps to find the document sets with key words will be found with the help of Information retrieval system. Group tagging could be used in order to verify the correct entities and relations during the retrieval of the documents. Also, an entity recognizer could help in finding the relevant "Person" or "Location" from the obtained documents. Ranking could be applied on the relevant textual paragraphs for getting accurate answers. Building a vector space model is one of the best strategies to classify candidate answers. This model helps in determining the correct answer type with reference the given question type. Validation of the correct candidate answers could be done by applying the inference technique. Every candidate answer is then assigned with a score which is determined based on the contained question words and how close those are to the candidate answer. The obtained answer is then transformed into meaningful representation for user understanding with the support of parsing. Considering the previous question, the best answer could be "26th February".

IV. QA SYSTEMS PRESENT AND THEIR FUTURE SCOPE

To serve additional domains of knowledge in recent years the QA systems have been improving rapidly[8]. At present, QA systems are able to cater automatic answers for temporal & geospatial questions, terminologies & definitions, biographical questions and content-based questions. Current QA research topics include:

- automatic reasoning
- response caching
- answer visualization
- interactiveness
- subject roles
- emotion analysis
- semantic analysis
- Use of linguistic resources and tools etc.

Modern information retrieval research focus in developing a well established methodology based on exhaustive laboratory experiments. Drawing a concrete conclusion is only possible when reliable and trustworthy retrieval methods using test collections are established. Text Retrieval Conference (TREC) is actively engaged in conducting workshops focusing various information retrieval research areas. TREC is associated with a question answering track since 1999; in each track the focus is defined in such a way that the systems are to fetch small snippets of text that contains an answer for open-domain, closed-class questions. TREC QA pools top n documents from each run, removes redundancies and hands over these filtered documents to human assessors for further evaluation. Jimmy Lin [9] has found that the literature available studies pertaining to evaluate QA systems are not similarly analyzed. He presented a meta-evaluation of the quality of answer patterns and lists which are employed in question answering evaluation. Based on this study he concluded that the underlying answer patterns and list of relevant documents are not suitable for post-hoc experimentation. Jimmy Lin has presented an evaluation of available resources which are intended to evaluate factoid question answering. Author has alerted the researchers that one must be cautious when employing existing resources to evaluate QA Systems. Author also suggested that the researchers should not make quantitative comparisons to find the effectiveness of different techniques.

Sanda et al. [10] presented a novel framework for answering complex questions that takes the support of question decomposition technique. The decomposition procedure works on Markov chain which subsequently follows a random walk on the bipartite graph of relations created between concepts of a complex question and sub-questions obtained from relevant passages which exhibit these relations. The decomposed question and answers identified during random walk are subjected as input to a clever QA system and thus derives a passages set. Later a Multi-Document Summarization (MDS) system is used to merge these passages to get complete answers. Authors by their decomposition and random walk technique proved that the relevance of summary length answers to complex questions is drastically enhanced.

4.1 Probabilistic Models for QA Systems

Ko et al. [11] applied a graphical model combined with probabilistic approach to assign ranking in QA system. This proposed model predetermines the collective probability considering all the potential answer entities which in turn estimates the correctness of every answer and also derives the correlations between answers. This process thus creates a base of most relevant answers. The proposed model is validated with a logistic regression model which could predict correctness of every potential answer. The major advantage of using such a combined model is that it gives the power of probabilistic inference. Authors pursued their experimental work with a set of 1818 questions chosen from the TREC8-12 QA evaluations which were considered as the test-bed, and along with a cross validation comprised of 5-fold was used to study the proposed model. Since the combined prediction model follows a graphical representation, thus its time complexity is denoted as $O(2^n)$ where n indicates the number of answer candidates. Thus authors validated their work only with top 10 candidate answers. Authors have demonstrated the improved efficiency of combined approach for answer ranking. Authors combined the proposed model with the logistic regression method and found a positive improvement in determining accurate answer ranking.

QA systems rapidly grow with new questions posted daily and this situation may be worsening the interest of users in picking out their choice of questions. Also, such rapid growth may result in unnecessary delay during answering of the newly posted questions. Hence, locating interesting questions can be treated as a potential problem of research. The solution for this kind of hurdle is to understand users' actual interest.

Development of techniques which recommend appropriate questions enable users to work with interesting questions. Mingcheng Qu et al. [20] adopted an integrated approach that performs semantic analysis with probabilistic inference. This model is termed as Probabilistic Latent Semantic Analysis (PLSA) designed to fetch out question recommendations. The work also proposed a novel metric to analyze the performance of PLSA. For a given question collection, Mingcheng Qu et al. have formulated an equation presented in 2.1, to represent the distribution of users with their attempts in answering questions as:

$$\Pr(T, x) = \sum_{z} \Pr(T|m)\Pr(x|m)\Pr(m) \quad \text{……(2.1)}$$

where $T \in t_1, t_2, ..., t_n$ denotes users, $x \in x_1, x_2, ..., x_m$ represents questions and $m \in m_1, m_2, ..., m_k$ represent k topic models, each associated with one topic T.

Authors refined this equation to cope with sparsity and thus made the user-word aspect model and devised the following equation as,

$$\Pr(T, x) = \sum_{z} \Pr(T|z)\Pr(A|z)\Pr(z) \quad \text{……(2.2)}$$

where $A \in a_1, a_2, ..., a_l$ are the words comprised in a question.

The distribution is then subjected to compute the local maximum in order to bring out the probability of the question base. Authors then computed the probability of word occurrence using Bayesian law as,

$$\Pr(T, x) = \left(\prod_{i} \Pr(T, wi)\right)^{1/|x|} \quad \text{…… (2.3)}$$

The developed question recommendation is validated with the help of newly devised accuracy metric as,

$$\text{Accuracy} = \frac{|P| - P_b - 1}{|P| - 1} \quad \text{……. (2.4)}$$

The quality of recommendation (P) has been experimentally improved and authors have opened a new direction to make question recommendation.

QA system predominantly works passively in which an asker puts a question and an answerer may give proper response at some other time. Once a question is posted, it will be put forwarded to all the users in the system. Sometimes answerers might not be interested to seek response from selected experts. Also, answerers may need to visit several questions while giving accurate response. Liu et al [30] have worked in eliminating such limitations and proposed a probability supported framework to find out the best answer for the posed questions. The proposed frame work keeps a track on the answerers' answering history and prepares a model with the support of Latent Dirchlet Allocation model. The work is experimented with both user activity and authority information and found that the proposed methodology can effectively work on newly posted questions for their best answerers.

