

# A Mobile based Intelligent Question Answering System for Education Domain

**Karpagam K**

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

**Saradha A**

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

Received: 21 July 2017; Accepted: 22 September 2017; Published: 08 January 2018

**Abstract**—The domain of intelligent question answering systems is leading the major role in fulfill the user requirement with specific answers stimulate QA research to the next level with machine learning techniques. In this paper, we present mobile based question answering system acts as a personal assistant in learning and for providing the user with information on computers, software and hardware, book reviews by using natural language for the communications. The proposed Mobile based QA models will accept the natural language query, analysis and match them with information stored in the knowledge base and display the optimized result. The knowledge base created from the benchmark data set such as Amazon book reviews, 20newsgroup and Yahoo! Answer data set clustered with content specific clustering and displays the outcome in the form of snippets as output. Sentiment analysis used to decrease the vocabulary gap among the user query and retrieved candidate answer solutions. The results of the proposed interface evaluated with standard metrics such as Precision, Recall, F1-Score, Inverse precision and Inverse recall for the appropriate return of relevant answer.

**Index Terms**—Natural language processing, Explicit Semantic Analysis, Tf-Idf, Cosine similarity, cuckoo search optimization, Mobile QA systems.

## I. INTRODUCTION

The widespread of web applications, smart phones and tablets are enhancing the data connectivity has recently enabled development of many web and mobile applications. Due to the need of information system at any instances within short span of response time, scenario for learning purpose has grown excessive and leads to solution of the mobile based learning model. Now days, mobile based intelligent interfaces have laid a surface to fulfill the requirements of the user with related content on context. People use their browsers to access search

engines i.e. search results appear from matched content on the web. The crowd sourcing paradigm has successfully adopted in web services that provide dynamic information sharing among the individual members of internet world through forums and blogs. The update information is connected to the web is automatically done will enhance the system performance and made it as a valuable resource. In providing the precise information instead of long answers to the learners, the question answering system plays a major component. It categorized into open-domain question answering and closed domain question answering; the open-domain question answer is capable of answering dynamic data irrespective of the domain nature. The closed domain Question Answering system uses datasets /knowledge base /database /dumps for retrieving the answers instead of using search engine. The learners who need static data use closed domain QA systems like E-learning platforms for contents, book reviews, old movie and product reviews, historical data, commercial, education, music, survey data, tourism, Medical health etc.

The conventional method to search the answer is through the search engine like Google, Bing, AltaVista etc is the primary way for information retrieval on the internet. To improve performance by providing the relevant answers with analysis of semantic similar meanings, semantic search engines have inducted. These social search engines group people with similar interests and refer to historically selected results of person's group members decide to person significant results. Although the search engine performs in answering factual queries for information in database, they are inappropriate for non-factual queries are more subjective, relative and multi-dimensional in advising research on social-based question and answer when the information is not available in database. Recently, emerging efforts have focused on question and answer systems based on social media networks for getting the precise information.

## II. RELATED WORKS

The foremost of all methods for performing the question and answer system in a decentralized manner as a social engine ask question either by instant message, the web I/P, text message (or) voice. The [13] proposed QA system which uses the surface pattern which automatically learns to check question and answer patterns and evaluated with the TREC 2005 and 2007 dataset to extract the answer to user questions. In paper [14], discussed on semantic similarity among mutual words analyzed for finding the semantic relationship between the words in the question, paragraph/sentences using the standard formulas and Wordnet dictionary.

The subsequent Mobile applications on question answering system based on the keyword based answer engine techniques for context-aware answer recommendations. However, such a system cannot directly use the previous centralized methods (or) broadcasting methods which generates the high cost of mobile internet access node overload, and high server bandwidth cost with the tremendous number of mobile users. Light mobile app Q&A systems, where mobile nodes access the Q&A systems through the internet, are promising considering the rapid increase of mobile users and the convenience of practical use. It is an Artificial intelligence bot that provides the answer to queries with the support of experts using NLP. It covers most of the current affairs, health, religion, travel, train schedules and counselors. The +ask mobile app developed to ask an opinion about the answer for a question to all generic things in 140 characters and get ratings. The advantage is getting answers by adding more users like friends, locality people and experts in the app. The Kiwi app used to ask and share information with the group of friends and community-based question and answers.

