International Journal of Evaluation and Research in Education (IJERE)
Vol.5, No.3, September 2016, pp. 227~234
ISSN: 2252-8822

Domain and Intelligence Based Multimedia Question Answering System

K. Magesh Kumar, P. Valarmathie
Computer Science and Engineering, Saveetha Engineering College, Chennai, Tamil Nadu, India

Article Info

Article history:

Received Jun 01, 2016
Revised Aug 27, 2016
Accepted Aug 30, 2016

Keyword:

Multimedia Answers
Naive Bayesian Ranking System
Question Answering System
Question Classification
Ranking
Textual Answers

ABSTRACT

Multimedia question answering systems have become very popular over the past few years. It allows users to share their thoughts by answering given question or obtain information from a set of answered questions. However, existing QA systems support only textual answer which is not so instructive for many users. The user's discussion can be enhanced by adding suitable multimedia data. Multimedia answers offer intuitive information with more suitable image, voice and video. This system includes a set of information as well as classification of question and answer, query generation, multimedia data selection and presentation. This system will take all kinds of media such as text, images, videos, and videos which will be combined with a textual answer. In a way, it automatically collects information from the user to improvising the answer. This method consists of ranking for answers to select the best answer. By dealing out a huge set of QA pairs and adding them to a database, multimedia question answering approach for users which finds multimedia answers by matching their questions with those in the database. The effectiveness of Multimedia system is determined by ranking of text, image, audio and video in users answer. The answer which is given by the user it's processed by Semantic match algorithm and the best answers can be viewed by Naive Bayesian ranking system.

Corresponding Author:

K. Magesh Kumar,
PG Scholar, Computer Science and Engineering,
Saveetha Engineering Collage,
Chennai, Tamil Nadu, India.
Email: kumar.magesh554@gmail.com

1. INTRODUCTION

Question Answering (QA) is a type of information retrieval processing. The system retrieve answers to questions posed in natural language. QA is considered as requiring more multifaceted natural language processing (NLP) techniques than other method of information retrieval such as document retrieval, and it is occasionally regarded as the next step beyond search engines. There are two types of QA system, which is closed domain QA system and Open domain QA system. The Closed domain question answering system works with questions under a specific domain but in an Open domain question answering system works with questions about any subject area and it can be founded on ontologies and world knowledge. In this project the Domain and Intelligence Based Multimedia Question Answering (DIMQA) system is focused as closed domain question answering system has to be developed for the student's education in order to enhance their knowledge. If user can have any doubts on subjects, then they can search the question in DIMQA system. If the answer is already in QA system, then it will return the answer automatically to their questions. If the answer is not found in the QA system means another option available, that is various faculty or users will give answer to the question in the form of text, image, voice and video.

Question answering (QA) is a system for automatically replying a question that is posted in natural language. Comparing to search systems based on keywords, it extremely facilitates the communication among computer systems and humans. It also avoids the aching job of browsing the very enormous amount of educational content which is returned as exact answers by search engines. However, fully computerized QA is still facing challenges which are not simple to handle, such as keen kind of complex questions and the sophisticated syntax, semantics and contextual processing to attain predictable answers. It is experimented that, mostly automated approach is not skilled of obtaining the consequences that are as good as those generated by human intelligence. The DIMQA system allows the users to answer in multimedia feature along with text. Sometimes, Textual answers may not provide sufficient and easily understandable information. The system approaches would help provide answer to users inform of multimedia. Because picture speaks a thousands of words, we are signifying a basic idea from this system that not only concise textual information but also other audio, video and image information can be teamed up with this textual answer to better emphasize it and thus provide a better experience to the all users.

The DIMQA system obtains information online along with particular question on some topic and obtains exact answer form other participant. In This system users share their knowledge according to their interest which is having dissimilar categories and users can search for answer to question from them. The result obtains from the DIMQA system forum are improved answer because that answer generated by human cleverness. Over the year, their huge amounts of answer and question have been accumulated to offer the facility like conservation and search of answered question. The problem with existing system is that they support only textual answer and which is not relevant for many times, if we add associated multimedia content such as image or video and audio to show the process which provide better result. The obtainable system of community based question answering system such as stack overflow, yahoo answer, wiki answer and ask.com provide answer only in textual form but a few question such as How to install Windows OS? .In this case, if provide answer in textual form which is not informative for many user . Associated video or images provide better result, in fact some community forum provide a balancing link to demonstrate the process. It confirms that multimedia content are important to show the process.

