

# Generation of Understandable Answer to a Query Using Multimedia Web Information

Y Srikanth<sup>1</sup>, M Thenmozhi<sup>2</sup>

<sup>1</sup>Department of IT, SRM University, Chennai, India

<sup>2</sup>Department of IT, SRM University, Chennai, India

<sup>1</sup>[sri.yenamandra@gmail.com](mailto:sri.yenamandra@gmail.com), <sup>2</sup>[thenmozhi.m@ktr.srmuniv.ac.in](mailto:thenmozhi.m@ktr.srmuniv.ac.in)

**Abstract:** Automated question answering (QA) still faces challenges such as processing and deep understanding of complex questions. In most cases human intelligence obtain better results than automated approach. As a result 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 provide textual answers, which are not informative enough for many questions. In this paper we propose a model that enriches textual answers with corresponding media data in CQA. Our model consists of three components; query analysis for multimedia search, answer pattern selection and multimedia data selection and presentation. This model automatically determines which type of media information should be added for textual answer by collecting data from web to enrich the answer.

**Index Terms:** Community question answering (CQA), Query analysis, Answer pattern selection, Multimedia data selection and presentation.

## I. INTRODUCTION

As rapid growth of internet information searching has become normal activity in people's daily life. The usage of search engine for information retrieval is increasing enormously over the years. Traditional search engines provides ordered list of web page results based on keyword based matching. This approach provides confusion to the users for vast quantity of information returned in response by search engines. In addition to this the user also finds it difficult to get the exact answer among the returned response by search engine. As a result Community question answering (CQA) emerged as an alternative to the users where the users can get the answers provided by other participants [1]. It not only provides answers to the users but also acts as a platform to users where they can share their answer, discuss their opinion and rate the answer. Moreover it generates flexibility to the user in order to get the best answer and also allow users to accumulate more question answer pairs for preservation and retrieval of answered questions in CQA repositories [2]. For example Wiki Answer, one of the most well-known CQA systems, hosts more than 13 million answered questions distributed in 7,000 categories.

But the disadvantage in the existing community question answering (CQA) forums mostly provides only textual answers which are not informative for many questions. Textual answer itself is not sufficient enough to the user for understanding and memorize the content for various questions. In fact users generally post URLs (Uniform resource locator) for linking textual answers with corresponding images and videos as shown in Figure.1. This confirms the importance of multimedia [3] content. On the other hand the increasing growth of multimedia content such as image and video over the web has been witnessed the importance of multimedia search. For example, YouTube serves 100 million distinct videos and 65,000 uploads daily, and the traffic of this site accounts for than 20% of all web traffic and 10% of whole internet, comprising 60% of video watched online. The photo-sharing containing site Flickr contained more than 4 billion images. This clearly shows the importance of multimedia content.

In this paper we propose a model that enrich the textual answers with the corresponding media data in CQA. The Figure .2 illustrates the schematic approach. It comprises of mainly three components:

1. Query analysis for searching the multimedia content. In order to find the multimedia content we need to find the informative word which helps to retrieve the corresponding content.
2. Answer pattern selection: For a given question it predicts which type of media data should be added. Here we will categorize mainly into four types they are: text, text+image, text+video, text+image+video. It means this approach will decides which type of data should be added for enriching the textual answers. Here we are not considering audio as most of the content can be represented in text form.
3. Multimedia data selection and presentation: Based on the generated informative word we will search and vertically collect corresponding media data such as video and image using multimedia search engines. Multimedia search engines are those which will provide only media data. For example YouTube for videos, Google images, Picasa and flickr for images.



