

## QUESTION ANSWERING SYSTEM FOR DISTANCE EDUCATION USING E-LEARNING AND COLLABORATING LEARNING ENVIRONMENTS IN INDIA

Karpagam K

Department of Computer Applications, Dr.Mahalingam College of Engineering and Technology, Pollachi,  
642003, India  
karpagam80@gmail.com

Dr. Saradha A

Department of Computer Science and Engineering, Institute of Road and Transport Technology, Erode, 638316,  
India  
saradha.irtt@gmail.com

### ABSTRACT

Today, e-learning in education & distance learning is the most popular way to gain knowledge and share information between learning communities. In India, formal learning phase is hard-hitting to lower middle class people due to economic background. To resolve this issue and improve the life standards of people, distance education based on e-learning and collaborative learning paves a way. Information retrieval domain supports in this to display the specific answer for the user query which makes the learning easier. Question answering system returns sentence segment rather than returning list of documents. We present a mobile based question answering system that serves as a personal assistant in learning and providing information about computers, software and hardware, book reviews using natural language. QA models accept the query in natural language, analyze and compare it with information stored in the knowledge base. It displays the optimized response snippets which improves the efficiency of e-learning retrieval responses. The proposed system was developed to help the users to learn the subject knowledge at times of examination, technical interview and share answer for other users. This paper proposes a technique to find the type of question that in turn leads to a correct answer. The knowledge base created from the benchmark dataset such as Amazon Book Reviews, 20newsgroup, Quora, and Yahoo! Answer. This paper also focuses on the Indian education scenario, eLearning tools for easy learning through content preparation and presentation. The system interface has been evaluated using standard metrics such as Precision, Fetch, F1 Score, Inverse Precision, and Inverse Recall for the appropriate relevant response.

**Keywords:** Natural language processing, E-Learning, Education domain, Cosine similarity, Cuckoo Search Optimization, Collaborative Learning.

### INTRODUCTION

Enduring learning for survival becomes essential in today's modern world. Electronic learning (or e-learning) and distance education is a technology-based education where the learning materials are available in internet where users can access and learn anywhere and anytime. According to the statics, distance education from the universities will approximately cost 35% of actual regular course university fee. In addition to that, E-learning provides a way to access high quality learning materials prepared by subject matter experts. The use of eLearning is to disseminate education in remote areas, the pros and cons of eLearning, and the future of eLearning in India in a cost saving mode. A few suggestions have been made to use e-learning for informal education, which is very effective for a developing country like India, where the majority of the population lives in rural / remote areas with less economical background. With low cost, the back warders have received neglected formal education. This e-learning also helps in rapid learning of the latest technical content in less span of time.

Due to the need of information at all instances within short response times, scenario for learning purpose has grown excessively and leads to solution of the mobile-based learning model. This leads to widespread mobile applications, smart phones and tablets for learning are being expanded with data connectivity. The question-and-answer system plays an important role in providing accurate information rather than long answers for learners. It categorized in an open domain question-answer and a closed domain question answer; the open-domain question-answer is able to answer dynamic data regardless of domain nature. The closed domain Question Answering system uses datasets /knowledge base /database /dumps for retrieving the answers.

Learners who need static data use closed domain QA systems such as content e-learning platforms, book reviews, product reviews, historical data, and education, etc. Because of the e-learning environment, students are more active and responsive to subject quizzes, assignments and tutorials, etc. The crowd-sourcing paradigm has been successful in evaluating answers and unanswered questions among the learning groups. The collaborative learning environment is used to exchange information with learning groups that consist of various technical and non-technical areas. It supports e-learning area to reach its level of efficiency.

### RELATED WORKS

G.Manoranjithm el (2013) explores the area of Mobile Question Answering system that used the social network like Facebook, Twitter, and LinkedIn and experts answers for answering the user queries and rates them according to the accuracy. D.Aravind Gosh (2015) discusses the cloud technologies to apply in mobile based QA system to increase the performance with the response time and reduced the storage space of information. This Waheeb Ahmed et al (2017) proposes a automatic web based Question Answering (QA) system is a used for improving e-learning and education field. The user query accepts in natural language, which returns precise answer instead of hyper links documents to improve e-learning scenario.

Dr. M. Sikandar Hayat Khiyal(2009),” converse about the usage of Natural Language Annotations in Mobile Question Answering with answer pattern analysis the parse tree used to process and stores a sentence into an easy-to-understand format subject relation object. Xiin-She Yang and Deb (2010) &Nitisha Gupta, Dr.Sharad Sharma (2016) authors converse about comparison on various optimization algorithms with their parameters and benefits like Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms, Artificial Intelligent and all available nature Inspired Algorithms.