2. LITERATURE SURVEY

2.1. From Textual QA to Multimedia QA

The early examination of QA systems started from 1961 and mainly focused on skilled systems in specific domains. Text based QA has gained its research reputation since the organization of a QA track in TREC in the late 1990s [1]. Based on the kind of questions and predictable answers, we can roughly summarize the sorts of QA into Open-Domain QA [2], Restricted-Domain QA [2], Definitional QA [3] and List QA [4]. On the other hand, in spite of the attainment as described above, automatic QA still has some difficulties in answering composite questions. All along with the blooming of Web 2.0, Community question answers becomes an alternative approach. It is a huge and various question-answer discussions, acting as not only a quantity for sharing technical knowledge but also a place where one can seek advice and opinions [3], [5]. Still, nearly all of the obtainable cQA (Community question answers) systems, such as Yahoo!Answers, WikiAnswers and Ask.com, Stack overflow, only support pure text-based answers, which may not give intuitive and enough information.

Some examine efforts have been put on multimedia QA, which is answer questions using multimedia data. Chua et al. [6] projected a comprehensive approach to extend text-based QA to multimedia QA for a range of factoid, definition and "how-to" questions. Their system was prepared to find multimedia answers from web-scale media resources such as Flicker and YouTube. However, article regarding multimedia QA is still moderately thin. Automatic multimedia QA only works in specific domains and can barely handle multifaceted questions. Different from these works, our approach is built based on cQA. As an alternative of directly collecting multimedia files for answering questions, our method only finds image, audio and video to enrich the textual answers provided by users. It makes our approach capable to deal with more common questions and to reach better performance.

2.2. Multimedia Search

Appropriate to the rising quantity of digital information stored over the web, penetrating for preferred information has become an essential task. The research in this area started from the 1980s [7] by addressing the common problem of decision images from a fixed database. With the quick development of content analysis technology in the 1990s, these efforts rapidly expanded to attempt the video and audio retrieval problems [7],[8]. In general, multimedia seeks efforts can be classified into two categories: text-based search and content-based search. The text-based search [9] approaches works with textual queries, a term-based requirement of the desired media entities, to search for media data by matching them with the

neighboring textual descriptions. To improve the performance of text-based search, some machine learning techniques that aim to mechanically annotate medium entities have been proposed in the multimedia community [5],[10],[11]. Additionally, a number of social media websites, such as Flickr and Facebook, have emerged to build up manually annotated medium entities by exploring the grass root Internet users, which also facilitate the text-based search. Conversely, user-provided text definition for media data are often biased towards individual perspective and context cues, and thus there is a break between these tags and the content of the medium entities that common users are interested in. To attempt this issue, content-based media retrieval [6] performs exploration by analyzing the contents of medium data rather than the metadata. Despite the marvelous improvement in content-based retrieval, still it has several limitations, such as high computational cost, trouble in finding visual queries, and the large break between low-level visual descriptions and user's semantic anticipation. As a result, keyword-based search engines are still broadly used for media exploration. However, the inherent limitation of text-based approaches build that all the present commercial media search engines tricky to link the gap between textual queries and multimedia data, particularly for wordy questions in natural languages.

2.3. Multimedia Search Re-ranking

As before mentioned, present media search engines are typically built upon the text information linked with multimedia entities, such as, ALT texts, and surrounding texts on multiple web page. But the text information typically does not exactly express the content of the images and videos, and this information can cruelly degrade search routine [12]. Re-ranking is a technique that improves seeks significance by mining the visual information of images and videos. Obtainable re-ranking algorithms can mostly be categorized into two methods, one is pseudo relevance feedback and the other is graph-based re-ranking.

The pseudo relevance feedback approach [9],[11],[13] regards top consequences as applicable samples and then it collects some samples that are unspecified to be irrelevant. A categorization or ranking model is educated based on the pseudo applicable and immaterial samples and the representation is then used to re-rank the original seek results. It is in distinguished to relevance feedback where users clearly provide opinion by cataloging the results as relevant or irrelevant.