A conventional QA system might be assigned to a particular domain. Still, if it caters only to a specific language then the development of such a system is worthless. Development of multilingual QA Systems is quite a worth. Ko et al. [11] have earlier developed a probabilistic graphical model to assign ranking for a Question-Answering System. An extension to this work, Ko et al. [32] have presented two probabilistic models for assigning ranks to the answers in a multi-lingual natural language QA system. Authors identified limited prior research using probabilistic models in traditional monolingual answer-ranking and he also found that only formal methods are employed to perform answering ranking in multilingual QA system. The work presents a probabilistic model which determines the probabilities of relevance for each answer individually. The work then progresses in devising another novel probabilistic method which predicts the relevance of answers individually as well as their mutual relations. The proposed framework has laid an initial leap in building a novel combination of probabilistic methods which positively drives in designing a flexible and scalable multilingual QA system. The work successfully addressed the criticism that this joint probabilistic approach is not thorough enough by conducting extensive set of experiments subjected to cross lingual translation of QA from English to Chinese and Japanese. The work also attempted in addressing monolingual QA of English, Chinese and Japanese with the support of TREC and NTCIR question base.

Major contribution to the Community Question Answering (CQA) system is given by a set of highly active users which usually exists in small in size, normally termed as experts who share high-quality useful answers. Identifying potential users and giving a concrete recognition to each expert is a good practice to make experts retain with the QA system. Several techniques are available to identify and user for locating his expertise. Early identification of an expert during initial participation will nurture and retain him in the system. Aditya Pal et al. [38] have addressed two problems in related to identify an expert. The first problem is identifying the currently associated experts in the QA system. The second problem is to identify those users who can become experts in near soon. Authors have presented a probabilistic model which captures the user's question selection properties according to their choice of answering. The proposed probabilistic model takes the support of machine learning methods to identify potential experts. The work considerably identified potential experts from normal users on the basis of selection preferences; allowing identification of potential experts with high accuracy when compared to the other standard models. Authors opined that the selection preferences may further integrated with standard baselines footsteps to retrieve improved predictive performance.

4.2 Ranking Frameworks for QA Systems

Open QA systems are of great interest to user community because of their voluminous source of knowledge base. Still a clever filtration is always to be done during fetching a right answer for a posed question because such a system is composed of at the most invaluable information.

Jiang Bian et al. [14] found that the majority of the content in an open QA system reflects individual specific, often unnecessary opinions. They inferred that a ranking that equally considers both quality and relevance is quite needed in developing factual QA system. Such a task poses a great challenge, since, the architecture design and the content base of open QA archives varies predominantly with the web dissemination setting. Jiang Bian et al. [14] have addressed this challenge and presented a universal ranking framework for retrieving expected facts from open online social media. Through experimental results, authors proved that their proposed methodology is quite effective in querying well-formed facts to seeker questions compared to a baseline factoid QA system. The proposed learning framework has the flexibility to fine tune with minimum effort of manual labeling. Authors provided the detailed analysis of their results to highlight those features which have high impact in searching over social media. The proposed system could be a decisive building unit to integrate responses obtained from heterogeneous social media systems with natural searching capabilities.

Public available QA systems at the most contain real-world facts which seem to represent as simple structure but may be complicated in matching with the base of facts because of their varying linguistics. The answer for the question “Who is the CEO of Google?” on publicly available knowledge base needs a match related to “leadership” entity comprised of three relations namely designation, company and person respectively. The search also should proceed in looking two relative entities namely “Google” and “Managing Director”. A surge of literature is found recently which present learning-based solutions for such a kind of problems. Hannah and Elmar further advanced the off the late methods by following ranking through learning methodology and also by attempting a problem of recognizing the entity, which has been ignored in the previous works. Authors coined their system as “Aqqu” and compared its functionality with two contemporary benchmark works, Free917 and WebQuestions respectively. These two standard benchmarks unleashed different challenges during evaluation. Aqqu outperformed considerably when compared with the previous best result obtained by each benchmark. Authors also considered the efficiency aspects which could be taken of answering all questions interactively.

Giovanni et al [76] carried a study to identify the significance of various kinds of models and features for determining question ranking in community QA system. These models include bag-of-words (BoW) and syntactic tree kernels (TKs) respectively. Authors made a note that structural kernels are never taken into account for pursuing question re-ranking task which requires modeling of paraphrase relations to locate question to question similarity. Online forums at the most comprised of text in informal form giving an additional challenge to use of tree kernels. The proposed learning to rank (L2R) algorithms is compared with one of the potential baseline work designed for Google rank (GR). Based on the experimental results it shown that the authors built shallow structures for using in tree kernels are quiet to noisy data and found that the existing Google Rank could be improved using the Bag-of-Words features and tree kernels.

Piero Molino et al after thorough examination of existing literature works on developing community question answering sites found that the there are several factors which show impact in determining accurate question-answer matching, but still, these factors are complex to deal with. Authors paradigm made a shift in perceiving the approach from application-oriented to a space of various dimensions which include distinctive attributes, footsteps and contents. Authors used machine learning based ranking framework which is most promising to date and included 225 features of five families in this rank framework. The work is tested over the largest dataset pertaining to Yahoo! Answers (40 Million Questions) to estimate the power of predicting the best answer with chosen 225 features. Authors proposed a new kind semantics measure of distributional nature that can flexibly replace most accepted linguistic similarity features to improve computational magnitude and thus provide greater prediction power. The work attained an improvement between 11% and 26% in predictive power of learning (P@1) which is quite promising than contemporary baseline methods.

4.3 Comparative Studies on QA Systems

Maxwell Harper et al. [15] carried a comparative study to investigate quality predictors by collecting a variety of user responses published on various available online QA systems. During the research authors has pointed two main questions namely,

- How do QA sites differ in functionality with respect to quality?
- What is the approach of seekers should follow to obtain most relevant answers from a QA system?

Authors have presented two high-level messages based on the question topic and their relationships. These messages are associated in demonstrating quantitative analysis with reference to the significance of factors. Authors observed that answer quality predominantly much more in Google Answers (a commercial QA system) compared to the open sites, also spending extra cost for a response will result in getting better output. Authors also found that the key to the success of QA system would be the maximum extent of user contribution. For Instance, Yahoo! Answers is an outperformed Q&A site because of it’s at the most provision given to users for library reference services.

4.4 Literature reported QA works focusing on Seeker Satisfaction

Yandong Liu et al. [16] experimented on Yahoo! Answers data in determining seeker satisfaction. The proposed research measures the seeker satisfaction based on the responses contributed by the user community. Authors designed an estimation model and developed a structure to support with diversified content. Authors employed baseline algorithms like support vector machines (SVM) and decision trees to extract features from the data. The work then proceeds with analysis of experimental results which are fetched from evaluation of numerous user ratings and real questions. During analysis it is identified that the proposed system could deliver better performance without having prior knowledge on seeker history.

