ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 2, Issue 3 (July - Sept. 2014)
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)

Multimedia Answer Generation by Harvesting Web Information

[I]Anjana Sekhar, [II]Shini S.T, [III]Pankaj Kumar
[I]Post Graduate Student, [II]Assistant Professor, [III]Assistant Professor
[I,II]School of Computer Sciences, M G University
[III]Federal Institute of Science and Technology

Abstract
There are many question answer sites are available now a days. Community question answering sites are efficient when compare with the automate question answering sites. The drawback of available community question answering system is that it can only provide textual answer. In this paper, we propose a scheme that enriches the textual answer with multimedia data. Our scheme consists of four components: qa pair extraction, answer medium selection, query generation and selection and presentation. The question answer pair is extracted from the available community question answering sites database. The type media information added with the textual data is determined. Query is generated forthe multimedia data. The resulting data is selected and present to the user.

Keywords
qa pair,medium selection,surf,openCV,QA,query generation

I.Introduction

The social media is changing as new trends rise,the popularity of varies platforms shifts and the market evolves. As per the studies in 2013, about 500 million persons are using social sites.The amount of information on the web is increasing day by day. When users search for a question in internet, he gets a list of documents and user need to browse through each document in order to get thee information. This information overloading problem can be solved with the help of Question Answering systems. In earlier the QA systems mainly focused in some specific domains but question answering systems provide precise answer to the question. Question answering system mainly divided into closed domain system and open domain question. In closed domain system extracting the data from the structured data and convert natural language question into database query. In open domain system, instead of using database, it uses large collection of unstructured data, which help to cover many subjects, information can be added and updated constantly and no manual work for building the database. However sometimes the information is not up to date, more irrelevant information etc demands a more complex system. Question Answering systems can efficiently handle the informative questions such as what,where,when,why, like that. But it can difficult to answer the questions like what is your opinion about,how it could be etc.Automatic QA still has difficulties in answering the complex questions. Community question answering systems can solve this problem. It is a large corpus for sharing technical knowledge but also a place where one can seek advice and opinion. In community question answering,when users post a question the answer is obtained from different sources with different participants. The problems in automate question answer can be replaces with answer that contain human intelligence [4]. So the gap between the question and answer is bridged by the crowed sourcing intelligence of community members. The existing community question answering systems such asyahoo!Answers, WikiAnswers, stack overflow,AskMetafilter etc.The problem with this available community question answering sites is that, it can only provide textual answers or urls that link to supplementary to the images or videos. These textual answers are not sufficient to answer some questions. Example "How the currency looks like?" The textual answers have some limitation to answer it correctly. If this question is answered with the image of currency, it will be more informative.So we introduce a multimedia question answering site.In multimedia question answering tried to enrich the textual answers obtained form available community question answering site and enriches those textual answers with multimedia data such as images and videos. There are mainly four components.

1. Question answers pair extraction.
2. Answer medium selection
3. Query generation
4. Multimedia search and presentation

II. Related Works

The investigation of QA systems started at 1960s, it mainly focused on some specific domains.In the late 1990s QA track in TREC[6] gain popularity in Text based QA.Based on the question and the expected answer, we classify QA into Open Domain QA[7], Restricted Domain QA[8], Definition QA[9] and Listed QA[10].Cqa is an alternative approach.It is a large and diverse question answer forum, which share technical knowledge as well as advices and opinion. However, the existing Cqa systems such as Yahoo!Answers, WikiAnswers, Ask Metafilter,only support pure text based answer which may not provide sufficient information.

Multimedia QA,which aims to answer question with multimedia data.An early system named VideoQA[11] extends the text-based QA technology to support factoid QA by leveraging the visual contents of news video as well as the text transcripts.Several video QA[11] systems were proposed and most of them relay on the use of text transcript derived from video OCR(optical Character Recognition) and ASR(Automatic Speech Recognition) outputs. An image-based QA [12] focused on finding information about physical objects.

In previous paper [1], they created a dataset for that contains two subsets.For the first subset randomly collect 5,000 questions and their corresponding answers from wikianswers. For the second subset,randomly collect 5,000

questions and their best answer from the dataset used in [2],which contains 4,483,032 questions and answers that determine by the asker or the community voting, and make it as a pool of question and answers.Classify all the questions with human labeling. Answer medium selection and query generation and the relevance of media data are checked with a ground truth labeling [5].Five volunteers, including two Ph.D. students and one faculty from computer science, one master student in information system, and

