

## **An efficient method for Multimedia Answer Generation for complex queries by Harvesting Web Information**

Manoba J P R <sup>1</sup>, Prathibha S Nair<sup>2</sup>

<sup>1</sup>Mtech CSE Student

<sup>2</sup>Assistant Professor, CSE, Department of Computer Science  
Mohandas College of Engineering & Technology, Trivandram

---

**Abstract-** QA(Question Answering) has emerged as an alternative method for retrieving information online. The information seekers can able to post their questions on any topic. The seekers obtain textual answers for their questions. This paper explains about MMQA, by which the users are able to get better answers including multimedia contents. The existing QA forums mostly support only textual answers or the URLs. But sometimes the textual answers may not provide sufficient information and it may be difficult to grasp the textual information. So, MMQA is a scheme that is able to enrich textual answers in cQA with appropriate media data which is more informative for users. But it fails to generate proper results for the complex large questions. Document Summarisation is the technique which is for summarising the large complex questions, from that it can generate appropriate multimedia answers based on the questions

**Keywords—** QA, MMQA

---

### **I. INTRODUCTION**

The amount of information on the Web has grown exponentially over the years. So as a result, when looking for information, users are often bewildered by the vast quantity of results from search engines. So, to look for a correct answer, users have to browse through a long list of results. Therefore Multimedia Question-answering (MMQA) is a technique to overcome this information-overload problem. Instead of returning a list of long textual results, it is able to retrieve correct results including multimedia contents such as video, image. For answering questions such as how to copy images to a computer, requires understanding the relevant content and often involves the composition and generation of specific answers. This is beyond the capability of current technologies except for narrow domains. Because of the strong demands for the services like community-based QA services, such as Yahoo! Answers (YA), have become popular. In YA services, people ask all types of questions and obtain the answers either searching for previously asked questions on their own or waiting for other users to provide the best answers. Even the best text-based answer is presented to users, the user might still have difficulty in grasping the answers. But if QA presents visual answers, it would be more intuitive and instructive for users. In addition to normal textual instructions, visual descriptions would be a complementary source of information for users. The video contents are evident in sites such as YouTube and other specialized video and image-sharing sites and blogs. This work does not aim to directly answer the questions, but instead, multimedia contents are enriched with the textual answers. So the MMQA strategy splits the huge gap between question and multimedia answer into two smaller gaps, i.e., the gap between question and textual answer and the gap between textual answer and multimedia answers. In this paper, a MMQA scheme which is used to enrich textual answers in QA for the complex queries with appropriate media data. It contains three main components: (1) Answer medium selection: For a query, it predicts which media data should be enriched along with the textual answers and for the complex queries, the textual summarization technique is applied and the summary is extracted and from the summary the answer type is selected. The answers will be categorized into one of the four classes: text, text +

image, text+ video, text + image + video.ie, MMQA will automatically collect images, videos, or the combination of images and videos to enrich the original textual answers.

## II. RELATED WORKS

The question answering schemes are of different types. Community question Answering(cQA) forums provide only textual answers, which are not informative enough for many questions. Yahoo! Answers [6] is a type of community portal that allows users to post questions and/or answer questions asked by other members of the community. Yahoo! Answers facilitates the retrieval of answered questions which is aimed at building an online knowledge base. Opinions expressed in the form of votes and comments. Yahoo! Answers provide only the textual answers.

An Interactive question answering (QA) [1] is another approach where a dialogue interface enables the clarification of questions. The InteractiveQA reports on design and implementation of YourQA, an open-domain, interactive QA system. YourQA is based on a Web search engine to obtain answers to both fact-based and complex questions, such as descriptions and definitions. The two main components of the system are- User Modelling (UM) component and Core QA module. User Modelling consists of creating a model of the target user's characteristics. The user modeling parameters are based on Age range, Reading level (basic, medium, good) and Interests. The Core QA component in YourQA provides both factoid and complex answers.

Content Based Image Retrieval (CBIR)[31] is a technology that helps to organize digital image by their visual content. The common form of CBIR is an image search based on visual contents. CBIR search images from large image databases according to users' requests which is in the form of a query image.

