ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 3, pp: (405-412), Month: July - September 2014, Available at: www.researchpublish.com

Generating Multimedia Information Using Web Data Mining

[1]SharathBabu.S, [2]Ravi.M

[1, 2] CMRIT CSE, Hyderabad, India

Abstract: **Question answering is the most common technique to collect information from different sources. One such source which gave rise in recent internet technology is Community Question Answering (cQA). Earlier the cQA gave information only in textual format which is not very much informative for many questions posted by the user, however it is natural to extend text based question answering research to Multimedia Question Answering which gave importance for many reasons. First most video contents are indexed with metadata, second many questions are better explained with the help of non textual medium, third media contents especially videos are now used to convey many types of information. In this project a technique is proposed which enriches textual answers with appropriate media data. This technique consists of three components to enrich textual data: Answer medium selection, Query generation for multimedia search, Multimedia data selection and Presentation. The multimedia question answering complements text QA with whole QA paradigm i.e., image video along with text. This approach or technique automatically selects media information for appropriate textual answer questioned by users or community members. Here Question answering languages leverages advance media content, linguistic analysis, and domain knowledge to return precise answers to questions posted by community members.**

Keywords: **Question Answering, Answer medium selection**

I. INTRODUCTION

Now a day's users using the web engines such as Google, Bing, and yahoo are over whelmed, to overcome the information overload problem lot of research is going on with respect to Question-Answering which leverages advanced media content, Linguistic analysis and domain knowledge to return precise answers to users natural language queries. However till to date Question –Answering has largely focused on text. The most information in the web is in the form of multimedia and is quite natural to extend text based Question-Answering research to multimedia Question-Answering. Here the answers other than pure text are identified as multimedia answers, including videos, video images, photographs, video images with text and so forth. Further multimedia Question-Answering research must have several key points in mind.

First we must clean noisy annotations and incomplete metadata, second appropriate multimedia answers are more intuitive for some questions, third multimedia answers are readily available for some questions based on the popularity of image and video sharing sites. Thus multimedia Question-Answering can give best answers with the combination of text and other mediums. Thus far few works have addressed multimedia

Question-Answering services.Current technology is still far from enabling us to benefit from MMQA. Furthermore, none of these works fully exploit the rich content on Web 2.0. As we know, Web 2.0 facilitates interactive information sharing, interoperability, and collaboration on the Internet. Therefore, an emerging question is how to leverage user-contributed data such as tagging, comments, and ratings for MMQA. Such information is rapidly becoming more abundant with the popularity of social media sites. For example, YouTube serves 100 million distinct videos and 65,000 uploads daily, and the traffic of this site accounts for more than 20 percent of all Web traffic and 10 percent of the whole Internet, comprising 60 percent of the video watched online. The photo-sharing site Flickr contained more than 4 billion images as of October 2009, and more than 3 billion photos are being uploaded every month on Facebook.

This article briefly surveys the progress of MMQA research and details its future directions. Although search is certainly one of the key techniques in the QA paradigm, here we focus on the problems introduced by MMQA.

II. FROM TEXT TO MULTIMEDIA

The collection of required information on the Web has been growing at an exponential rate. Retrieval of multimedia information from Web is an emerging research topic. When looking for information on the Web, users are often bewildered by the vast quantity of information returned by the search engine, such as the Google or Yahoo. Users often have to painstakingly browse through large ranked lists of results in order to look for the correct answers. Hence question-answering (QA) research has been evolved in an attempt to tackle this information-overload problem. Instead of returning a ranked list of documents as is done in current search engines, QA aims to leverage on deep linguistic analysis and domain knowledge to return precise answers to users' natural language questions.

Research on text-based QA has gained popularity following the introduction of QA in TREC (The Text Retrieval Conference) evaluations in the late 1990s [1]. There are many types of QA, depending on the type of questions and the expected answers. They include: factoid, list and definitional QA, and more recently, the "how-to", "why", "opinion" and "analysis" QA. Typical QA architecture include stages of: question analysis, document retrieval, answer extraction, and answer selection and composition. In factoid and list QA, such as *"What is the most populous country in Africa?"* and *"List the rice producing countries"*, the system is expected to return one or more precise country names as the answers. On the other hand, for definitional QA, such as *"What is X?"* or *"Who is X?",* the system is required to return a set of answer sentences that best describe the question topic. In a way, definition QA is equivalent to query-oriented summarization, in which the aim is to provide a good summary to describe a topic. These three types of QA have attracted a lot of research in the last 10 years. They provide fact-based answers, often with the help of resources such as the Wikipedia1 and WordNet. In fact, Factoid QA has achieved good performance and commercial search engines have been developed, such as the Powerset that aims to return mainly factoid answers from Wikipedia.