The paper 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 [1].

The following paper 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 [2]. The authors work in the area of Question Answering on Smartphone with mobile app and Utilization is classified based on the user knowledge level and need [3].

In this paper, converse about the usage of Natural Language Annotations in Mobile Question Answering with answer pattern analysis the parse tree used to process and store a sentence into an easy-to-understand format subject relation object [4]. This paper indicates the usage of the text message processing based on the context of the user demands in mobile Phone Question Answering and discuss on their advancements in future [5].

The authors have been discussing on the question answering systems need to be developed for the mobile-

based platform with their flexibility towards the user demands and persuade the requirements [6] & [7].

The authors converse about comparison on various optimization algorithms like Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms, Artificial Intelligent and all available nature Inspired Algorithms [8], [10] & [11].

In this paper, is proposed that the technique for a relevant document clustering and apply the new meta-heuristic algorithm cuckoo search optimization using Levy flight. In this paper, 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 [12]. In this paper, for a question Q and a sentence or paragraph that contains the answer to Q, an answer selection module is to select the "exact" answer [13].

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 [24].

In this paper, 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 cluster the need sentences [25].

In this paper, 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 [26].

In this paper, 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 [27].

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. And we interested in building scalable solutions in different application domains for text data related problems.

## III. PROPOSED SYSTEM

The proposed model is a mobile based QA to help the user to get the precise answer on education based book reviews, computer hardware & software, general information instead of long length answers with low node overhead and system cost with quick response to questions askers. This app is novel in that achieves lightweight answer search with the knowledge base and optimizes results using natural inspired algorithm. Recently, there are numbers of optimization techniques

based on evolutionary and natural inspired algorithms introduced to obtain the global optimum solution to maximize the accuracy of results and minimize the response time. These optimization algorithms include Particle Swarm Optimization, Cuckoo Search, Firefly, Bat algorithm, Bees algorithm etc. Optimization algorithms suffer from low-quality results, low convergence rates, complex structures and deciding the values for the number of input parameters.

The question can be answered by answer engine with information retrieval concept. The returned answer has verified for appropriateness and ratings through standard metrics and experts, and the feedback rating stored in the knowledgebase for future processing. If the answer found in knowledgebase it asks for the rating for the accuracy appropriate of answer. In the case of the answer not found the system uses the learning model with reinforcement technique to ask the user for the answer to get updated knowledge for the future usage. In this paper, answering system based on the user behavior and with analysis for the question recommendation for the improvement of the system [22]. It leverages the answer snippets with less time complexity by reducing the number of loops, optimization techniques and computation costs. The feedback from the users shows the light weighted knowledge app in question and answer can provide high-quality answers.

Table 1. Question types

Question type	Description class	Sample Question
Who/What	Person /place /values / Organization	What is the formula of time complexity of the algorithm?
When	Time/date	When was Java developed?
Where	Location/place	Where is the Microsoft office is located
Why	Reason	Why is ROM volatile in nature?
How many	Number	How many derived classes can be getting from base class?

