Fusing Textual and Visual Ontology with "K-means Algorithm" for Improving Retrieval Accuracy using Voting Annotation Technique

Miss. Pallavi H. Dhole
Department of CSE,
P.R.Patil College of Engg. & Tech. Amravati

Prof. Ajay B. Gadicha
Department of CSE,
P.R.Patil College of Engg. & Tech. Amravati

Abstract

Now a days the popularity of digital images increases due to the improving digital imaging technologies and convenient availability facilitated by the internet. So to find user intended images from internet is very difficult. It is an important issue that how to retrieve the images accurately on the world-wide web. This paper presents an advanced framework for web image retrieval search engine which relay not only on ontology to discover the semantic relationship between different keywords inside the web page but also propose a new voting annotation technique extract the shared semantically related keywords from different web pages to eliminate and solve the problem of subjectivity of image annotation of traditional approaches and enhance the performance of the retrieval results by taking the semantic of the correlated data into consideration. The proposed approach is not used only to enhance the retrieval accuracy of web images; but also able to annotated the unlabelled images and try to narrow the semantic gap problem and enhance the retrieval by fusing two basic modalities of web images.

1. Introduction

There has been a growing interest in implementing online Web image retrieval in the semantic level, because of the volume of information continues to increase, there is growing interest in helping people better find, filter and manage these resources [4]. The rapid growth in the volume of such digital images can make the task of finding and accessing image of interest are difficult for users Therefore, some additional and advanced processing is needed in order to make such collections searchable in a useful manner. Google image search, Lycos and AltaVista photo finder

are the current web image retrieval search engines, uses textual information, such as file name, surrounding text, URL, etc., according to keywords input by users to look for images, without considering image content [5,22]. However, when the surrounding words are ambiguous or even irrelevant to the image, the search based on text only will result in many unwanted result images.

Alternatively in Content-based approaches the search engine extracts semantic information from image content features, such as color, shape, texture, spatial location of objects in images, which are low-level features[5]-[15]. The extracted visual information is natural and objective, but completely ignores the role of human knowledge in the interpretation process. The bottleneck to the efficiency of Content-based approaches is the semantic gap between the high-level image interpretations of the users and the low-level image features stored in the database for indexing and querying. Hybrid approaches combines both the visual content of images and textual information obtained from the Web for the WWW image retrieval [16]-[18]. Such methods exploit the usage of the visual information for refining the initial text-based search result. Especially, through user's relevance feedback, i.e., the submission of desired images or visual content-based queries, the re-ranking for image search results can achieve a significant performance improvement

This paper presents an advanced voting technique that combines the two basic modalities of Web images textual and visual features of image in a re-annotation and search based framework. This advanced framework considers each web page as a voter to vote the relatedness of keyword to the web image, the proposed approach is not only pure combination between image low level feature and textual feature but it take into consideration the semantic meaning of each keyword

that expected to enhance the retrieval accuracy. This approach try to narrow the semantic gap problem and enhance the retrieval precision by fusing the two basic modalities of Web images, i.e., Textual context (usually represented by keywords) and visual features for retrieval. The proposed approach is not used only to enhance the retrieval accuracy of web images; but also able to annotated the unlabelled images. They evaluated the proposed model using Purity for K-Means clustering algorithm. The results demonstrated a performance improvement compared to the traditional vector space model and latent semantic indexing model [10]. More NLP techniques may be included to enhance the performance of the text extraction.

2. Previous and proposed work

Recent research in web image retrieval suggested a combine use existing textual context and visual features can provide better web image retrieval results [1, 2, 7, 8]. The simplest approach for this method is based on counting the frequency of occurrence of words for semantic automatic indexing. This simple approach can be extended by giving more weights to the words which occur in the alt or src tag of the image or which can occur inside the head tag or any other important tags of the HTML document, because the combination of traditional text-based retrieval and content-based retrieval is not adequate to deal with the problem of image retrieval on the WWW. The first reason is that there is already a lot of clutter and irrelevant information on the web pages. These semantic features are less accurate than annotating text. The second reason is due to the mismatch between the page author's expression and the users understanding and expectation. This problem is similar to the subjectivity of image annotation. The third reason is due to the difficulty to find out the relationship between low-level features and high-level features.