Sweta P. Lende and M. M. Raghuwanshi (2016) which describes the different methodology and implementation details of question answering system for closed domain QA System for handling documents related to education domain using NLP techniques to retrieve more precise answers. Tannaz Alinaghi et al(2013)discussed on validating the content of questions with all available resources including course materials, FAQs and rating from other learners using a recommender system. The most appropriate answer(s) with respect to several conditionals such as learner’s knowledge, research background, history of previous questions, and the candidate answers relevant to the question will be preserved.

In this paper Tilani Gunawardena (2015), a mobile based QA system for satisfying the user requirements in education field related to book through reviews and computers proposed. Thus the main challenge of the proposed scheme is to reduce the response time and negative ratios for smart devices. Liu, S., Liu, F., Yu, C., & Meng, W (2004) &Christos Bouras, Vassilis Tsogkas (2012), discussed on semantic similarity among mutual words are analyzed for finding the semantic relationship between them using the empirical formulas and Wordnet dictionary. Mansaf Alam, Kishwar Sadaf (2016) discussed the role and efficiency of k-means Clustering based on Cuckoo Search and Consensus Clustering for attainment of web search result and the results compared with the cuckoo search, k-means and Bayesian Information Criterion (BIC) method.

Richard Khoury (2012) a new clustering algorithm is proposed to cluster similar sentences automatically based on the sentences’ part-of-speech. It demonstrates question type classification with positive or negative impact, syntactic similarity metric and clusters the need sentences. Prof. Kohei Arai (2013) proposed a collaborative learning with domain knowledge and answer quality predictor is used with the target information as a reference. The knowledge base will be enriched for future question answering by system and others users.

Waheeb Ahmed et al (2017) propose the goal of an intelligent answering system is that the system can respond to questions automatically. For developing such kind of system, it should be able to answer, and store these questions along with their answers. Our intelligent QA (iQA) system for Arabic language will be growing automatically when users ask new questions and the system will be accumulating these new question-answer pairs in its database. The various mobile applications are developed in IT industry are analyzed for further processing such as Light mobile app Q&A systems, +ask mobile app and Kiwi app.

Our proposed system combines several hybrid areas such as text and data mining, machine learning, information retrieval and natural language processing for building the intelligent interface for question answering system. And we interested in building scalable solutions in different application domains for text data related problems.

### PROPOSED METHODOLOGY

The proposed model is a mobile based QA which gets user query in natural language and displays precise answer on education based book reviews, computer hardware & software, general information with quick response. Recently, there are numbers of optimization techniques such as Particle Swarm Optimization, Cuckoo Search, Firefly, Bat algorithm, Bees algorithm etc are available. Optimization algorithms suffer from low-quality results& convergence rates, complex structures due to its nature in deciding input parameters. To overcome this proposed application achieves in extract answer from knowledge base and applies cuckoo optimization.

The system accepts the user query preprocess it to identify the type of question and judges the answer to be extracted. The knowledgebase is created with the bench mark datasets in a defined tree structure. In return helps

to show it in easy retrieval of answers from the huge sets. The relevant appropriate answer sentences are extracted and evaluate for the accurate answers. The accuracy has been calculated with the similarity score between the question and the answer generated.

The learner community will be benefited by the collobarative learning method with accurate answers and shares informations. Also learner can rate the answer generated by the other learners and system for promoting the correct answers to the learning community. The crowdsourcing concpet is applied for distance education for answer generation for systems unanswered question and to train the system for answer generation by inputing the human generated answers. The answer is displayed to the learners in the user interface. The propsoed system architecture is shown in Figure 1.

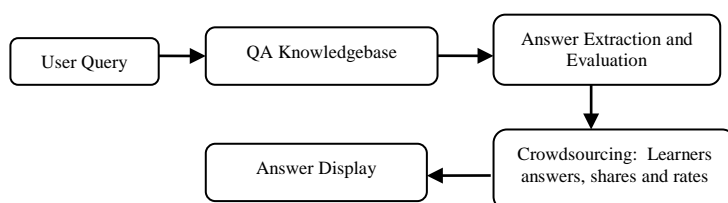


Figure 1: System Architecture

**A. Question Analysis**

Due to users with fat fingers, a chance occurring spelling mistakes like reciept with receipt, theater with theatre and cast with cats can solved with automatic corrections. The query pre-processing is done for removing stop words and stemming keywords. The query type is identifying with Stanford POS tagger which gives the grammatical structure of sentences. There are many Question types as Factoidal and non Factoidal. Question type pattern is framed to identify the question type to map exact candidate answer in the knowledge base. The Table1 gives the sample of the question.

