International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-12 Issue-7, June 2023

Image Retrieval Through Free-Form Query using Intelligent Text Processing

S. A. Angadi, Hemavati C. Purad

***Abstract:** Image Retrieval is the process of retrieving images from the image/multimedia databases. Retrieval of images are carried out with various types of queries, free-form query is a text-query that consists of single or multiple keywords and/or concepts or descriptions of images with or without the inclusion of wild-card characters and/or punctuations. This work aims to handle image retrieval based on free-form text queries. Simple & complex queries of conceptual descriptions of images are explored and an intelligent processing system with free-form queries based on the Bag-of-Words model is modified and built for natural scene images and on Diverse Social Images using the Damerau-Levenshtein edit distance measure. The efficacy of the proposed system is evaluated by testing 1500 free-form text queries and has resulted in a recall accuracy of 91.3% on natural scene images (of Wang/Corel database) and 100% on Diverse Social Images (of DIV400 dataset). These results show that the system proposed has produced satisfactory performance compared to published results such as the harmonic mean of precision and recall (i.e. F1-Score) of 76.70% & 63.32% at retrieval of 20 images etc in reported works.*

***Keywords:** Free-text query; Damerau-Levenshtein edit distance; Bag-of-Words Model*

I. INTRODUCTION

Text-Query based Image retrieval (TQBIR) works date back to late 1970s, in which the database of images are annotated with the captions or descriptions. Image annotation can be performed manually or automatically. Though there is an advancement in querying methods, still most of the current IR systems and users rely on text-query methods because they need to have database with sufficient images to query and again requires searching of images carrying the required information or semantics to query. For this reason, TQBIR is storage efficient. Normally in TQBIR systems the images are indexed using words associated with the images. The text-query will be formulated via keyword or descriptions and thus helps the user to pose their query to represent the semantics of the images and avoids multiple image submissions as in image-based querying.

Manuscript received on 24 May 2023 | Revised Manuscript received on 26 May 2023 | Manuscript Accepted on 15 June 2023 | Manuscript published on 30 June 2023.
*Correspondence Author(s)
S. A. Angadi, Department of Computer Science and Engineering, VTU, Belagavi, (Karnataka), India. E-mail: saangadi@vtu.ac.in, ORCID ID: 0000-0001-9756-9786
Hemavati C. Purad*, Department of Computer Science and Engineering, VTU, Belagavi, (Karnataka), India. E-mail: hemacpurad@gmail.com, ORCID ID: 0000-0001-7032-1780

However, image based querying systems do not really capture the semantics or meanings of images well. When user submits the text-query, it will be processed to obtain the base words/keywords to match with the image concepts/captions and/or descriptions. Such text preprocessing involves removal of stop-words, stemming, erasing punctuations etc. Further retrieval of the images is performed by employing the matching process which involves string matching or similarity matching to retrieve the relevant images from the database. Searching based on text-query is much faster compared to image-based search. A new methodology for retrieval of images based on Bag-of-Words model for the free text queries is proposed in this paper. Bag-of-Words are constructed for each image category and natural scene images of the Wang database (20 image categories) and Diverse Social Images of DIV400 dataset (132 image categories) are used for experimentation. Free text-query is formulated by having one or more key-concepts or labels of 20 & 135 image categories separately. Intelligent processing of text-query is performed to improvise the performance of the TQBIR system. One thousand and five hundred free text queries, involving concepts and image category names of 20 and 132 image categories are tested from the two datasets, namely Wang(/Corel) and DIV400 to evaluate the efficacy of the proposed system. The system has achieved recall accuracy of 91.3% and 100% on Wang and DIV400 dataset respectively.

The contribution of this work includes, Intelligent processing of free-form text queries using word-length matching to spot the keywords/concepts and concepts mapping using Damerau-Levenshtein edit distance measure and modified Bag-of-Words model in retrieval of images.

The performance of the proposed system outperforms with the recall accuracy (or precision at retrieval of 100 images) of 91.3% and 100% on two datasets, compared to 100% precision at retrieval of 5 images, 38.7% precision at retrieval of 20 images, and harmonic mean of precision and recall (or F1-Score) of 76.70% in scene text retrieval of (Mafla et al. 2020, [1]); 77.68% of precision at retrieval of 20 images, harmonic mean of precision and recall of 63.32% at retrieval of 20 images in Visual and textual based image retrieval of (Yuan et al. 2019, [2]); and 41% precision at retrieval of 5 images, 34.8% precision at retrieval of 10 in medical keyword based medical image retrieval of (Torjmen Khemakhem et al. 2019, [3]) of the state-of-the-art text-query based image retrieval (TQBIR) systems.

The paper is organized as follows. Section 2 describes the literature survey, section 3, presents the proposed method. The experimental setup is presented in Section 4. The section 5 reports the results and discussion. Finally, the conclusion is given in Section 6.

II. LITERATURE SURVEY

The TQBIR system comprises of image annotation, text-query analysis and processing, similarity matching and retrieval of the images. Image annotation plays significant role in the image analysis and in other applications such as image retrieval. There are mainly 3 types in annotation of an image, namely free text annotation, keyword annotation and annotation using ontologies. Annotation can be performed either manually or automatically. Manual image annotation is time-consuming and expensive and exceeds the capacity of manual control and management and hence automatic image annotation yields better results. Some of the Automatic Image Annotation (AIA) methods, such as the generative model, nearest neighbor model, discrimination model, tag completion based model, deep learning based models are described and compared in terms of underlying idea, main contribution, model framework, computational complexity, annotation accuracy in (Q. Cheng et al 2018, [4]). Similarly insights into seven algorithms for the automatic image annotation and evaluation of these algorithms are provided by leveraging different image features, such as color histograms and Gist descriptor in (Yilu. Chen et al 2020, [5]). Manual annotation does not always provide all relevant tags and because of image features, AIA may produce variable length and/or unnecessary labeling and very challenging as well. Considering the fusion of manual and automatic methods results in best image annotation and can reduce the semantic gap in retrieval of the images.

