ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 178-183 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Query Generation for Multimedia Data Selection and Presentation

M Suresh[1] Y Sowjanya Kumari[2]

[1]M.Tech Student, Dept of CSE, St. Ann's College of Engineering Technology, Chirala, Prakasam Dist, A.P, India
[2] Associate Professor, Dept of CSE, St. Ann's College of Engineering Technology, Chirala, Prakasam Dist, A.P, India

***Abstract*—Community Question Answering (CAQ) services have acquired popularity over the past years. This service not only allows community members to post and answer questions but also enables general users to seek information from a comprehensive set of well-answered questions. However, current community question answering forums usually produce only textual answers, which are not informative enough for many questions. In this paper, we propose a scheme that is able to enrich textual answers in community question answering with appropriate media data. Our scheme consists of three components: answer medium selection, query generation for multimedia search, and multimedia data selection and presentation. This approach automatically determines which type of media information should be added for a textual answer. It then automatically collects data from the web to enrich the answer. By processing a large set of question answer pairs and adding them to a pool, our approach can enable a novel multimedia question answering (MMQA) approach as users can find multimedia answers by matching their questions with those in the pool. Different from a lot of multimedia question answering research efforts that attempt to directly answer questions with image and video data, our approach is built based on community-contributed textual answers and thus it is able to deal with more complex questions.**

***Index Terms*—Community Question Answering, medium selection, Multimedia Question Answering, Reranking**

I. Introduction

Question-Answering (QA) is a technique for automatically answering a question posed in natural language {?}, {?}. Compared to keyword-based search systems, it greatly enables the communication between humans and computer by naturally stating users' intention in plain text sentences. It also avoids the intend browsing of a huge quantity of information contents returned by search engines for the correct answers. Hence, fully automated question answering still facing challenges that are not easy to handle, such as the deep understanding of complex questions and the sophisticated syntactic, semantic and contextual processing to generate answers. It is found that, in most cases, automated approach cannot obtain results that are as good as those generated by human intelligence.

Community Question Answering has emerged as an extremely popular alternative to acquire information online, possess to the following reasons. *First*, information seekers are able to post their specific questions on any topic and obtain answers provided by other participants. By leveraging community efforts, they are able to get better answers than simply using search engines. *Second*, in comparison with automated question answering systems, community question answering usually receives answers with better quality as they are generated based on human intelligence. *Third*, over times, a tremendous number of question answering pairs have been accumulated in their repositories, and it facilitates the preservation and search of answered questions.

Despite their great success, existing community question answering forums mostly support only textual answers, as shown in Figure 1.Unfortunately textual answers may not provide sufficient natural and easy-to-grasp information. Figure 1 (a) and (b) illustrate two examples. For the questions "*Who is the winner of cricket world cup in 2011?*" and "*FIFA world cup winner in 2014?*" the answers are described by long sentences. Clearly, it will be much better if there are some accompanying videos and images that visually demonstrate the process or the object. Therefore, the textual answers in community question answering can be significantly enhanced by adding multimedia contents, and it will provide answer seekers more comprehensive information and better experience.

Who is the winner of cricket world cup in 2011?	FIFA world cup winner in 2014?
ANSWER India won the tournament, defeating Sri Lanka by 6 wickets in the final in Mumbai, thus becoming the first country to win the Cricket World Cup final on home soil	ANSWER The 2014 FIFA world championship which took place at several venues across Brazil. Germany won the tournament and took its fourth title by defeating Argentina in final.

(a) (b)

Who is the winner of cricket world cup in 2011?	FIFA world cup winner in 2014?
ANSWER https://www.youtube.com/watch?v=mnN_UPCpWhg https://www.youtube.com/watch?v=LSP6maR5NZ0	**ANSWER** https://www.youtube.com/watch?v=r-7fnTK7QYw https://www.youtube.com/watch?v=3aqAjhoJ-6s
(c)	(d)

Fig.1. Examples of QA pairs (a), (b) An example with textual answers; (c), (d) an example that only contains links to two videos.

In fact, users usually post URLs that link to supplementary images or videos in their textual answers. For example, for the questions in Figure 1 (c) and (d), the best answers contain video URLs. It further confirms that multimedia contents are useful in answering several questions. But existing community question answering forums do not provide adequate support in using media information.