The second approach design a different stand and treats images and texts as equivalent data. It attempts to discover the correlation between visual features and textual words on an unsupervised basis, by estimating the joint distribution of features and words and posing annotation as statistical inference in a graphical model. For example image retrieval system based on decision trees and rule induction was presented in [11,14] to annotate web image using combination of image feature and metadata and employ the state of the art of machine learning technology to learn semantic image concepts from image contents so as to make images be indexed and retrieved like text, while in [15], a system that automatically integrate the keyword and visual features for web image retrieval by using technology of association rule mining. These approaches usually learn

the keywords correlations according to the appearance of keywords in the web page, and the correlation may not reflect the real correlation for annotating Web images or semantic meaning of keywords such as synonym [16]. In semantically rich ontology explain the need for complete descriptions of web image retrieval and improves the precision of retrieval. But, the lack of text information which affects the performance of keyword approach is still a problem in text ontology approach. While Ontology-based image retrieval is an effective approach to bridge the semantic gap because it is more focused on capturing semantic content which has the potential to satisfy user requirements accurately[17,18]. Ontology works better with the combination of web image features [19].this paper presents a new framework for web image retrieval search engine which relay not only on ontology to discover the semantic relationship between different keywords inside the web page but also propose a new voting annotation technique extract the shared semantically related keywords from different web pages to eliminate and solve the problem of subjectivity of image annotation of traditional approaches and enhance the performance of the retrieval results by taking the semantic of the correlated data into consideration.

3. The proposed architecture

There are many unrelated keywords are associated to the web image, in order to improve the retrieval process of web image. So it is necessary to decrease or remove these keywords from the web page. The enhance approach trying to solve this problem of current web image retrieval system by proposing a voting annotation technique which depends on web mining to get the relation between high-level features and the low-level features of web image. The frequency of occurrence of keywords is the key of measuring keywords correlation. The keywords are highly relevant to each other if the keywords appearing together frequently with an image. So considering visually similar images as a transaction and its associated keywords as the items in the transaction, it is very natural process to discover correlation between image low-level features and keywords by applying association rules mining model [14].

In this proposed web image retrieval system consists mainly two phases which are the very important system components, pre-processing phase and semantic retrieval phase. The first phase pre-processing phase is responsible for data and image collection and semantic annotations, while semantic retrieval phase is responsible for actual image retrieval. The next following sections explain in details each phase and its important components.

3.1. Pre-processing phase

Figure 1 shows the diagram of pre-processing phase of web image retrieval. The pre-processing phase has the following main modules: These modules are: (1) Web Crawler module, (2) HTML Parser module, (3) Image Processing module, (4) Natural Language module, (5) Voting Module. Figure 1 shows these modules. Each module from these modules will be composed to a set of functions in terms of system functionality. The following section of the research contains the description of each module and its functions in details.
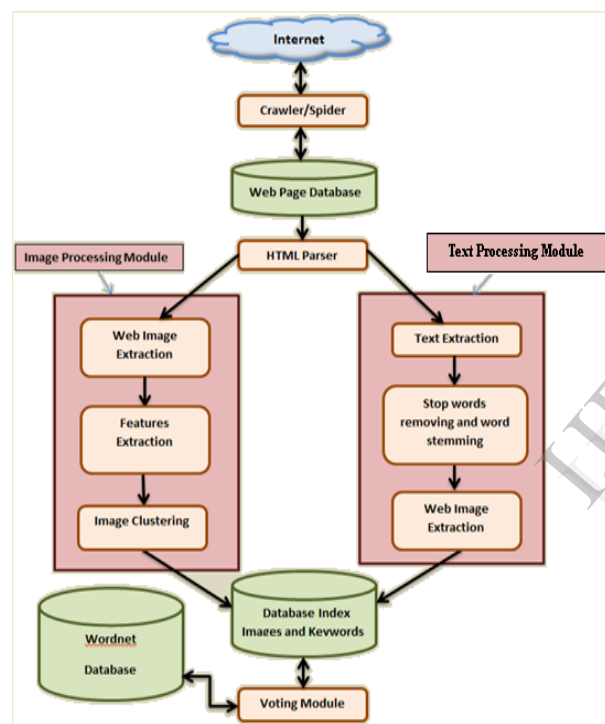

