

# Informative Multimedia QA using Web based Approach

Darshana D. Ambatkar  
Department of Computer Engineering  
G. H Rasoni Institute of engineering &  
Technology, Pune

Vaishali Pujari  
Department of Computer Engineering  
G.H Rasoni Institute of Engineering & Technology  
Pune,India

## ABSTRACT

Automated question answering (QA) still faces challenges such as processing and deep understanding of complex questions. In some cases, human intelligence obtains better results than automated approach. The result shows that community question answering (CQA) emerged as an extremely popular alternative to obtain information in which users are able to obtain better answers provided by other participants. But existing CQA forums mostly support textual answers, which are not informative enough for many questions. In this paper we propose a system that enriches textual answers with corresponding media data in Community QA. Our model consists of three components; query analysis for multimedia search, answer medium selection and multimedia data selection and presentation. This system automatically determines which type of media information should be added for textual answer by collecting data from web to enrich the answer. Multimedia QA scheme uses diversification methods to collect the best suitable answers based on questions and make the enriched media data more diverse. It determines the type of medium to be used by adding Nave Bayes Classifier, which helps to generate queries based on existing QA dataset pool and performs multimedia search. This scheme also performs query adaptive re-ranking and redundancy removal to obtain a set of images and videos for presentation accompanying textual answer. It uses Page Ranking algorithm which result shows that it provides more satisfactory and effective results

## Keywords

CQA, medium selection, question answering, Reranking

## 1. INTRODUCTION

Over the year, the change in social media arise a new trends with popularity in platform shifts and market evolution. As per the survey in 2014, about 700 million people are using social sites. The quantity of information on web has increased over the year. To search information on web, user needs to type a set of keyword in the form of unstructured queries and search engine display a vast quantity of data. In order to get information from web user need to browse each and every document. This type of problem solved by using Question Answering technique. Question answering is a method for automatically answering a question in a natural language. QA avoids browsing of vast quantity information from web i.e. return by keyword based search engine system to find correct answer. The problems with fully automated QA are that lack of deep understanding of complicated question, sophisticated semantic, syntactic and conceptual process to generate precise answer. But we found that automated question answer cannot generate better result as compared to human intelligence [1].

Community question answer obtains information online along with specific question on any topic and obtains precise answer form other participant. In This system people share their knowledge according to their interest which is having different categories and people can seek answer to question from them. The result obtain from community based

question answering forum are better answer because that answer generated by human intelligence.

Over the year, their vast amounts of answer and question have been accumulated to provide the facility like preservation and search of answered question. The problem with Existing system is that they support only textual answer and which is not relevant for many times, if we add accompanying multimedia content such as image or video to demonstrate the process which provide better result [2].

The existing system of community question answer such as yahoo answer, wiki answer and ask Metafilter provide answer only in textual form but some question such as How to install SQL server? .In this case, if we provide answer in textual form which is not informative for many user .Accompanying video or images provide better result, in fact some community forum provide a complementary link to demonstrate the process .It confirm that multimedia content are important to demonstrate the process[1]. In this paper, we propose a system that is rich with multimedia content such as images and video to enrich the textual answer. The proposed system contain three main component such as

- Answer medium selection
- Query analysis for multimedia search
- Multimedia data selection and presentation

## 2. RELATED WORK

In 1960s ,The research on QA has been started and mainly focused on expert system in appropriate domain .The establishment of QA track in TREC has gained its research popularity in 1990s [3]. The QA Can be categories in to open domain QA [4], Restricted domain QA [4], Definitional QA [5], and List QA [6]. As we have seen the automated QA has some difficulties in answering complex question. The existing community QA forum such as yahoo answer, wiki answer and ask Metafilter has a large and diverse community forum which share their knowledge where one can take advice and opinion from other community member .But the problem with existing community QA forum is support only textual answer which may not be provide informative and relevant information. The aim of multimedia QA is to provide precise answer with multimedia content for that video QA system was introduced. This system provide a facility by adding multimedia content to support factoid QA such as video ,images along with text based answer .Multimedia QA system relies on video optical character(VOOCR) and automatic speech recognition (ASR)[7]. To search information on web has become difficult due to vast quantity of digital information over the web. In 1980s, the research and analysis effort has been started to address the problem while searching multimedia content from a fixed

type of database [8] .The text based and content based search categories the effort made by multimedia content .The search has based on the text queries and term based classification where it matches the text with media in the web to retrieve media data. Also, in order to improve the performance of

search techniques, some machine learning methods are used which automatically annotate the entries for gathering information [9].