Figure 1 illustrates a traditional QA framework, which consists of three main components: document retrieval, question analysis, and answer extraction (candidate answer retrieval, answer selection, and composition). Research in these three types of QA has achieved some success. For example, the commercial QA service Powerset is a factoid QA that focuses on returning factual answers by mining Wikipedia information.

One of the major problems in question answering (QA) is that the queries are either too brief or often do not contain most relevant terms in the target corpus. Most users are interested in searching for *information*, while the current search engines are designed to retrieve only *documents*. There are many simple

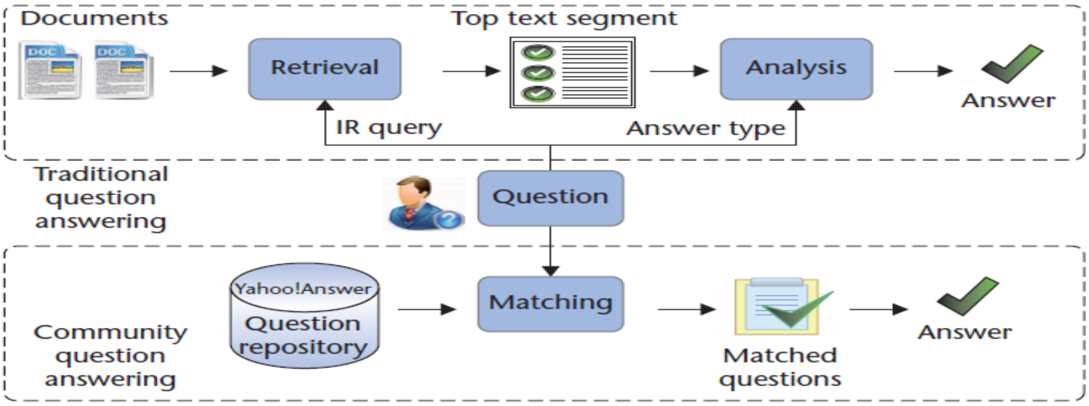

Figure 1. A conceptual framework for traditional and community question answering (QA). The three main components are document retrieval, question analysis, and answer extraction.

Factoid questions like: "*Who is Tom Cruise married to?*" or "*How many chromosomes does a human zygote have?*" While the users expect short precise answers to such questions, typical search engines would return thousands of documents where the actual answer is embedded in some of the documents. This has resulted in a gulf between the expectation of the users and that retrievable by the systems. This gulf will become more severe as we are increasingly being overwhelmed by information overloads.

To address this problem, a Question-Answering (QA) task was initiated in TREC conference series. This has in turn motivated much of the recent research on open domain QA focusing on short, factoid questions. Most modern QA systems combine the strengths of traditional Information Retrieval (IR), natural language processing (NLP) and information extraction (IE) to provide an appropriate way to retrieve concise answers for open-domain natural language questions against a large text collection (termed the QA corpus).

In the TREC-11 evaluation [2], successfully illustrated the power of extensive WordNet to perform logic proof relied on knowledge reasoning. In contrast to the linguistic-based approaches, the use of the Web for QA is an emerging topic of interest among the computational linguistic communities. Several studies suggested that using the Web as additional knowledge resource could improve the performance of a QA system by 25-30%.

In TREC-11, we explored the use of external resources like the Web and WordNet to extract terms that are highly correlated with the query, and use them to perform linear query expansion. While the technique has been found to be effective, we found that there is a need to perform structured analysis on the knowledge obtained from the Web/WordNet in order to further improve the retrieval performance. In this work, we model the TREC-style QA task as *QA entities or Events*. Each QA event comprises of elements describing different facets of the event like *time, location, object, action* etc. For most QA events, there are inherent associations among their elements. Thus if we know a portion of the elements, we can use this information to narrow the search for the rest of unknown event elements, which are likely to contain the answers. In order to incorporate more knowledge of event elements from the external resources and to use event structure systematically for more effective QA, we analyze the external knowledge to discover the event structure. The integration and structured use of both linguistic and web knowledge for the TREC-style QA, which we call ***Event-based QA***. In particular, we describe a high performance question answering system called **QUALIFIER (QUestion Answering by LexIcal FabrIc and External Resources)** and analyze its effectiveness using the TREC-11 benchmark. Our results show that the Event-based approach, together with the use of *Successive Constraint Relaxation,* gives rise to an effective question answering system