Figure: The Pre-processing phase of web image retrieval

3.1.1 Web crawler module. There are lots of images available on the Web pages. To collect these images, a crawler (or a spider, which is a program that can automatically analyse the Web pages and download the related pages hyperlinked to the analysed Web pages) is used to collect images from many web sites. There are lots of images available on the Web pages. In order to collect these images, a crawler (or a spider, which is a program that can automatically analyse the web pages and download the related pages). Instead of creating a new ontology from scratch, we extend Wordnet, the well-known word ontology, to word-image ontology, Wordnet is one of the most widely used, commonly supported and best developed ontologies. In Wordnet,

different senses and relations are defined for each word. We will use Wordnet to provide a comprehensive list of all classes likely to have any kind of visual consistency. We do this by extracting all non- abstract nouns from the database, 75,062 of them in total, by collecting images for all nouns, we have a dense coverage of all visual forms.
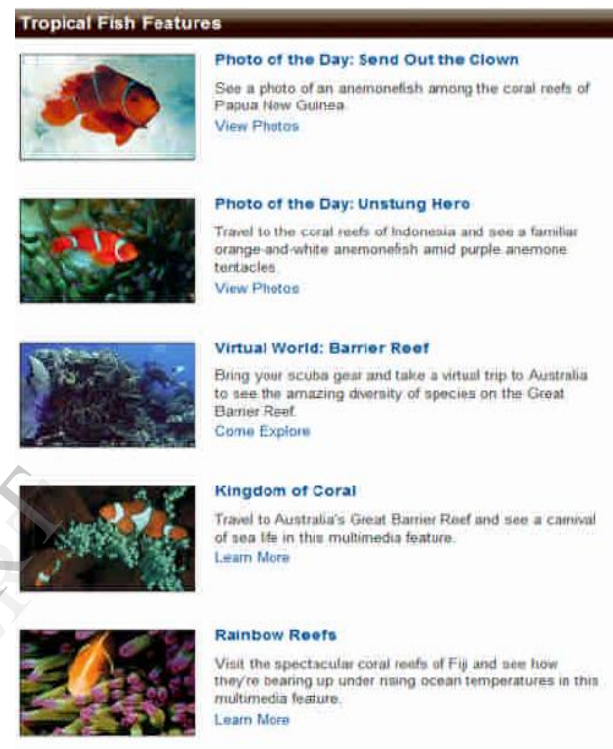

Figure 2: An example of blocks annotation in URL:http://animals.nationalgeographic.com/animals/fish/ clownanemonefish. html

3.1.2 HTML parser module. In this proposed system, we use HTML parser to transform html documents into DOM tree. The DOM Tree is based webpage segmentation algorithm that automatically segments web pages into sections, with each section consisting of a web image and its contextual information (i.e. image segment), and then extract the text and images by traversing through the DOM tree. First, we generate the DOM tree for each web page containing the web images. From the bottom, visual objects like image, text paragraph are identifier as basic elements. The tags such as <TABLE>, <TD>, <TR> and <HR> are used to separate the different content passages. Second, we extract the semantic text information for the images. The extracted information includes: ALT (Alternate) text and the closest title with the images in the DOM tree. Figure 2 is example of structural blocks from the URLhttp://animals.nationalgeographic.com/animals/fis