Photo-based question Answering [] is another way of finding information about physical objects. Current question answering (QA) systems are text-based. It is difficult to answer when a question involves an object with specific visual features. Photo-based QA system directly use photo to refer to the object.

Even-though the best text-based answer is presented to users. The user might have difficulty for grasping the answers. Because , from the textual answers the users might have no idea about answers that involve technical terms. The Multimedia answering is provided for the question of how USB cable to connect a digital camera to a PC. For getting the community contributed web videos, along with the textual descriptions, images and videos are included for the better understanding.

Multimedia Answer Generation by harvesting web information [] mainly focuses on generating multimedia answers along with the textual descriptions. From the question type, the system generates multimedia contents. But it fails to produce appropriate answers for complex queries.



*Fig 1: Extending textual question answering to multimedia question answering.*

[2] Elena Lloret, Text Summarization explains about the summarization of the document. However the MMQA generate appropriate media data with the textual answers, it fails to generate proper answer for the complex question. By text summarization the summary of the complex questions are obtained by using statistical sentence scoring method.

### III. SYSTEM ARCHITECTURE

Multimedia Question Answering System is a technique of automatically answering for a question with appropriate media contents such as image, video or combination of image and video MMQA contains three components. They are:

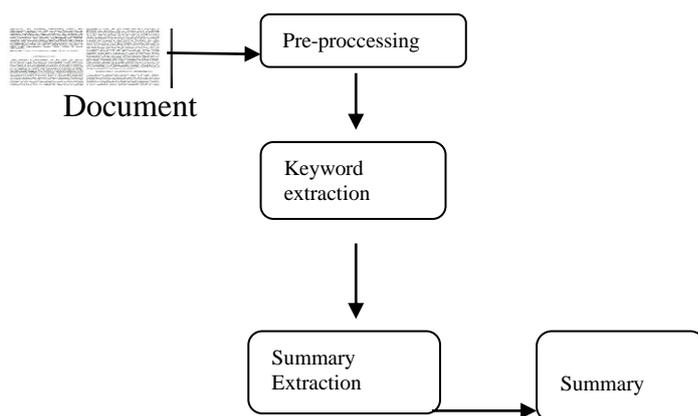
- Answer Type selection
- Query generation
- Multimedia search and Presentation

The first component of MMQA scheme is answer medium selection. It determines which type of medium should be added to enrich the textual answers based on the questions. Recently QA have placed more emphasis on generating answers for questions such as “How to” and “Why?”. Multimedia answers are especially applicable to these types of questions.

#### a. Answer Type Selection

For a Question or answer, it determines which type of medium should be added to enrich the textual answers for the given question. Each QA pair is categorized into anyone of the four predefined classes: text only; text and image; text and video, or text, image, and video. This scheme then automatically selects the more informative query for multimedia search. For the questions, such as “*When did first bomb blast happened in Japan*”, pure textual answers are sufficient. But for some other questions image or video information is added. For example, for the question “*Who is Sachin Tendulkar*”, it is better to add images to the textual answer, and add the videos for answering the question “*How to install Bluestack app in computer*” . The classification for Answer medium selection can be done in different ways: Question based classification, Answer based classification, Media resource analysis.

If any complex query is arised by an user, for example a document, the summary of the document should be obtained for selecting the medium of answer. The document summarization can be done by statistical sentence scoring method. This can be done by calculating a score for each sentences in the document.



*Fig: Statistical scoring method for document summarization*

The input complex query or a document is fed to preprocessing. In pre-processing, sentence breaking is done then stop words and stemming is performed. The next step is to find keywords or words which occur frequently in the document and for the summary extraction, a score is calculated for each sentences.

- **Keywords:** Keywords are the words which occur more frequently in the document.