Query formulated as a free text will fall under the scenario called Query-by-Strings (QbS), and formed by keywords, phrases, or sentences. Queries of earlier Image Database (IDB) systems were based on command-oriented language (H. Tamura et al. 1984, [6]). Later systems were implemented to use free text queries, processed by natural language processing (Vijayarajan et al. 2016, [7]; P.M. Ashok Kumar et al. 2021, [8]) or ontology based processing to retrieve the images which are indexed by captions or labels. One such example which uses triples (ARG1-REL-ARG2) a pragmatic relation to index and retrieve the crime scene images is described in (Pastra Katerina et al. 2004, [9]). Normally in ontology based image retrieval systems the free text queries will be first simplified and transformed into SPARQL queries and into RDF triples and these RDF triples are used to search the images (Sohail Sarwar et al. 2013, [10]; Vijayarajan et al. 2016, [7]). The free text queries will avoid the semantic gap problem in retrieval of images (Hui Hui Wang et al. 2010, [11]). Similarity matching between text-query and concepts of the images can be performed by various similarity measures to retrieve the images. The scenario in which text-query containing the keywords, the word-to-word or string similarity or unigram or N-gram similarity is performed based on edit distance metrics such as Levenshtein distance (Levenshtein 1966, [12]), Damerau-Levenshtein distance (Damerau 1964, [13], Levenshtein 1966, [12]), Longest Common Subsequence (LCS) of the two strings, Dice coefficient, and XXDICE (Grzegorz Kondrak, 2005, [14]) etc. Concept similarity is obtained based on ontology (Taxonomic similarity) and/or concept name based similarity measures in (Htet Htet Htun et al. 2017, [15]). (Mihai Lintean et al. 2012, [16]) gives survey on sentence level similarity based on capturing semantics of the sentence and (Mamdouh Farouk 2019, [17], summarizes sentence similarity based on word, structure and vectors and describes datasets used for the same. Most of the current matching strategies are ontology matching and are based on lexicon or structural or instance or semantic or combination of these matching techniques (Thi Thuy Anh Nguyen and Stefan Conrad 2015, [18]). Further these ontology matching strategies require one-to-one or one-to-many alignment of ontologies and machine learning techniques to obtain best matching results. Further text-query based image retrieval is found in the recent works as the cross-modal image-text matching and retrieval process. For instance, images are processed as graphs, and reasoning is performed on the graphs by employing a graph convolutional network (GCN) with attention mechanism and Bidirectional Encoder Representations is adopted to learn distinctive textual representations for matching and cross retrieval of text and image modalities in (Jin Zhang et al. 2021, [19]). Another example describes extracting the features of the image and text modalities and fusing the features and performing similarity matching between the desired image and the source image plus modified text is proposed via deep metric learning in (Chunbin Gu et al. 2022, [20]). Another couple of examples explain usage of vision and language pretrained models such as DenseNet and BERT models in (Zafran Khan et al. 2022 [21]) and ResNet and VLN-BERT models in (Zheyuan Liu et al. 2022, [22]) , to extract image and text features in retrieval of images etc. Polysemy is one of the common issues in handling semantics of the text queries. In this paper an effective approach is proposed to intelligently process the free-form text-query by having modified Bag-of-Words model incorporated with the word-length and N-gram information to quickly spot all the query-words in Bag-of-Words model, and Damerau-Levenshtein edit distance to map the text query-words to the respective image category in order to handle polysemy and semantics of the query-words. Image annotations in this work are made available apriori as Bag-of-Words in the knowledge-base, related to specific 20 image categories and 132 image categories chosen from the Wang and DIV400 databases respectively. The following sections will give detailed explanation of the proposed system.

III. PROPOSED METHODOLOGY

The proposed method handles the free-form queries using the Query-by-Strings scenario with query containing image concepts, and/or keyword/s for retrieval of images in TQBIR using modified Bag-of-Words model. The proposed method consists of two main steps

1) Pre-processing of simple or complex free-form text queries & obtaining Bag-of-Qwords

2) Intelligent processing of free-form text-query to retrieve and display the images,

These are described in detail in the following sub sections. The flow chart of proposed work is given in Figure 1.

Retrieval Number: 100.1/ijitee.G96180612723
DOI: 10.35940/ijitee.G9618.0612723
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

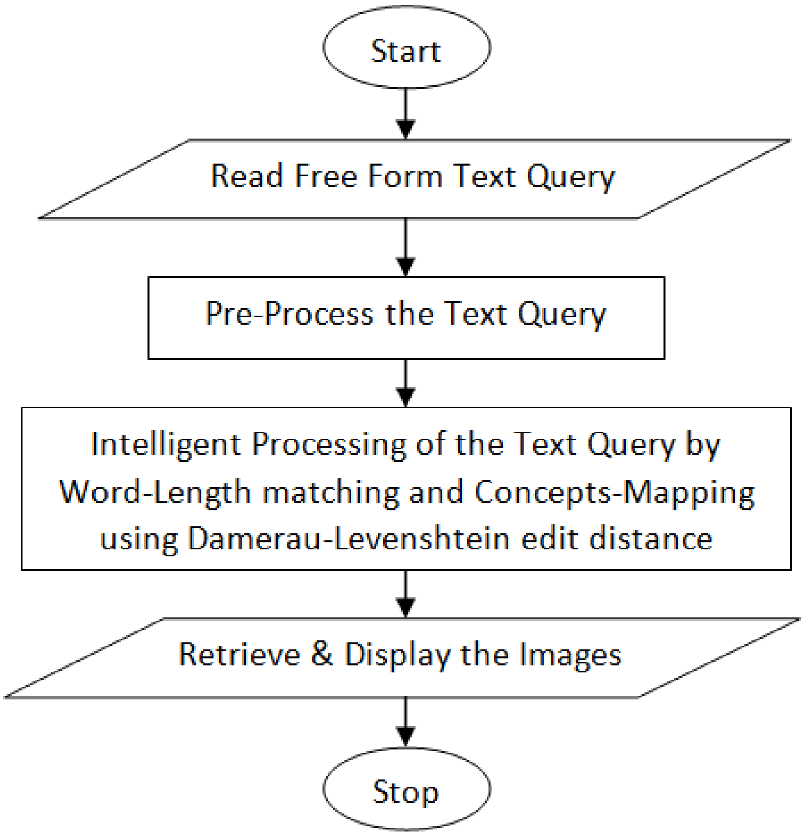

Fig. 1 Flow chart of the proposed work

The flow chart of the proposed work in Figure 1 gives the main steps of the retrieval of images and the detailed explanation of each step of the flow chart is presented in the following.

3.1. Pre-Processinng Of Free-Form Text-Query & Obtaining Bag-Of-Qwords

Pre-processing involves identification and processing of single and/or multiple words or concepts present in the given text-query when the input text-query is given in its simple (single) or complex (multi-word) form that comprises a keyword/s or concept/s with special symbols and/or punctuations. And the following subsections give the detailed explanation of the process.

3.1.1. Simple (Single-Word) Query Processing

Text-query in the simplest form is presented as single keyword and is normalized in the first step. Normalization involves bringing the word to its base or root form by applying Porter's stemming algorithm. Let the query containing single word "α" be an array of characters, such that {c1,…ck, ck+1,…cn} where {c1…ck} is a base word and {ck+1,… cn} is suffix, and the value of "n" is between 2 and 15 ($2 \leq n \leq 15$) characters, which is processed by Porter stemming algorithm to bring into base word "β" such that it is a character array of {c1, … ck} or {c1,…ck,ck+1} sometimes with the additional character to complete the word. Similarly, if the word "α" is a base-word with the characters {c1…cn} is submitted then the stemmer leaves the query word unaffected and is given in equation (1), and is treated as Bag-of-Qword.

$$\alpha \longrightarrow (\beta \text{ or } \alpha) \quad (1)$$

Where LHS "α" is a query word and RHS "β" indicates the base word with some additional characters and RHS "α" indicates the unaffected base word, which is same as query word.

3.1.2. Free-Form (Multi-Word) Text-Query Preprocessing & Obtaining Bag-Of-Qwords

A complex free-form text-query consists of one or more keywords or concepts and may consist of numbers, operators or any symbols. Preprocessing of the free-from (multiword) text-query and obtaining Bag-of-Qwords is detailed in Algorithm 1.