Authors complemented their results with a thorough walkthrough over answer seeking patterns in QA systems correlated to asker satisfaction. The developed models and predictions are useful in applications like rank labeling, query modeling and customized user interface. The proposed work ensures a promising direction of building practical environments with improved quality for answer serving communities.

QA systems may or may not produce required response within expected timelines. Sometimes they may take several hours to several days for bringing out satisfactory response. If answer seeker waits for beyond his patience then it results in deterioration of credibility on the QA system. Agichtein et al. [19] pursued a study to address this challenge by estimating the seeker satisfaction for his posted question very prior to the responses submitted by the answering community. Authors devised a generic structured prediction model included with organized content and chosen features for community collaboration. This model is termed as WEKA [49] framework possessing both decision making and classification capabilities. For every satisfied class of question answers, authors computed precision, recall and f1 metrics. The work is evaluated experimentally using huge collection of real-time questions and their associated user ratings. The work successfully unleashed the generic methodology of finding out seeker satisfaction. Authors also worked in exploring customized models of seeker satisfaction, and demonstrated that if adequate user interaction history exists then the proposed model may work as a universal model for prediction.

The paradigm shift is observed in QA system where conventional textual based questions are slowly been replaced with visual question-answers. Similar to conventional QA systems, visual question answering systems provide users to make a query on an image and can expect a valid answer from users. The biggest challenge is ambiguousness that a visual question may lead to several answers with different contexts. This results in a mixture of agreements and disagreements pertaining to its subject. Danna and Kirsten[77] proposed a model termed as CrowdVerge for automatically predicting an answer's user agreement from a visual question. The work then efficiently collects all valid answers of visual questions and draws a prediction. Authors with an assertion that positive answer agreement will imply less human responses compared to negative answer agreement. The work is experimented on 1,21,811 visual questions queried by visually challenged users, and show that the proposed system fetches out diversity for the same answer with reduced user contribution.

4.5 Studies on QA works focusing on User Participation Behavior

Kevin et al. [21] attempted to analyze the South Korea's leading QA system Naver Knowledge – iN for understanding the process of knowledge creation and accumulation as well as human interaction within the system. A big of bunch of 2.6 million Question-Answer pairs pertaining to 15 categories posted during the years 2002 and 2007 are collected by the authors. This data is thoroughly analyzed and identified 26 experts to capture their patterns of interactions connected with their back ground motivations at different contexts. Authors found self learning, well being of others, and showing capability are the key underlying intentions of top contributors, still, these contributors are found to be irregular in their participation. In the literature, Guru score is identified as a significant measure to assess user performance while answering to a question. Authors employed Guru Score combined with a user performance model to estimate the user's participation. Authors revealed that the more participation of a user results in improved performance of system and thus acts as a motivation factor to other users.

The primary aim of QA system might give a proper answer to the answer seeker; however, after certain refinements to the best answer whose final product could be durable, might help a lot to a broad group of users. Such a shift of enhancing the answer to cater community-driven is positively accepted. To make into end product it needs well versed hands on experience and several community QA systems follow voting mechanism as their best practice to serve users in determining the genuineness and relevance of the response. Ashton Anderson et al [37] studied QA systems to understand this shift in focus as community knowledge-base generation process. Authors have considered a question and its every possible answer associated as a basic unit of analysis, instead of focusing to a individual question-answer pair. The outcome of the investigation is quite interesting because of the constant change in the users' participation over the period of time in both answering and voting. For example, authors observed significant involvement of the reputations of co-answerers, their attentiveness and the likelihood of answer selection, but still the most satisfied answer completely relies on the dynamics of answer appearances. Authors then presented the abstraction of those characteristics which are generally been appropriate predicting various significant qualities like durability of a question, its corresponding responses along with the need of the giving a better answer. The implications of the results with the proposed design in QA sites is also been presented in the work.

4.6 Use of reasoning to detect right answer in QA systems

Xudong Tu et al. [22][23] presented a comparative reasoning study to reduce the semantics gaps among similar question pairs and to subsequently to detect suitable answers in community question-answering sites. With the support of baseline methodology presented in Bayesian Analogical Reasoning (BAR) framework [24], authors transformed the posted question answer pairs into relational data, also to assign ranks for answers by their relevance and accuracy to the query. The relevance and accuracy of an answer is determined with the help of analogical reasoning. The answer which relates much homologous to the newly posted question can get higher score and thus be treated as the best answer. Authors experimented their proposed methodology with nearly 30 million questions from Yahoo! Answers QA base. The proposed approach is compared with three contemporary popular methods. The work found to be out performed when compared with existing baseline methods.

Numerous questions expect answers of a particular kind in open QA System, for example "who is the successive president of Pranab Mukherjee?" The answer lies within the instances of Indian President and could never go beyond. Grappy and Grau [28] attempted in developing a method to verify an answer for its relativeness with posted question.

To pursue this method authors have combined criteria obtained from various literature available methods. Statistical methods, entity recognizers and support of Wikipedia are three major ways that authors have employed to verify the appropriateness between answer and its kind. Such a methodology can be perceived as a learning paradigm that makes use of different features. The work can be used to improve a QA system by checking all returned answers. The work alone can't be used to locate good answer for a posed question. Authors in their earlier work [29] have presented a validity method through learning and this proposed work is an extension to make the validation system a complete one. The proposed work also finds its place in an answer validation module which decomposes the given question into various kinds of information for checking.

4.7 Multimedia QA systems

The rapid increase of information over the Internet in different forms creates a new challenge in developing QA system which supports to cater multimedia information. For the past 2 decades QA systems are at the most responded in text-based manner. Yeh et al. [26] have done promising work in designing visual QA system. Earlier Yeh et al. in [25] have presented a structure for designing visual QA system which comprised of 3 layers. The initial layer looks for an appropriate match for the query picture with the help of modeled textual questions and domain specific keywords. Middle layer strive to search in its internal base for matching response answer. Third layer takes the support of manual expertise if needed. Tat-Seng Chua et al. [26] are the early researchers who extended the fact-based QA systems towards designing a common framework for developing multimedia QA system. The proposed framework is aimed in locating multimedia answers over Internet public content such as Flicker and YouTube. The work can be considered as a preliminary and could be extended to several directions.