However, current QA systems are based on text alone and can be difficult to use when questions are centered on physical objects with distinctive visual attributes. For example, a person who has just seen an interesting poster may want to ask the question “where can I buy this poster?” A text based QA system would require the person to meticulously describe the visual details of the poster in order to identify it. The difficulty of this task stems from the fact that such questions are inherently dual-modal: it involves a verbal component that states its intent (where to buy) and a visual component that identifies the object (a specific poster). Unfortunately, with text as the only available input modality, users of current QA systems are often forced to express in words what would be best expressed visually.

We propose photo-based QA as a solution to the limitation of current text-based QA systems [3]. By taking advantage of recent advances in QA and image matching technologies, photo-based QA supports direct use of photos in phrasing questions and finding answers. In contrast, current text-based QA system is hard to use when visual objects are involved. These problems highlight the unique usability benefits of photo-based QA. Two factors play in our favor in developing useful photo-based QA systems. First, many online multimedia (i.e., image and text) data sources can supply a photo-based QA system with structured information to handle a variety of common photo-based questions automatically. Second, many community human users are willing to look at photos and answer questions that the automatic process fails to find answers for.

We describe three-layer system architecture for photobased QA. It draws inspiration from three popular QA approaches: template-matching, information retrieval, and human computation (Figure 2).

In recent years, community generated video collections on the web have grown rapidly, and video search engines have been utilized to help people to access these resources, such as YouTube and Yahoo Video. Millions of video available on the web make them an ideal source for visual reference. However,

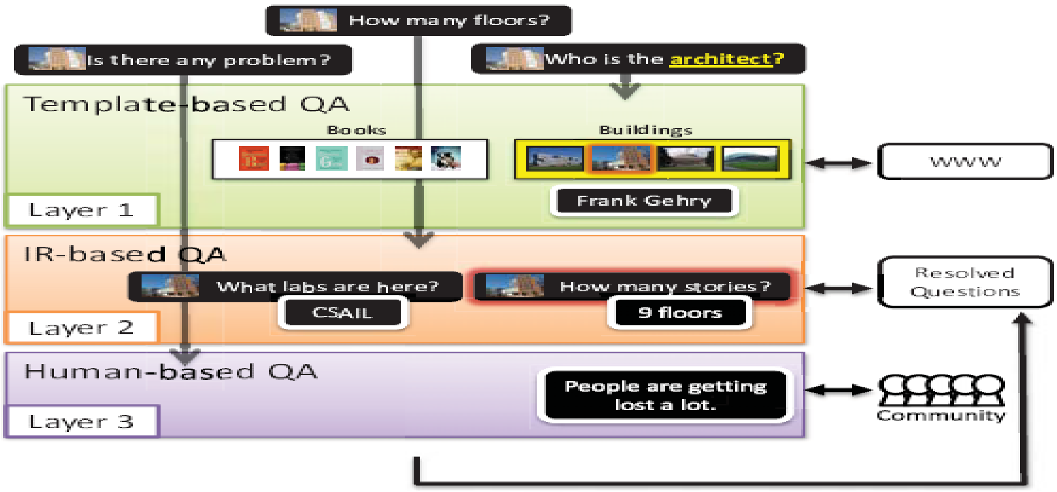

Figure 2: Three-layer system architecture for photobased QA. For a simple question, the template-based layer identifies its category (e.g., building), finds a matched image within the category, and forms a template to extract an answer from the Web. For a harder question, the IR-based layer searches for relevant photo-based questions already resolved. The human-based layer handles the rest of the questions too difficult for the first two layers.

Web video search engine cannot be directly used to find visual answers. First, most of web video search engines do not work well with verbose query. Second, manually identifying some phrases from the original question as query may help to obtain related videos, but it is still inconvenient for people to try different query combinations before finding a satisfactory video as reference. Third, even if some related videos are retrieved, it is possible that the most related video is still not at the top position.