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

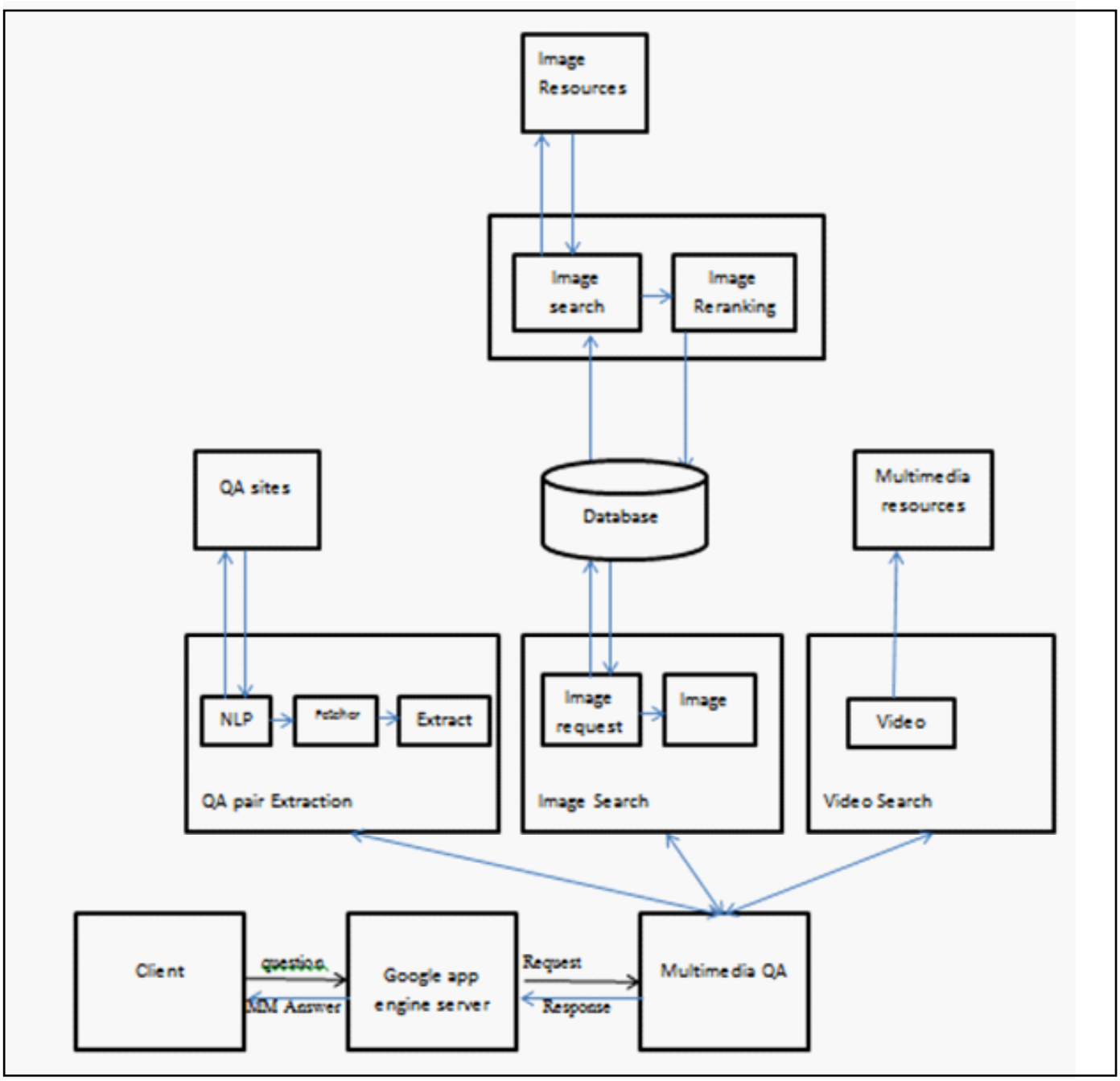

Fig.1: Architecture of MQA

one software engineer. Labelers are trained with short tutorial. The multimedia search selected and eliminate the duplicate and irrelevant pictures with the help of multimodal graph based reranking [3]. Finally the data is present to the user.

III. Architecture

The Fig 1: shows the overall structure of the system. The main functions of this system is to provide well defined interfaces for modules and tools, manage the collaborations of modules and handle all web related aspects.This frame work provides all the requirements such as modularity, flexibility, configurability, scaling, initialization and synchronization.That is, it minimizes the dependency between the modules and between the modules and the framework.Flexibility is achieved by dynamically load modules into the framework and allowsto pass data in any format between the modules.

This proposed framework is an online MQA, user can post his question in website interface, developed with the help of Google app engine server. Our application named MQA will accept the user request and control all activities on top level. MQA act as a work manager to process the question. This work manager sends the requests to each and every module and accepts the responses from that module. This work manager instantiate the main modules such as qa pair extraction, answer medium selection, query generation, multimedia search and presentation. The main function of qs pair extraction is that, the question asked by the user is in natural language, the question is processed and the key words are extracted. These key words are used to search the answer on the available community sites database. The searching result sends the data for the given question. From the give data extract the correct answer and present it back into the work manager MQA.