4.8 Applications of Prediction in QA Systems

In a community based QA system (CQA), users post their question and wait for appropriate answers from other users. Till the year 2012, no researcher has attempted in addressing the unanswered questions in a QA system. Lichun Yang et al [33] are the early researchers who attempted in analyzing not-answered questions. Authors presented a formal prediction model to analyze the not-answered questions which is comprised of a supervised learning task along with the support of selected question features. The work is positively evaluated with real questions collected from Yahoo! Answers.

Fetching the most relevant answer is the prime task of a conventional QA system. But this problem is categorized as a complex task because of matching numerous relevant responses. Several researchers have contributed their efforts in laying a path of extracting the right answer but very few have addressed the treatment of voluminous question-answer data prior to perform actual analysis. Baichuan Li et al [35] presented two research analyses to understand the scope of the question quality issue as a process of preliminary screening and thus picking and ruling out the bad quality questions. The initial study is focused on identifying question quality factors and determines the user participation as a quality measure between seekers and relevant topics. This study gives the difference of the quality of the question according to the topic. The second study is comprised of a mutual reinforcement-based label propagation (MRLP) model which predicts the quality of question with support of question relevancy and seeker participation. Authors successfully separated best quality questions with the poor quality questions but still opined that since the extracted features are no salient, the proposed methodology may not yield satisfactory performance.

Answer seekers may get frustrated if their questions remain unanswered in QA systems. Gideon Dror et al [57] are the initial contributors to estimate whether the posted does really get answered or not. In order to attain this objective, authors associated a kind of "heads up" to the askers giving an estimated count of answers they may likely get. Such a feature reduces the frustration to the answer seeker and if possible seeker can rephrase the posted question to increase the chances of getting proper response. Authors introduced a novel estimation model which is especially customized to build tree structured QA systems. Authors evaluated their proposed work with one year data collected from Yahoo! Answers which is comprised of 10 million questions posted in the year 2009. Authors unleashed an unexpected assertion that accumulation of more number of questions never deteriorates the performance of a QA system, until community of answerers is well preserved. Experimental evaluations of proposed work show significant improvement than potential contemporary methods because of its novel structure built.

Every community QA system frames certain guidelines of framing a question in order to reduce completely irrelevant or poor quality of questions. Certain questions like "hi question about mathematics", "NDTV iPad app screen design", "plzz can u help me" etc seems annoying and these type of questions will be down voted and be deleted from QA systems. Experienced community members and moderators regularly watch and delete these kinds of questions. Denzil Correa and Ashish Sureka attempted in analyzing deleted questions from Stack Overflow QA system [58]. Authors as a first step created a database of deleted questions which were deleted during 2008 to 2013. The characterization of these deleted questions is then prepared. Secondly, authors developed a prediction framework to estimate whether the newly posted question is likely to be deleted or not. Authors brought out multiple insights over the phenomena of question deletion. The down voting takes a substantial time and once score of the question reaches then community takes immediate action. Query makers often delete their questions if they find down voting in order to preserve their reputation. Good quality questions which are deleted accidentally can get back to be undeleted by up voting process. Authors represented a pyramidal tree structure every node (question) assigned with quality and made poor quality questions to hang at the end of the tree. Authors build a predictive model with 47 significant quality properties taken from user background history, user participation in the group, semantics of text and justification to label the question for its deletion during question creation time. Authors experimented with 270,604 deleted questions from Stack Overflow and reported an accuracy of 66%. The work is the first of its kind to analyze on deleted questions on a large scale on Stack Overflow.

The work unleashed several potential suggestions to preserve quality in community QA systems.

Very limited attempt is presented in the literature to identify an insightful question that could motivate huge number of users for further discussions in a community QA system. In parallel, detecting a potential answer that could help many users is also a major challenge. The impact measure of a question-answer pair over a long term possibly gives answer to these two challenges. Yuan Yao et al [60] pursued their research in predicting the significance durability of a Question-Answer pair immediately after their entry in community QA systems. Authors located a collection of algorithms to predict the significance of question-answer pair by configuring three crucial aspects namely non-linearity, Interoperability and temporal variation. The proposed work has 3 basic advantages. Firstly, the chosen set of algorithms is comprehensive in nature that naturally captures the expected three aspects. Second, thesealgorithms are quite flexible and works positively at certain instances where these aspects can't perform significantly. Third, these algorithms are quite adaptive and scalable in nature. Authors analyzed the proposed algorithms for their correctness, optimality and time complexity subjected to real-time community QA data available in Yahoo! Answers.

Imrul et al. [92] investigated the impact of users' social behavior in QA systems. Authors analyzed the social behavior of 2,00,000 worldwide users collected from yahoo QA database. The cultural behavior analytics presented by Robert Levine and Geer Hofstede is considered as the baseline methodology for the proposed work. Authors empirically measured the cultural metrics with the users of community QA Systems. The work presented different national cultures in Yahoo Answers unleashing their temporal predictions, contribution-relations, privacy concerns, and power in equality.

4.9 Hybrid Approaches

The prior preservance of inter-relationships between question answers in a QA system will sure result in getting best quality answer. Mohan et al. [13] presented a framework which estimates the values of potential features in bringing out the best possible response. The specialty of the work lies in considering both actual-textual and associated-textual features during estimation. Authors empirically tested the framework with Yahoo QA base and opined that textual features like completeness and relevancy with its proper justification predominantly drive in suggesting the best quality answer. Associated-textual features include popularity of the question and administrative control over a question to its seeker etc may not play prominent role in getting top answer still they do partially support. Length of the text and the its language are found least significant features. Authors opined that the proposed framework could help researchers in identifying significant features to pick out top answers.

Answerers accessing with modern QA systems often include URLs along with the actual answer with the intent of providing complete and further information. Long Chen et al [54] attempted analyzing the answers associated with navigational links for estimating the capability of searching systems. Authors evaluated the automated process of identifying navigational references with primary-textual features and supporting metadata. Authors presented a hybrid approach which is a blend of several language modeling techniques like likelihood model, language conversion model and context based language model respectively. Authors evaluated the performance of their proposed mixture language model using Wiki Answers [55] and Yahoo! Answers [56] question database. The proposed intent-based language model show significant performance compared to contemporary language modeling approaches. How big the literature may grow accumulating the works in fetching relevant question-answers, there exist a "gap" among the originally asked query and the system recommended responses. This challenge is termed as lexical chasm or word mismatch problem. Ming Chen et al have attempted to address this challenge by improving the traditional Topic Inference based Translation Language Model (T2LM) through the preserved systematic topics base. Authors specifically take support of user adoption answers for further refining their work. Both these models make a bride to fill the gap by grouping related words. To reduce the gap between different semantics, the proposed work takes the support of user information. The combination of both these methods resulted in a significant improvement in the performance during question retrieval. The work is experimented on real-time Chinese QA system and found improvement in retrieval performance over T2LM base line work.