We propose a novel system named *Video Reference* as a solution to the above problem [4]. We use community generated video website YouTube as the video source. YouTube is a growing video sharing website that contains a large number of "how-to" videos that depict how the video owners solve some specific problems. Each video is also commented by other active viewers. The combination of video and user comments offers great advantage over other video collections, making YouTube a good video source to support *Video Reference*.

As depicted in Figure.3, Video Reference works as follows. User enters a question, such as *"how do I get my camera to put pics. on the computer?"*. The first is a recall-driven related video search step that finds similar questions posed in different forms from Yahoo! Answers site. Similar questions have been found to increase the coverage of the topic. However, source video site like YouTube can only take in precise queries; we extract key phrases from these questions as multiple search queries. The second step is the precision- driven video re-ranking, where related videos based on their relevance to the original questions are re-ranked. We manually select training images for certain concept using Google Image Search. We then perform salient object recognition based on image matching techniques to recognize the visual relatedness of the video to certain concepts. In addition, community viewers' comments play the role of opinion voting for the video's popularity. Finally, a rank fusion scheme is adopted to generate a new ranking list based on evidences from visual cues, opinion voting and video redundancy. Our initial test shows that our approach is effective.

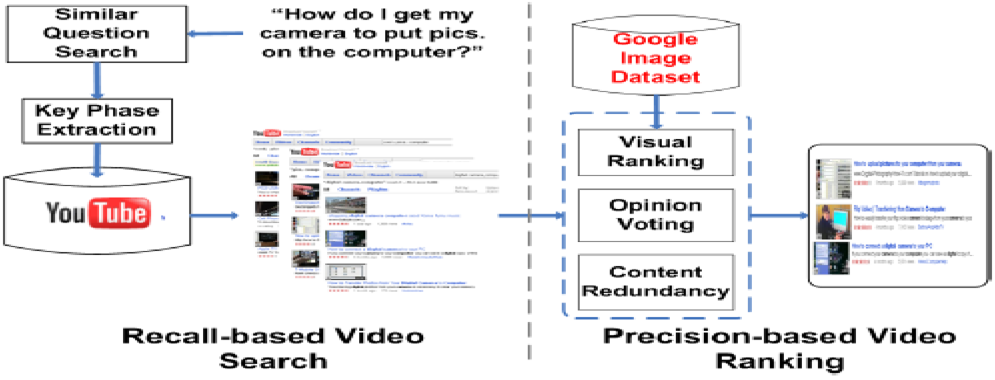

Figure 3: System Flowchart of Visual Reference

Many users are interested in searching for *information*, while the current video retrieval engines are designed to return only *video sequences*. We perform question answering (QA) to support personalized news video retrieval [5]. Users interact with our system, VideoQA, using short natural language questions with implicit constraints on contents, duration, and genre of expected videos. VideoQA returns short precise news video fragments as answers. The main contributions of this work are: (a) the extension of question answering technology to support QA in news video; and (b) the use of external knowledge and visual content analysis to help correct speech recognition errors and to perform precise question answering. The system has been demonstrated to be effective.

The development of a video-based QA system requires the solution to three fundamental problems in video and text processing. The first is to segment the video sequence into story units with correct genre classification. Several works have been done on this, including (Chaisorn et al 2002) and (Hsu & Chang 2003). These approaches perform multi-modal analysis using a combination of visual, audio and textual features based on HMM or entropy techniques and reported accuracies of about 90% in story segmentation and genre classification. The second problem is that the users' questions are normally short and assume previous context. In order to extract relevant sentences in the video's speech track that answer the query precisely, we need the ability to analyze the text transcript at sentence level. Thus a third problem is to overcome the recognition errors in the speech accompanying the video. Text from speech (or transcript) is a major source of semantic information for news video. The conventional video retrieval based on transcript suffers a lot from the numerous speech recognition errors. This is especially severe for substitution errors that cause many names of person, location and organization to be wrongly recognized.

Video QA aims to provide precise video answers to simple factoid questions posed over the news video collection. It handles natural language questions by unearthing the answers embedded in the video collection and presenting the video segments in the form of video summary. It is naturally used in a personalized video setting in which a user may request for details of certain aspects of news or summary of latest news. It will be an essential component of future information systems.

During the pre-processing stage, VideoQA performs video story segmentation and classification, as well as video transcript generation and correction. During question answering, VideoQA employs modules for: question processing, query reinforcement, transcript retrieval, answer extraction and video summarization. Figure 4 gives the system architecture of VideoQA.

