

# Web Image annotation using probabilistic correlation and graph mining

K.SashiRekha

*Department of Computer Science & Engineering  
United Institute of Technology, coimbatore*

B.A.Sabarish

*Department of Information Technology,  
Amrita School of Engineering, coimbatore*

**Abstract-** As the increase the number of images in the digital world makes the need for proper annotation for every image for proper identification. To facilitate the image identification major research has been carried out on automatic annotation to bring a common way of representation and unique identification in the huge image database. Because a single image may related to many label. It can be classified a search engine application also. A threshold multiple approaches can be applied to improve the searching process efficient and reduce the searching time. Annotation can be done with the help of semantic analysis comparison with text in the web page and probability string matching can be applied to search the images. Proposed approach involves the process of creating a graph with weights for each node and making a prediction for the images. Probabilistic approach reduces the number of miss hits by extended training.

## I. INTRODUCTION

In Recent development and rapid increase in the size of data but retrieving a small analysis report and identification of image from that huge data has been a big challenge. As the popularity of digital images many represent the information using images to increase the ability to convey and interpret the information correctly. In the data explosion environment organizing and utilizing the information effectively and efficiently become critical. To satisfy the information requirement need of the customer and improve experience of the users.

In web there is no constraint on type of resources as well as the source for the data. The data can be generated from heterogeneous sources. Have to model a system framework which supports the heterogeneous representation of data and making use of automatic annotation and probabilistic matching.

To increase the reliability of the searching process can be increased by choosing the good quality image data set for the training process. An automatic selection process can be done to choose the training dataset from the image data and candidate images also chosen from the image database. Candidate images are selected from the historical data on the image searching process by various type of user. Query can also be used as an indicator for the type of user, researcher, student or normal user. Based on the query type a dynamic searching process can be chosen and initial stage itself irrelevant images can be removed from the search.

## II. PROPOSED ALGORITHM

In the proposed system uses the probabilistic based approach along with semantic correlation to improve the efficiency of image search. The system consists of three phases: query classification, selection of keyword and calculate the probability of annotation correlation, finally search through the sub-tree to reduce the number of images in the result dataset.

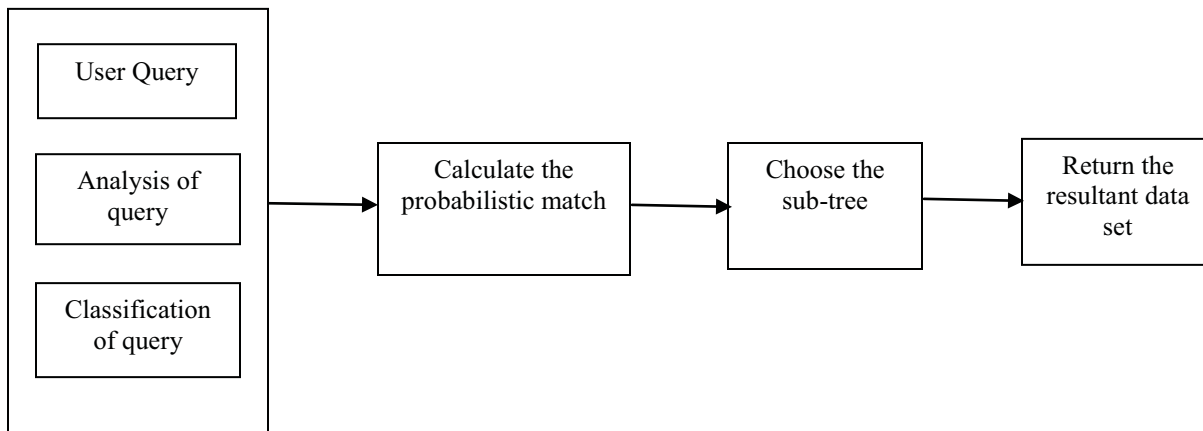


Fig.1. Proposed Architecture

For the training process we select the images from the huge image collection in the web. Images are categorized and annotation is provided based on the content of the image. Tree will be created with the clustered based on the content and weighted tree is created from the probabilistic calculation to match the annotation string, Based on the keyword provided by the user in the query the match is calculated using the correlation and string matching analysis. The clustering of images using annotation reduces the effort in the search process around 23% from the available mechanisms.

### III. EXPERIMENT

Experimental system is designed with the dataset of 2000 images from www from the 10 different broad categories. The categories include vehicles, human, flowers, birds, water reservoirs, fire, animals, painting, plants and building. The experiment can be expanded by narrowing down the classification further that’s from human to men, women and children etc.

Image Classification	Available System	Proposed System
Vehicles	230	190
Human	280	175
Flowers	300	192
Birds	180	178
Water reservoirs	220	180
Fire	210	200
Animal	180	180
Painting	300	190
Plants	280	220
Building	240	200

The dataset consists 200 images in each classification . Systems available and proposed gives the various results. The result shows that the proposed system gives a better reduced set of results which improves the efficiency of search

#### IV.CONCLUSION

The proposed system is tested with the dataset and the result shows the process improves the efficiency. The system efficiency can be increased by increasing the dataset size and making the training process effective. The results of the system show the reduction of resultant data set and increased semantic correlation to improve the efficiency.

#### REFERENCES

- [1] Yiyao Lu, Hai He, Hongkun Zhao, Weiyi Meng , Clement Yu, “Annotating Search Results from Web Databases”, IEEE transactions on knowledge and data engineering, vol. 25, no. 3, march 2013.
- [2] Keiji Yanai, “Automatic Web Image Selection with a Probabilistic Latent Topic Model”, WWW 2008 / Poster Paper April 21-25, 2008 · Beijing, China
- [3] Weifeng Su, Jiying Wang, Frederick H. Lochovsky, Yi Liu, “Combining Tag and Value Similarity for Data Extraction and Alignment “,IEEE transactions on knowledge and data engineering, vol. 24, no. 7, july 2012
- [4] Hao Ma, Irwin King, Michael Rung-Tsong Lyu, “Mining Web Graphs for Recommendations”. IEEE transactions on knowledge and data engineering, VOL. 24, NO. 6, JUNE 2012
- [5] Yi Yang, Fei Wu, Feiping Nie, Heng Tao Shen, Yueting Zhuang, and Alexander G. Hauptmann, “Web and Personal Image Annotation by Mining Label Correlation With Relaxed Visual Graph Embedding, IEEE transactions on image processing, vol. 21, no. 3, march 2012
- [6] Chu-Hong Hoi, Michael R. Lyu, “Web Image Learning for Searching Semantic Concepts in Image Databases”, The Chinese University of Hong Kong
- [7] Zheng Chen, Liu Wenyin, Feng Zhang, Mingjing Li, Hongjiang Zhang, “Web Mining for Web Image Retrieval “Microsoft Research China