ENHANCING THE PERFORMANCE OF INFORMATION RETRIEVAL IN QA USING SEMANTIC WEB TECHNOLOGY

S.Lakshmi Prabha[1],C.Meenaa [1], S.Pugazhl Vendhan [1] and S.Raja Ranganathan[2]

[1]BE Computer Science and Engineering

[2]Assistant Professor Computer Science and Engineering

SNS College of Technology

Abstract-Question Answering is a technique for automatically answering a question in natural language.This question answering services have attained great success over the past years.Despite their great success,existing cQA forums mostly support only the textual answers.These textual answers are not provide the sufficient information.In this paper,we propose a scheme that is able to enhance the textual answers in QA with appropriate media data.For retrieving the media information such as image and video we introduce a semantic web Technology.This approach automatically determines which type of media data should be added for the textual answer.It then automatically collects data from the web to enrich the answer. By processing a large set of QA 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 MMQA 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. We have conducted extensive experiments on a multisource QA dataset. The results demonstrate the effectiveness of our approach.

Index Terms—semantic web ,Question answering, cQA, medium selection, 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 greatlyfacilitates the communication between humans and computer by naturally stating users' intention in plain sentences. It also avoids the painstaking browsing of a vast quantity of information contents returned by search engines for the correct answers. However, fully automated QA still faces challenges that are not easy to tackle, 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.Along with the proliferation and improvement of underlying communication technologies, community QA (cQA) has emerged as an extremely popular alternative to acquire information online, owning to the following facts. 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 QA systems, cQA usually receives answers with better quality as they are generated based on human intelligence.Third, over times, a tremendous number of QA pairs have been accumulated in their repositories, and it facilitates the preservation and search of answered questions. For example, WikiAnswer, one of the most well-known cQA systems, hosts more than 13 million answered questions distributed in 7*;* 000 categories (as of August 2011).Despite their great success, existing cQA forums mostly support only textual answers. Unfortunately,textual answers may not provide sufficient natural and easy-to-grasp information. Figure 1 (a) and (b) illustrate two examples. For the questions "*What are the steps to make a weather vane*" and "*What does* $*1 Trillion Look Like*" ,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 cQA can be significantly enhanced by adding multimedia contents, and it will provide answer seekers more comprehensive information and better experience.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 on Y!A both contain video URLs. It further confirms that multimedia contents are useful in answering several questions. But existing cQA forums do not provide adequate support in using media information. In this paper, we propose a novel scheme which can enrich community-contributed textual answers in cQA with appropriate media data. It contains three main components:(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 reranking and duplicate removal to obtain a set of accurate and representative images or videos to enrich the textual answers. It is worth mentioning that there already exist several research efforts dedicated to automatically answering questions with multimedia data, i.e., the so-called Multimedia Question Answering (MMQA). For example, Yang et al. [?] proposed a technology that supports factoid QA in news video. Yeh et al. [?] presented a photo-based QA system for finding information about physical objects. Li et al. [?] proposed an approach that leverages YouTube video collections as a source to automatically find videos to describe cooking techniques.But these approaches usually work on certain narrow domains and can hardly be generalized to handle questions in broad domains. This is due to the fact that, in order to accomplish automatic MMQA, we first need to understand questions,which is not an easy task. 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 ancomputer. Figure 3 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. For example, our approach can provide a set of candidate images and videos based on textual answers, and answerers can manually choose several candidates for final presentation.This framework was first explored in our previous work [?].Compared to the preliminary version [?], we have a lot of improvements in this work. For example, for answer medium selection, we add a media resource analysis component. The results of the media resource analysis are also regarded as evidences to enable a better answer medium selection. For multimedia data selection and presentation, we propose a method that explores image search results to replace the original text analysis approach in judging whether a query is person-related or not. We introduce a new metric to measure how well the selected multimedia data can answer the questions in addition to the simple search relevance. We also investigate the cases that textual answers are absent.