4.10 Use of Graphical Representations

The structure of data over the Internet is constantly changing thus it become a potential challenge to the user community in accessing the public data. Resource description framework (RDF) enables in describing as standard knowledge base represented as a triplet comprised of subject, predicate and object. Such a base possibly been constructed as a graph in which objects takes the role of nodes and predicates are denoted as edges. SPARQL is the baseline query language for accessing RDF data, but it is quite complicated to work for end users because of its syntax and RDF schema. The goal of an ideal system should give maximum benefit to end users abstracting the internal complexity of both SPARQL and RDF [61]. Thus, RDF gained wide popularity in building natural language processing (NLP) supported QA systems [62, 63]. The QA system with natural language support follows both understanding underlying meaning of question and delivering its proper response. Prime target should go into grabbing out the actual meaning of question with no ambiguity in natural language phrases. Joint disambiguation is the most accepted technique with exponential search space. Lei Zou et al [64] proposed a systematic framework using graphical representation to deal with natural language questions supported on RDF repository. Authors devised this two-stage task as a subgraph matching problem. The problem of ambiguity is resolved when no further matching could be progressed. Thus the cost of disambiguation is saved when there is no matching found. The work is compared with contemporary DEANNA algorithm [63] and those systems running in QALD-3 QA datasets. Authors utilized the relationship phrases provided in Patty [64] system to construct the paraphrase dictionary.

Authors evaluated their methodology with other equivalent benchmark Yago2 [65] system.

Based on the experimental results authors successfully proved that their work significantly improves the precision along with speed and performance.

4.11 QA Systems as Online Educational Resources

During recent years massive open online courses (MOOCs) received wide much attention among educators and learners [74]. Due to the rapid accumulation of digital content, educators are in search of supporting tools that can cater the answers automatically for a posed question without human effort. Chase Geigle and ChengXiang Zhai [75] proposed to exploit the historical or archived MOOCs discussion forum to fetch an answer automatically using IR methods. The proposed methodology is much beneficial because of rapid accumulation of repeated offerings of the course and thus will likely fetch more relevant question-answers. Authors further evaluated their work by preparing a test collection and found that the proposed work is quite better than contemporary techniques. The work also pursues error analysis as a further refinement.

4.12 Addressing Class Imbalance Problems in QA Systems

Class imbalance problem has rapidly gained attention in the machine learning research community [69, 70]. This challenge encounters if the machine learning classifier show bias during prediction of majority class. Such a problem is often encountered in several applications of machine learning. During training the classifier, in a conventional scenario, a user predominantly focuses in training with more positive examples than negative examples. The implication of this paradigm results in biased approximations with respect to specificity and sensitivity. Sub-optimal models can be refined by using existing performance measures to obtain accurate classification. Min-Yuh Day and Cheng-Chia Tsai presented an empirical study on recognizing inference extracted from text. Authors pursued an evaluation of existing NTCIR gold standard linguistic features and found that antonym, negation and 7 linguistic phenomena play key role in N labeling. Experimental results infer that the recognizing inference with analyzed linguistic phenomenon may change with the capability of classifier.

V. RESULTS

Upon evaluating the literature on various topics the below are the contributions done in machine learning approaches with an accuracy of 80-85%. The below works contributed are on answer quality and QA systems with variant approaches.

Table 5.1 List of Literature for QA systems in Machine Learning and Deep Learning.

Contributions using machine learning and deep learning approaches	**Authors**
Attempted to investigate the performance of machine learning techniques and text processing algorithms especially with reference to defence knowledge management discussion forums	Lange et al. (2008)
Proposed three-level question type taxonomy with the base support of publicly available questions captured from community QAs	Zhang et al. (2009)
Presented a new label ranking method developed using a machine learning algorithm which can determine a rank for a question-label to classify. Once the question is classified then authors evaluated the performance of the system to find out whether the inclusion of user's intention really supports or not	Wei Wang et al (2011)
Automatically identifying local questions with a framework of PU learning from positive and un labelled examples along with the standard textual phrases	Long Chen et al (2012)
Developed Sibyl, a new factoid QA system, especially configured for spoken-word documents	Pere et al (2012)
Attempted the initial analysis of "closed" questions on the technical QA platform Stack Overflow. Authors downloaded the archived publicly available 34 lakh questions accumulated during a span of four years	Denzil Correa and Ashish Sureka (2013)
A novel methodology has been developed to test the performance of automated reading systems through reading comprehension tests	Anselmo Peñas et al (2013)
Proposed a framework which predicts users' response with respect to time and location features	Nikolay Burlutskiy et al (2015)
Attempted to address certain issues of translation of Arabic question-answers belonging to medical domain	Yassine et al (2016)

Proposed a tri-modal deep boltzmann machine (tri-DBM) to capture a scholastic integration of query-response for a posted inquiry statement	Baolin Peng et al (2014)
Constructed a study of two-phase question retrieval approach that is comprised of retrieving similar questions and re-ranking the retrieved questions	Ghosh et al (2017)
Designed the modern non-factoid noiseless QA system, work in analysing models which deal in accessing question answering comprised of lengthy descriptions and explanations	Kateryna Tymoshenko et al (2016)
Pursued a research work in health care supported QA system in which their work is of two-fold	Hong Cai et al (2017)

Table 5.2 List of literature om Answer Quality

Contributions	Authors
Addressed a collaborative question answering task by experimenting on huge collection of real question answers available with Yahoo! Answers	Maggy et al (2009)
Attempted to handle the assessment of answer quality by determining the quality of an answer in a QA system	Chirag and Jefferey (2010)
Presented a novel topic modelling framework with Dirichlet forest priors (LDA-DF) to analyse questions and answers	Horowitz et al (2010)
Proposed work determines the correctness of the answers through a process of cross validation done by the fellow learning community and thus gives accurate feedback to the answer providers	Šimko et al (2013)
Băidù zhidao is a collaborative Web-based collective intelligence QA system in Chinese language.	Robin Li
Attempted to address two potential challenges pertaining to the quality of contents and the popularity of the answers	Lifan Guo and Xiaohua Hu (2013)
Presented a general model to determine the quality of information using supporting features like star rating, supporting answer referencing, positive votes, contributor profile, possible count of best answers, acceptance ratio etc.	Kohei Arai and ANIK Nur Handayani (2013)
Constructed a framework which automatically determines the quality of the newly responded answers, based upon several distinct groups of features namely user profile, user access to the community, user textual interactions and the background thrust respectively.	Long et al (2016)

Table 5.3 List of Literature on QA Systems in Various topics.