Table 1: Question types

Question type	Description class	Sample Question	Tagging
Who/What	Person /place /values / Organization	What is the formula of calculating area of a cube?	Positive tag
When	Time/date	When was Groovy and Grail developed?	Positive tag
Where	Location/place	Where the Google office is located India?	Positive tag
Why	Reason	Why not RAM volatile in nature?	Negative tag
How many	Number	How many derived classes can be getting from base class?	Positive tag

Rita Wordnet used to provide various utility functions like semantic meaning of terms such as same word with different context meaning, different words gives same meaning like good, better, nice and best etc. It is used for finding similarity among the words like sing and sang. The questions analyzed for the positive and negative tagging for the production of the appropriateness of the resultant answers.

**B. Knowledge base building and optimized information retrieval**

The knowledge base is building using content based on Amazon book reviews like book title, star rating from reviewer, review and voting of book .The answers related to the computer software and hardware from Yahoo answer dataset & all technical and non technical from Quora dataset. It confers that building knowledge base through keyword matching and semantic similarity between query and document clustering for information searching Sanglap Sarkar (2015).

After finding the related sentence for the user query, the answer list if optimized with cuckoo search optimization. The cuckoo search optimization is population-based optimization techniques with limited parameters such as number of nests, number of eggs, fitness function value and number of iteration. The advantage of choosing the cuckoo search optimization by its stochastic nature, non-deterministic and fewer number of parameter for answer convergence. The extracted sentence has optimized using the cuckoo search optimization algorithm to reduce the mean square error rate. The fitness function is for average number of

mutual keywords among each sentence of paragraph and question. It also calculates and performs the search on retrieved sentences for best global best and local best.

The answer generation phase has 2 cases:

Case i) IF answer for user query is found in knowledge base, the system requests the user to select, rank the sentences for generating summaries according to the highest relevance score.

Case ii) IF answer for user query is not found in knowledgebase, the system accepts the significant answer from crowd user by applying their subject knowledge and experience.

Collaborative learning refers to methods and environments in which learners engage in a common task in which each individual relies on each other and is accountable to each other. It involves the use of small groups so that all students can maximize their learning and that of their peers. Peer learning supports to inculcate the knowledge sharing habit among the people with learning thrust. The teacher plays a role of facilitator in the peer collaborative learning environment to facilitate the learning. In developing country like India, this type of learning culture is needed to grow the education field, where maximum cannot afford the higher degree education.

### EXPERIMENTAL RESULTS

The quality of answer validation is to review the returned answer by QA system for accurateness and related ones. In paper Ming Tan, Cicero dos Santos, Bing Xiang & Bowen Zhou (2016), discuss on the improved question and answer matching model is coined by developing a language learning model with the pattern identification.

The proposed learning model trained with 500 questions to learn question type and tested with 100 questions asked by learners. The test set consists of each 20 questions on “Wh” and “How” to type questions, which capable of retrieve candidate answers from relevant documents. Proposed system performance is analyzed in real world scenario by conducting pre-test with 50 students group (of 20-25 age) involved for acquiring user experience. The questionnaire list is framed to get the initial feedback for further improvement consists of questions on relevant and irrelevant answers for the user queries, User interface experience and overall user rating on a 4 point scale. In appendix A, the feedback form on user experience is shown for reference. At beginning of survey, the students/learners provided with a brief handout about

- What and how to interact with the system
- Domain area of QA knowledge base
- Total time duration 15mins

The result of query and response from above said datasets are shown in Table2:

Table 2: Query Request from User and Response

Type of Occurrences	Question Type with response ratio					
	What	When	Where	Why	How	Accuracy
Related Domain	83.2%	78.5%	85.4%	80%	75.4%	80.5%
unrelated domain	60%	45%	63%	52.5%	42.8%	52.66%
Repeated questions	67%	75%	68.5%	58.2%	73.5%	68.44%
Same Question asked in different Methods	56.2%	77%	62.5%	58.6%	48.6%	60.58%

### EVALUATION METRICS

Gunnar Schröder et al(2011), discusses on various evaluation metrics used for retrieving the information along with the Precision, Recall, Inverse precision, Inverse recall and F1.Terms represents are TP as true positive, FP as false positive, FN as false negatives and TN as true negative used or calculating the accuracy of the generated answers.