- **Title words:** Sentences containing words that appear in the title have more chance of contained in the summary.
  - **Location:** Sentences in the first and last paragraphs are have greater chances for contained in the summary.
  - **Cue words:** The words which denotes the importance of a sentence. For example, the word 'important' is considered to be a cue word.
  - **Quotations:** Sentences which contain quotations are considered important.
- The score for each sentence is the weighted sum of the various features present in the sentence.

$$\text{Score}(S_i) = E(w_j) * (iij)$$

Where,

$S_i$  - ith Sentence

$W_j$  - Weight of jth feature

$iij$  - Number of times jth feature occurs in ith Sentence

The summary obtained from the document is classified into type of answer should be presented to user. If the query is a simple question, the question is classified based on query. The classification can be done by:

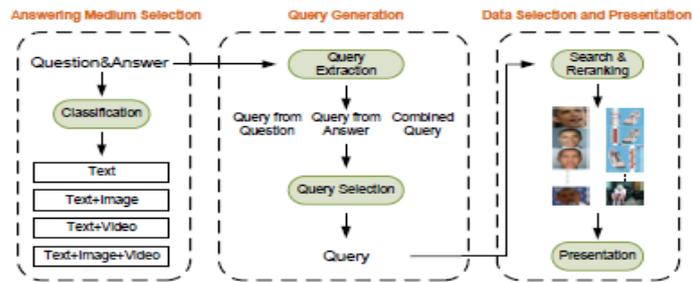
#### i) Question based classification

For the question based classification the questions can be categorized based on interrogatives (some starting words and ending words), and in this way one can directly find questions that should be answered with text. The Questions can be mainly categorized based on interrogative words: yes/no class (such as “Does Film star Ajith is married?”), choice class (such as “Which state is bigger, UP or Rajasthan ”), quantity class (such as “When was the first computer invented ”), enumeration class (such as “Name the neighbouring countries of India ”), and description class (such as “What are the ways of reducing the soil pollution”). For example, if the interrogative is “how + adj /adv” or “when” a question will be categorized into the “quantity” class. For the “yes/no”, “choice” and “quantity” questions, it is categorized into the class of answering with only text, whereas the “enumeration” and “description” questions need “text”, “text + image” or “text+ video” answers. ie, for a given question, first judge whether it should use only textual answer based on the interrogative word

Along with these classes, for the summary obtained from the complex queries, head words are considered for classification. Head words are the words specifying the object of the question. The head words also play an important role in determining answer medium. The textual answers are obtained based on clarity score of the document. Clarity score is the ratio of count of words in a given query to the total count of words in a document. The document with lowest clarity score has highest priority.

#### ii) Answer based classification

Besides question, answer can also be an important information clue. For example, for the question “how do you make cake in pressure cooker”, the textual answer is as “mix the sugar with maida powder, put in oven proof dish”. Thus, from the answers one can judge that the question can be better answered with a video clip. The verbs in an answer will be useful for judging whether the answer can be enriched with video content.