Figure.1 Examples of QA from several popular CQA forums

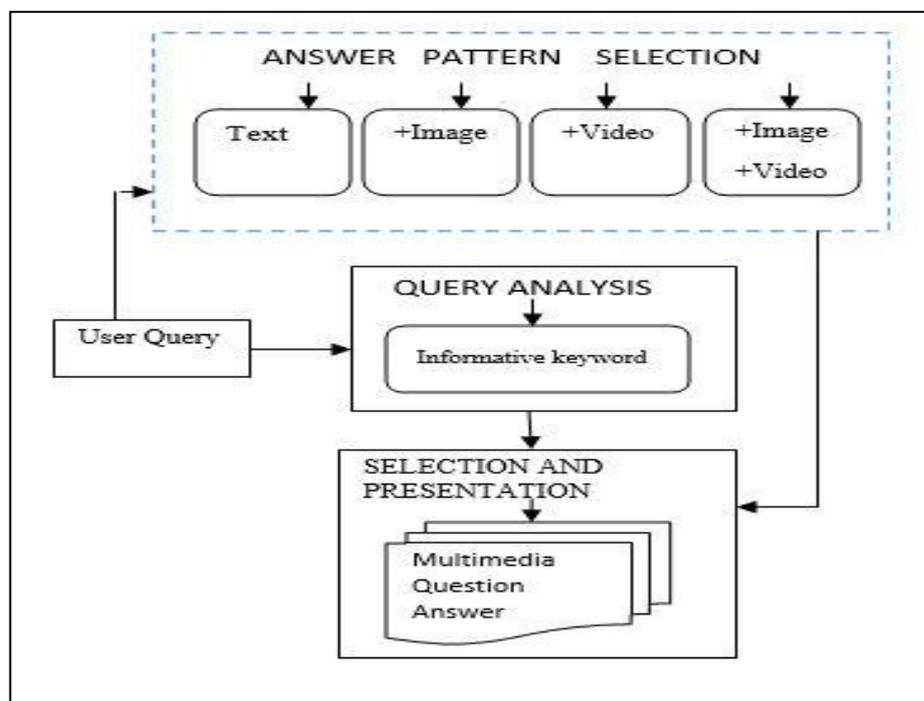


Figure. 2 Schematic Approach

## II. System Description

### Query Analysis

As mentioned in Section I, the first component of our model is query analysis. It helps to find the informative keyword for searching corresponding media data using multimedia search engines. The main objective of this process is to find the stem word which is considered as the informative keyword [4]. Here we are using an algorithm called stemming algorithm. Stemming algorithm is generally used to remove the stop words which can be applied as follows

1. The first step is to consider the given query and initialize the empty variable of string data type
2. Split the query based on the space between them and pass them into array list of string type.
3. Initialize for loop and remove the stop words i.e., a, and, an, in, be, for and soon by passing the words in the array list .continue the process until length of the array list.
4. Pass the remaining words into empty string variable initialized in step 1.
5. Finally use the obtained words as informative keyword for search and vertically collect the media data.

### Answer pattern selection

The second component of our model is answer pattern selection. In this model the given question is judged whether it requires any media data or it requires only textual answer. Here we will categorize mainly into four types such as text, text+image, text+video, text+image+video based on the given question [5]. It is not sufficient for the user to understand if we provide only textual answers. For example, for the question “how to cook fish” we may find the answer as “clean it and cut it into ...” This clearly indicates that it can be better understandable if we provide video answers. Similarly for the question “who is ajith” it is better

understandable if we provide images along with text. It is not necessary to provide media data for all the questions only textual answers is sufficient.

Table 1. Representation of four classification model.

Categories	Class-Specific Related Word List
<b>Text</b>	Name, population, period, times, country, height, website, birthday, age, date, rate, distance, speed, religions, number, etc.
<b>Text+Image</b>	Color, pet, clothes, look like, who, image, pictures, appearance, largest, band, photo, surface, capital, figure, what is a, symbol, whom, logo, place, etc.
<b>Text+Video</b>	How to, how do, how can, invented, story, film, tell, songs, music, recipe, differences, ways, steps, dance, first, said, etc.
<b>Text+Image+Video</b>	President, king, prime minister, kill, issue, nuclear, earthquake, singer, battle, event, war, happened, etc.

For example, for the questions like “distance between vellore and chennai “ and “population of India” so on. These type of question will be understandable by the textual answer itself. The verb in the question acts as a clue for judging whether the answer can be enriched with video. Based on the verbs we also build Naïve Bayes classifier on set of data and then produce a four classification model as shown in the following table 1.