h/clownanemonefish.html. In fact, we can easily build the DOM tree for a web page through open source tools, the DOM tree will systematically represents the web into a tree view. By counting the repeating sequence of HTML tags, we can easily find out the repeated structural blocks. And we can find out the annotation for each image in the structural blocks easily by referencing the DOM tree which is shown in Figure 3 which is very small portion of DOM tree from Figure 2. We can notice that the text and image are always located in the leaves of the DOM tree. The ALT text in a web page is used for displaying to replace the associated image in a text-based browser. Hence, it usually represents the semantics of the image concisely. We can obtain the ALT text from ALT tag directly. A feasible way is to analysis the context of the web image to obtain the semantic text information of the images.
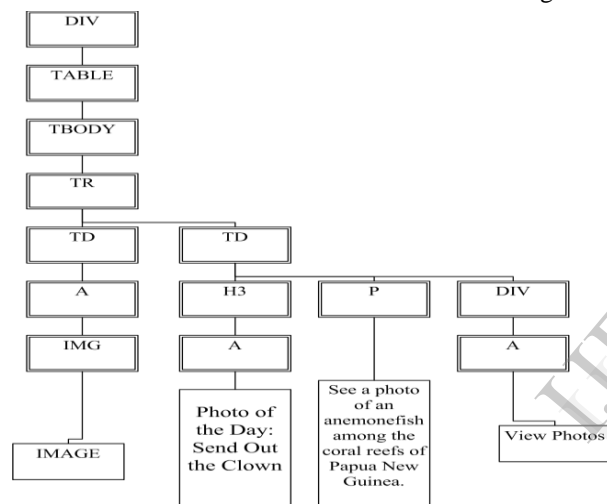

Figure 3: HTML DOM tree example

3.1.3. Text processing module. This module is also called as natural language processing module (NLP). In order to extract terms from text, classic Text Processing techniques are applied, because the raw text of the document is treated separately as well[3,10]. This module is responsible for these functions.

3.1.3.1. Text extraction. The first step of in text processing is extracting textual data from the web pages. Then convert each page into individual text document to apply text pre-processing techniques on it. This step is applied on input Web documents dataset by scanning the web pages and categorizing the HTML tags in each page. Then exclude the tags that contain no textual information like formatting tags and imaging tags. Then extract the textual data from other tags (like paragraphs, hyperlinks, and metadata tags) and store it into individual text documents as input for next steps.

3.1.3.2. Stop words removing and word stemming. Stop words, i.e. words thought not to convey any meaning, are removed from the text. Removal of non-informative words is a commonly used technique in text retrieval and categorization. In this scenario, a predefined static list which consists of hundreds of less meaningful high frequency words (e.g., prepositions and conjunctions), is employed to eliminate irrelevance information in text documents. Such a method is quite advantageous for improving the accuracy of the search results and reducing the redundancy of the computation. By stemming of word, it will be changed into the word's basic form. The documents are first parsed into words. Second the words are represented by their stems, for example 'walk', 'walking' and 'walks' would be represented by the stem 'walk'.

3.1.3.3. Keywords extraction. After stemming each word are then weighted using a normalized (tf-idf). At the end, the text part of the document is represented simply by a set of keywords and weights.

3.1.4. Image processing module. This module is responsible for performing the function that are related to the image, the next section explain these functions in details.

3.1.4.1. Features extraction. The Features extraction is the very important and critical step in preprocessing phase. To extract patterns and derive knowledge from large collections of images, deals mainly with identifications and extraction of unique features for a particular domain. The color feature is extracted using color histogram method. Color histogram is popular because they are trivial to compute, and tend to be robust against small changes in object rotation and camera viewpoint. The color histogram represents an image by breaking down the various color components of an image and extracts the three histograms of RGB colors; Red (HR),Green (HG) and Blue (HB), one for each color channel by computing the occurrences of each color (histogram). After computing the histogram, the histogram of each color is normalized because the images are downloaded from different sites which maintain images with different size.

3.1.4.2. Image clustering using K-means algorithm. Image Clustering is the unsupervised classification of images into groups. The fundamental objective of clustering is to acquire content information the users are interested in from the image group label associated with the image. There is a wealth of clustering techniques available like hierarchical clustering algorithm, partition-based algorithm, mixture-resolving

and mode seeking algorithm, nearest neighbour clustering, fuzzy clustering and evolutionary clustering.