ALGORITHM 1 PREPROCESSING OF FREE-FORM TEXT-QUERY & OBTAINING BAG-OF-QWORDS

Input: A free-form (multiword) text-query "y"
Output: Bag-of-Qwords "**δ**"
Begin
Step-1: Read the free-form (multiword) text-query "y"
Step-2: Convert the text-query "y" into lower case and tokenize
Step-3: Remove the stop-words from the text query
Step-4: Normalize the remaining words of text-query using Porter's algorithm
Step-5: Remove the punctuation
Step-6: Remove short words (w<2) and long words (w>15) (as these types of words are unusual and generally will not convey much information)
Step-7: Output the Bag-of-Qwords "**δ**"
End (begin)

The Algorithm 1 depicts the steps for pre-processing of the free-form (multiword) text-query & obtaining Bag-of-Qwords.

Retrieval Number: 100.1/ijitee.G96180612723
DOI: 10.35940/ijitee.G9618.0612723
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The goal of this procedure is to make a text-query ready for further processing by obtaining Bag-of-Qwords. As a first step the query will be converted to lowercase and tokenized by chopping the multiword query into words, called tokens. Removal of stop-words is applied to remove low information words such as “a”, “the”, “is”, “are” etc. from the tokenized query. Lemmatization and/or stemming is performed to normalize the previously processed text-query words using Porter’s algorithm (M.F. Porter, 1980, [23]) to bring back the words into canonical form of the original word. Punctuations are erased. Very short words containing less than two characters (let ‘w’ denote the word and w<2) which may contain low information or may be a symbol and very long words containing greater than fifteen characters (w>15), which are unusual and may be meaningless and hence such words are removed. The remaining words are Bag-of-Qwords and are saved in datastore. Let “y” be the free-form (multiword) text-query that consists of the strings (let’s say, $S_1, S_2, \ldots, S_k, \ldots, S_n$) of finite length “n” and is converted into lower case and tokenized (as “S_1”, “S_2”, …, “S_k”, …“S_n”) and the stop-words are removed. Finally stemming is applied as in equation (1) to obtain the base words and hence noise in the text-query is eliminated and Bag-of-Qwords “**δ**” are obtained as given in equation (2) which are noise free query strings and are saved for further use.

$$y \longrightarrow \{\boldsymbol{\delta}\} \qquad (2)$$

Where “y” indicates free-form (multiword) text-query and “**δ**” indicates the Bag-of-Qwords (processed text-query words).

For example,

Let the input query be: *“retrieve animal images and red-roses”*

For the input query, the Bag-of-Qwords generated are: *{“retrieve”, “animal”, “ image”, “red”, “rose”}*

3.2.Intelligent Processing Of Text Query

The intelligent processing of text-query comprise of two main steps, namely, generating Qstrings and concepts mapping using modified Bag-of-Words model.

3.2.1. Generating Qstrings

Generating Qstrings is the key step in intelligent processing. Qstrings includes both the Bag-of-Qwords and Bag-of-Ngrams from the free text query, and thus preserves the spatial information of the text-query and presents the possible words from the given text-query for matching and retrieval of the images. The definition of Bag-of-Qwords and Bag-of-Ngrams are given below.

Bag-Of-Qwords:First, Bag-of-Qwords are generated by deploying Algorithm 1 of Section-3.1.2, depending solely on the given text-query and extracts all possible unigrams. Once Bag-of-Qwords extraction is carried out, Bag-of-Ngrams are generated in order to restore actual N-grams.

Bag-Of-Ngrams:Bag-of-Ngrams are generated from the Bag-of-Qwords (noise free query strings). This helps in preserving actual N-grams and spatial information present in the text-query (which might have converted into individual words while obtaining the Bag-of-Qwords) and provides spatial information among text-query words by linking unigrams. The value of ‘N’ is positive and length of N-gram lies within 15 characters (words with more than 15 characters are unusual hence number of characters in the word are limited to 15). *Let Bag-of-Qwords be {“S_1”, “S_2”, …“S_k”,…”S_n”} and let ‘N’ value be two i.e. N=2 the Bag-of-Ngram becomes Bag-of-Bigram such that {“S_1S_2”, “S_2S_k”, “S_kS_n”} are generated from successive words of the given Bag-of-Qwords.*

For instance,

Let previously generated Bag-of-Qwords be: {“retrieve”, “animal”, “image”, “red”, “rose”}

For N=2, the Bag-of-Bigrams are obtained: {“retrieve animal”, “animal image”, “image red”, “red rose”}

Thus the Bag-of-Ngrams will restore the actual N-grams & spatial information present in the input query such as “red rose” in the above example.

Qstrings: are generated by applying set union function over Bag-of-Qwords and Bag-of-Ngrams as in equation (3).

$$Qstrings=\{Bag\text{-}of\text{-}Qwords\} \cup \{Bag\text{-}of\text{-}Ngrams\} \qquad (3)$$

Where Qstrings are the combination of unigram to Ngram-words generated by Bag-of-Qwords and Bag-of-Ngrams, respectively. Qwords are all unigram-words present in the text-query and QNgrams are all Ngram-words generated from Bag-of-Qwords.

For instance the Qstrings = {“S_1”,“S_2”, “S_k”, “S_n”, “S_1S_2”, “S_2S_k”, “S_kS_n”}

for the given Bag-of-Qwords = { “S_1”, “S_2”, … “S_k”,…”S_n” } and

for the given N=2 the Bag-of-Bigrams = { “S_1S_2”, “S_2S_k”, “S_kS_n” }.

Example Qstrings: {“retrieve”, “animal”, “image”, “red”, “rose”, “retrieve animal”, “animal image”, “image red”, “red rose”}

3.2.2. Concepts Mapping

It consists of word-length matching, and concept mapping, that helps to interpret the semantics of the free text-query and maps to the possible images categories in the knowledge-base using modified Bag-of-Words model, to retrieve the images belonging to most relevant and possible image categories from the knowledge-base.

Word-Length Matching: First step of concepts mapping is word-length matching. It operates on numbers and speeds up the process of keyword and /or concepts spotting in the modified Bag-of-Words model. It performs the equality between word-length of each word in Qstrings (query concepts) to word-length of image concepts available in the modified Bag-of-Words model of the knowledge-base to spot the keyword/concepts of text query, as given in the equation (4).

$$W_{Q,I} \longrightarrow (|Q| == |I|) \qquad (4)$$

Where ‘W’ indicates word-length matching function, |Q| and |I| indicates positive length of Qstrings and image concepts (in Bag-of-Words model) greater than zero (because zero indicates absence of the word).

Concepts Mapping: The Qstrings generated from equation (3) from the text-query (preserving the spatial information), which are spotted in the modified.

Bag-of-Words model of the knowledge-base using equation (4), are mapped to most relevant & possible image categories, in the knowledge base, in order to handle the semantics and polysemy of text-query in retrieving most relevant images from the possible image categories of knowledge-base, by employing edit distance measure named Damerau-Levenshtein edit distance as given in equation (5). The polysemy issue is addressed by mapping the Qstrings to all possible, most relevant image categories from the knowledge-base, covering and retrieving wide range of most relevant images for the users.