Now, there is tremendous improvement in search criteria with having some limitation such as cost, multimedia content and exceptions regarding visual description and semantic exceptions. Therefore, there is need of keyword based search engine for finding multimedia content [10]. The multimedia search systems are based on the text based information such as tag of entity provided by user such as title, alternative text and around web pages on texts. But sometime it may not provide accurate and appropriate information of images and videos. This type of problem solved by using reranking algorithm which mine visual information of multimedia content. The reranking algorithm uses two approaches such as pseudo relevance feedback and graph based reranking [11].

In pseudo relevance approach, showing top results which consider as a relevant answer and later collected sample which may consider irrelevant pseudo relevance. It has learned a classification or ranking model [12]. The graph based reranking approach made two assumptions. First, variance between the initial ranking list and clarify ranking list must be small. Second, visually similar data ranked to be close. This method creates a graph where vertices are images or videos and edges are their similar pair. The conventional method count the similarities from multimedia content such as colour, texture shape and visual word which extract feature from fixed set of similar multimedia data [13]. The query adaptive method used to estimate the similarities of facial characteristic queries which are classified into two classes namely related query or non related query [14].

### 3. ARCHITECTURE FOR PROPOSED SYSTEM

The fig 1 shows the overall architecture of the system. The aim of this system is to provide well defined multimedia content with textual data. This framework is a full of all packages such as modularity, flexibility, configurability, scaling, initialization and synchronization. This proposed system is dividing in to three parts defined as bellow add figure here

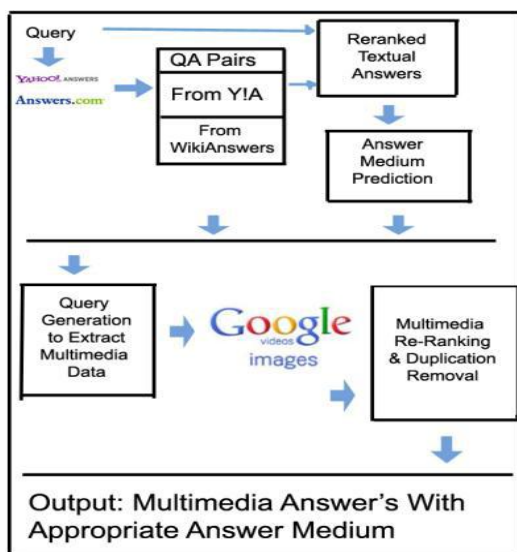


Fig 1:Architecture of proposed system

### 3.1 Answer medium selection

Answer medium selection is help to find out which type of medium will require to enrich the textual answer for Ex. How many times India won cricket world cup? Such question only needs simple textual answers. But some question will require textual answer with an image say for Ex. Who is Abdul Kallam? , So it will be more informational. Sometimes the question may be like, how to get install Linux operating system? Such example needs to explain with textual answer in addition of video that shows the complete procedure of Linux installation and it will help to understand easily. It means every question will require a different medium to enhance textual data .On the basis of such analysis here we can classify the answer based on medium as Text,Text+Image, Text+Video,Text+Image+Video. Since many question contain multiple sentences for example Y!Answer! shows ratio of at least 1/5 of the question contains two sentences and similarly 1/10 for Wiki answers and some of the sentences are uninformative. Here needs to employ a method which extracts core sentences from every question. The classification is determined by two steps. First, we discriminate questions based on interrogative words which include the words like shown in Table I. In this way we can directly find question that only need textual answer. Second, for the rest question we performed classification using Naïve biased classifier shows in Table II. It creates class specific related word and their corresponding category. The question that is posted by user is in the form of a natural language so we need tokenize the sentence and assigned parts of speech to the words. If the question contains the words that are given in the class specific related word list then appropriate categories will search for the answer medium selection.

Table1. Illustrative Questioning Words

Interrogative Word	Category
Be, there, have, when, will, can, how + adj /adv, Why, how,	Text
where, who, to, which, what	

Table2. Illustrative Class Precise Associated Words

Category	Class specific related word list
Text	Distance,Speed,Height,Weight,Age,Date,Birth day,Religion,Caste,Number,Population,Name, Website
Text + Image	Symbol, Figure, Logo Picture, Photo, Place, Color, Look like, Who, Image, Appearance, Pet
Text+ Video	How to, How do, How can, Story, Recipe, Song, Music, Dance, Film Steps
Text + Image +Video	President, Prime Minister, Singer, Battle, Issue, Earthquake, Tsunami, event, War, Happened

### 3.2 Query Analysis for multimedia search

To collect relevant image and video data from the web having community question answering site, we need to generate

appropriate queries from text QA pairs before performing search. The task accomplish with two steps. First step is query extraction and second step is query selection. Textual QA are generating complex sentences. Sometime a search engine on web do not perform well for queries which are long and verbose .Therefore we need to extract a set of informative keywords from the QA pairs for querying. In query extraction we generate three queries from each QA pair. First, we convert question to a query i.e. correct interrogative sentence into meaningful phrases. Second, we identify key concepts from verbose answers. And third, we merge two queries that are generated from question and answer respectively. Therefore we obtain three queries. The query selection is formulated as three class classification tasks; here we need to select one from three queries that are generated from question, answers and combination of QAs. It has forced us to adopt following features. POS histogram, it reflects characteristics of query. We used POS tagger to assign a part of speech to each word of QA. It generates 36 dimensional histogram. Search performance predictions, it measure clarity score for each query based on KL divergence between query and collection model. Therefore for each QA pair, we can generate 42 dimensional features.