The k-means algorithm is used to perform the clustering process. This choice was mainly motivated by the comparably fast processing of the k-means algorithm compared to other unsupervised clustering. K-means can be described as a partitioning method. It is an unsupervised clustering method that provides k clusters, where k is fixed a priori. K-means treats each observation in data as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. First k points are chosen as centroids (one for each cluster). The next step is to assign every point from data set to the nearest centroid. After that, new centroids of the clusters resulting from the previous step are calculated. Several iterations are done, in each the data set is assigned to the nearest new centroid.

3.1.5. Voting module. In the previous modules all images have been initially annotated and visually clustered. Due to the primarily annotation error, the target image may be primarily annotated with error keyword. The underlying problem that we attempt to correct is that annotations generated by probabilistic models present poor performance as a result of too many "noisy" keywords. By "noisy" keywords, we mean those which are not consistent with the rest of the image annotations and in addition to that, are incorrect. Our assumption in this module is that if certain images in the database are visually similar (located in one cluster) and semantically related to the candidate annotations, the textual descriptions of these images should also be related together. If there is in the cluster image label does not have similarity in semantic to annotations to other images, this means that this image was initially annotated with error keyword. In this module, we attempt to find the proper label for each cluster based on semantic analysis of the different candidate image labels inside the cluster and then using data mining technique to find association rule between each different keyword and the cluster to select most appropriate keyword to this cluster. This module consist these functions:

3.1.5.1. Keyword extraction. In this function the concept of the keyword will be extracted from Wordnet, then the similarity between the different keywords are measured, this process performed because the image may be described by different keywords and different web pages but the meaning is related. Two main relationships between keyword are analysed; taxonomy and Partonomy. Taxonomy divides

a concept into species of kinds (e.g. Car and bus are types of vehicle), while Partonomy divides the concept as a whole into different parts (e.g. Car and wheel). For example, such analysis might show that car is more like a bus than it is a tree, due to the fact that car and bus share vehicle as an ancestor in the Wordnet noun hierarchy

3.1.5.2. Voting image label. Our assumption is that if certain images in the database are visually similar to the target image and semantically related to the candidate annotations, the textual descriptions of these images should also be correlated to the target image. If the target image label does not have similarity in semantic to the candidate image label, this means that the target image was annotated with error keyword. One of the typical data mining functions is to find association rules among data items in a database. To discover the Association between the high-level concept and low-level visual features of images, we need to quantify the visual features by clustering, because the concept space is discrete while the visual feature space is continuous in general. Therefore, we aim to associate the concepts and the visual feature.

Dentitions
1) imageset: A set of one or more images
2) k-images $X = \{x_1, \ldots, x_k\}$
3) count of X: Frequency or occurrence of instance of image X
4) support, s, is the fraction of transactions that contains X (i.e., the probability that a transaction contains X)
5) An imageset X is frequent if X's support is no less than aminsup threshold
6) support, $X \to Y$, probability that a transaction contains $X \to Y$
7) confidence, c, conditional probability that a transaction having X also contains Y
8) An association rule is a pattern that states when X occurs, Y occurs with certain probability.

3.1.5.3. Association rule for annotation process.

(1) Scan the transaction DB to get the support S of each concept K_i and visual cluster C_i, and select those concepts and clusters with support greater than user specified minimum support.
(2) Construct the transaction database D and the basic candidate 2-itemsets based on the existing inverted file. We do not start from 1-itemset because the visual features are very high dimensional and the associations between concepts which are single modality association rules are much stronger than the associations between concepts and low-level features or low-level visual

clusters. If starting from 1-itemset, the concepts and visual feature cluster are equally treated, and then most of the created 2-itemsets based on 1-itemset are concept and concept, but few of concept and visual feature cluster. Our goal is not the association between concept and concept. We are interested in the association between concepts and visual feature clusters. Therefore, only the imageset containing one concept and one visual feature cluster are considered. The existing inverted file relates the concepts to their associated images.