The Damerau-Levenshtein edit distance (Damerau 1964, [13], Levenshtein 1966, [12]) ensures minimal number of insertions, deletions, substitutions, and swaps and specifically used to handle misspellings (Gregory V. Bard, 2007, [24]) with at most one edit operation. And hence enables the mapping of concepts (even with the misspellings, recoverable with one edit operation) to image categories.

$$d_{Q,I}(i,j) = min \begin{cases} 0 \quad if\ i = j = 0, \\ d_{Q,I}(i-1,j)+1 \quad if\ i > 0, \\ d_{Q,I}(i,j-1)+1 \quad if\ j > 0, \\ d_{Q,I}(i-1,j-1)+1_{(Q_i \neq I_j)} \quad if\ i,j > 0, \\ d_{Q,I}(i-2,j-2)+1 if\ i,j > 1\ and\ a_i = b_{j-1}\ and\ a_{i-1} = b_j \end{cases} \tag{5}$$

Where ‘d’ indicates Damerau-Levenshtein edit distance, ‘Q’ indicates Query strings that is Qstrings, where as ‘I’ indicates word-length matched image concepts in Bag-of-Words of the Knowledge-base. And

$d_{Q,I}(i-1,j)+1 \quad correspondstodeletion(fromQtoI)$

$d_{Q,I}(i,j-1)+1 \quad correspondstoinsertion(fromQtoI)$

$d_{Q,I}(i-1,j-1)$
$+ 1_{(Q_i \neq I_j)}$corresponds to match or mismatch,
depending on whether the respective symbols are the same

$d_{Q,I}(i-2,j-2)$
$+ 1$
$correspondstotranspositionbetweentwosuccessivesymbols$

The Damerau-Levenshtein distance between Q and I is then given by the function value for the full strings y, where i=|Q| denotes the length of string Q and j=|I| is the length of I.

And the details of intelligent processing of the free-form text-query is given in Algorithm 2.

ALGORITHM 2 INTELLIGENT PROCESSING OF TEXT QUERY

Input: Knowledge-base and Bag-of-Qwords “**δ**”

Output: Predicted image categories or query-class(Qc)

Begin

Step-1: Read the Knowledge-base and Bag-of-Qwords “**δ**”

Step-2: Generate the Bag-of-Ngrams using Bag-of-Qwords “**δ**”

Step-3: Generate the Qstrings as in equation(3)

Step-4: Find the Length of each string in Qstrings

do

Step-5: Perform Word-Length matching by applying equation(4) to spot the Qstrings in the modified Bag-of-Words model of the knowledge-base

Step-6: Perform Mapping of Concepts to general image categories of the knowledge-base for all the Qstrings spotted in Step-3 using Damerau-Levenshtein edit distance as in equation(5)

Step-7: Extract image category for Qstrings, matching with the Bag-of-Words in knowledge-base

While(end of modified Bag-of-Words models in the knowledge-base)

Step-8: Finally Output the extracted image categories for the given text-query or query-class(Qc)

End (begin)

Generating Qstrings and Concepts Mapping are the two pivotal steps in Intelligent processing of free-form text-query described in Algorithm 2 that leads to extraction of query-class (Qc) and in turn helps in retrieval of images. Word-Length matching enables speeded up matching process in locating the query-concepts in the modified Bag-of-Words model. And Damerau-Levenshtein edit metric will enable the mapping of concepts to possible image categories of the knowledge-base. The process of mapping is continued until the concepts array in the modified Bag-of-Words of Knowledge-base exhausts. And finally, all the matched query classes will be extracted and saved in the datastore for later use.

3.3. Retrieve And Display The Images

Image retrieval is performed in the query-by-text method, by submitting the free-form text query. The proposed method pre-processes the given text-query as described in Algorithm 1 and obtains the query-class (Qc) by employing Algorithm 2. The similarity matching is performed against the extracted query-class (Qc) to the image-category(Ic) names associated with the images in the database using the Levenshtein edit distance metric (Levenshtein 1966, [12]) given in equation (6). The smallest distance indicates the closest match. And the indices of the matched image categories are obtained and saved. Finally, the specified number of images are retrieved and displayed from the database.

The Levenshtein distance between Query-class (Qc) to Image Category (Ic) (of length |Qc| and |Ic| respectively) is given by lev(Qc,Ic) as in equation(6),

$$lev(Qc,Ic) = \begin{cases} |Qc| \quad if\ |Ic| = 0, \\ |Ic| \quad if\ |Qc| = 0, \\ lev(tail(Qc),tail(Ic)) \quad if\ Qc[0] = Ic[0], \\ 1 + min \begin{cases} lev(tail(Qc),Ic) \\ lev(Qc,tail(Ic)) \\ lev(tail(Qc,tail(Ic)) \end{cases} \quad otherwise, \end{cases} \tag{6}$$

IV. EXPERIMENTAL SETUP

This section describes the Knowledge-base, modified Bag-of-words model and the dataset used in the proposed system.

Knowledge-Base:The knowledge-base is constructed with the modified Bag-of-Words model for specific 20 image categories of the Wang database, and 132 image categories of DIV400 dataset selected for experimentation and can be constructed on the fly as well, for more image categories as needed, by creating modified Bag-of-Words model for each of the new image category.

Modified Bag-Of-Words Model:The general Bag-of-Words model, represents the text as the bag-of its words. In which each key is the word and each value is the frequency of that word in the given text document.

This general Bag-of-words model is modified to have table and is populated with possible adjectives, verbs and/or conceptual descriptions and corresponding word-length and Ngram (N-consecutive but not exceeding 15 characters) information for each of the image category. And hence each row of the table contains {word, word-length, N-gram} information which represents an object, and provides spatial-information to interpret each image category. This can handle complex textual semantic queries and polysemy problem by helping them to map text-query words to cover possible image categories and helps to retrieve images belonging to most relevant image categories to the user.

Image Dataset: The proposed method is experimented with two datasets namely Wang(or Corel) and DIV400. The 20 classes, made up of 2000 images' dataset, obtained from the Wang database (also called as Corel database) which contains natural scene images (Wang et al. 2001, [25] & Li et al. 2003, [26]). And 132 classes of total 13200 images' dataset obtained from DIV400 image database, which contains diverse social images (B. Ionescu et al. 2014, [27]). And the details of the datasets are given in Table 1.

Table 1 image Dataset Chosen for Experimentation

Dataset	No of Classes (N)	No of Images in each Class (M)	Chosen Image Categories
Wang/Corel Database	20	100 & Total Images (NXM) = 2000	Beach, Building, Bus, Butterfly, Car, Dinosaur, Elephant, Food, Forest, Horse, Map, Mountain, People, Planet, Rose, Spacecraft, Sport, Sunrise, Text and Yacht
DIV400	132	100 & Total Images (NXM) = 13200	Admiralty Arch London, Allen County Courthouse Indiana Fort Wayne, Angel Falls Venezuela, ANZAC War Memorial Sydney, etc. and so on 132 classes from DIV400 image database

Free-form text queries are formed by using concepts and/or keywords belonging to these 20 and 132 categories of two datasets and intelligently processed using modified Bag-of-Words model to retrieve most relevant images. And the results discussed are given below.

V. RESULTS AND DISCUSSION

This section presents, experimental results obtained for the free-form query based image retrieval through intelligent text-query processing on natural scene images of the Wang dataset and on diverse social images of the DIV400 datasets respectively. The proposed method is executed on free text queries containing the image category names and/or concepts. Total 1500 queries are executed and the system correctly predicts the image category names but misses (do not display any) image category names when normalization (lemmatization or stemming) affects the query strings. Once image category name is predicted correctly, the image retrieval is performed. Figure 2 and Figure 3 gives the performance of the proposed method by considering the number of queries containing single/multiple keywords and/or concepts belonging to 20 and 132 image categories respectively.