Contributions of Literature	Authors
Proposed a simple technique of getting a good response to newly posted similar questions by looking at their “equivalent” stored questions which are already been responded earlier	Tianyong et al (2012)
Addressed grabbing the context from the mind of answer seekers by proposing a supervised trained topic modelling for QA systems	Kai Zhang et al (2014)
Attempted in refining question retrieval system to perceive the actual meaning and intention of question creator working with community question answering (CQA)	Kai Zhang et al (2016)
Attempted to understand the actual cause of why a question remains unanswered help answer seekers to refine their question in order to increase the chances of addressing their question	Muhammad Asaduzzaman et al (2013)
Presented a novel answer summarization approach to deal with the challenge “Incomplete Answer”	Vinay et al (2013)
Attempted in early detection of answerers who could become potential contributors of community QA systems in near future	Dijk et al (2015)
Focused on picking out the top contributors in QA systems are primarily been estimated according to user community participation	Yao-Ming Yang et al (2016)

Proposed a combined methodology to find experts who make top contribution in answering. Authors chose questions tags, content and answer's votes as significant parameters to locate the experts	Congfu Xu et al (2016)

After the evaluation of the literature, it is observed that there are many challenges in the QA systems to be addressed further to increase the accuracy rate up to 90-95%. Few of them are listed below.

i) Rise in the levels of spam,

ii) Inequality in participation

iii) Declination in the responses

iv) Managing the team of volunteers and members.

v) Technology on demand.

V. CONCLUSION

This paper attempted in bringing out various research works and their scope focused on question answering systems. Yahoo! Answers, Baidu Knows, Stack Overflow, Wiki Answers, Brainly, Quora, PubMed etc. gained lot of attention in user community. At the most researchers experimented their works on these advanced community QA systems, in which works on Yahoo! Answers take the major stake of research. Estimating answer quality in a QA system is always a challenge and several researchers addressed this issue in their works. Apart from addressing answer quality, finding the question quality, similar questions and response time are equally significant challenging tasks of research. Probabilistic modeling, machine learning, deep learning and hybrid approach of solving have presented promising results when compared to traditional approaches in QA systems. The thrust of configuring a community QA system to a potential educational resource is highly needed and is attempted by Chase Geigle and ChengXiang Zhai during recent years.

REFERENCES

1. *Weizenbaum, Joseph, Computer Power and Human Reason: From Judgment to Calculation. New York, W.H. Freeman and Company, Pp. 2,3,6,182,189, 1976.*
2. *Colby Kenneth Mark, Watt James B, Gilbert John P, "A Computer Method of Psychotherapy", The Journal of Nervous and Mental Disease, Vol. 142, No. 2, Pp. 148–52, 1966.*
3. *GREEN JR, Bert F, BASEBALL: an automatic question-answerer, Western joint IRE-AIEE-ACM computer conference: Pp. 219–224, 1961.*
4. Woods, William A, Kaplan, R, "Lunar rocks in natural English: Explorations in natural language question answering". Linguistic structures processing, Vol. 5, Issue 5, Pp. 521–569, 1977.
5. Terry Winograd, "Procedures as a Representation for Data in a Computer Program for Understanding Natural Language", MIT AI Technical Report 235, February 1971.
6. Lindsay, Robert K., Bruce G. Buchanan, E. A. Feigenbaum, Joshua Lederberg, DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation, Artificial Intelligence, Vol. 61, No. 2, Pp: 209 – 261, 1993.
7. Robert Wilensky, David N. Chin , Marc Luria , James Martin , James Mayfield , Dekai Wu, The Berkeley UNIX Consultant Project, Artificial Intelligence Review, Vol.14, No. 1-2, Pp. 43 - 88, April 2000.
8. Maybury M. T, New Directions in Question Answering, Vol. 31, No. 3, AAAI/MIT Press 2004.
9. Jimmy Lin, Evaluation of Resources for Question Answering Evaluation, SIGIR'05, ACM, August 15–19, Salvador, Brazil, 2005.
10. Sanda Harabagiu, Finley Lacatusu and Andrew Hickl, Answering Complex Questions with Random Walk Models, SIGIR'06, ACM, August 6–11, Seattle, Washington, USA, 2006.
11. Jeongwoo Ko, Luo Si, Eric Nyberg, A Probabilistic Graphical Model for Joint Answer Ranking in Question Answering, ACM, SIGIR'07, Amsterdam, The Netherlands, July 23–27, 2007.
12. Douglas S Lange, Text Classification and Machine Learning Support for Requirements Analysis Using Blogs, B. Paech and C. Martell (Eds.): Monterey Workshop 2007, LNCS 5320, pp. 182–195, Springer-Verlag, Berlin, Heidelberg 2008.
13. Mohan John Blooma, Alton Y.K. Chua, Dion Hoe-Lian Goh, A Predictive Framework for Retrieving the Best Answer, SAC'08, ACM, March 16-20, 2008, Fortaleza, Ceará, Brazil, 2008.
14. Jiang Bian, Yandong Liu, Eugene Agichtein, Hongyuan Zha, Finding the Right Facts in the Crowd: Factoid Question Answering over Social Media, International World Wide Web Conference Committee (IW3C2), April 21–25, 2008, Beijing, China. ACM, 2008.
15. F. Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, Joseph A. Konstan, Predictors of Answer Quality in Online Q&A Sites, ACM CHI 2008 Proceedings, Exploring Web Content, Florence, Italy, April 5-10, 2008.
16. Yandong Liu, Jiang Bian, Eugene Agichtein, Predicting Information Seeker Satisfaction in Community Question Answering, ACM SIGIR'08, Singapore, July 20–24, 2008.
17. Maggy Anastasia Suryanto, Ee-Peng Lim, Aixin Sun, Roger H. L. Chiang, Quality-Aware Collaborative Question Answering: Methods and Evaluation, ACM WSDM '09 Barcelona, Spain, 2009.
18. J. M. Ponte and W. B. Croft. A Language Modeling approach to Information Retrieval. In SIGIR, 1998.
19. Eugene Agichtein and Yandong Liu, Modeling Information-Seeker Satisfaction in Community Question Answering, ACM Transactions on Knowledge Discovery from Data, Vol. 3, No. 2, Article 10, April 2009.
20. Mingcheng Qu , Guang Qiu , Xiaofei He , Cheng Zhang , Hao Wu , Jiajun Bu , Chun Chen, Probabilistic Question Recommendation for Question Answering Communities, ACM WWW 2009, Madrid, Spain, April 20–24, 2009.
21. Kevin K. Nam, Mark S. Ackerman, Lada A. Adamic, Questions in, Knowledge iN? A Study of Naver's Question Answering Community, ACM CHI 2009 QA Systems, Boston, Massachusetts, USA, April 4 - 9, 2009.
22. Xudong Tu , Xin-Jing Wang , Dan Feng , Lei Zhang, Ranking Community Answers via Analogical Reasoning, ACM WWW 2009, Madrid, Spain, April 20-24, 2009.
23. Xin-Jing Wang , Xudong Tu, Dan Feng , Lei Zhang, Ranking Community Answers via Analogical Reasoning, ACM SIGIR'09, Boston, Massachusetts, USA, July 19–23, 2009.
24. Jeon, J., Croft, W., et al. A Framework to Predict the Quality of Answers with Non-Textual Features. In Proc. SIGIR, 2006.
25. T. Yeh, J. J. Lee, T. Darrell. "Photo-based Question Answering", ACM Multimedia, 2008.
26. Tat-Seng Chua, Richang Hong, Guangda Li, and Jinhui Tang, From Text Question-Answering to Multimedia QA on Web-Scale Media Resources, ACM LS-MMRM'09, Beijing, China, October 23, 2009.
27. Yaoyun Zhang, Xuan Wang, Member, IEEE, Xiaolong Wang, Member, IEEE, Shixi Fan, Daoxu Zhang, Using Question Classification to Model User Intentions of Different Levels, IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, Pp. 1153 – 1158, USA - October 2009.
28. Arnaud Grappy, Brigitte Grau, Answer type validation in question answering systems, Proc. RIAO—Adaptivity Personalization and Fusion of Heterogeneous Information, pp. 9-15, 2010.
29. A. Grappy, A.-L. Ligozat, and B. Grau, Evaluation de la r´eponse d'un syst`eme de question-r´eponse et de sa justification. In CORIA, 2008.
30. Mingrong Liu, Yicen Liu, and Qing Yang, Predicting Best Answerers for New Questions in Community Question Answering, L. Chen et al. (Eds.): WAIM 2010, LNCS 6184, pp. 127–138, Springer-Verlag, Berlin, Heidelberg 2010.
31. Chirag Shah, Jefferey Pomerantz, Evaluating and Predicting Answer Quality in Community QA, SIGIR'10, Pp. 411 – 418, July 19–23, Geneva, Switzerland, 2010.