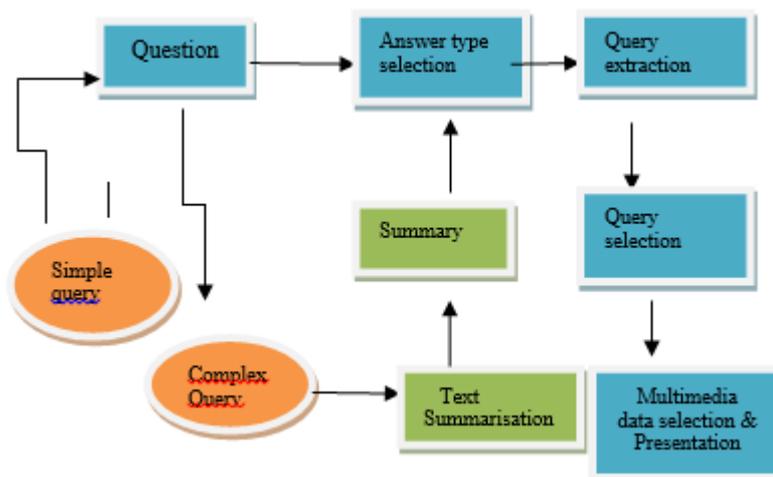
*Fig: MMQA Architecture*

### b. Extracting Queries from Question

In order to collect multimedia data, informative queries are extracted from questions. For a QA pair, and extracts three queries from the question, the answer, and the QA pair, respectively. The most informative query will be selected, and it is important to generate proper queries from text QA pairs before searching on multimedia search engines to generate appropriate answers. Extracting query involves two steps. The first step is query extraction. If the textual questions and answers are complex sentences, it is important to extract a set of informative keywords from questions and answers for querying. The second step is query selection. This can generate different queries: one from question, one from answer, and one from the combination of question and answers. Some QA pairs embed the useful query terms in their questions, such as “*What does train tickets look like*”; some will hide the key words in their answers, such as the QA pair “*Q: What is the capital of Tamilnadu; A: Chennai*”; and some should combine question and answer to get proper answers.

### c. Multimedia data selection and Presentation

The search is performed using the generated queries to collect image and video data with Google image and video search engines respectively. The commercial search engines, such as Google, Yahoo usually index their web images using textual information, such as titles. Re-ranking is an approach to improve the search relevance by mining the visual information of images and videos. After re-ranking, visually similar images or videos may be ranked together. So it is necessary to perform a duplicate removal to avoid redundancy in images. The ranking list is checked from top to bottom. If an image or video is similar to a sample that appears above it, remove that particular image or video.



#### IV. CONCLUSION

The MMQA (Multimedia Question Answering) enrich textual answers with media Information. For a given QA pair, the MMQA first predicts which medium is appropriate to enrich the textual answer. It then generates a query based on the QA and retrieves relevant image and video from search engines. MMQA aims to automatically generate multimedia answers with simple questions. If any complex question came, it failed to achieve good results, so text summarization is a method used for summarizing the complex text for getting the summary of the complex text. Text summarization can be done using statistical sentence scoring method ie.,by calculating scores of each words based on some features of words. After obtaining the summary, this can be used for medium type selection and multimedia search and thus it can deal with complex questions and achieve better performance.

#### REFERENCES

- [1]. T. S. Chua et al., “From Text Question- Answering to Multimedia QA on Web-Scale Media Resources,” Proc. ACM Multimedia Workshop Large-Scale Multimedia Retrieval and Mining (LS-MMRM), ACM Press, 2009, pp. 51\_58.
- [2]. H. Yang et al., “Structured Use of External Knowledge for Event-based Open-Domain Question- Answering,” Proc. 26th Ann. Int’l ACM SIGIR Conf. Research and Development in Information Retrieval, ACM Press, 2003, pp. 33\_40
- [3]. T. Yeh, J.J. Lee, and T. Darrell, “Photo-Based Question Answering,” Proc. 16th ACM Int’l Conf. Multimedia, ACM Press, 2008, pp. 389\_398.
- [4]. R. Datta et al., “Image Retrieval: Ideas, Influence, and Trends of the New Age,” ACM Computing Surveys, vol. 40, no. 2, 2008, article no. 5.
- [5]. H. Cui, M. Y. Kan, and T. S. Chua, “Soft Pattern Matching Models for Definitional Question Answering,” ACM Trans. Information Systems, vol. 25, no. 2, 2007, article no. 8.
- [6]. K. Wang et al., “Segmentation of Multi-Sentence Questions: Towards Effective Question Retrieval in cQA Services,” Proc. 33rd Int’l ACM SIGIR Conf. Research and Development in Information Retrieval, AMC Press, 2010, pp. 387\_394.
- [7]. R. Hong et al., “Beyond Search: Event Driven Summarization for Web Videos,” ACM Trans. Multimedia Computing, Comm., and Applications, vol. 7, no. 4, 2011, article no. 35.
- [8]. R. Hong et al., “Mediapedia: Mining Web Knowledge to Construct Multimedia Encyclopedia,” Advances in Multimedia Modeling, LNCS 5916, Springer, 2010, pp. 556\_566.
- [2] Elena Lloret, Text Summarization : An Overview, University of Alicante, 2007