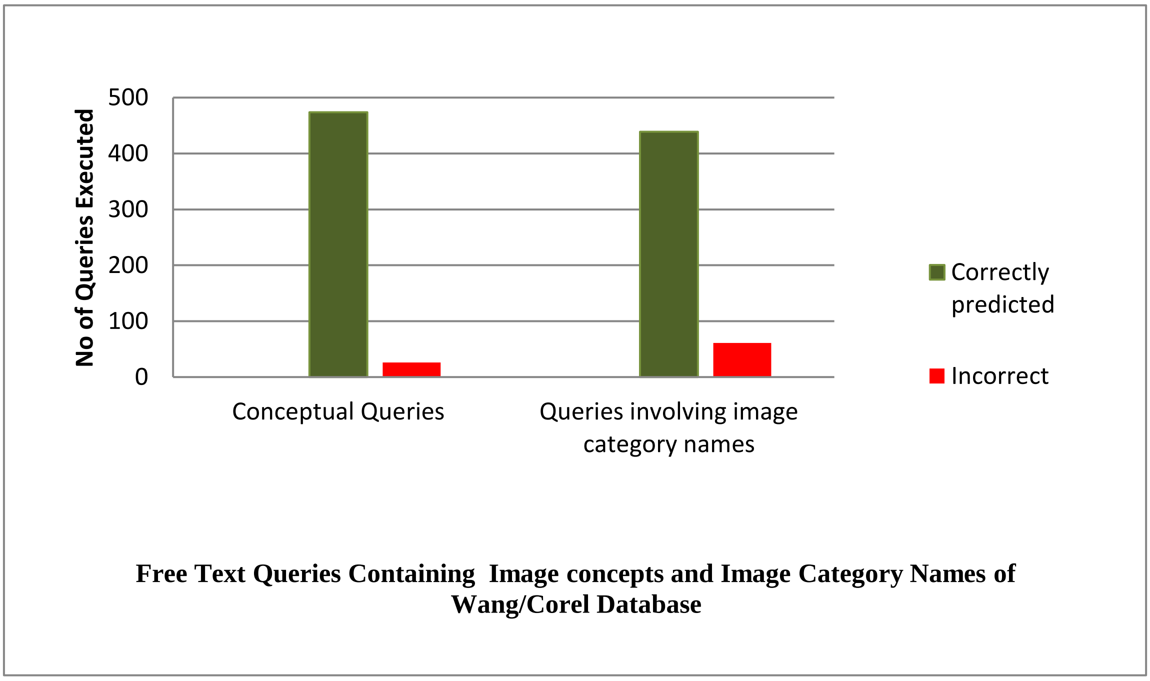

Fig. 2 The performance of the free text queries containing concepts & image category names of Wang/Corel Database

Retrieval Number: 100.1/ijitee.G96180612723
DOI: 10.35940/ijitee.G9618.0612723
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

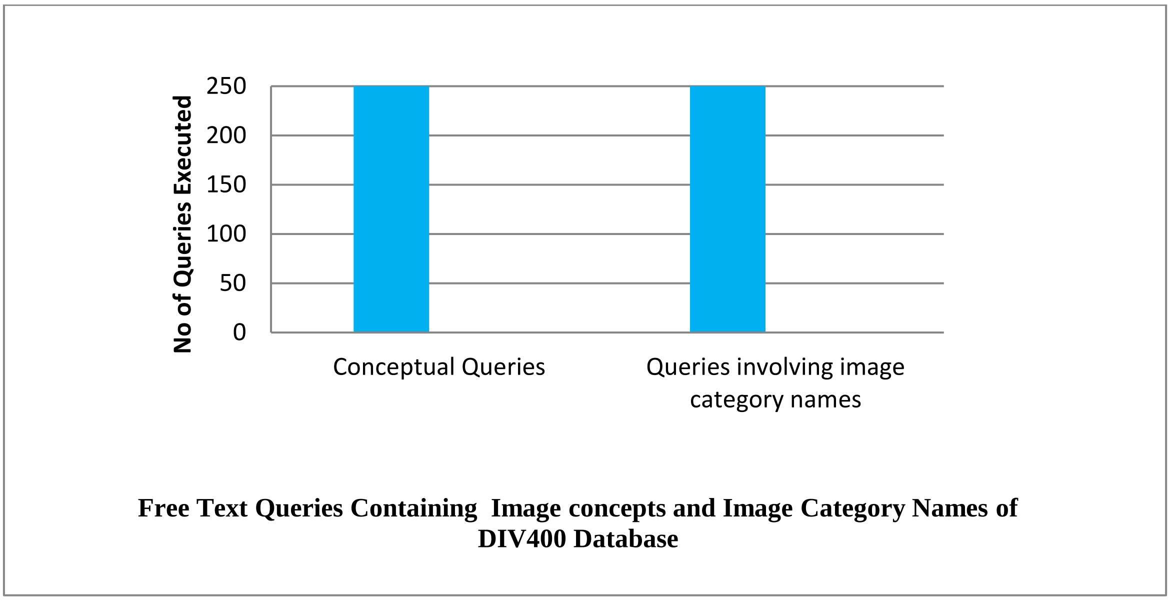

Fig. 3 The performance of the free text queries containing concepts & image category names of DIV400 database

The graph in Figure 2 and Figure 3 shows the performance of the system with the free text queries containing number of image categories and/or concepts evaluated on two datasets namely Wang/Corel and DIV400 datasets respectively. The system performance is evaluated by executing 750 queries involving direct image category names and another 750 are conceptual queries. The queries are executed with misspellings in the query strings to test the robustness of the proposed system. Misspellings of the query strings include interchange of characters, adding new characters, or deleting the characters in text-query words etc. The methodology guarantees the robustness and resulted in recall accuracy of 91.3% and 100% on Wang/Corel dataset and DIV400 dataset respectively, when simple and/or complex free text queries are submitted to retrieve the images. The snap shot of the retrieved images for the free text-query containing image category names from Wang dataset is given in Figure 4. And the query read as: “Retrieve the images belonging to planets, maps, sports, roses and butterflies” and resultant images are displayed with the heading of matched image categories as follows: Matched image categories are: "butterfly" "map" "planet" "rose" "sport"