#### A. Question Analysis

In analysis phase, the user can able to type any questions in natural language through user interface. Due to users with fat fingers, a chance occurring spelling mistakes like receipt with receipt, theater with theatre and cast with cats can solved with automatic corrections. The query processed by the pos tagger for removing the stop words and stemming the keywords. The pre-process of the query with Stanford POS tagger which gives the grammatical structure of sentences like the subject, object. And provides question type pattern identification of the input query by using the verb, noun, adverb etc to map candidate answer in the knowledge base. Rita Wordnet used to provide various utility functions for annotating the corpus and positive and negative word separation. If the word is a preposition found using the POS else [19] & [20] used for the analysis of different words with same meaning like good, better, nice and best etc. The reason for choosing the POS tagger and Wordnet than NLTK

toolkit is 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. The Table 2 gives the sample of the question

Table 2. Question types

Input Query	Question type	Tagging
Has Xcode reached all technical people?	Choice	Positive tag
Has not Facebook become boring these days?	Yes /No	Negative tag
Is the Apple watch really a breakthrough user interface?	Yes /No	Positive tag
Is not smartphone used for the technical purpose?	Yes /No	Negative tag

#### B. Knowledge base building and 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& business, politics, sports, entertainment from 20newsgroup dataset. It confers that building the structure knowledge base through keyword inference and semantic similarity to construct the document clustering for information searching in easier manner [21].

The knowledge base built from the documents in aligned dataset grouped with semantic similar keyword in the document based on content. For e.g. Amazon Book review dataset contains multi-variant data of book review details such as review score, review title, URL tail and review text in Html format, Html tags was removed in preprocessing phase and converted to text format. The retrieving process of the document from the knowledge is done with mapping all along with the question keyword. The information retrieval uses Explicit Semantic Analysis (ESA), a vector-based representation of documents/words in the document corpus.

The documents in the dataset converted into the bag of words and stored in the form of inverted index with document ID. The result of an inverted index calculation of document is for the query keyword that is mapped to the list of indexed documents is collected .The highest score is given to the document with the maximum number of matched keywords. The documents are cluster with the semantic similar keywords and found the mean for the grouping of the final dependent cluster of documents. If the keyword phrase exists in final cluster of documents, it will be marked as value one, for not available marked with value zero. The zero value documents are removed to reduce searching time and the documents with value one are grouped together .The term frequency (TF) & Inverse Document Frequency (IDF) calculated to find the relevant document by using the formula (1) & (2) .

Words represent in column vector in the tf-idf matrix and document act as the centroids of the words representation in the vector.

$$tf_{ij} = \frac{freq_{ij}}{\max_i freq_{ij}} \quad (1)$$

$$idf_i = \log \frac{N}{n_i} \quad (2)$$

Where  $freq_{ij}$  is the number of frequency of  $i^{th}$  words in the  $j^{th}$  sentences,  $maxfreq_{ij}$  is the maximum number of frequency of  $i^{th}$  words in the  $j^{th}$  sentences,  $N$  is the total number of sentences,  $n_i$  is the number of sentences in which word  $i$  occurs.

The cosine similarity calculated for finding the semantic relatedness between the words with the summing of the vectors of all words in the text. The formula for finding the cosine similarity with comparing vectors ( $u$  &  $v$ ) is

$$sim(u, v) = \frac{\sum_{i=1}^N u_i * v_i}{\sqrt{\sum_{i=1}^N u_i^2} * \sqrt{\sum_{i=1}^N v_i^2}} \quad (3)$$

The set of documents are retrieved for analysis for the suitable answers and as a subsequent activity the related document to be ranked for the extraction, there is a chance of more than one document contain the keyword in group. From the retrieved relevant documents, the appropriate paragraphs are extracted for the answer sentence and by matching degree between a paragraph and question keywords are calculate by using

$$keyword_{sim} = \frac{Keyword(Q) \cap Keywords(C)}{Keywords(Q)} \quad (4)$$

The top ranked 10 documents are taken for consideration for the resultant answers; the paragraphs are extracted by the adding the individual sentence score from [17] as

$$S_K = \sum_{i=1}^N S_j^k \quad (5)$$

Average paragraph score is calculated using a threshold of all the paragraphs. The paragraph is extracted by using

$$S_l \geq \frac{1}{P} \sum_{k=1}^p S_k \quad (6)$$

The total score of the sentence is calculated by summing up the weighted score as given below

$$S_j^k = \sum_{i=1}^n W_i S_i (0 \leq w_i \leq 1) \quad (7)$$

### C. Answer Extraction

In the answer extraction phase, the representation of the question and candidate answer bearing texts are matched against each other to give the specific and correct answer. The multiple retrieved sentences with high coherence and score from the related documents from the knowledge base in the information retrieval phase are taken into consideration as the initial population for the cuckoo search optimization. The overall steps for cuckoo search is by