32. Jeongwoo Ko, Luo Si, Eric Nyberg, Teruko Mitamura, Probabilistic Models for Answer-Ranking Multilingual Question-Answering, ACM Trans. Inform. Syst. Vol. 28, No. 3, Article 16 June 2010.
33. Lichun Yang, Shenghua Bao, Qingliang Lin, Xian Wu, Dingyi Han, Zhong Su, Yong Yu, Analyzing and Predicting Not-Answered Questions in Community-Based Question Answering Services, Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011.
34. Wei Wang1, Baichuan Li, Irwin King, Improving Question Retrieval in Community Question Answering with Label Ranking, IEEE Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011.
35. Baichuan Li, Tan Jin, Michael R. Lyu, Irwin King, Barley Mak, Analyzing and Predicting Question Quality in Community Question Answering Services, WWW 2012 – CQA'12 Workshop, Lyon, France, April 16–20, 2012.
36. Tianyong Hao and Eugene Agichtein, Bootstrap-Based Equivalent Pattern Learning for Collaborative Question Answering, A. Gelbukh (Ed.): CICLing 2012, Part II, LNCS 7182, pp. 318–329, Springer-Verlag Berlin Heidelberg 2012.
37. Ashton Anderson Daniel Huttenlocher Jon Kleinberg Jure Leskovec, Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow, KDD'12, August 12–16, , Beijing, China 2012.
38. Aditya Pal F, Maxwell Harper, Joseph A Konstan, Exploring Question Selection Bias to Identify Experts and Potential, Experts in Community Question Answering, ACM Transactions on Information Systems, Vol. 30, No. 2, Article 10, Publication date: May 2012.
39. Long Chen, Dell Zhang, and Mark Levene, Identifying Local Questions in Community Question Answering, Y. Hou et al. (Eds.): AIRS 2012, LNCS 7675, pp. 306–315, Springer-Verlag Berlin Heidelberg 2012.
40. Pere R Comas, Jordi Turmo, Llu Is, Marquez, Sibyl, a Factoid Question-Answering System for Spoken Documents, ACM Transactions on Information Systems, Vol. 30, No. 3, Article 19, Publication date: August 2012.
41. Muhammad Asaduzzaman, Ahmed Shah Mashiyaty, Chanchal K. Roy, Kevin A. Schneider, Answering Questions about Unanswered Questions of Stack Overflow, IEEE 97 MSR, San Francisco, CA, USA, 2013.
42. Jakub Šimko, Marián Šimko, Mária Bieliková, Jakub Ševcech, Roman Burger, Classsourcing: Crowd-Based Validation of Question-Answer Learning Objects, C. Bădică, N.T. Nguyen, and M. Brezovan (eds.): ICCCI 2013, LNAI 8083, pp. 62–71, Springer-Verlag Berlin Heidelberg, 2013.
43. https://stackoverflow.com/ (Last accessed on 21.01.2019)
44. Denzil Correa, Ashish Sureka, Fit or Unfit: Analysis and Prediction of 'Closed Questions on Stack Overflow, COSN'13, , Boston, Massachusetts, USA, October 7–8, 2013.
45. https://zhidao.baidu.com/ (Last accessed on 21.01.2019)
46. Lifan Guo and Xiaohua Hu, Identifying Authoritative and Reliable Contents in Community Question Answering with Domain Knowledge, J. Li et al. (Eds.): PAKDD 2013 Workshops, LNAI 7867, pp. 133–142, Springer-Verlag Berlin Heidelberg 2013.
47. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. In: Proceedings of the 19th International Conference on World Wide Web, pp. 431–440, 2010.
48. Kohei Arai and Anik Nur Handayani, Predicting Quality of Answer in Collaborative QA Community, International Journal of Advanced Research in Artificial Intelligence, Vol. 2, No. 3, 2013.
49. I. Witten and E. Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufman, 2nd edition, 2005.
50. Anselmo Peñas1, Eduard Hovy2, Pamela Forner3, Álvaro Rodrigo1, Richard Sutcliffe4, and Roser Morante5, QA4MRE 2011-2013: Overview of Question Answering for Machine Reading Evaluation, P. Forner et al. (Eds.): CLEF 2013, LNCS 8138, pp. 303 –320, Springer-Verlag Berlin Heidelberg 2013.
51. Chan, W., Zhou, X., Wang, W., Chua, T.-S.: Community answer summarization for multi-sentence question with group l1 regularization. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers. Association for Computational Linguistics, vol. 1, pp. 582–591, 2012.
52. Shen, D., Sun, J.-T., Li, H., Yang, Q., Chen, Z: Document summarization using conditional random fields. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence vol. 7, pp. 2862–2867, 2007.
53. Vinay Pande, Tanmoy Mukherjee, and Vasudeva Varma, Summarizing Answers for Community Question Answer Services, I. Gurevych, C. Biemann, and T. Zesch (Eds.): GSCL 2013, LNAI 8105, pp. 151–161, Springer-Verlag Berlin Heidelberg 2013.
54. Long Chen, Dell Zhang, and Mark Levene, Understanding and Exploiting User's Navigational Intent in Community Question Answering, R.E. Banchs et al. (Eds.): AIRS 2013, LNCS 8281, pp. 392–403, Springer-Verlag Berlin Heidelberg 2013.
55. http://answers.yahoo.com/ (Last accessed on 21.01.2019)
56. http://wiki.answers.com/ (Last accessed on 21.01.2019)
57. Gideon Dror, Yoelle Maarek, and Idan Szpektor, Will My Question Be Answered? Predicting "Question Answerability" in Community Question-Answering SitesH. Blockeel et al. (Eds.): ECML PKDD 2013, Part III, LNAI 8190, pp. 499–514, Springer-Verlag Berlin Heidelberg 2013.
58. Denzil Correa, Ashish Sureka, Chaff from the Wheat: Characterization and Modeling of Deleted Questions on Stack Overflow, International World Wide Web Conference Committee (IW3C2), WWW'14, April 7–11, 2014, Seoul, Korea, ACM 978-1-4503-2744-2/14/04, 2014.
59. Baolin Peng, Wenge Rong, Yuanxin Ouyang, Chao Li, Zhang Xiong, Learning Joint Representation for Community Question Answering with Tri-modal DBM, WWW'14 Companion, April 7–11, 2014, Seoul, Korea, ACM 978-1-4503-2745-9/14/04, 2014.
60. Yuan Yao, Hanghang Tong, Feng Xu and Jian Lu, Predicting Long-Term Impact of CQA Posts: A Comprehensive Viewpoint, KDD'14, New York, NY, USA, August 24–27, 2014.
61. V. Lopez, C. Unger, P. Cimiano, and E. Motta. Evaluating question answering over linked data. J. Web Sem., 21:3–13, 2013.
62. I. Androutsopoulos and P. Malakasiotis. A survey of paraphrasing and textual entailment methods. J. Artif. Intell. Res. (JAIR), 38:135–187, 2010.
63. M. Yahya, K. Berberich, S. Elbassuoni, M. Ramanath, V. Tresp, and G. Weikum. Natural language questions for the web of data. In EMNLP-CoNLL, pages 379–390, 2012.
64. A.-M. Popescu, O. Etzioni, and H. Kautz. Towards a theory of natural language, nterfaces to databases. In Proceedings of the 8th international conference on Intelligent user interfaces, pages 149–157. ACM, 2003.
65. Johanees Hofart, Fabian M Suchanek, Klaus Berberich, Gerhard Weikum, YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia, ACM Journal of Artifical Intelligence, Vol. 194, Pp. 28 – 61, January 2013.
66. S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. Okapi at trec-3. In TREC, 1994.
67. C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. ACM Trans. Inf. Syst., 22(2):179 - 214, 2004.
68. Kai Zhang, Wei Wu, Haocheng Wu, Zhoujun Li, Ming Zhou, Question Retrieval with High Quality Answers in Community Question Answering, ACM CIKM '14, November 03 - 07, Shanghai, China, 2014.
69. R. Barandela, J. S. Sanchez, V. Garcia, and F. J Ferri, "Learning from imbalanced sets through resampling and weighting," Pattern Recognition and Image Analysis, Proceedings, vol. 2652, pp. 80-88, 2003.
70. M. M. Haque, M. K. Skinner, and L. B. Holder "Imbalanced Class Learning in Epigenetics," Journal of Computational Biology, vol. 21, pp. 492 - 507, Jul 2014.
71. R. Batuwita and V. Palade, "Adjusted Geometric-Mean: A Novel Performance Measure for Imbalanced Bioinformatics Datasets Learning," Journal of Bioinformatics and Computational Biology, vol. 10, Aug 2012.
72. Min-Yuh Day and Cheng-Chia Tsai, Machine Learning for Imbalanced Datasets of Recognizing Inference in Text with Linguistic Phenomena, IEEE International Conference on Information Reuse and Integration, Pp. 562 – 568, 2015.
73. Imrul Kayes, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, Francesco Bonchi, Cultures in Community Question Answering, ACM HT '15, September 1–4, Guzelyurt, Northern Cyprus, 2015.
74. Kaplan Andreas M, Haenlein Michael, Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster". Business Horizons. 59 (4): 441–50, 2016.
75. Chase Geigle, ChengXiang Zhai, scaling up Online Question Answering via Similar Question Retrieval, The Third ACM Conference, Pp. 257 – 260, April 2016.