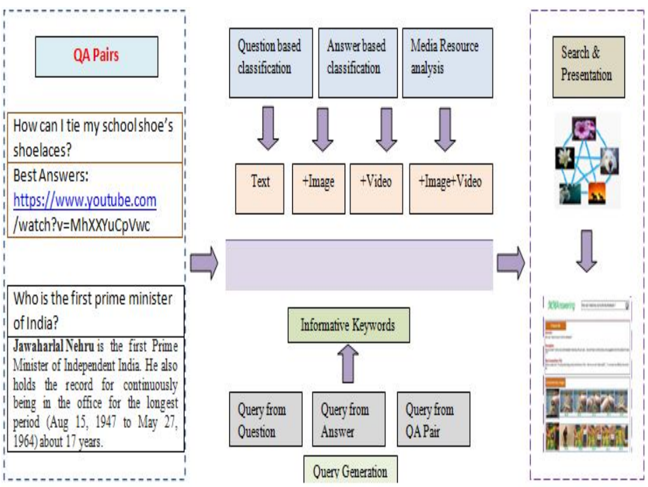


Fig.2. The schematic illustration of the proposed multimedia answering scheme. The scheme mainly contains three components, i.e., answer medium selection, query generation, and data selection and presentation.

In this paper, we propose a novel scheme which can enrich community-contributed textual answers in community question answering with appropriate media data. Figure 2 shows the schematic illustration of the approach. It contains three main components: *Answer medium selection, Query generation for multimedia search, Multimedia data selection and presentation.*

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+video. It means that the scheme will automatically collect images, videos, or the combination of images and videos to enrich the original textual answers.

Query generation:

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.

Multimedia data selection and presentation:

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

Our proposed approach in this work does not aim to directly answer the questions, and instead, we enrich the community-contributed answers with multimedia contents. Our strategy splits the large 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 answer. In our scheme, the first gap is bridged by the crowd-sourcing intelligence of community members, and thus we can focus on solving the second gap. Therefore, our scheme can also be viewed as an approach that accomplishes the MMQA problem by jointly exploring human and computer. This approach demonstrates the difference between the conventional MMQA approaches and an MMQA framework based on our scheme. It is worth noting that, although the proposed approach is automated, we can also further involve human interactions.

II. Related Work

i. Textual Answer to Multimedia Answer:

The amount of information on the Web has been growing at an exponential rate. An huge amount of increasingly multimedia contents are now available on almost any topics. When looking for information on the Web, users are often confused by the vast quantity of information returned by the search engine, such as the Google or Yahoo. Users often have to attention ally 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

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 178-183 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

analysis and domain knowledge to return precise answers to users' natural language questions.

ii. Multimedia Search:

Because of the increasing amount of digital information stored over the web, searching for desired information has become an essential task. The research in this area started by addressing the general problem of finding images from a fixed database. Rapid development of content analysis technology these efforts quickly expanded to tackle the video and audio retrieval problems.

Multimedia search efforts can be categorized into two categories: *text-based search and content-based search*.

Text-based search:

This approaches use textual queries, a term-based specification of the desired media entities, to search for media data by matching them with the surrounding textual descriptions. To boost the performance of text-based search, some machine learning techniques that aim to automatically annotate media entities have been proposed in the multimedia community. Other, several social media websites, such as Flickr and Face book, have emerged to accumulate manually annotated media entities by exploring the common data. Since, user-provided text descriptions for media data are often biased towards personal perspectives and context cues, and thus there is a gap between these tags and the content of the media entities that common users are interested in.