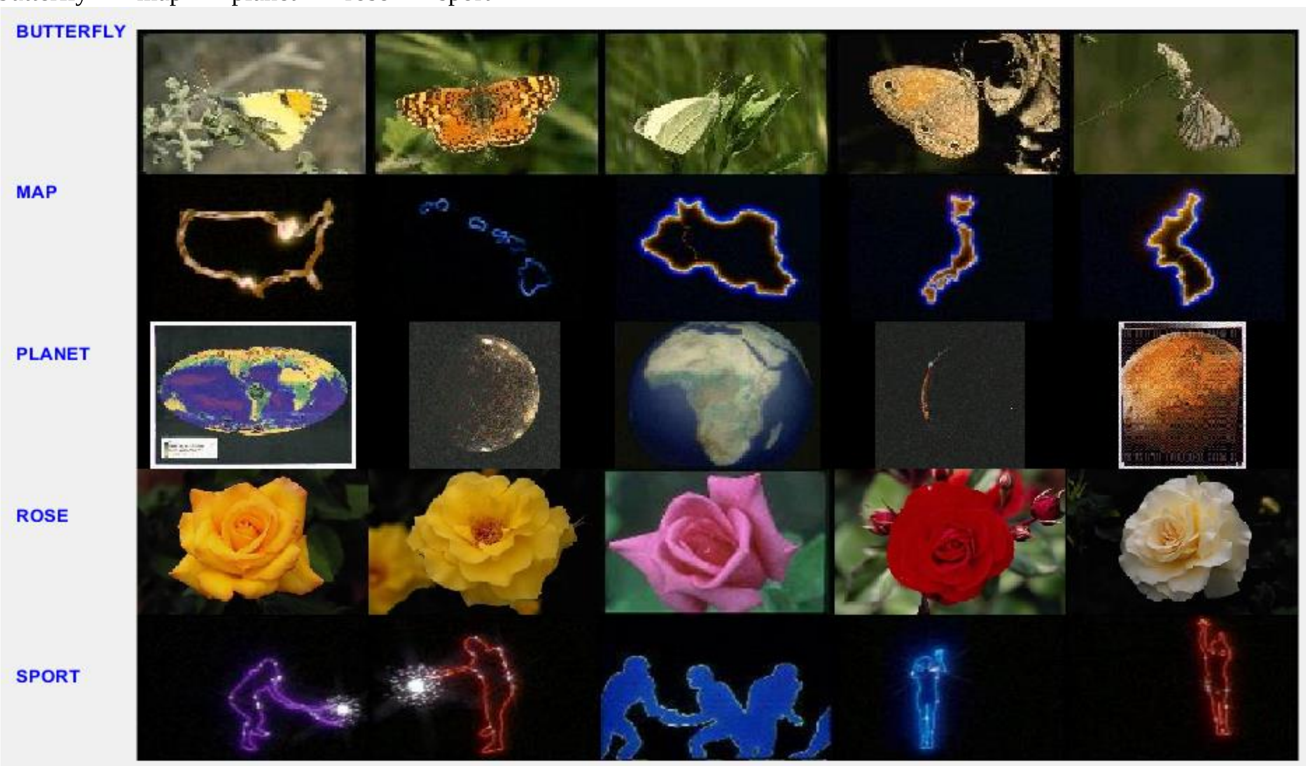

Fig. 4 Images retrieved for free text-query containing image category names from Wand/Corel Database

Similarly the snap shot of the retrieved images for the free text-query containing image concepts from the Wang database is shown in Figure 5. And the query read as: “Retrieve animal images” and results are as follows,
MATCHED IMAGE CATEGORIES ARE: "DINOSAUR" "ELEPHANT" "HORSE"

Fig. 5 Images retrieved for free text-query containing image concepts

The snap shot of the retrieved images for the free text-query containing image category names from DIV400 dataset is given in Figure 6. And the query read as:
"Retrieve the images belonging to admiralty arch London"
and resultant images are displayed with the heading of matched image category as follows:
MATCHED IMAGE CATEGORIES ARE: "ADMIRALTY ARCH LONDON "

Fig. 6 Images retrieved for free text-query containing image category names from DIV400 image dataset

For the free text-query containing totally irrelevant image-concept/category the system outputs "No Match Found" error message. For instance, there is no image-category "BIRD" included in the chosen Wang dataset, when queried for it, has resulted in "No Match Found" error message.

5.1 ANALYSIS

This section details the comparative analysis made between proposed TQBIR system to other state-of-the-art TQBIR systems. The Table 2 illustrates the various state-of-the-art methodologies, databases and distance metrics proposed in literature for text based image retrieval.

Table 2 Various Methodologies, Databases & Distance Metrics Used in The State-Of-The-Art of Tqbir

Sl. No.	Paper	Methodology	Database	Distance Metric
1	Real-time Lexicon-free Scene Text Retrieval (Mafla et al. 2020) [1]	The task of scene text retrieval is performed for the given text-query & the system returns all images containing the queried text. The model uses a single shot CNN architecture(YOLOv2-820-d) that predicts bounding boxes and builds a compact representation of spotted words.	The STR dataset, Sports-10k dataset, Street view text (SVT) dataset, Multi-lingual scene text (MLT) datasets, Text in videos (TiV) dataset	A nearest neighbor search of the textual representation of a query is performed over the outputs of the CNN collected from the totality of an image database.
2	Diversified Textual Features Based Image Retrieval	A framework for retrieval of images, combines visual and textual information. Firstly, Latent Dirichlet Allocation (LDA) is applied on visual features to construct topics. Then textual	DIV 400 database, MediaEval 2013 database	constructing a textual similarity graph cluster and removing the outliers, and

	(Yuan et al. 2019)[2]	information is used to improve the coherence of the topics. To extract diversified textual features, this method focused on collecting the unique words. For topic improvement outliers are removed from the constructed graph. Finally, the representative images are selected as retrieval results.		clustering-based image retrieval.
3	Document /Query Expansion Based on Selecting Significant Concepts for Context Based Retrieval of Medical Images (Torjmen-Khemakhemet al. 2019)[3]	A context-based medical image retrieval is performed by using the queries composed of medical keywords and the documents of knowledge-base succinctly describing the medical images. And a mapping of the medical text of queries and documents into concepts and then applying a concept-selection method to keep only the most significant concepts is proposed for retrieval of medical images.	Image Clef 2009 and 2010.	A hybrid distance metric (combination of hierarchical (word with various meanings) and content distance (a word with the central meaning)) is used.
4	Combination of Term Weighting and Integrated Color Intensity Co-occurrence Matrix for Two-Level Image Retrieval on Social Media Data (Siradjuddin et al. 2019)[28]	The two-level image retrieval that combines the text and image-based query for image retrieval is performed. In the text-based retrieval, The text and document pre-processing, weighting the word from the dictionary, based on TF-IDF Model & the similarity is calculated. In the image-based approach for image retrieval, feature extraction using Integrated Color Intensity Co-occurrence Matrix & the similarity is calculated. First, the images are filtered based on the text query, and second, the images are retrieved from filtered images, using an image as a query.	Data from Twitter social media are collected, since March 2018. tweet data are filtered based on the keywords in the Indonesian language	Cosine Similarity(is calculated between a text-query and the text document) & Manhattan Distance(is calculated between image query features to features of images in the db)
5	DenseBert4Ret: Deep bi-modal for image retrieval (Zafran Khan et al. 2022)[21]	Use of Vision and Language pretrained models such as image pretrained model DenseNet and Language pretrained model BERT in retrieval of images. Query contains both image and a text	Fashion200k, MIT-states and FashionIQ	Triplet loss function is used in similarity matching.