- Step 1: Initialize the CSO parameters such as Number of nests, number of eggs, fitness value, and number of iteration.
- Step 2: Random assignment of sentences to birds.
- Step 3: Birds assessment based on cost function.
- Step 4: Evaluate with fitness function.
- Step 5: Updating the bird's position for the nearest optimized result.
- Step 6: Generate new solution.
- Step 7: Terminate the process till it reaches the number of iterations, number of iteration and algorithm converge.

#### Pseudo code for Cuckoo search algorithm

---

```

Define Objective function f(x), x=(x1, x2,...,xd)
Define cuckoo search parameters
Initial a population of n host nests xi (i=1, 2... d)
While (t < Maxgen) or (stop)
Get a cuckoo (say i) randomly and generate a new solution
by Levy flights;
Evaluate quality/fitness Fi
[For maximization Fi α F(xi)];
Choose a nest among n (say j) randomly;
If (Fi > Fj)
Replace j by the new solution;
End
Abandon a fraction (Pa) of worse nests
[And build new ones at new locations];
Keep the best solutions (or nests) with quality solutions;
Rank the solutions and find the current best;
Pass the current best solutions to the next generations;
End;
    
```

---

Fig.1. Pseudo code for Cuckoo search algorithm

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 it balances local and global optima efficiently with fewer number of parameter. As the results of the cuckoo search optimization sentences scrutinized for the similarity among them using scoring techniques

like sentence similarity, sentence length, word order, word relevance and sentence to sentence relations using cosine and Levenshtein distance. The extracted sentence has optimized using the cuckoo search optimization algorithm to reduce the mean square error rate. It brings the answer nearest to the question asked and satisfies the requirement. The fitness function is for average number of mutual keywords among each sentence of paragraph and question.

From this set of such candidate answers produced and ranked according to likelihood of correctness. For example, the sentences retrieved as answers for the “Where Mahatma Gandhi is born?” mapped with the correct answer based on the context. And concentrate on the ambiguous words i.e. same word with different meaning; is resolved by analyzing the context of question and user behavior model with machine learning technique for the answer extraction.

- T train – How to train the slow learning student?
- T train – When the train is arriving at Delhi?

As a result, the sentence is displayed to the users about the book review about the features, computer related answers candidates and generic information on major areas. The fitness function calculated and performs the search on the retrieved sentences for the best global best and local best. The sentences with the maximum score have chosen for top best answer candidates.

IV. EXPERIMENTAL RESULTS

The objective of answer validation task is to review the answer returned by QA system for accurateness and relative to query asked. The result of the proposed system calculates using the standard measurements such as precision, recall and f-measure for the accuracy of the defined inference answers. In paper [23], 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 with different benchmark dataset to answer and evaluated with test set of 100 questions. The test set consists of each 30 questions on “Wh” and “How” to type questions, which capable of retrieve candidate answers from relevant documents.

The mean average precision is one of the popular performance measures in the field of information retrieval which used to evaluate the ranked relevant documents retrieved with average precision values. The Mean average precision calculated for various benchmark datasets like Amazon book reviews, 20newsgroup and Yahoo! Answer documents has capacity for the given user query exhibited in tables.