Content-based search:

Content based media retrieval performs search by analyzing the contents of media data rather than the metadata. Despite the tremendous improvement in content-based retrieval, it still has several limitations, such as high computational cost, difficulty in finding visual queries, and the large gap between low-level visual descriptions and users' semantic expectation. Therefore, keyword-based search engines are still widely used for media search. However, the intrinsic limitation of text-based approaches make that all the current commercial media search engines difficult to bridge the gap between textual queries and multimedia data, especially for verbal questions in natural languages.

iii .Multimedia search reranking

The explosive growth and widespread accessibility of community contributed media content on the Internet have led to a surge of research activity in multimedia search. Some approaches that apply text search techniques for multimedia search have achieved limited success as they entirely ignore visual content as a ranking signal. Multimedia search re-ranking, which reorders visual documents based on multimodal cues to improve initial text-only searches, has received increasing attention in recent years. Such a problem is challenging because the initial search results often have a great deal of noise. Discovering knowledge or visual patterns from such a noisy ranked list to guide the re-ranking process is difficult. Numerous techniques have been developed for visual search re-ranking. The purpose of this multimedia search reranking is to categorize and evaluate search results based on the ranks provided by the internet users while searching using multimedia enabled search.

III. ANSWER MEDIUM SELECTION

The component of our scheme is answer medium selection, as we first need to determine whether we need to and if so which type of medium we should add to enrich the textual answers. For some questions, such as "*what day is Mahatma Gandhi's birthday*", using pure textual answers is sufficient. But we need to add image or video information for some other questions. For example, for the question, "*who is Mahatma Gandhi*", it is better to add images to complement the textual answer, whereas we should add videos for answering the question "*how can I tie my shoe's shoelaces? ".* We regard the answer medium selection as a QA pair classification task, i.e., given a question and textual answer, we classify it into the following four classes of answer medium combinations: (1) only text; (2) text + image; (3) text + video; and (4) text + image + video. For the "only text" class, we do not need to perform more operations, and for the other cases we will need to collect appropriate data by the other two components: *Question based classification and Answer based classification.*

Question based classification:

In this question based classification many questions contain multiple sentences and some of the sentences are uninformative, we first employ the method to extract the core sentence from each question. The classification is accomplished in two steps. First, we categorize the questions based on the interrogatives i.e.; some starting words and ending words, and for some questions we can directly derive that they should be answered with text. Second, for the rest questions, we perform

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 178-183 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

classification using a naive bayes classifier. We first introduce the categorization based on interrogative words. Questions can mainly be categorized into the following classes based on interrogative words: yes/no class (such as "*Does Logic involve in designing processors* "), choice class (such as "*Which is concurrent language C or Verilog*"), quantity class (such as "*When is the IOS developed* "), enumeration class (such as "*Name of the wise man who came from America and developed SOC developed in India*"), and description class (such as "*What are the steps to form a government*"). For example, a question will be categorized to the "quantity" class if the interrogative is "how+adj/adv" or "when". For the "yes/no", "choice" and "quantity" questions, we categorize them into the class of answering with only text; while the "enumeration" and "description" questions can need "text+image", "text+video" or "text+image+video answers". Therefore, given a question, we first judge whether it should use only textual answer based on the interrogative word. If not, we further perform classification with a Naive Bayes classifier. For building the Naive Bayes classifier, we extract a set of text features, including bigram text features, head words and a list of class-specific related words.

Answer based classification:

Besides question, the answer with rich text can also be an important information clue. For example, for the question "*how do you run a c program*", we may find a textual answer as "*first compile it, and execute it...*" Then we can judge that the question can be better answered with a video clip as the textual answer contains many verbs and describes a dynamic process. The verbs in the answer will be useful for judging whether the question can be answered with video content. Intuitively, if a textual answer contains many complex verbs, it is more likely to describe a dynamic process and thus it has high probability to be well answered by videos. Therefore, verb can be an important clue. Based on the bigram text features and verbs, we again build a Naive Bayes classifier with a set of training data, and then perform a four-class classification with the model. The classification results are linearly combined with those of question-based classification.

IV. MULTIMEDIA SEARCH QUERY GENERATION

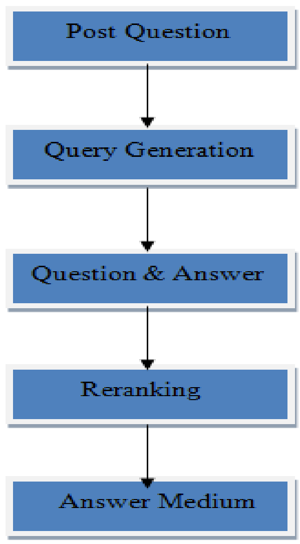
To collect relevant image and video data from the web, we need to generate appropriate queries from text QA pairs before performing search on multimedia search engines. We accomplish the task with two steps: query extraction and query selection.