The graph-based re-ranking approach [12],[14]-[16] regularly follows two assumptions. First, the disagreement between the first ranking list and the refined ranking list should be small. Second, the ranking positions of visually related samples should be close. Usually, this approach constructs a graph where the vertices are images or videos and the edges imitate their pair-wise similarities. A graph-based learning process is then formulated based on a regularization structure.

Both of the two approaches rely on the visual similarities between medium entities. Conservative methods usually calculate the similarities based on a fixed set of features extracted from medium entities, such as color, texture, shape and bag-of-visual words. However, the resemblance estimation actually should be question adaptive. For example, if we want to find a person, we should calculate the similarities of facial skin texture instead of the features extracted from the whole images [17]. It is sensible as information seekers are future to find a person rather than other objects.

3. PROPOSED SYSTEM

The domain and intelligence based Multimedia QA system having the following features; these are Search and Post Questions, Document Retrieval, Answers Extraction, Answers Evaluations, Answering Mode and Ranking.

The multimedia QA system consoles and helps the students and professors by providing their needs. In this system students and professors are consider as users. If users require answer for any question, they can seek the answer in QA system.

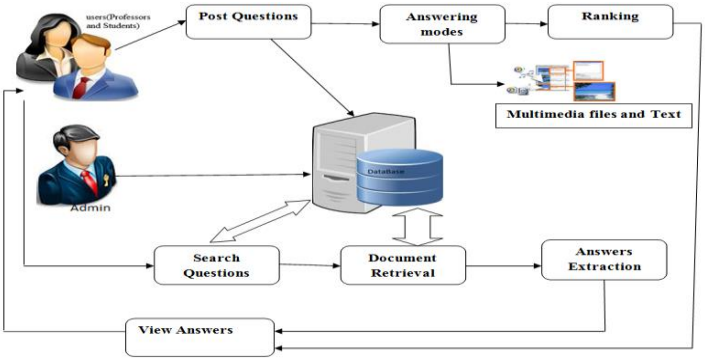

Figure 1. Architecture of student question answering system

If the answer is already in the database the users can retrieve the documents and extract the answer from which they can evaluate it. At last the users can view and utilize the answer. If in case no answer is found in database other user can post the answer to their question in the form of text, image, voice and video. The best answers are ranked by ranking methodology. At last the user evaluates and views the best answers. The following factors are used to develop a DIMQA (Domain and Intelligence based Multimedia Question Answering) system. There are several answering modes available to users, especially this system support multimedia features in Answering mode. Textual answers may not always offer sufficient natural and simply acceptable information. The system's approach would help give answer gainers more concise, comprehensive information and enhanced experience.

As image speaks a thousands of words, this system that not only concise textual information but also other multimedia information can be teamed up with this textual answer to better highlight it and thus provide a better experience to the common users. Answering mode is help to determine which type of medium is required to improve the textual answer. For example, "What is mean by Ring topology?" this question only needs pure textual answers. But some questions may be like, "How to connect the systems into Ring Topology?" provide the textual answer with an image of Ring topology, it will be more informative. Sometimes the questions may be like this," How the systems communicate in ring topology?" The answer is explained with a video that shows how the system communicates with each other, and then it will be easier to understand. So each question needs different medium to improve the textual data. Based on this analysis we can classify the answers based on the medium as:

a. Text
b. Text + Image
c. Text + Video
d. Text +Image +Video

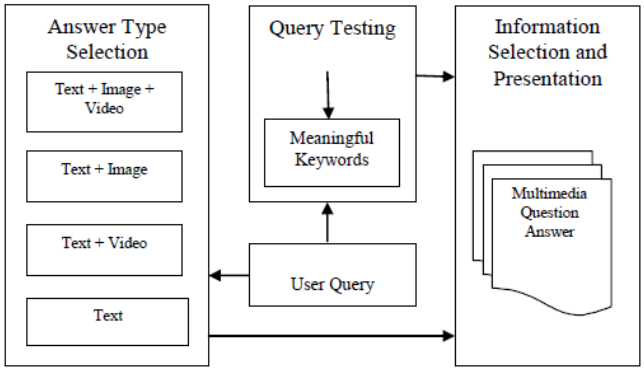

Figure 2. Mode of Answering and Processing

a. Text: it means that unique textual answers are enough.
b. Text+image: it means that textual information is not enough to user so image information must be added.
c. Text+video: it means that textual information and video information must be added.
d. Text+image+video: it means that we add both image and video information along with textual information.