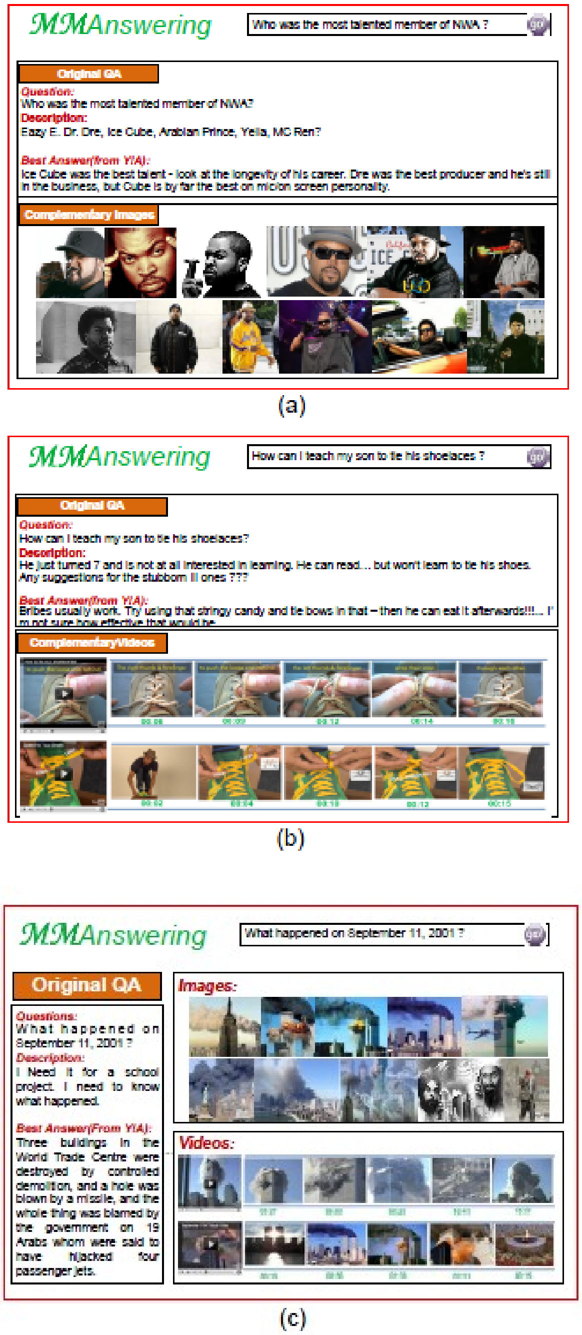
TableI
REPRESENTATIVE CLASS-SPECIFIC RELATED WORDS.

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

II. RELATED WORK

A. From Textual QA to Multimedia QA

The early investigation of QA systems started from 1960s and mainly focused on expert systems in specific domains.Text-based QA has gained its research popularity since the establishment of a QA track in TREC in the late 1990s [?].Based on the type of questions and expected answers, we can roughly summarize the sorts of QA into Open-DomainQA [?], Restricted-Domain QA [?], Definitional QA [?] and List QA [?]. However, in spite of the achievement as described above, automatic QA still has difficulties in answering complex questions. Along with the blooming of Web 2.0, cQA becomes an alternative approach. It is a large and diverse question-answer forum, acting as not only a corpus for sharing technical knowledge but also a place where one can seek advice and opinions [?], [?]. However, nearly all of the existing cQA systems, such as Yahoo!Answers, WikiAnswers and Ask Metafilter, only support pure text-based answers, which may not provide intuitive and sufficient information. Some research efforts have been put on multimedia QA, which aims to answer questions using multimedia data. An early system named VideoQA was presented in [?]. This system extends the text-based QA technology to support factoid QA by leveraging the visual contents of news video aswell as the text transcripts. Following this work, several video QA systems were proposed and most of them rely on the use of text transcript derived from video OCR (Optical Character Recognition) and ASR (Automatic Speech Recognition) outputs [?], [?], [?], [?]. Li et al. [?] presented a solution on "how-to" QA by leveraging community-contributed texts and videos. Kacmarcik et al. [?] explored a non-text input mode for QA that relies on specially annotated virtual photographs. An image-based QA approach was introduced in [?], which mainly focuses on finding information about physical objects. Chua et al. [?] proposed a generalized approach to extendtext-based QA to multimedia QA for a range of factoid, definition and "how-to" questions. Their system was designedto find multimedia answers from web-scale media resources such as Flicker and YouTube. However, literature regarding multimedia QA is still relatively sparse. As mentioned in Section I, automatic multimedia QA only works in specific domains and can hardly handle complex questions. Different from these works, our approach is built based on cQA. Instead of directly collecting multimedia data for answering questions, our method only finds images and videos to enrich the textual answers provided by humans. This makes our approach able to deal with more general questions and to achieve better performance.