MQA will select an appropriate answer medium for the given question. If it needs images to enrich the textual answer, the work manager sends a request to the database.If the question is already available in the database then we can directly fetch the image url from the database. Otherwise, if the question is appear in very first time then we search the images on the available resources, download the images and perform matching function for duplicate elimination. After the removal of duplicate and irrelevant pictures, the url of the image are inserted into the database together with the question and its rank. This will help for the future reference; hence we can reduce the time cost.

We can enrich the textual answer with videos also. For that the work manager sends the request to search for the video. After searching on the multimedia resource, it sends the response back to work manager. The textual answer together with image and video is give back to the server and finally server will present it to the client.

IV. Experiments

A. Question answer pair extraction

YahooAnswers,wikianswers,answerbag.,ask.com are the available community question answering sites. In this paper, Question answer pair is selected dynamically from the database of available community question answering sites with the help of api keys of their respective applications.Commonly we fetch question answer pairs from yahoo answers and select the best answer that is provided by the participants. Before searching onthe database, the question that is asked by the user in natural language is need to process. That is the question which contain unwanted words and verbs, so need to eliminate that type of words and select the key word of that question .These keywords are used to perform searching.

B. Answer medium selection

Answer medium selection is help to determine which type of medium is need to add to enrich the textual answer. For example, "When India gets freedom?" this question only needs pure textual answers. But some questions may be like, "Who is KalpanaChawla", provide the textual answer with an image of KalpanaChawla, it will be more informative. Sometimes the questions may be like this, how to cook Sambar. The recipe of Sambar is explained with a video that shows how to cook Sambar, and then it will be easier to understand. So each question needs different medium to enrich the textual data. Based on this analysis we can classify the answers based on the medium as

i. Text
ii. Text + Image
iii. Text + Video
iv. Text +Image +Video

In this paper, this classification is done with the help of monitoring the starting and ending words of the question. If that words include the words like be, can, will, have, when, be there, how+adj/adv, then that type questions only need textual answers. Otherwise it needs further classification with Naïve Bayes classifier as shows in the Table I. In Naïve Bayes classifier, we create class specific related word and their corresponding category.

The question that is post by the user is in the form of natural language.so we need to tokenize the sentence and assign parts of speech to the words. If the question contains the words that are given in the class specific word list, then corresponding category will select for the answer medium search.

Category	Class specific related word list
Text	Distance,Speed,Height,Weight,Age,Date, Birthday,Religion,Caste,Number,Population,Name,Website
Text + Image	Symbol, Figure, Logo, Picture, Photo, Place, Color, Look like, Who, Image, Appearance, Pet
Text+ Video	How to, How do, How can, Story, Recipe, Song, Music, Dance, Film Steps
Text + Image + Video	President, Prime Minister, Singer, Battle, Issue, Earthquake, Tsunami, event, War, Happened

C.Query generation

In this phase, based on the answer medium selected, we search for the multimedia data on the web. Due to the increase in the amount digital information stored on the web, searching for desired information has become an essential task. With the rapid development of content analysis technology, it helps 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 [13] approach uses a term-based specification of the desired media entities, to search for media data by matching them with the surrounding textual description. To boost the performance of text based search, some machine learning techniques that aims to automatically annotate media entities has been in multimedia community[14]-[15]. 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 solve this problem, content based media retrieval [16] 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 user's sematic 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 engine difficult to bridge the gap between textual queries and multimedia data, especially for verbose questions in natural languages. For to search images and videos on the net we need to generate a query. For that first of all select key words from question answer pair and generate a query that contain api key for the search of images in Google image and api key for to search video in the YouTube.

D. Multimedia search and presentation

After searching on the web for the multimedia data with api keys, we get lot of images and multimedia data. Most of the current commercial search engines are built upon text-based indexing and usually return a lot of irrelevant results. So we need to eliminate the duplicate and irrelevant data by reranking the explored visual information to reorder the initial text-based search results. This duplication and irregular images are finding with the help of

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

SURF [17] in openCV. SURF (Speeded Up Robust Features) approximates Laplacian of Gaussian with Box Filter and it relay on determinant of Hessian matrix for both scale and location. SURF uses wavelet responses in horizontal and vertical direction for neighborhood of size 6s.Adequate Gaussian weight are also applied to it. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of angle 60 degrees.