(3) For each concept ki in the cluster cj, calculate the support between concept ki and cluster cj. Supp (Ki, Cj)= Count(Ki,Cj) / Size(Cj). Where count(ki) is the frequency of occurrence of conceptki in the cluster cj; while size(cj) is the total number of visual images in the cluster.

(4) All imageset that have support above the user specified minimum support are selected. These imageset are added
to the frequent imageset.

(5) For each imageset in the frequent imageset, calculate the confidence between concept ki and cluster cj. conf (Ki, Cj)= Count(Ki,Cj) / count(ki) Where count(ki) is the frequency of occurrence of conceptki in the database;

(6) The rules that have confidence >= minimum Confidence are selected to strong rule.

(7) Order all frequent imagset in the strong rule according to their confidence, and then select the concept with highest confidence as a label to the associated cluster.

3.2. Semantic retrieval phase

This semantic retrieval phase framework not only support the text based image retrieval, but also try to enhance the retrieval result by taking into account the semantic meaning of the user's query. When the user provide his/her query of text, the system understands the syntax and meaning of an users query and uses a linguistic ontology to translate this into a query against the visual ontology index and any metadata or keywords associated with the image. Figure 4, shows in details the semantic retrieval phase and its functions. This phase contains two modules, NLP module and ontology reasoning module. The NLP module was discussed in details in section 3.1. The next section explains the ontology reasoning module.

3.2.1. Visual ontology reasoning. The digital image does not tell what the image is about. It is possible to retrieve images from a database using pattern matching techniques, but usually textual descriptions attached to

the images are used. Semantic web ontology and metadata languages provide a new way to annotating and retrieving images. In this proposed system ontology reasoning is the cornerstone of the semantic web, a vision of a future where machines are able to reason about various aspects of available information to produce more comprehensive and semantically relevant results to search queries. Rather than simply matching keywords, the web of the future will make use of ontology to understand the relationship between disparate pieces of information in order to more accurately analyse and retrieve images. Most image retrieval method always assumes that users have exact the mind searching goal in mind. However, in the real world application, the case is that users do not clearly know what they want. Most of the times, they only hold a general interest to explore some related images. The ontology reasoning is based on the semantic associations between keywords. This is achieved by finding which concepts in the ontology relate to a keyword and retrieving information about each of these concepts. By this module the ontology is used for quickly locating the relevant semantic concept and a set of images that are semantically related to the user query are returned.
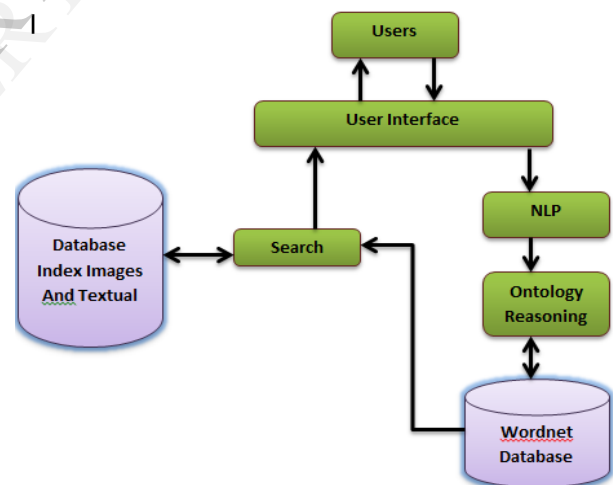

Figure 4: Semantic retrieval phase

4. Conclusions

After a review of existing techniques related to retrieval of web images, we note that these methods are not powerful enough to efficiently retrieve relevant images including the concept of semantic web.

We propose an enhance architecture that combines semantic annotation ,textual features and visual features which are collected from different web pages that share visually similar images which help to widely increase the amount of annotation not from single web page as

traditional approaches. This system use visual ontology, which is a concept hierarchy, is built according to the set of annotations. In the retrieval process to suggest more results that is related to the user's query. Currently, we look for the best appropriate algorithms and methods to deal with interesting relevant descriptive metadata and visual ontology for generating more accuracy results for the users.