Who was the most talented member of NWA ?

Question:
Who was the most talented member of NWA?
Description:
Eazy E. Dr. Dre, Ice Cube, Arabian Prince, Yella, MC Ren?

Best Answer(from Y!A):
Ice Cube was the best talent - look at the longevity of his career. Dre was the best producer and he's still in the business, but Cube is by far the best on mic/on screen personality.

(a)

How can I teach my son to tie his shoelaces ?

Question:
How can I teach my son to tie his shoelaces?
Description:
He just turned 7 and is not at all interested in learning. He can read… but won't learn to tie his shoes. Any suggestions for the stubborn ll ones ???

Best Answer(from Y!A):
Bribes usually work. Try using that stringy candy and tie bows in that – then he can eat it afterwards!!!... I' m not sure how effective that would be

(b)

What happened on September 11, 2001 ?

Questions:
What happened on September 11, 2001 ?
Description:
I Need it for a school project. I need to know what happened.

Best Answer(From Y!A):
Three buildings in the World Trade Centre were destroyed by controlled demolition, and a hole was blown by a missile, and the whole thing was blamed by the government on 19 Arabs whom were said to have hijacked four passenger jets.

(c)

Fig 1.Results of multimedia answering for 3 example queries, "the most talented member ofNWA", "tie shoelace", and "September 11". Our scheme answers the three questions with text + image, text +video, and text + image + video, respectively.

B. Multimedia Search

Due to the increasing amount of digital information stored over the web, searching for desired information has becomean essential task. The research in this area started from the early 1980s [?] by addressing the general problem of finding images from a fixed database. With the rapid development of content analysis technology in the 1990s, these efforts quickly expanded to tackle the video and audio retrieval problems [?], [?]. Generally, multimedia search efforts can be categorized into two categories: text-based search and content-based search. The text-based search [?] 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 [?][?][?]. Further, several social media websites, such as Flickr and Facebook, have emerged to accumulate manually annotated media entities by exploring the grassroot Internet users, which also facilitates the textbased search. However, 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. To tackle this issue, 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 ashigh computational cost, difficulty in finding visual queries, and the large gap between low-level visual descriptions andusers' 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 verbose questions in natural languages.

C. Multimedia Search Reranking

As previously mentioned, current media search engines are usually built upon the text information associated with multimedia entities, such as their titles, ALT texts, and surrounding texts on web pages. But the text information usually does not accurately describe the content of the images and videos, and this fact can severely degrade search performance [?]. Reranking is a technique that improves 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 is graph-based reranking.The pseudo relevance feedback approach [?], [?], [?] regards top results as relevant samples and then collects some samples that are assumed to be irrelevant. A classification orranking model is learned based on the pseudo relevant and irrelevant samples and the model is then used to rerank the original search results. It is in contrast to relevance feedback where users explicitly provide feedback by labeling the results as relevant or irrelevant. The graph-based reranking approach [?], [?], [?], [?] usually follows two assumptions. First, the disagreement between the initial ranking list and the refined ranking list should be small. Second, the ranking positions of visually similar samples should be close. Generally, this approach constructs a graph where the vertices are images or videos and the edges reflect their pair-wise similarities. A graph-based learning process is then formulated based on a regularization framework. Both of the two approaches rely on the visual similarities between media entities. Conventional methods usually measure the similarities based on a fixed set of features extracted from media entities, such as color, texture, shape and bag-of-visualwords. However, the similarity estimation actually should be query adaptive. For example, if we want to find a person, we should measure the similarities of facial features instead of the features extracted from the whole images [?]. It is reasonable as information seekers are intended to find a person rather than other objects. In this paper, we categorize queries into two classes, i.e., person-related and non-person-related, and then we use the similarities measured from different features according to the query type.