For feature description, SURF uses Wavelet response in horizontal and vertical direction. A neighborhood of size 20sX20s is taken around the key points, where s is the size. It is divided into 4X4 subregions. For each subregion, horizontal and vertical wavelet responses are taken and a vector is formed like this εdx, εdy, ε|dx|, ε|dy|. This when represented as a vector gives SURF feature description with total 64 dimensions. Lower the dimension, higher the speed of computation and matching, but provide better distinctiveness of features. For more distinctiveness, SURF feature descriptor has an extended 128 dimension version. The sum of dx and |dx| are computed separately for dy<0 and dy>0.Similarly, the sums of dy and |dy| are split up according to the sign of dx, thereby doubling the number of features. OpenCV supports both by setting the value of flag extended with 0 and 1 for 64-dim and 128-dim respectively. The sign of the Laplacian distinguishes bright blob on dark background from the reverse situation. In the matching stages, we only compare features if they have the same type of contrast. This minimal information allows for fast matching, without reducing the descriptor's performance.

We compare each image with other images with some feature points. The images with exact equal matching means they are duplicate images. By this way we can find the duplicate image and eliminate them. After the elimination, the urls of the images are store in a database at every time a new question is arises. This is for to reduce the waiting time for the search and presentation of images. If the question asked for the second time then the image urls are fetched from the database instead of search it once more on the web. Finally present it to the user

V. Conclusion and Future Work

In this paper, we describe the motivation and implementation of Multimedia question answering system. For a question, retrieve question answer pair from the available question answering sites database and select an answer medium to enrich the textual answer. Then generate a query for the multimedia search, resulting data are undergoes duplicate elimination and irrelevant data removal. Finally present the answer that contains textual data, images and videos.

In our study, we find out that this image and video data provided with the textual answer will take some seconds than the normal access. So in future these problems need to solve.

References

[1] *LiqiangNie, MengWang,Yuegao,Zheng-Jun Zha, and Tat-Seng Chua, "Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information" IEEE Trans. Multimedia, vol. 15, no. 2, Feb. 2013.*

[2] *M Surdeanu, M. Ciaramita, and H Zaragoza, "Learning to rank answer on large online QA collections." in proc. Association for Computational Linguistics, 2008.*

[3] *Meng Wang, Hao Li, Dacheng Tao, Ke Lu and XindongWu,"Multimodal graph–based reranking for web image search". IEEE Trans.ImageProc vol. 21,no. 11,Nov.2012.*

[4] *L.A Adamic, J. Zhang, E. Bakshy, and M.S. Ackerman, "Knowledge sharing and yahoo answers: Everyone knows something," in Proc. Int. World Wide Web Conf., 2008.*

[5] *L. Nie, Wang, Z. Zha, G. Li and T –S Chua," Multimedia answering: Enrich text QA with media information," in Proc. ACM Int. SIGIR Conf.,2011.*

[6] *Trec: The Text Retrieval Conf. [Online]. Available: http://trec.nist.gov/.*

[7] *S.A. Quarteroni and S. Manandhar,"Designing an interactive open domain question answering system," J.Natural Lang. Eng.,vol. 15,no. 1, pp.73-95,2008.*

[8] *D. Molla and J.L. Vicedo,"Question answering in restricted domains:An overview," Computat. Linguist., vol. 13, no. 1, pp.41-61,2007.*

[9] *H. Cui, M.-Y.Kan, and T.-S.Chua,"soft pattern matching models for definitional question answering," ACM Trans. Inf., vol. 25, no.2,pp. 30-30,2007.*

[10] *R. C. Wang, N. Schlaefer, W. W. Cohen, and E. Nyberg,"Automatic set expansion for list question answering,"inProc.Int. Conf. Empirical Methods in Natural Language Processing,2008*

[11] *H. Yang, T.-S.Chua, S. Wang, and C.-K.Koh, "Structured use of external knowledge for event-based open domain question answering," in Proc.ACM Int. SIGIR Conf., 2003*

[12] *T. Yeh,J. J. Lee, and T. Darrell, "Photo-based question answering,"in Proc. ACM Int. Conf. Multimedia, 2008.*

[13] *I.Ahmad and T. –S.Jang, "Old fashion text-based image retrieval using FCA," in Proc. ICIP, 2003.*

[14] *J. Tang, R. Hong, S. Yan, T. S. Chau, G. J. Qi, and R. Jain, "Image annotation by KNN-sparse graph based label propagation over naisy-taggesd web images," ACM Trans. Intell.Syst.Technol., vol. 2, no. 2, pp.1-15, 2011.*

[15] *J.Tang, X. S. Hua, M. Wang, Z. Gu, G.J. Qi, and X.Wu, "Correlative linear neighborhood propagation for video annotation," IEEE Trans. Syst., Man, Cybern. B, vol. 39, no. 2,pp. 409-416, 2009.*

[16] *S. K. Shandilya and N. Singhai,"Article: A survey on: Content based image retrieval systems," Int. J. Comput. Appl., vol. 4,no. 2,pp. 22-26, 2010.*