The above answering modes are used to give brief answers to the faculty and student to gain more knowledge the video answers are explained in detail manner. When comparing with the textual answering mode the multimedia answering modes provide the essential information to the users.

4. METHOD

4.1. Semantic Match Algorithm

The Semantic match (S-match) algorithm is trying to close the gap between user command and the need for hyperlink accessibility. The DIMQA system starts in one document and then move through an unending set of documents, which connected by topic. This information networking is based on the proposal of semantic associations, where one unit (node) is connected to another unit (node) by means of a relationship (an edge). Most search engines retrieve information accurately by exploiting key content of associations in Semantic resources, or relations. The Semantic based search engines which rely on information that could be extracting from user query and the ontology for a given document. The idea is to use surviving relations in the ontology names "virtual links" along with apply them to a set of documents to increase the probabilities of finding the inherent associations made by the user at the time of the query. The idea of exploiting ontology-based annotations for information is not latest; semantic search engine would consider keyword concept associations and would return a document only if keywords (or synonyms, homonyms, etc.)Are found within the document and relate to associate concepts. The semantic algorithm used to produce the minimal resulting similar answers in effective QA system.

Within natural language we use a vocabulary of tiny expressions and a grammar to build well-formed and meaningful expressions and sentences. In the framework of an ontology language the vocabulary is called signature. It can be defined as follows.

4.1.1. Definition of Signature

A signature K is a quadruple K = (C, P, R, I) where C is a set of concept names, P is a set of object property names, R is a set of data property names, and I is a set of individual names. The union P ∪ R is referred to as the set of property names.

4.1.2. Definition of Similarity Search Algorithm σ

Given two ontologies O1 and O2 and their signatures S1 = {C1, P1, R1, I1i} and S2 = {C2, P2, R2, I2}, a similarity search algorithm σ is defined as σ(S, SimlrString) → T where S = C2 | P2 | R2 | I2 is the search space such that T ∈ S. SimlrString ∈ S1 is a search string. T type should be same as SimlrString, i.e. SlrString ∈ C1 will lead to T ∈ C2 and so on. By reducing the problem with just considering one name from S1 as SimlrString, we tried to keep the algorithm more general, so it could be used by other applications such as search engines, which need to find a concept in ontology similar to a search text. For the sake of the simplicity, in the followings, we only refer to concepts but similar methods could be applied to search for other parts of signatures.

4.1.3. Algorithm of Similarity Search Algorithm

FINDINGSIMILARITY (SimlrString, OntoSearchList)
1: First tries to find resource that are similar to SimlrString directly
2: SimlrOntRes ←FINDLEXICALSIMILAR (SlrString, OntoSearchList, IsubT hrshld)
3: if SimlrOntRes 6= NIL then
4: if SEMANTICFILTERACCEPTS (SlrOntRes.LocalN ame, SimlrString) then 5: return SimlrOntRes
6: end if
7: end if
8: B Creating Search Matrix
9: M ← WORDNETNUMBEROFMEANING (SimlrString)
10: SimlrMatrix ← BUILDEMPTYSIMILARITYMATRIX (M)
11: for i ← 0 to M − 1 do
12:ADDTOROW(SimlrMatrix,i,WORDNETGETSYNONYMS(SimlrString,i))
13:APPENDTOROW(SimlrMatrix,i,WORDNETGETHYPERNYMS(SimString ,i))
14: end for
15: B Calculate Most Similars
16: CALCULATESIMILARITIES (Ont0SearchList, SearchMatrix)
17: CandidateArray ← BUILDARRAY (M)
18: for i ← 0 to M − 1 do
19: CandidateArray[i] ← FINDCANDIDATE (SearchMatrix, i)
20: end for
21: B Word Sense Disambiguation
22: preferredMeaning ← WSD (SearchMatrix[i])
23: if CandidateArray [preferredMeaning][i] 6= NIL then
24: return CandidateArray [preferredMeaning].MostSimilarOntRes

25: end if 26: B If WSD failed 27: for i ← 0 to M − 1 do
28: if CandidateArray[i] 6= NIL then
29: return CandidateArray[i].MostSimilarOntRes
30: end if
31: end for
32: B Not found
33: return NIL