Further the Figure 7 gives the comparison of performance of proposed system evaluated on DIV400 dataset with the performance of TQBIR evaluated on the same dataset (DIV400 dataset) presented in (Yuan et al. (2019), [2]). The precision at retrieving specified number of images are symbolically represented as “Pr@#”

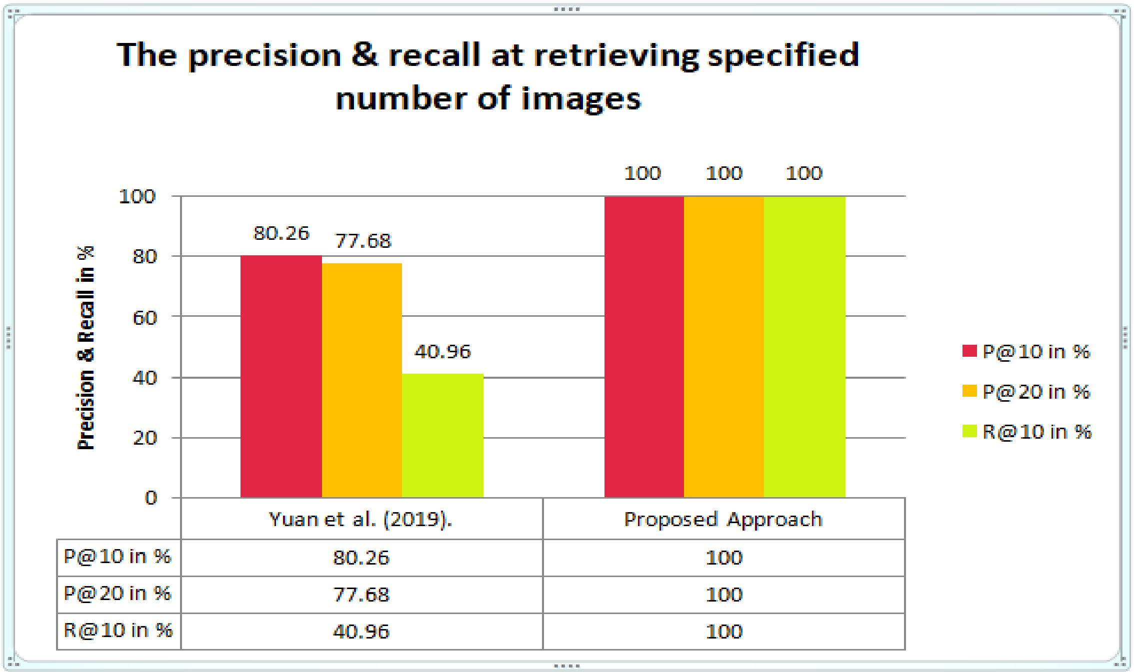

Fig. 7 Performance of the proposed method compared to state-of-the-art TQBIR using DIV400 dataset

And the Figure 8 gives the Recall performance of the proposed approach evaluated on Wang/Corel dataset and DIV400 datasets respectively.

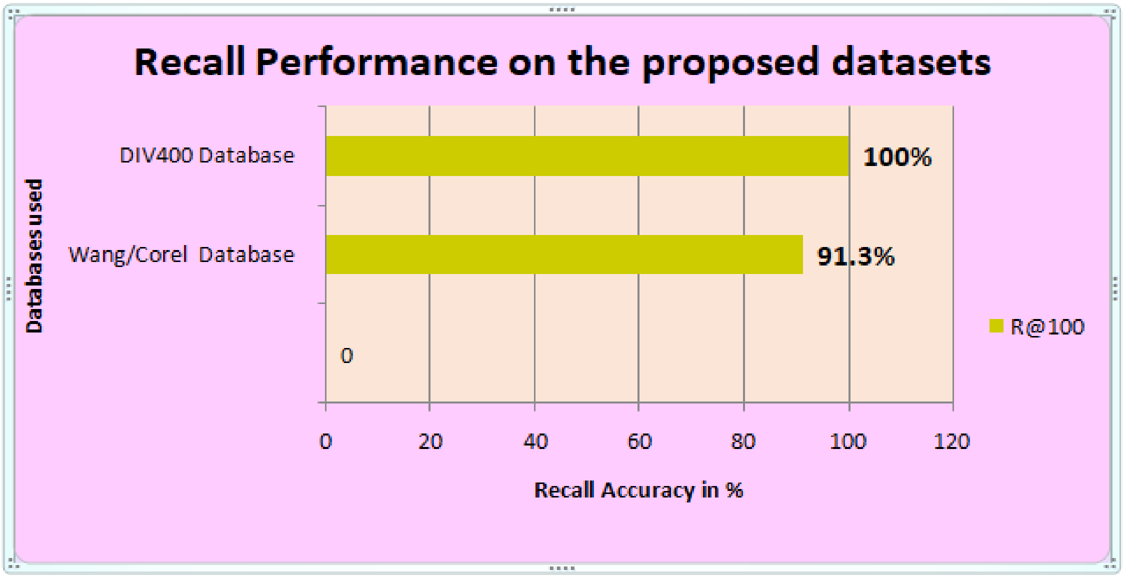

Fig. 8 Recall Performance of the proposed method evaluated on Wang/Corel and DIV400 datasets

The comparison of Recall performance of the proposed method on two datasets is presented in Figure 8. The proposed methodology is tested over 1000 text queries on Wang/Corel dataset and 500 text queries on DIV400 dataset. The system is robust in spotting and mapping image categories to handle polysemy, in turn it can handle semantics of the free-form text queries pertinent to the underlying image category and/or concepts and is outperformed resulting in recall accuracy at retrieving 100 images of 91.3% and 100% over two datasets (Wang and DIV400) respectively.

VI. CONCLUSION

In this paper a new methodology for intelligent processing of free form text query based, image retrieval is proposed using Modified Bag-of-Words model on natural scene images and diverse social images. The specific twenty image categories of 2000 images are selected for experimentation from the Wang database and 132 image categories of 13200 images are selected from DIV400 dataset. The free text queries are formulated using image categories and/or concepts. Images are indexed by image category labels. Spotting and mapping the image categories and/or concepts in the free text query, belonging to a specific image category is performed using {word, word-length, Ngram} information of the modified Bag-of-Words model in the knowledge-base. The 1500 free text queries are processed and evaluated to predict the image categories for retrieval of images.

The methodology is very robust in spotting the keywords in the free text query for retrieval of the images and has resulted in recall accuracy (total images=100 available in each image category) of 91.3% and 100% on Wang/Corel dataset and DIV400 dataset respectively.

ACKNOWLEDGMENT

This work is supported by department of CSE, VTU-Belagavi and partially supported by all other departments of VTU Belagavi.

DECLARATION

Funding/ Grants/ Financial Support	No, I did not receive any financial support for this article.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material/ Data Access Statement	Yes, Wang and Div400 are publicly available databases
Authors Contributions	All authors have equal participation in this article.

REFERENCES

1. Mafla, A., Tito, R., Dey, S., Gómez, L., Rusiñol, M., Valveny, E., & Karatzas, D. (2020),"Real-Time Lexicon-Free Scene Text Retrieval", Pattern Recognition, 107656. Vol. 110, ISSN 0031-3203. doi:10.1016/j.patcog.2020.10765. [CrossRef]
2. Bo Yuan and Xinbo Gao. (2019),"Diversified Textual based imge retrieval", Neurocomputing, vol.357, pp 116-124.Sep 2019,doi:10.1016/j.neucom.2019.03.048. [CrossRef]
3. Torjmen-Khemakhem, M., & Gasmi, K. (2019),"Document/query expansion based on selecting significant concepts for context based retrieval of medical images", Journal of Biomedical Informatics, vol. 95, ISSN 1532-0464, 103210. doi:10.1016/j.jbi.2019.103210 [CrossRef]
4. Qimin Cheng, Qian Zhang, Peng Fu, Conghuan Tu &, Sen L. (2018), "A survey and analysis on automatic image annotation", Pattern Recognition, Vol. 79, pp. 242-259. doi:10.1016/j.patcog.2018.02.017 [CrossRef]
5. Yilu Chen, Xiaojun Zeng, Xing Chen and Wenzhong Guo. (2020), "A survey on automatic image annotation", Applied Intelligence, vol. 50, pp. 3412–3428. doi:10.1007/s10489-020-01696-2 [CrossRef]
6. Tamura, H. and Yokoya, N. (1984), "Image database systems: A survey", Pattern Recognition, Vol.17, No.1, pp. 29–43, 1984. doi:10.1016/0031-3203(84)90033-5. [CrossRef]
7. V. Vijayarajan, M. Dinakaran, Priyam Tejaswin and Mayank Lohani. (2016),"A generic framework for ontology-based information retrieval and image retrieval in web data", Human-centric Computing and Information Sciences, vol.6, no.1, pp. 1-30, 2016. DOI 10.1186/s13673-016-0074-1 [CrossRef]