III. EXPERIMENTS

In this section, we introduce the empirical evaluation of the proposed scheme. We first introduce the experimental settings, such as dataset and ground truth labeling. Then, we present two kinds of evaluation. One is local evaluation which tests the effectiveness of the components in the scheme, such as answer medium selection, query selection, and multimedia search reranking. The other one is global evaluation which tests the usefulness of the enrichment of media data for question answering.

A. Experimental Settings

Our dataset for experiments contains two subsets. For the first subset, we randomly collect 5; 000 questions and their corresponding answers from WikiAnswers. For the second subset, we randomly collect 5; 000 questions and their best answers from the dataset used in [?], which contains 4; 483; 032 questions and their answers from Y!A. Here we use the best answer that is determined by the asker or the community voting7. Inspired by [?], [?], we first classify all the questions into two categories: conversational and informational.Conversational questions usually only seek personal opinions or judgments, such as "*Anybody watch the Bears game last night*", and informational questions are asked with the intent of getting information that the askers hopes to learn or use via fact-oriented answers, such as "*What is the population of Singapore*". There are several automatic algorithms for the categorization of conversational and informational questions, such as the work in [?] and [?]. But since it is not the focus of our work, we perform the categorization with human labeling. To be specific, each question is labeled by at least two volunteers independently.

In the case that the first two volunteers have different decisions about the question type, we solicit two additional volunteers to label this question again. The question will be viewed as ambiguous if the four voters cannot come to a majority classification. It is worth noting that each volunteer was trained with the question type definition as well as corresponding examples before labeling. This question type labeling process is analogous to [?]. In this way, we extract 3; 333 informational questions from the Y!A subset and 4; 000 from the WikiAnswers set. The QA pairs in our dataset cover a wide range of topics, including travel, life, education, etc.

The answer medium selection and query selection components need to learn classifiers based on several training data, and thus we split the 7; 333 QA pairs into two parts, a training set that contains 5; 866 QA pairs and a testing set of the remaining 1; 467 QA pairs. The testing set consists of 800 QA pairs from WikiAnswers and 667 from Y!A. Classification models are trained with the whole training set, i.e., 5; 866 QA pairs. They are tested on the 800 QA pairs from WikiAnswers, 667 QA pairs from Yahoo!Answers, or the both. For ground truth labeling (including the ground truths for answer medium selection, query generation, and the relevance of media data), the five volunteers that have been involved in the labeling task of [?] were involved again, including two Ph.D. students and one faculty in computer science, one master student in information system, and one software engineer. The labelers are trained with a short tutorial and a set of typical examples. We need to admit that the ground truth labeling is subjective. But a majority voting among the five labelers can partially alleviate the problem. We have also analyzed the inter-rate relability of the labeling tasks with the fixed marginal kappa method in [?], and the results demonstrate the there are sufficient inter-rater agreements. As an example, we illustrate the labeling analysis results on the answering medium selection ground truths of the 1; 467 testing points in Table III. The Kappa value is greater than 0:7, and it indicates a sufficient inter-rater agreement.

B. Evaluation of Reranking

To evaluate the method of judging whether a QA pair is person-related or non-person-related, we select 500 QA pairs from each dataset randomly. Table XI illustrates the statistics of the person-related and non-person-related classes based on these two subsets. Then, we learn an SVM model with RBF kernel based on 7-dimensional facial characteristics. The parameters are turned by 10-fold cross-validation. The results are illustrated in Table XII. We can see that our approach achieves fairly good performance. In the table, we also illustrate the performance of the method in [?] for comparison. The method in [?] mainly relies on the analysis of the text content of questions and answers, and thus we denote it as "text-based method". In order to evaluate our query-adaptive strategy, we first randomly selected 25 queries from the person-related ones. For each query, the top 150 images or videos are collected for reranking. We adopt NDCG@10 as our performance evaluation metric, which is estimated by

$$NDCG@n = \frac{DCG}{IDCG} = \frac{(rel_1 + \sum_{i=2}^{n} \frac{rel_i}{\log_2 i})}{IDCG}$$