Table 3. Query Request from User and Response

Type of Occurrences	Question Type					
	what	when	where	Why	How	Accuracy
Related Domain (CONTEXT -SPECIFIC)	80.2%	75.5%	85.4%	80%	75%	79.2%
unrelated domain (GENERIC)	30%	35%	28%	25%	28%	29.2%
Repeated questions	80%	75%	85%	80%	75%	79%
Same Question asked in different Methods	20%	27%	25%	28%	48%	29.6%
Positive Tag	50%	45%	52%	26%	52%	41%
Negative Tag	42%	27%	35%	38%	28%	34%

The result of proposed system has measured with the standard metrics for correct answer by top 1 accuracy .Threshold values analyzes point of change for retrieving the answers at various levels of abstraction. The accuracy ranges of the answers from the retrieved resources 1, 5,10,15,50 have calculated for the threshold value t. The top1 accuracy is calculated by the formula

$$Top1Accuracy = \frac{No.of\ rightanswers + No.of\ right\ nilanswer\ questions}{No.of\ questions} \tag{8}$$

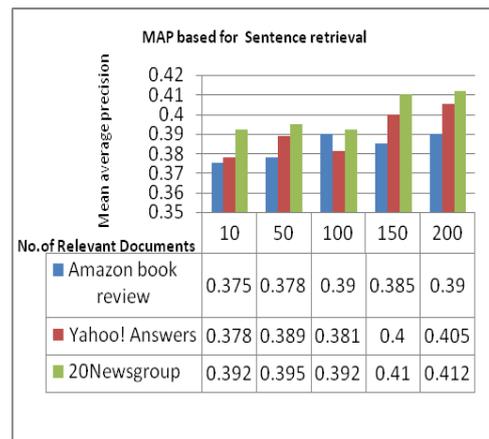


Fig.2. MAP based for Sentence retrieval

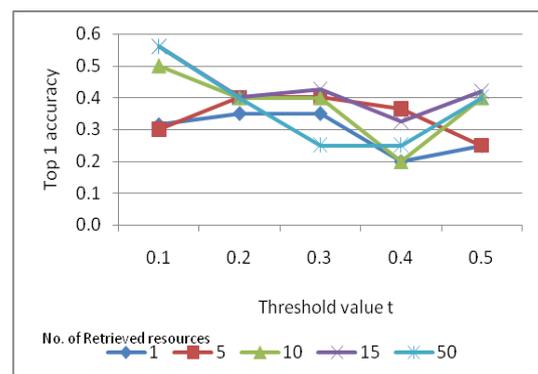


Fig.3. Number of retrieved sentences with threshold value t

The performance analysis of cuckoo search algorithm's measured with time complexity i.e. number of loops running time for each population with iterations and denoted by Big O notation. The best case scenario is at finding the exact solution in first iteration with the initial population 10 and population 20 with which reduces the search space. Worst case scenario is at finding accurate answer at n<sup>th</sup> iteration with the population 10 and population 20. The algorithm has implemented and test with the population of 10 and 20 with 5 iterations each to get the best fitness value of the appropriate candidate solution. The fitness function tests with the initial population of 10 and increased in population of 20 with equality constraint of 2 top sentences with 5 iterations.

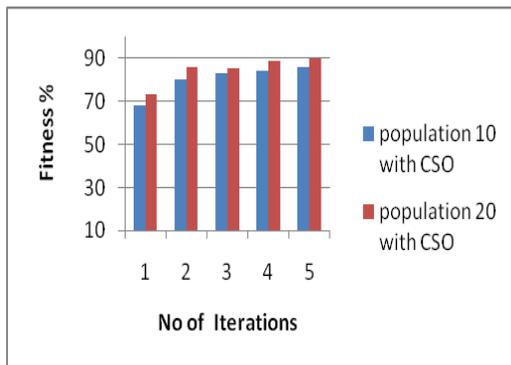


Fig.4. Compare CSO fitness function for various populations

To analysis, the nature of proposed system deployed in real world, the groups of 50 students involved in this pre-test for getting user experience and performance with age group of 20-25. Beginning of the survey with the end users, the user provided with a brief handout about domain area in QA systems knowledge base, what and how to interact with the system. They requested to interact with a system for about 15mins through the console based entry on education-based questions. The list of questionnaire 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 5 point scale where the high value for the positive side.