### 3.3 Multimedia Data selection and Presentation

Text based search method which returns huge quantity of data

Sometime that data may not satisfy the question raised by users. Therefore, there is a need to reorder the data accordingly to close to that query. Graph based reranking method helps to reorder the data and remove the redundant data which are irrelevant to the user. The following equation helps to adopt reranking method as

$$r^j_{(k)} = \alpha \sum_{i \in B_j} r^i_{(k-1)} P_{ij} + (1-\alpha)r^j_{(0)} \dots\dots\dots (1)$$

The initial score for the relevant sample at jth position is calculated as

$$r^j_{(0)} = \frac{N-i}{N} \quad i = 1,2,\dots,N \quad (2)$$

Where  $r^j_{(k)}$  is a state probability of node j with  $k^{th}$  round of iterations is a parameter which satisfies  $0 < \alpha < 1$ , and  $P_{ij}$  is transition probability from data point i to j .The problem with existing graph based reranking method is that they are depends upon query with global features .If query is person related then we have to use facial feature so we used query adaptive reranking method. The categorization is based on person related and non person related query. For that we used different types of features to re-rank the search result.

## 4. EXPERIMENTAL RESULT

The Data set for experiments contain two subsets. First we randomly collect some question from wiki Answer and for second subsets we collect some question and their relative answers from Y!A .We select best answers by using ranking method that is vote by the user. The question can classify in to two categories i e conversational and informational .Conventional question take opinion from community users such as anybody watch dance show on TV last night? The informational questions are asked to get information for the intention to learn something such as what is the population of India? We perform categorization with human labeling. We classify QA Pairs in our dataset such as travel, life, education and others. Next we will see evaluation of each module.

### 4.1 Answer Medium Evaluation

To evaluate our answer medium selection approach we mentioned the label to involved in the ground truth labeling process .They are used to investigate web information from answer medium information for example How can I extract the juice from swat lime at home? The community user will expected multimedia answer .The answer will investigate after the labeler’s practical investigation on the web. If they find insufficient images and video resources related to this topic and this question labeled as text. We can see that more than 50% of the question can be answered by adding multimedia contents instead of purely text. So we can conclude that multimedia approach highly desired.

### 4.2 Query Generation Evaluation

Here we evaluate query generation and selection approach. As previously mentioned .five labelers participate in the ground truth labeling process. Each labeler selects the most informative one. Query generation allowed performing search on the web to compare better search result. The final ground truths are obtained by majority voting.

### 4.3 Reranking Evaluation

Here we evaluate a method of judging whether a QA pair is person related or non person related .For this we randomly collect each dataset and they are depends on two subsets. Here we used SVM model with RBF kernel based on 7 dimensional facial characteristics. Our approach achieve good performance

### 4.4 Results

According to study the comparison between original textual answer and media answer generation shows without the actual Assistance of textual answer. The experimental settings present the user study result in table 3. According to study more than 50% people prefers media answer along with textual answer. It is more important community member to provide the answer with media data to better understand the question.

**Table.3. Illustrative Comparison of textual answer with multimedia answer.**

Prefer answer	media	No answer	Perefer original textual answer
52%		16%	35%

It is interesting to see that although the media answer are not as informative as those generated with the assistance of textual answer .They are still informative in comparisons with pure textual answer. The graph shows average informativeness of textual answer and without textual answer.

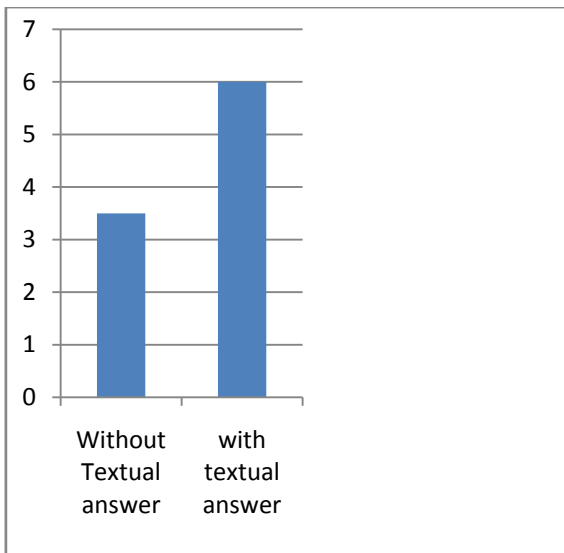


Fig 2: Comparison of community answer between textual answer and without textual answer