where *reli* is the relevance score of *i*-th image or video in the ranking list, *IDCG* is the normalizing factor that equals to the *DCG* of an ideal ordering. Each image or video is labeled to be very relevant (score 2), relevant (score 1) or irrelevant (score 0) to a query by the voting of the five human labelers. Figures 5 and 6 illustrate the average performance comparison of our approach and the conventional method that uses only global features for the 25 person-related queries. Here we illustrate the performance with different values of the parameter _. Smaller _ means more initial text-based ranking information is taken into consideration. We can see that, our approach consistently outperforms the method that uses global features. This demonstrates that, in image or video reranking, it is more reasonable to use facial features for person-related queries. We then randomly select 100 queries from image and video class, respectively. We compare the following methods by conducting experiments on these queries:

(1) The conventional method that only uses global features. It is denoted as "conventional".
(2) Query-adaptive reranking with the text-based query classification strategy in [?]. That means, we use the methodin [?] to perform query classification and then adopt queryadaptive reranking accordingly. It is denoted as "text-based query-adaptive".

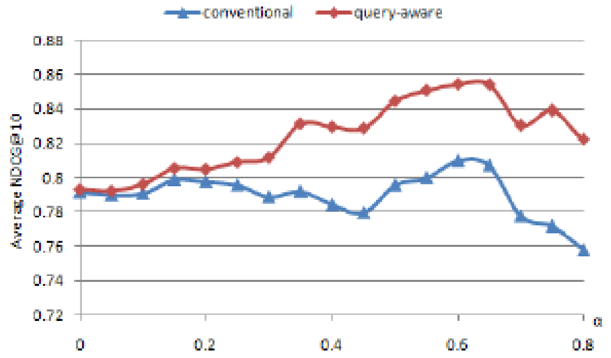


Fig 2.The image search reranking performance comparison of using our method and using the conventional method.

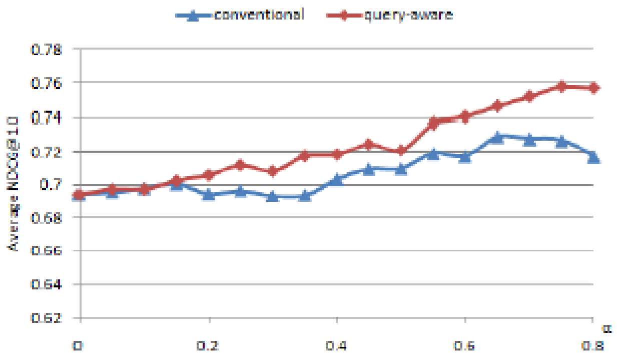


Fig 3.The video search reranking performance comparison of using our method and using the conventional method.

(3) Query-adaptive reranking with the proposed query classification strategy. That means, we use the proposed facial analysis method to perform query classification and then adopt query-adaptive reranking accordingly. It is our proposed approach and it is denoted as "proposed". Figures 7 and 8 illustrate the average performance comparison. From the results we can see that the two query-adaptive methods consistently outperform the conventional method that uses only global features. The proposed approach performs better than the "text-based query-adaptive" method due to the more accurate classification of person-related and non-personrelated queries. Throughout our rest experiments, we set_ as 0:65 and 0:8 for image and video reranking, respectively. After reranking, we perform duplicate removal and present the images or/and videos together with the textual answers, depending on the results of answer medium selection.

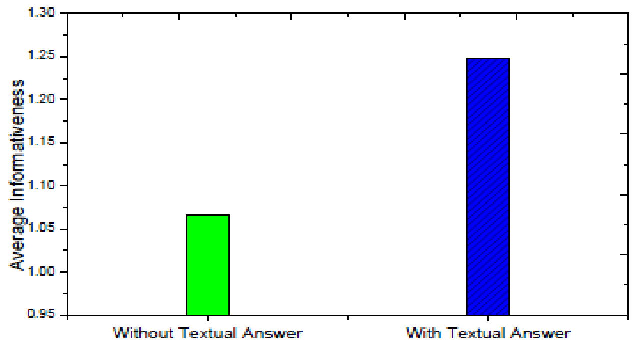


Fig 4.Comparison of overall average informativeness scores between with and without textual answers.