76. Giovanni Da San Martinoy, Alberto Barrón-Cedeñoy, Salvatore Romeoy, Antonio Uva, Alessandro Moschitti, Learning to Re-Rank Questions in Community Question Answering Using Advanced Features, ACM CIKM ’16, October 24–28, Indianapolis, IN, USA, 2016.
77. Danna Gurari and Kristen Grauman, CrowdVerge: Predicting If People Will Agree on the Answer to a Visual Question, CHI 2017, May 6–11, 2017, Denver, CO, USA.

AUTHORS PROFILE

Mr. Venkateswara Rao P, a Research Scholar in JNTU College of Engineering Anantapur, has completed his B. Tech from JNTU Hyderabad, Master of Technology from JNTU Hyderabad. Currently working as an Assistant Professor in VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad. He has 10 years of teaching experience, two years of research experience and one year of industry experience. His areas of interest include Deep Learning, IOT, Natural Language Processing and Image processing. He has many publications in significant international journals.

Dr. A. P. Sivakumar, has completed his B. Tech from JNTU Hyderabad, Master of Technology from JNTU Anantapur, Doctor of Philosophy from JNTU Anantapur. Currently he is working as an Assistant Professor in Department of CSE, JNTU Anantapur. He has 13 years of teaching experience whose areas of interest includes Natural language processing, Information Retrieval and Cross Lingual Systems. He has many publications in national/international journals. He was awarded as an outstanding Faculty for the Contribution and Achievement in the field of Cross Lingual Information Retrieval, by Centre for Advanced Research and Design, VIF, Chennai.