Given the news video collection, the pre-processing stage "prepares" the video for later answer retrieval. We analyze the raw video using a two-level story segmentation scheme as proposed in Chaisorn et al (2002). The basic unit of analysis is the shots, and we employ multi-modal analysis involving visual, audio and textual features. Briefly, we model each shot using high-level object-based features (face, video text, and shot type), temporal features (background scene change, speaker change, motion, audio type, and shot duration), and low-level visual feature (color histogram). At the shot level, we employ the Decision Tree to classify the shots into one of 13 genre types of: *Intro/ Highlight*, *Anchor-person*, *2-anchor-person*, *Meeting/ Gathering*, *Speech/Interview*, *Live-reporting*, *Still-image*, *Sports*, *Text-scene*, *Special*, *Finance*, *Weather*, *and Commercials*. We then perform HMM analysis to detect story boundaries using the shot genre information, as well as time-dependent features such as the speaker change, scene change and key phrases. The resulting video story may contain shots of different genre types. For example, a general news story typically contains shots of type *Anchor-person*, *Live-reporting* and *Speech/Interview*; while a sports story includes shots of type *Sports* and *Text-scene*.

Given that the vast amount of Web content is nontextual media, it is natural to extend the text-based QA research to MMQA. MMQA is important for several reasons.

First, although most media contents are indexed with text metadata, most such metadata, such as those available on YouTube, is noisy and incomplete. As a result, much multimedia content will remain unretrievable without advanced media content analysis techniques.

Second, many questions are better explained with the help of a nontextual medium. For example, in providing textual answers to a definitional question such as ''what is a thumb drive?'' it would help to provide an image or video of a thumb drive with a textual description. Figure 5 illustrates how MMQA differs from other retrieval approaches.

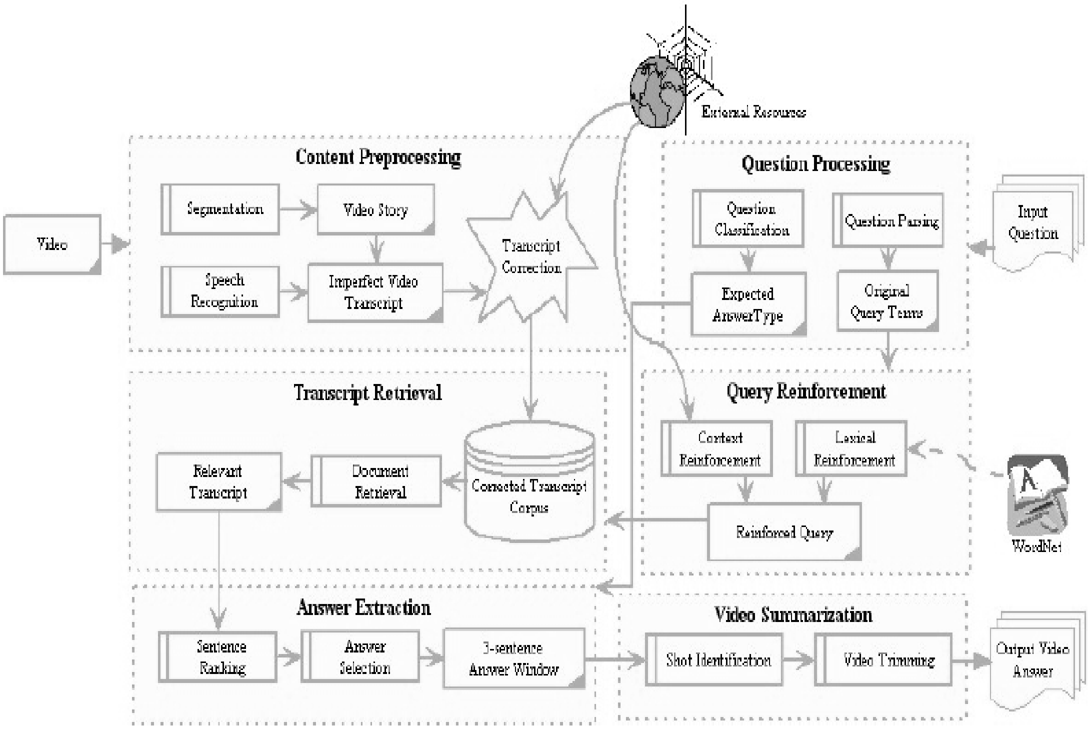

Figure 4: System Architecture of VideoQA

Third, media content, especially videos, are now used to convey many types of information, as sites such as YouTube and other specialized video-/image-sharing sites and blogs have shown. Thus, many questions already have available answers in the form of video. This is especially true for the more challenging analysis and how-to questions. Answering such questions using traditional text-based framework is difficult because further analysis and composition is often necessary. It would be much clearer and more instructive to answer the question ‘‘How do I transfer photos from my camera to computer?’’ with a readily available how-to video