Finally, we conduct a user study to compare the original textual answers and the media answers generated withoutthe assistance of textual answers. We adopt the experimental settings introduced in Section VI-*F* and present the user study results in Table XVI. It is interesting to see that, although the media answers are not as informative as those generated with the assistance of textual answers, they are still very informative in comparison with pure textual answers. Therefore, we can draw several conclusions from the investigation. First, there will be informativeness degradation for the obtained media data if there is no textual answer. Second, the performance of answer medium selection will also degrade. Third, the obtained media answers can still be useful for many questions.

IV. CONCLUSION AND FUTURE WORK

In this paper, 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 reranking 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 communitycontributed 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.

V.REFERENCES

[1] Trec: The text retrieval conference. see http://trec.nist.gov/.

[2] 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.

[3] 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.

[4] E. Agichtein, S. Lawrence, and L. Gravano. Learning search engine speci_c query transformations for question answering. In *Proceedings of International World Wide Web Conference*,2001.

[5] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. *Proceedings of European Conference on Computer Vision*, 2004.

[6] J. Arguello, F. Diaz, J. Callan, and J. F. Crespo. Sources of evidence for vertical selection. In *Proceedings of ACM International SIGIR conference*, 2009.

[7] M. Bendersky and W. B. Croft. Discovering key concepts in verbose queries. In *Proceedings of ACM International SIGIR conference*, 2008.

[8] J. Cao and J. Jay F. Nunamaker. Question answering on lecture videos: A multifaceted approach. *Proceedings of International Joint Conference on Digital Libraries*, 2004.

[9] T.-S. Chua, R. Hong, G. Li, and J. Tang. From text question-answering to multimedia qa on web-scale media resources. In *Proceedings of ACM workshop on Large-Scale Multimedia Retrieval and Mining*, 2009.

[10] S. Cronen-Townsend, Y. Zhou, and W. B. Croft. Predicting query performance. In *Proceedings of ACM International SIGIR conference*, 2002.

[11] H. Cui, M.-Y. Kan, and T.-S. Chua. Soft pattern matching models for de_nitional question answering. *ACM Transactions on Information Systems*, 2007.

[12] Z. Gyongyi, G. Koutrika, J. Pedersen, and H. Garcia-Molina. Questioning yahoo! answers. Technical report, Stanford InfoLab, 2007.

[13] F. M. Harper, D. Moy, and J. A. Konstan. Facts or friends?: distinguishing informational and conversational questions in social qa sites. In *Proceedings of International Conference on Human Factors in Computing Systems*, 2009.

[14] F. M. Harper, D. Raban, S. Rafaeli, and J. A. Konstan. Predictors of answer quality in online qa sites. In *Proceedings of International Conference on Human Factors in Computing Systems*, 2008.

[15] W. H. Hsu, L. S. Kennedy, and S.-F. Chang. Video search reranking through random walk over document-level context graph. In *Proceedings of ACM International Conference on Multimedia*, 2007.

[16] Z. Huang, M. Thint, and Z. Qin. Question classi_cation using head words and their hypernyms. In *Proceedings of International Conference on Empirical Methods in Natural Language Processing*, 2008.

[17] G. Kacmarcik. Multi-modal question-answering: Questions without keyboards. Asia Federation of Natural Language Processing, 2005.

[18] Y.-S. Lee, Y.-C. Wu, and J.-C. Yang. Bvideoqa: Online english/chinese bilingual video question answering. *American Society for Information Science and Technology*, 2009.

[19] B. Li, Y. Liu, A. Ram, E. V. Garcia, and E. Agichtein. Exploring question subjectivity prediction in community qa. In *Proceedings of ACM International SIGIR conference*, 2008.

[20] G. Li, R. Hong, Y.-T. Zheng, S. Yan, and T.-S. Chua. Learning cooking techniques from youtube. In *Advances in Multimedia Modeling*. 2010.

[21] G. Li, H. Li, Z. Ming, R. Hong, S. Tang, and T.-S. Chua. Question answering over community contributed web video. *IEEE Multimedia*, 2010.

[22] X. Li and D. Roth. Learning question classi_ers. In *Proceedings of International Conference on Computational Linguistics*,2002.