Query Extraction- In query extraction text questions and answers are usually complex sentences. But frequently search engines do not perform well with queries that are long and verbose. Therefore, we need to extract a set of informative keywords from questions and answers for querying.

Query Selection-This is because we can generate different queries: one from question, one from answer, and one from the combination of question and answer. Which one is the most informative depends on the question and some should combine question and answer to generate useful query. For each QA pair, we will generate three queries. First, we convert question to query, i.e., we convert a grammatically correct interrogative sentence into one of the potential syntactically correct declarative sentences or meaningful phrases. We directly employ the method.. Second, we identify several key concepts from verbose answer which will have the most impact on effectiveness. Here we employ the method. Finally, we combine the two queries that are generated from question and answer respectively. Therefore, we obtain three queries, and the next step is selecting one from them. The query selection is formulated as a three-class classification task, since we need to choose one from the three queries that are generated from the question, answer and the combination of question and answer.

ISSN 2278-3091
International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 178-183 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Architecture:



Multimedia data selection and presentation:

We perform search using the generated queries to collect image and video data with Google image and video search engines respectively. However, commercial search engines, such as Google, Yahoo and Bing, usually index web images and videos using textual information, such as titles, alternative text and surrounding texts on web pages. But frequently the text information does not fully describe content of the images and videos, and this fact can severely degrade the search relevance of web images and videos. Reranking is an approach to improving search relevance by mining the visual information of images and videos. Existing reranking algorithms can mainly be categorized into two approaches, one is pseudo relevance feedback and the other one is graph-based reranking. The pseudo relevance feedback approach regards top results as relevant ones and then collects some samples that are assumed to be irrelevant. A classification or ranking model is learned based on the pseudo relevant and irrelevant samples and the model is then used to rerank the images.

However, a problem with existing reranking methods is that they usually use features extracted from the whole images or video frames and they overlooked that many queries are actually person-related. Clearly, for person-related queries, it is more reasonable to use facial features instead of global visual features for reranking. For question-answering, our statistics show that more than 1/4 of the QA pairs in our data set are about person. Therefore, in this work we propose a query-dependent reranking approach. We first decide whether a query is person-related or not, and then we use different features for reranking. We establish the following rules to judge whether a query is person-related:

a. If the given question starts by interrogative words who or whom, then it is categorized as person-related.
b. We perform morphological analysis on the given question to extract information on Part-of-Speech, verb-phrase and noun-phrase. We then extract the main core terms followed by the other possible key terms.

V. CONCLUSION

In this paper, we describe the motivation and evolution of MMQA, and it is analysed 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 community question answering. 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 re ranking and duplicate removal are 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.

In our study, 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. In our future work, we will further improve the scheme, such as developing better query generation method and investigating the relevant segments from a video. We will also investigate multimedia search diversification methods, such as the approach in [?], to make the enriched media data more diverse.

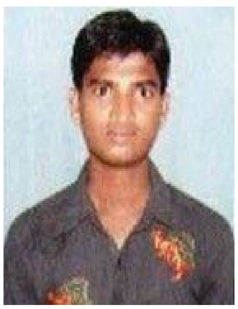
REFERENCES

E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *Proceedings of International Conference on Web Search and Web Data Mining*, 2008.

Liqiang Nie, Meng Wang, Member, IEEE, Yue Gao, Zheng-Jun Zha, Member, IEEE,and Tat-Seng Chua, Senior Member, IEEE "Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information"- IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 2, FEBRUARY 2013.

L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of International World Wide Web Conference*, 2008.

AUTHOR PROFILE



M.Suresh Received B.tech degree from QIS college of engineering and Technology which Affiliated to JNTU Kakinada, Currently he is pursuing M.Tech in St. Ann's college of engineering and technology which is affiliated to JNTU Kakinada.

Y. Sowjanya Kumari presently working as Associate Professor, Dept of computer science & Engineering at St.Ann's College of Engineering . She guided many UG and PG students. She has more than 11 years of teaching experience. She received her B.tech degree from NBKRIST, vidyanagar, india in 2002. She received her M.tech degree from JNTU Kakinada in 2004..