4.2. Naive Bayesian Ranking Algorithm

The Naive Bayesian Ranking Algorithm helps to rank the audio and video by the user based review. This is one of the most effective algorithms to rank the audio and video files. The described problem of ranking and suggesting things are arise in a variety of applications include interactive computational scheme for helping people to power social information; in technological these systems are called social navigation systems. These social navigation systems help each individual and their performance and their decision making over selecting answers. Based on the each personalities reply the ranking and suggesting of popular items was done. The person's opinion might be obtained by displaying a set of suggested answers, where the selection of answers is based on the liking of the entity. The plan is to propose accepted items by rapidly studying the true popularity ranking of answers. By this method proposed in, which defines a score for a query based on the relative entropy between the query and collection language models.

$$Clarity_q(C_i) = \sum_{w \in V_{ci}} P(w|\theta_q) log_2 \frac{P(w|\theta_q)}{P(w|\theta_{C_i})} \tag{1}$$

Where V_{ci} is the entire vocabulary of the collection C_i, and i = 1*;* 2*;* 3 represent text, image and video, respectively. The Terms $P(w|\theta q)$ and $P(w|\theta C_i)$ are the query and collection language models, respectively. The Clarity value becomes smaller as the top ranked documents approach a random sample from the collection. The query language model is estimated from the top documents,
R, as the following formula,

$$P(w|\theta_q) = \frac{1}{\mathcal{Z}} \sum_{d \in \mathcal{R}} P(w|D) P(q|D) \tag{2}$$

and $\mathcal{z}$ is defined as,

$$\mathcal{Z} = \sum_{D \in \mathcal{R}} P(q|D) \tag{3}$$

Where $P(q|D)$ is the query likelihood score of document D.
We apply this method to calculate,

$$P(q|D) = \prod_{w \in q} P(w|D) \tag{4}$$

In this work, for a query generated from a given QA pair, we use multiple documents (for several complex queries, there may be less than 20 results returned) to estimate the retrieval effectiveness for each medium type, including text, image and video. The Naive Bayesian Approach represents the class-specific related words in multiple formats.

Table 1. Representative Class-Specific Related Words

Categories	Class-Specific Related Work 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, singerm battle, event, war, happened, etc

5. EXPERIMENTAL RESULTS AND ANALYSIS

The Data set for experiments have two subsets. First, arbitrarily collect some question from wiki answers and for second, collect a few question and their relative answers from Y!A .Ranking method is used to find/select the best or finest answer from the database and here for ranking the vote is collected from the user based on Naive Bayesian algorithm.

To calculate our answer, the medium selection approach is used here it mentions the label that involved in the ground truth labeling process .They are helpful in answer medium information ,for example How to connect a system using ring topology. For this question, if the answer for relevant question is available in the database then the QA system easily retrieve the answer from the database. Otherwise the question is posted in the QA system by the user for this posted question the user can post the answers in the way of text, video and images. In this system a set of user will expect multimedia answer because when comparing to the textual answers the video answers are more informative .In this table it shows that more than 50% of the question can be answered by adding multimedia contents instead of purely text. So we can conclude that multimedia approach highly preferred. According to the result, the comparison between original textual answer and media answer the multimedia answering system is more useful because the textual answer have only text and it is less preferable by the user but the answers with the text and multimedia features are more effective and understandable . The table shows the actual Assistance of only textual answer and textual answers with multimedia features. The experimental settings present the user study result in Table 1. According to study more than 70% people prefers media answer along with textual answer. It is more important community member to provide the answer with media data to better understand the question.

Table 2. Comparison of System Usability

Prefer media answer	No answer	Prefer original textual answer
70%	5%	25%

The following table (Table 3) compares the multimedia feature with Y!A and Wiki Answers, the Y!A and Wiki Answers does not support multimedia answers but the DIMQA provide the 95% of effective answers in multimedia format.

Table 3. DIMQA with Multimedia Feature

Method	Y!A	WikiAnswer	DIMQA
Text-Base method	82.17%	85.26%	92.45%
Multimedia approach	NIL	NIL	95%

The following chart (Figure 1) shows the result of DIMQA system using only textual answer and textual with multimedia answer.