#### Multimedia selection and presentation

The third component of our model is multimedia data selection and presentation. For those question which require media data such images and videos we will vertically collect media data by using multimedia search engines such as YouTube for videos and Google images API(Application programming interface). Sometimes these multimedia search engine provides some irrelevant results. In order to overcome this we have to rearrange the result. Here we adopt [6] the reranking method based on graph which is based on exploring visual information. We re-state the equation as

$$r_{(K)}^j = \alpha \sum_{i \in B_j} r_{(k-1)}^i P_{ij} + (1 - \alpha) r_{(0)}^j$$

Where  $r_{(K)}^j$  stands for the state probability of node  $j$  in the  $k$ th round of iterations,  $\alpha$  is a parameter that satisfies  $0 < \alpha < 1$ , represents the number of edges connected to  $j$  node and  $P_{ij}$  is the transition probability from data-point  $i$  to  $j$ . Here  $\mathbf{P}$  is a row-normalized transition matrix obtained from similarity matrix  $\mathbf{W}$ , and  $r_{(0)}^j$  is the initial relevance score of the sample at the  $j$ th position, which is heuristically estimated as

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

For images, each element of the symmetric similarity matrix  $\mathbf{W}$  is measured based on  $K$ -nearest-neighbor ( $K$ -NN) graph,

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) & \text{if } j \in N_k(i) \text{ or } i \in N_k(j) \\ 0 & \text{Otherwise} \end{cases}$$

Where  $N_k(i)$  denotes the index set for the  $k$  nearest neighbors of an image computed by Euclidean distance. In our work, we empirically set  $k=0.3 \times N$ , where  $N$  is the number of images collected for a query. The parameter  $\sigma$  is simply set to the median value of the Euclidean distance of all images pairs.

We perform shot boundary detection for videos and then extract a key-frame from each shot using method for example two videos ( $V_{i,1}, \dots, V_{i,m}$ ) and ( $V_{j,1}, \dots, V_{j,n}$ ), which contain  $m$  and  $n$  key-frames respectively, we employ average distance of all cross-video key-frame pairs for similarity estimation, i.e.,

$$W_{ij} = \begin{cases} \exp\left(-\frac{\sum_{q=1}^m \sum_{p=1}^n \|v_{i,q} - v_{j,p}\|^2}{MN\sigma^2}\right) & \text{if } j \in N_k(i) \text{ or } i \in N_k(j) \\ 0 & \text{otherwise} \end{cases}$$

Similarly,  $N_k(i)$  denotes the index set for the  $K$  nearest neighbors of a video measured by Euclidean distance and the parameter  $\sigma$  is simply set to the median of the Euclidean distance of all video pairs

### III. RESULTS

Using java server pages in java programming language, net beans integrated development tool(IDE) and glassfish server we had obtain the results as follows (Figure.3)