#### 4. CONCLUSION AND FUTURE WORK

In this paper, we are elaborating the information about “A multimedia Answer Generation System”. Here our approach is built based on the community contributed answers instead of conventional MMQA research that can automatically generate answers. During research on it we have analyzed some drawback & limitations of the existing approaches mainly focus on narrow domains, that’s why looking for more general & wider approach to get more accurate results to the world. While working on cQA our scheme first determine the type of appropriate medium to enrich & enhanced the queries generated by answer selection method. Here we have shown the generation of queries in three different types by extracting it from questions, answers & question-answers posted by users. These queries help to extract the core sentence from the bunch of complex sentences. On the basis of this core sentence which is nothing but having some keywords which we have been proposed here, it is easy to understand & search the relevant & precise answer. During answer generation it accumulates & accompanies the images, videos with pure textual answer from dataset pool.

During analysis have seen some limitations & failures. For example, the system may not be able to provide reasonable multimedia data if the questions are complex & verbose. Sometimes for questions videos, images are enriched but only parts of them are informative. Then presenting the whole video can be unnecessary & misleading. We have seen lack of diversity as well; to overcome it we have adopted a method to remove redundant data. Here in this paper we would like to propose some evolutionary approaches which help to mine & extract the data for more accurate results. Subsections

#### 5. ACKNOWLEDGMENTS

I express true sense of gratitude towards my project guide Prof. V.R.Pujari, for her invaluable co-operation and guidance that she gave me throughout my project. I specially thank our P.G coordinator Prof. Amit Dravid, H.O.D. Prof. D. Bharadwaj for inspiring me and providing me all the lab facilities, which made this cPGCON and project work very convenient and easy. I would also like to express my appreciation and thanks to GHRIET principal **Dr. R.D. Kharadkar** and all my friends who knowingly or unknowingly have assisted me throughout my hard work

#### 6. REFERENCES

- [1] Nie, Liqiang, et al. "Beyond text QA: Multimedia answer generation by harvesting Web information." *Multimedia, IEEE Transactions on* 15.2 (2013): 426-441.
- [2] Chua, Tat-Seng, et al. "From text question-answering to multimedia QA on web-scale media resources." *Proceedings of the First ACM workshop on Large-scale multimedia retrieval and mining*. ACM, 2009.
- [3] Nie, Liqiang, et al. "Multimedia answering: enriching text QA with media information." *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011.
- [4] Moschitti, Alessandro, and Silvia Quarteroni. "Linguistic kernels for answer re-ranking in question answering systems." *Information Processing & Management* 47.6 (2011): 825-842.
- [5] Hsu, Chih-Hao, et al. "Using domain ontology to implement a frequently asked questions system." *Computer Science and Information Engineering, 2009 WRI World Congress on*. Vol. 4. IEEE, 2009.
- [6] R. C. Wang, N. Schlaefler, W. W. Cohen, and E. Nyberg, "Automatic set expansion for list question answering," in *Proc. Int. Conf. Empirical Methods in Natural Language Processing*, 2008.
- [7] E. Parzen and F. Hoti, "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, vol. 33, no.3, 1962.
- [8] M. Wang and X. S. Hua, "Active learning in multimedia annotation and retrieval: A survey," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, pp. 10–31, 2011
- [9] Y. Gao, M. Wang, Z. J. Zha, Q. Tian, Q. Dai, and N. Zhang, "Less is more: Efficient 3d object retrieval with query view selection," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 1007–1018, 2011.
- [10] I. Ahmad and T.-S. Jang, "Old fashion text-based image retrieval using FCA," in *Proc. ICIP*, 2003.
- [11] D. Liu et al., "Tag Ranking," *Proc. 18th Int'l Conf. World Wide Web*, ACM Press, 2009, pp. 351-360.
- [12] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua, "Bayesian video search reranking," in *Proc. ACM Int. Conf. Multimedia*, 2008.
- [13] Liu, Shaowei, et al. "Social visual image ranking for web image search." *Advances in Multimedia Modeling*. Springer Berlin Heidelberg, 2013. 239-249.
- [14] H. Feng, A. Chandrashekhara, and T.-S. Chua, Tamra." *An Automatic Temporal Multiresolution Analysis*
- [15] Framework for Shot Boundary Detection, *Proc. Int'l Conf. Multimedia Modeling (MMM)*, ACM Press, 2003
- [16] .S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural SceneCategories," *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR)*, IEEE CS Press, 2006.
- [17] D.R. Radev et al., "Evaluating Web-based Question Answering Systems," *Proc. Int'l Conf. Language Resources and Evaluation*, 2002