Precision or true positive accuracy is calculated as the number of correct positive predictions to the total number of true positive predictions with false positive predictions of sentences. They evaluate and answers on how relevant are the retrieved results?

$$Precision = \frac{TP}{TP + FP} \text{---(1)}$$

Recall or true positive rate is the used to find the relevant sentence from the retrieved one to the relevant sentence. They evaluate and answer on retrieving many of the truly relevant documents. They evaluate and answers on did the system retrieve many of the truly relevant documents?.

$$Recall = \frac{TP}{TP + FN} \text{---(2)}$$

The F1-Score is the combination of precision and recall into a single score by calculating different types of means on both metrics.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \text{---(3)}$$

The inverse precision or true negativity accuracy is to compute for the ratio of certainly irrelevant sentences that are not selected to the total number of not selected sentences. This is the probability of the absolutely irrelevant sentence which is not selected is given by the formula

$$Inverse Precision = \frac{TN}{FN + TN} \text{---(4)}$$

The inverse recall or true negativity rate is to compute for the ratio of actually irrelevant sentences not selected to the total number of irrelevant sentences.

$$Inverse Recall = \frac{TN}{FP + TN} \text{---(5)}$$

From the top 50 retrieved sentences of cuckoo search optimization technique applied for various questions types has analyzed with standard metrics for accuracy results. The results analyzed for the accuracy of the result with above metrics for different question and are shown in given table 3.

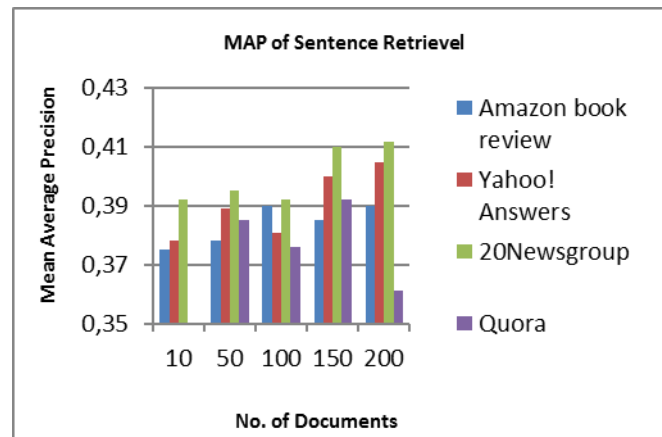
Table 3: Top N Sentence Retrieval

No. of Queries	Precision	Recall	F1	Inverse precision	Inverse Recall
10	0.8	0.64	0.7111	0.7	0.84
20	0.7	0.5185	0.5957	0.5666	0.7391
30	0.75	0.5	0.6	0.5	0.75
40	0.45	0.45	0.45	0.6333	0.6333
50	0.6	0.75	0.6666	0.8666	0.7647
60	0.9	0.5294	0.6666	0.4666	0.875
70	0.55	0.3793	0.4489	0.4	0.5714
80	0.5	0.625	0.5555	0.8	0.7058
90	0.55	0.6875	0.6111	0.8333	0.7352
100	0.5	0.4761	0.4878	0.6333	0.6551

Mean average precision is one of the popular performance measures of information retrieval field. It uses to evaluate the rank of retrieved relevant documents with the average precision values.

$$MAP = \frac{1}{n} + \sum_{Q_i} \frac{1}{R_i} \sum_{D_j \in R_i} \frac{1}{r_{ij}} \text{---(6)}$$

Where n is the number of test questions, r is the rank of the jth relevant document D<sub>j</sub> in Q<sub>i</sub> and R<sub>i</sub> is the relevant document for Q<sub>i</sub>. The result of sentence retrieval is shown in figure 2.



**Figure 2 : Mean Average Precision of Sentence retrieval**

### CONCLUSION

The role of distance education, e-learning and collaborative learning methodologies is very high in supporting and educating learners of developing nations. This work motivated to develop an intelligent answering system as a mobile application that helps learning community for answer rating and sharing in e-learning and collaborative learning. The proposed framework evolves for optimized answer generation with cuckoo search optimization due to less complexity and best global optima. The benchmark datasets Quora, 20newsgroup, Amazon Book review and Yahoo! Answers are used to train and test performance of the proposed methodology. A pre-test is conducted with group of learners with general instructions for getting user experience for improvement. The collaborative learning and crowdsourcing concepts supports for system efficiency in answer sharing and learn together. From the above discussion and overall assessment, it is proved experimentally that the proposed approach is well accomplished for QA system in providing precise answers when compared to state of art works. The restriction on the proposed system is not much investigate and deep analysis on medical terminologies.