5. References

[1] Chang, H.T.,"Web Image Retrieval Systems with Automatic Web Image Annotation Techniques",2008.

[2] Riad, A.,Elaminir, H. and Elghany, S.A. "Web Image Retrieval Search engine based on Semantically Shared Annotation", IJCSI International Journal of Computer Science Issues, 2012, pp. 1694-0814.

[3] Kannan, A., Mohan, V. and Anbazhagan, N. "Image Clustering and Retrieval using Image Mining Techniques", IEEE International Conference on Computational Intelligence and Computing Research, 2010.

[4] Torralba, A., Fergus, R. and Freeman, W.T., "80 Million Tiny Image: A Large Data Set for Nonparametric Object and Scene Recognition",IEEE, 2008, pp. 1958-1970.

[5] Kherfi, M.L., Ziou, D. and Bernardi, A., "Image Retrieval From the World Wide Web: Issues, Techniques, and Systems", ACM Computing Surveys, 2004, pp. 35-67.

[6] Alemu, Y., Koh, J. and Ikram, M., "Image Retrieval in Multimedia Databases: A survey",2009.

[7] He, R., Xiong, N., Yang, L.T. and Park, J.H.," Using Multi-Model Semantic Association Rules to Fuse Keywords and Visual Features Automatically for Web Image Retrieval", 2010.

[8] Hou, J., Zhang, D., Chenl, Z., Jiang, L., Zhang, H. and Qin, X., "Web Image Search by Automatic Image Annotation and Translation", 2010.

[9] Chen Z., Wenvin L., Zhang F. and Zhang H., "Web Mining for Web Image Retrieval",2001.

[10] Fouad, K.M. and Hassan, M.O., "Agent for Documents Clustering using Semantic-based Model and Fuzzy", International Journal of Computer Applications, 2013, pp. 0975-8887.

[11] Wang, J.Z., German, D., Luo, J. and Gray, R.M., "Real World Image Annotation and Retrieval: An Introduction to the Special Section",IEEE Transactions on Pattern Analysis and Introduction to the Special Section, 2008.

[12] Syed, A. and Kanya, B.N., "Extracting Images From The Web Using Data Mining Technique", IJATER, 2012.

[13] Wu, Q., Iyengar, S.S. and Zhu, M., "Web Image Retrieval Using Self-Organizing Features Map", 2001.

[14] Wong, R.C.F. and Leung, C.H.C., "Automatic Semantic Annotation of Real-World Web Image", IEEE transaction on Pattern Analysis and Machine Intelligence, 2008.

[15] He, R., Xiong, N., Kim, T.H. and Zhu, Y.," Mining Cross-Model Association Rules for Web Image Retrieval", 2008.

[16] Xu, H., Zhou, X., Lin, L., Xiang, Y. and Shi, B., "Automatic Web Image Annotation viaWeb-Scale Image Semantic Space Learning", 2009, pp. 211-222.

[17] Mezaris, V., Kompatsiaris, I. and Strintzis, M.G., "An Ontology Approach to Object-Based Image Retrieval", IEEE ICIP, 2003.

[18] Murdoch, O., Coyle, L. and Dobson, S.," Ontology-Based query Recommendation as A Support to Image Retrieval", 2008.

[19] Wang, H., Liu, S. and Chia, L.T.," Does Ontology Help in Image Retrieval?: A Comparison Between Keyword, Text Ontology and Multi-Modality Ontology Approaches", 2006.

[20] Yang, Y., Huang, Z., Shen, H.T. and Zhou, X.," Mining Multi-Tag Association for Image Tagging", 2011.

[21] Agrawal, R., Imielin, T.S. and Swami, A.T.," Mining Association Rules Between Sets of Items in Large Databases", 1993, pp. 207-216.

[22] Gong, Z. and Liu, Q.," Improving Keywords Based Web Image Search with Visual features Distribution and Term Expansion", 2009.