V. EVALUATION METRICS

In papers [15], [16] & [18] discusses on various evaluation metrics used for retrieving the information along with the Precision, Recall, Inverse precision, Inverse recall and F1. 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?

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

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?

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

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

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{11}$$

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

$$InversePrecision = \frac{TN}{FN + TN} \tag{12}$$

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.

$$InverseRecall = \frac{TN}{FP + TN} \tag{13}$$

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. The statics of the outcome are given below.

Table 4. Top N list possible Sentence ranking for 20newsgroup dataset

No. of Question	Precision @N	Recall @N	F1@N	Inverse Precision @N	Inverse Recall @N
50	0.9	0.6	0.72	0.6	0.9
100	0.85	0.53	0.65	0.5	0.83
150	0.6	0.42	0.5	0.46	0.63
200	0.65	0.5	0.56	0.56	0.70
250	0.8	0.48	0.60	0.43	0.76
300	0.75	0.53	0.62	0.56	0.77
350	0.85	0.47	0.60	0.36	0.78
400	0.55	0.34	0.42	0.3	0.5
450	0.45	0.31	0.36	0.33	0.47
500	0.5	0.38	0.43	0.46	0.58

## VI. CONCLUSION

This work motivated by the problem of developing an intelligent answering system that supports for the learning community. The proposed framework composed of a learning model for question pattern analysis, term frequency in documents, semantic similarity relation based answer extraction. The cuckoo search algorithm used for optimizing the results due to less complexity nature and best in find the global optima. It outperforms than other algorithms due to the nature of easy implementation, efficiently handling large population and less number of parameters such as number of nests, number of eggs, fitness value and number of iteration. The answers analyzed with the question keywords to the extracted sentences by word forms, word order, distance & string similarity, semantic similarity among words. The benchmark datasets 20newsgroup, Amazon review and Yahoo! Answers used to penalize the performance of the proposed methodology. The system evaluated with the standard information retrieval metrics such as mean average precision, recall and accuracy. The experiment result shows that the cuckoo search optimization with semantic similarity outperforms. The limitation of the proposed system is not contemplated on deep examination of single word questions and long sentence questions.