### REFERENCES

- G.Manoranjithm el (2013),”Mobile Question Answering system based on social network “, International Journal of Advanced Research in computer and communication Engineering “, September.
- D.Aravind Gosh (2015),” Literature survey on mobile Q & A system in the cloud based environment “,International Journal of Innovative research in computer and communication Engineering“, January.
- Waheeb Ahmed et al (2017),“An Automatic Web-Based Question Answering System For E-Learning”, Information technology and learning tools ,volume 58, April.
- Dr. M. Sikandar Hayat Khiyal(2009),” Mobile System Using Natural Language Annotations for Question Answering”, International Conference on Computer Technology andDevelopment.
- Xiin-She Yang and Deb (2010), “Engineering optimization byCuckoo Search “, Journal of Mathematical Modeling and Numerical Optimization, vol. 1, no. 4.
- Nitisha Gupta, Dr.Sharad Sharma (2016), “Nature-inspiredTechniques for Optimization: a brief review”, International Journal of Advance Research in Science and Engineering, Vol no.5.
- Sweta P. Lende and M. M. Raghuwanshi (2016),”Question answering system on education acts using NLP techniques”, IEEE explorer,October.
- Tannaz Alinaghi etal(2013),”A Multi-Agent Question-Answering System for E-Learning and Collaborative Learning Environment”, System and Technology Advancements in Distance Learning.
- Tilani Gunawardena(2015),“Performance Evaluation Techniques for an Automatic Question Answering System”, International Journal of Machine Learning and Computing, Vol. 5, No. 4, August.
- Sanglap Sarkar (2015) ,”NLP Algorithm Based Question and Answering System”, Seventh International Conference on Computational Intelligence, Modeling and Simulation.
- Gunnar Schröder et al(2011), ”Setting Goals and Choosing Metrics for Recommender System Evaluations “5th ACM Conference on Dresden University of Technology Recommender Systems Chicago, October 23th.
- Liu, S., Liu, F., Yu, C., & Meng, W (2004), “An effective approach to document retrieval via utilizing WordNet and recognizing phrases”, In Proceedings of the annual international ACM SIGIR on research and development in information retrieval ,pp.266–272, ACM.
- Christos Bouras, VassilisTsogkas (2012),”A clustering technique for news articles using WordNet,” Elsevier-Knowledge-Based Systems,.
- Tunstall-Pedoe, William (2010),”True Knowledge: Open-Domain Question Answering Using Structured Knowledge and inference”, AI Magazine, Vole 31 Number 3,pp.80-92.

Gang Liu and Tianyong Hao (2012), "User-based Question Recommendation for Question Answering System", International Journal of Information and Education Technology, Vol. 2, No.3, June.

Ming Tan, Cicero dos Santos, Bing Xiang & Bowen Zhou (2016), "Improved Representation Learning for Question Answer Matching", Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, August pp. 7-12.

Mansaf Alam, Kishwar Sadaf (2016), "Web Search Result Clustering based on Cuckoo Search and Consensus Clustering", Indian Journal of science and Technology, Volume 9, Issue 15, April.

Richard Khoury (2012), "Sentence Clustering Using Parts-of-Speech", International Journal of Information Engineering and Electronic Business, Feb.

Prof. Kohei Arai (2013), "Collaborative Question Answering System Using Domain Knowledge and Answer Quality Predictor", International Journal of Modern Education and Computer Science, Nov 2013.

Waheeb Ahmed et al (2017), "Developing an Intelligent Question Answering System", International Journal Education and Management Engineering, Nov.

### Appendix A

#### Sample Mobile Question Answering System Application Survey Phase

Name of student / learner :	Date :
User rating (*****)	
Name the Similar apps you come across	
Time duration spent on application	
Functionality (UI Experience)	Excellent/ well accomplished / Good/ Poor
Performance (Responsiveness)	Excellent/ well accomplished / Good/ Poor
Load balancing of Mobile App	Excellent/ well accomplished / Good/ Poor
Deal with unrelated queries	Excellent/ well accomplished / Good/ Poor
Consistent way of content representation	Excellent/ well accomplished / Good/ Poor
Response time	Excellent/ well accomplished / Good/ Poor
Appropriateness of answer	Excellent/ well accomplished / Good/ Poor
Self assessment on subject	Excellent/ well accomplished / Good/ Poor
Network Balancing	Excellent/ well accomplished / Good/ Poor
Answer delivery	Excellent/ well accomplished / Good/ Poor
Is enough time duration for response?	Excellent/ well accomplished / Good/ Poor
With in small quantity of questions able to reach the expected answer?	Excellent/ well accomplished / Good/ Poor
Bottlenecks Faced	
Willingness to share with peers	Yes/ No
Suggestions for Feature Enhancements	