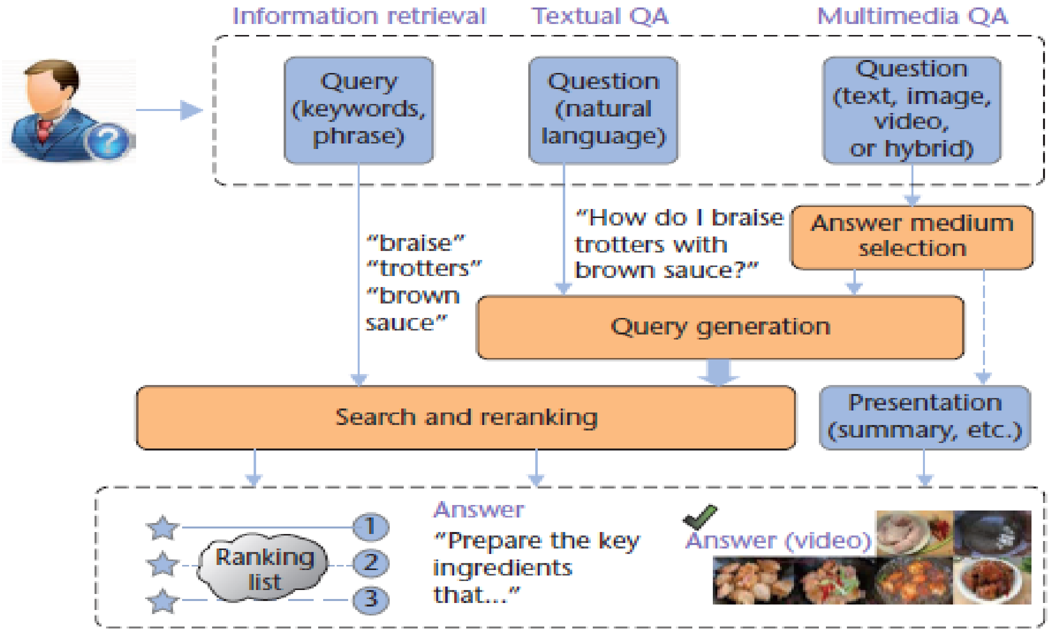
Information retrieval
Textual QA
Multimedia QA
Query (keywords, phrase)
Question (natural language)
Question (text, image, video, or hybrid)
"braise" "trotters" "brown sauce"
"How do I braise trotters with brown sauce?"
Answer medium selection
Query generation
Search and reranking
Presentation (summary, etc.)
Ranking list
1
2
3
Answer
"Prepare the key ingredients that..."
Answer (video)

Figure 5: Differences among information retrieval, textual QA, and MMQA. Search and reranking are common components among them, but MMQA differentiates itself with answer medium selection and answer presentation.

Than to direct people to a textual instruction on how to do it in a step-by-step manner

In this paper, we propose a scheme which can enrich community-contributed textual answers with appropriate media information. Figure 6 shows the schematic illustration of the approach. It contains three main components:

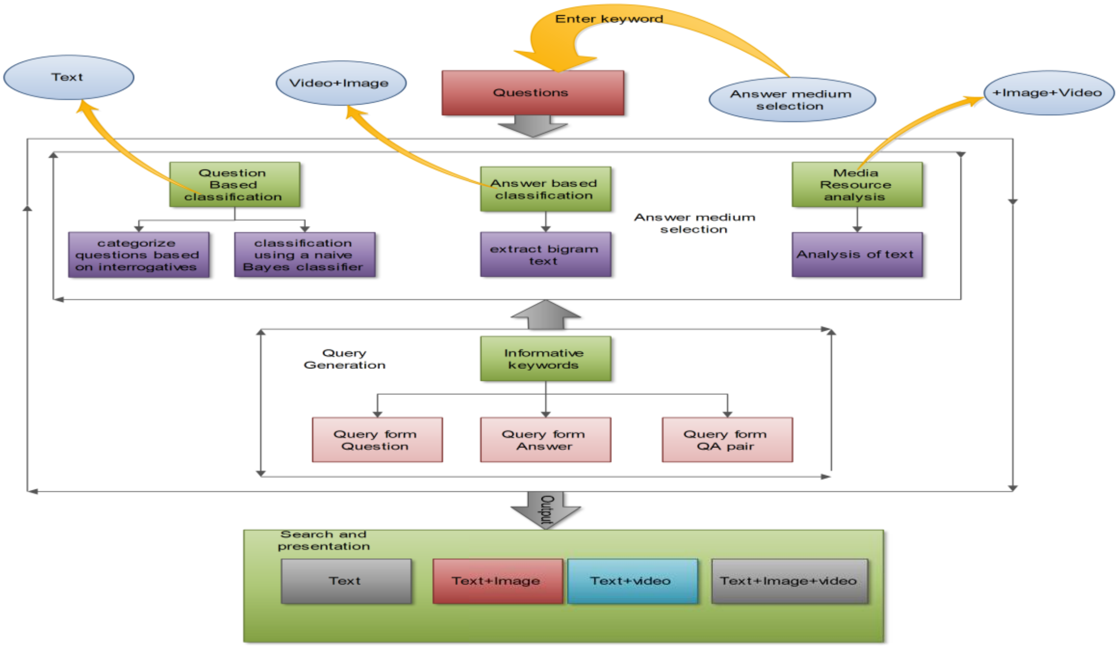

Figure 6: Schematic illustration of proposed multimedia answering scheme

1) Answer medium selection. Given a QA pair, it predicts whether the textual answer should be enriched with media information, and which kind of media data should be added. Specifically, we will categorize it into one of the four classes: text, text+image, text+video, and text+image+video1.It means that the scheme will automatically collect images, videos, or the combination of images and videos to enrich the original textual answers.

2) Query generation for multimedia search. In order to collect multimedia data, we need to generate informative queries. Given a QA pair, this component extracts three queries from the question, the answer, and the QA pair, respectively. The most informative query will be selected by a three-class classification model.

3) Multimedia data selection and presentation. Based on the generated queries, we vertically collect image and video data with multimedia search engines. We then perform re ranking and duplicate removal to obtain a set of accurate and representative images or videos to enrich the textual answers.

III. CONCLUSION AND FUTURE ENHANCEMENT

We describe the motivation and evolution of MMQA, and it is analyzed that the existing approaches mainly focus on narrow domains. Aiming at a more general approach, we propose a novel scheme to answer questions using media data by leveraging textual answers in cQA. For a given QA pair, our scheme first predicts which type of medium is appropriate for enriching the original textual answer. Following that, it automatically generates a query based on the QA knowledge and then performs multimedia search with the query. Finally, query-adaptive duplicate removal is performed to obtain a set of images and videos for presentation along with the original textual answer. Different from the conventional MMQA research that aims to automatically generate multimedia answers with given questions, our approach is built based on the community contributed answers, and it can thus deal with more general questions and achieve better performance.

We have also observed several failure cases. For example, the system may fail to generate reasonable multimedia answers if the generated queries are verbose and complex. For several questions videos are enriched, but actually only parts of them are informative. Then, presenting the whole videos can be misleading. Another problem is the lack of diversity of the generated media data. We have adopted a method to remove duplicates, but in many cases more diverse results may be better.

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] G. Li et al., ''Video Reference: Question Answering on YouTube,'' Proc. 17th Int'l ACM Conf. Multimedia, ACM Press, 2009, pp. 773_776.

[5] VideoQA: Question Answering on News Video Hui Yang, Shi-Yong Neo, Lekha Chaisorn, Tat-Seng Chua School of Computing, National University of Singapore Singapore 117543

[6] 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.

[7] 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.

[8] K. Wang et al., ''Segmentation of Multi-Sentence uestions: 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.

[9] 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.

[10] R. Hong et al., ''Mediapedia: Mining Web Knowledge to Construct Multimedia Encyclopedia,'' Advances in Multimedia Modeling, LNCS 5916, Springer, 2010, pp. 556_566.