## REFERENCES

- [1] G.Manoranjith el, "Mobile Question Answering system based on social network", International Journal of Advanced Research in computer and communication Engineering, September, 2013.
- [2] D.Aravind Gosh, "Literature survey on mobile Q & A system in the cloud based environment", International Journal of Innovative research in computer and communication Engineering, January 2015.
- [3] Yavuz Selim Yilmaz, "Targeted Question Answering on Smartphone Utilizing App Based User Classification", IEEE, 2014.
- [4] Dr. M. Sikandar Hayat Khiyal, "Mobile System Using Natural Language Annotations for Question Answering", International Conference on Computer Technology and Development, 2009.
- [5] Jilani, Aisha, "Mobile Phone Text Processing and Question Answering. In: Future Technologies in Computing and Engineering": Proceedings of Computing and Engineering Annual Researchers' Conference, 2010.
- [6] Daniel Ortiz-Arroyo, "Flexible Question Answering System for Mobile Devices", IEEE, 2008.
- [7] Fernando Zacarias F, "mQA: Question Answering in Mobile devices", 2009.
- [8] P.Civicioglu, and E. Besdok, "A conceptual comparison of the Cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms", Artificial Intelligent Reviews, Springer, 2011.
- [9] Xin-She Yang and Deb, "Engineering optimization by Cuckoo Search", J. Mathematical Modeling and Numerical Optimization, vol. 1, no. 4, 2010.
- [10] Nitisha Gupta, Dr.Sharad Sharma, "Nature-inspired Techniques for Optimization: a brief review", International Journal of Advance Research in Science and Engineering, Vol no.5, 2016.
- [11] M. Bhuvaneswari, "Nature Inspired Algorithms: A Review", International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 12 Issue, DECEMBER, 2014.
- [12] Iztok Fister Jr., Xin-She Yang, Janez Brest, Dušan Fister, "A Brief Review of Nature-Inspired Algorithms for Optimization", Elektrotehniški Vestnik 80(3): 1-7, 2013.
- [13] Abdessamad Echihabi, Ulf Hermjakob, Eduard Hovy, Daniel Marcu, Eric Melz, and Deepak Ravichandran, "How to Select Answer String", Springer Netherlands, 2006.
- [14] Jeon, J., Croft, W., & Lee, J, "Finding semantically similar questions based on their answers", In Proceedings of the annual international ACM SIGIR conference on research and development in information retrieval, 2005.
- [15] Lomiyets, Oleksandr & Moens, Marie-Francine, "A survey on question answering technology from an information retrieval perspective", Information Sciences, 2011.
- [16] Tilani Gunawardena, "Performance Evaluation Techniques for an Automatic Question Answering System", International Journal of Machine Learning and Computing, Vol. 5, No. 4, August, 2015.
- [17] Sanglap Sarkar, "NLP Algorithm Based Question and Answering System", Seventh International Conference on Computational Intelligence, Modeling and Simulation, 2015.
- [18] Gunnar Schröder et al, "Setting Goals and Choosing Metrics for Recommender System Evaluations", 5th ACM Conference on Dresden University of Technology Recommender Systems Chicago, October 23th, 2011.
- [19] Liu, S., Liu, F., Yu, C., & Meng, W, "An effective approach to document retrieval via utilizing WordNet and recognizing phrases", In Proceedings of the annual international ACM SIGIR conference on research and development in information retrieval (pp.266-272). ACM, 2004.
- [20] Christos Bouras, Vassilis Tsogkas, "A clustering technique for news articles using WordNet", Elsevier- Knowledge-Based Systems, 2012.
- [21] Tunstall-Pedoe, William, "True Knowledge: Open-Domain Question Answering Using Structured Knowledge and inference", AI Magazine, Volume 31 Number 3, pp.80-92, 2010.
- [22] Gang Liu and Tianyong Hao, "User-based Question Recommendation for Question Answering System", International Journal of Information and Education Technology, Vol. 2, No.3, June 2012.
- [23] Ming Tan, Cicero dos Santos, Bing Xiang & Bowen Zhou, "Improved Representation Learning for Question Answer Matching", Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, August pp. 7-12, 2016.
- [24] Mansaf Alam, Kishwar Sadaf, "Web Search Result Clustering based on Cuckoo Search and Consensus Clustering", Indian Journal of science and Technology, Volume 9, Issue 15, April, 2016.
- [25] Richard Khoury, "Sentence Clustering Using Parts-of-Speech", I.J. Information Engineering and Electronic Business, Feb 2012.
- [26] Prof. Kohei Arai, "Collaborative Question Answering System Using Domain Knowledge and Answer Quality Predictor", I.J.Modern Education and Computer Science, Nov 2013.
- [27] Waheeb Ahmed et al, "Developing an Intelligent Question Answering System", I.J. Education and Management Engineering, Nov 2017.

**Authors' Profiles**

**K. Karpagam** is working as Asst. Professor in Department of Computer Applications, Dr.Mahalingam College of Engineering and Technology, Pollachi. She has 12 years of academic experience and currently pursuing research in Anna University, Chennai. The area of research interests includes text mining, Question Answering system and Human

computer Interaction.



**A. Saradha** is working as Professor and HOD, Department of Computer Science and Engineering, IRTT, Erode. She has 25+ years of academic experience. Her research interests are Semantic Web, Human Computer Interaction, Image processing.

**How to cite this paper:** Karpagam K, Saradha A," A Mobile based Intelligent Question Answering System for Education Domain", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.10, No.1, pp. 16-23, 2018. DOI: 10.5815/ijieeb.2018.01.03