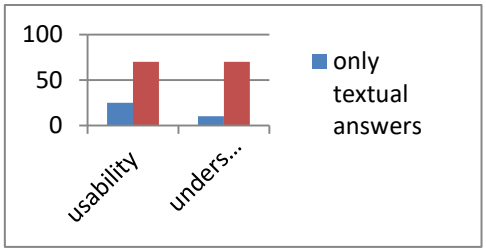

Figure 3. DIMQA system performance evaluation

The answering modes are used to give brief answers to the faculty and student to gain more knowledge and the video answers are explained in a detailed manner. When comparing with the textual answering mode, the multimedia answering modes provides the essential information to the users.

6. CONCLUSION AND FUTURE ENHANCEMENT

The QA system developed by the semantic match algorithm and Naive Bayesian ranking algorithm which allows the multiple users to share their answer in the way of text, image, audio and video. The effective way of answering modes is evaluated with semantic match and naïve Bayesian algorithms and it provides the best answer for the users. With the help of this system every student and faculty will get good knowledge on various subjects. This system can be further extended to generate the automatic Question bank for a student and generate the question papers for college examination. Mainly, the system further extends with e-professor mode to get immediate solution for the question which means it will design for online chats.

REFERENCES

[1] L. Nie, *et al.*, "Beyond text QA: Multimedia answer generation by harvesting Web information," *Multimedia, IEEE Transactions on*, vol/issue: 15(2), pp. 426-441, 2013.
[2] A. Moschitti and S. Quarteroni, "Linguistic kernels for answer re-ranking in question answering systems," *Information Processing & Management*, vol/issue: 4(7,6) pp. 825-842, 2011.
[3] C. H. Hsu, *et al.*, "Using domain ontology to implement a frequently asked questions system," *Computer Science and Information Engineering, 2009 WRI World Congress on.*, vol. 4, 2009.
[4] R. C. Wang, N. Schlaefer, W. W. Cohen, E. Nyberg, "*Automatic set expansion for list question answering,"* in Proc. Int. Conf. Empirical Methods in Natural Language Processing, 2008.
[5] E. Parzen and F. Hoti, "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, vol/issue: 33(3), 1962.
[6] T. S. Chua, *et al.*, "From text question-answering to multimedia QA on web-scale media resources," *Proceedings of the First ACM workshop on Large-scale multimedia retrieval and mining, ACM*, 2009.
[7] M. Wang and X. S. Hua, "Active learning in multimedia annotation and retrieval: A survey," *ACM Trans. Intell. Syst. Technol.*, vol/issue: 2(2), pp. 10–31, 2011.
[8] Y. Gao, M. Wang, Z. J. Zha, Q. Tian, Q. Dai, N. Zhang, "Less is more: Efficient 3d object retrieval with query view selection," *IEEE Trans. Multimedia*, vol/issue: 13(5), pp. 1007–1018, 2011.
[9] I. Ahmad and T. S. Jang, "Old fashion text-based image retrieval using FCA," in *Proc. ICIP*, 2003.
[10] D. Liu, *et al.*, "*Tag Ranking*," Proc. 18th Int‟l Conf.World Wide Web, ACM Press, pp. 351-360, 2009.
[11] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, X. S. Hua, "*Bayesian video search reranking*," in Proc. ACM Int. Conf. Multimedia, 2008.
[12] S. Liu, *et al.*, "Social visual image ranking for web image search," Advances in Multimedia Modeling, Springer Berlin Heidelberg, pp. 239-249, 2013.
[13] H. Feng, A. Chandrashekhara, T. S. Chua, "*Tamra: an Automatic Temporal Multiresolution Analysis Framework for Shot Boundary Detection*," Proc. Int‟l Conf. Multimedia Modeling (MMM), ACM Press, 2003.
[14] S. Lazebnik, C. Schmid, J. Ponce, "*Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS Press, 2006.
[15] H. Bay, *et al.*, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol/issue: 110(3), pp. 346-359, 2008.
[16] J. L. Song, "*Scable Image Retrieval Based on Feature Forest*," Proc. Asian Conf. Computer Vision, Springer Press, 2009.
[17] D. R. Radev, *et al*., "*Evaluating Web-based Question Answering Systems*," Proc. Int‟l Conf. Language Resources and Evaluation, 2002.