Figure 3(a). Text +image

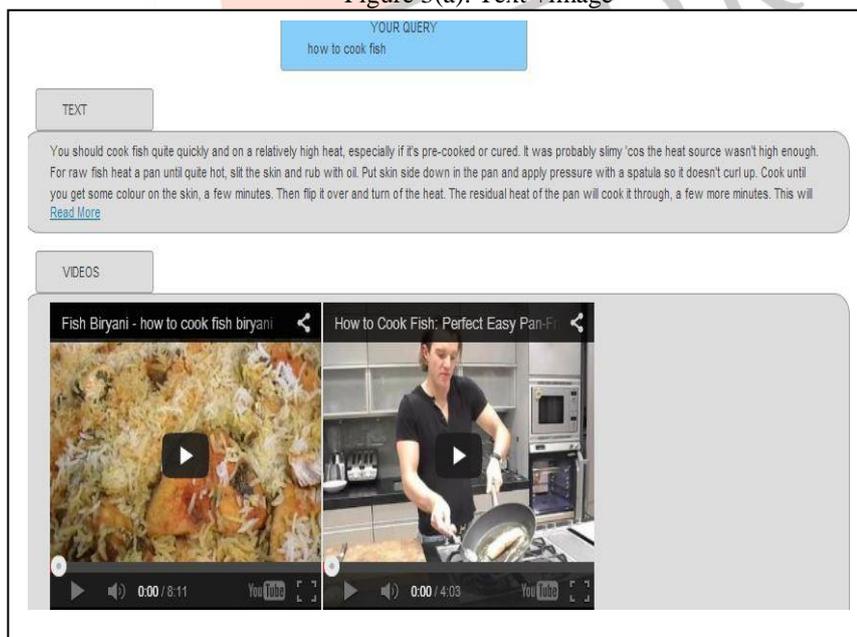


Figure 3(b). Text +video

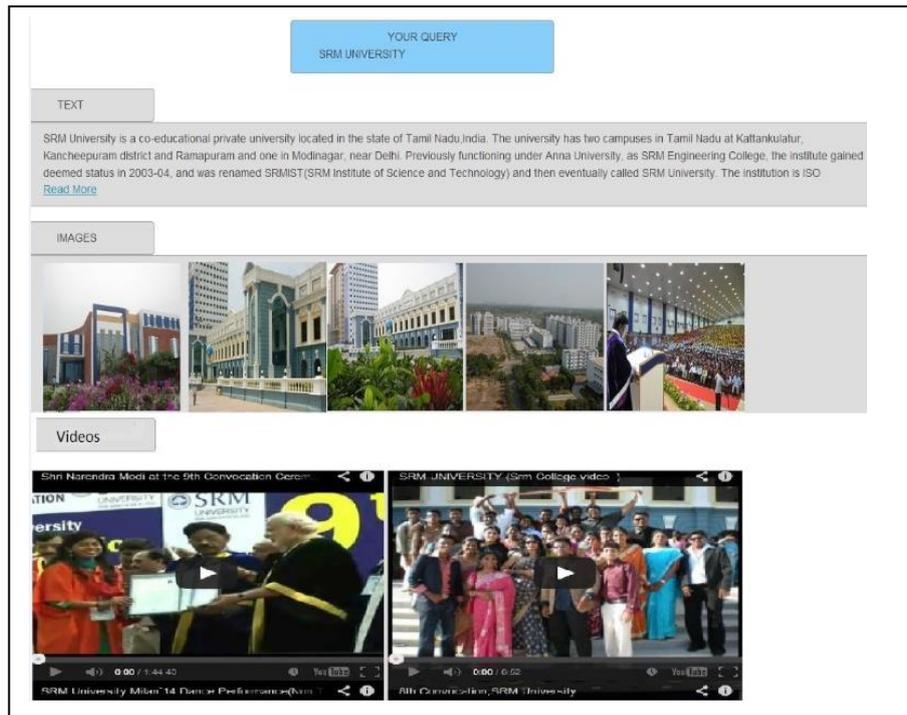


Figure 3(c). Text +image+Video

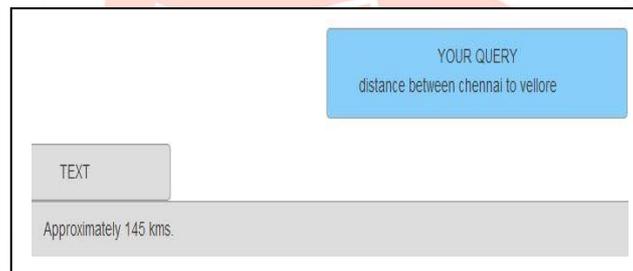


Figure 3(d). Text

#### IV. CONCLUSION

In this paper we describe a schematic approach that enrich the textual answer in CQA with corresponding media data . It automatically decides what type of medium answer should be added to present answer in order to make user more understandable. In addition to that we performed reranking based on graph for images and videos to make it more relevant . In this approach we also find some failure results, the system fails to generate multimedia answer for complex question such as containing more words. we also used a method to reduce irrelevant results but sometimes it is better to produce diverse results which is observed in [7] to produce more diverse media data.

#### REFERENCES

- [1] L.A. Adamic, J. Zhang, E. Bakshy and M.S. Ackerman, "Knowledge sharing and Yahoo answers: Everyone Knows something" in *Proc. Int. World Wide Web Conf.*, 2008.
- [2] G. Zoltan, K. Georgia, P. Jan and G-M Hector, "Questioning Yahoo! Answers," Stanford InfoLab, 2007, Tech. Rep.
- [3] R. Hong, M. Wang, G. Li, L. Nie, Z. J. Zha, T. S. Chau "Multimedia Question Answering," *IEEE, Multimedia*, pp. 72-78, 2012.
- [4] Yi Chen, Wei Wang, Ziyang Liu "Keyword-based search and exploration on databases" *IEEE conf. ICDE*, pp. 1380-1383, 2011.
- [5] L. Nie, M. Wang, Y. Gao, Z. J. Zha, T. S. Chau "Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information", *IEEE, Multimedia*, vol 15, no. 2, pp. 426-441, 2013.
- [6] W.H. Hsu, L.S. Kennedy, S-F Chang, "Video search reranking through random walk over document-level context graph," in *Proc. ACM Int. Conf. Multimedia*, 2007.
- [7] M. Wang, K. Yang, X-S Hua, and H-J Zhang, "Towards a relevant and diverse search of social images," *IEEE Trans. Multimedia*, vol. 12, no. 8, pp. 829-842, 2010.