8. P. M. Ashok Kumar, T. Subha Mastan Rao, L. Arun Raj, and E. Pugazhendi. (2021),"An Efficient Text-Based Image Retrieval Using Natural Language Processing (NLP) Techniques", Intelligent System Design, Advances in Intelligent Systems and Computing, vol. 1171, https://doi.org/10.1007/978-981-15-5400-1_52 [CrossRef]
9. Pastra, K., Saggion, H., & Wilks, Y. (2003), "NLP for indexing and retrieval of captioned photographs", Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics - EACL '03. Vol.2, pp. 143-146. doi:10.3115/1067737.1067769 [CrossRef]
10. Sohail Sarwar, Zia Ul Qayyum, Saqib Majeed. (2013),"Ontology Based Image Retrieval Framework using Qualitative Semantic Image Descriptions", Procedia Computer Science, Vol. 22,pp.285 – 294, 2013. doi: 10.1016/j.procs.2013.09.105 [CrossRef]
11. Hui Hui Wang, Dzulkifli Mohamad, N.A. Ismail. (2010),"Approaches, Challenges and Future Direction of Image Retrieval", Journal of Computing, vol. 2, issue 6, June 2010, ISSN 2151-9617
12. Levenshtein, V. (1966),"Binary codes capable of correcting deletions, insertions, and reversals", Soviet Physics-Doklady, Vol.10, No. 8, pp. 707-710, February 1966.
13. Damerau, F. (1964),"A technique for computer detection and correction of spelling errors", Communications of the ACM, Volume 7, Issue 3, March 1964 pp. 171–176. https://doi.org/10.1145/363958.363994 [CrossRef]
14. Grzegorz Kondrak. (2005),"N-gram similarity and distance", SPIRE'05: Proceedings of the 12th international conference on String Processing and Information Retrieval, November 2005,vol. 3772, Pages 115–126.https://doi.org/10.1007/11575832_13 [CrossRef]
15. Htet Htet Htun and Virach Sornlertlamvanich. (2017),"Concept Name Similarity Measure on SNOMED CT", vol. 780, pp.76-90, DOI: 10.1007/978-981-10-6989-5_7 [CrossRef]
16. Mihai Lintean and Vasile Rus. (2012),"Measuring Semantic Similarity in Short Texts through Greedy Pairing and Word Semantics", Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference.
17. Farouk, M. (2019, "Measuring Sentences Similarity: A Survey", Indian Journal of Science and Technology, vol.12, issue 25, pp. 1–11. doi:10.17485/ijst/2019/v12i25/143977 [CrossRef]
18. Thi Thuy Anh Nguyen and Stefan Conrad. (2015),"Ontology Matching using Multiple Similarity Measures", In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015), Volume. 1 KDIR, pages 603-611.
19. Jin Zhang;Xiaohai He;Linbo Qing;Luping Liu;Xiaodong Luo. (2021). "Cross-modal multi-relationship aware reasoning for image-text matching", Multimedia Tools and Applications, vol. 81, pp. 12005-12027. doi:10.1007/s11042-020-10466-8 [CrossRef]
20. Chunbin Gu, Jiajun Bu, Xixi Zhou, Chengwei Yao, Dongfang Ma, Zhi Yu, Xifeng Yan. (2022),"Cross-modal image retrieval with deep mutual information maximization", Neurocomputing, Volume 496, 2022, Pages 166-177, ISSN 0925-2312. https://doi.org/10.1016/j.neucom.2022.01.078 [CrossRef]
21. Zafran Khan, Bushra Latif, Joonmo Kim, Hong Kook Kim, Moongu Jeon. (2022), "DenseBert4Ret: Deep bi-modal for image retrieval", Information Sciences, Volume 612, 2022, Pages 1171-1186, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2022.08.119. [CrossRef]
22. Zheyuan Liu, Cristian Rodriguez-Opazo, Damien Teney, Stephen Gould. (2022), "Image Retrieval on Real-life Images with Pre-trained Vision-and-Language Models", https://doi.org/10.48550/arXiv.2108.04024
23. Porter, M.F. (1980), "An algorithm for suffix stripping", Program: electronic library and information systems, vol.14, issue 3, pp. 130–137. doi:10.1108/eb046814 [CrossRef]
24. Bard, Gregory V. (2007),"Spelling-error tolerant, order-independent pass-phrases via the Damerau–Levenshtein string-edit distance metric", Proceedings of the Fifth Australasian Symposium on ACSW Frontiers : 2007, Ballarat, Australia, January 30 - February 2, vol. 68, pp. 117–124, 2007. ISBN 978-1-920682-49-1.
25. Wang James Z., Li Jia, Gio Wiederhold. (2001),"SIMPLIcity: Semantics-sensitive Integrated Matching for Picture Libraries", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 23, no.9, pp. 947-963, 2001. [CrossRef]
26. Li Jia, Wang James Z. (2003),"Automatic Linguistic Indexing Of Pictures By A Statistical Modeling Approach", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1075-1088, 2003. [CrossRef]
27. B. Ionescu, A.-L. Radu, M. Menéndez, H. Müller, A. Popescu, B. Loni, "Div400: A Social Image Retrieval Result Diversification Dataset", ACM Multimedia Systems - MMSys2014, 19-21 March, Singapore, 2014. [CrossRef]
28. Siradjuddin, I. A., Bahruddin, R. P., & Sophan, M. K. (2019), "Combination of Term Weighting and Integrated Color Intensity Co-occurrence Matrix for Two-Level Image Retrieval on Social Media Data", Procedia Computer Science, vol. 157, pp. 329–336. doi:10.1016/j.procs.2019.08.174[CrossRef]

AUTHOR PROFILE

Shanmukhappa A. Angadi is a Professor in the Department of Computer Science and Engineering at Visvesvaraya Technological University, Belagavi, India, and is currently on deputation as Registrar Evaluation, at Bagalkot University, Jamkhandi India. He was the founder chairperson of Department of Computer Science and Engineering VTU Belagavi. He was the Director (i/c) of the Information Technology Infrastructure and Services Management Unit (ITISMU) at Visvesvaraya Technological University, Belagavi, India. He is the Chief Architect and mentor for the team responsible for the development of a bunch of software for VTU, including pre-exam, post-exam, and digital valuation systems. He has published more than 108 articles in journals and conferences having Scopus H Index: 10, Google Scholar H Index: 14. He has guided five students to their Ph.D. His areas of interest include Image Processing and Pattern Recognition, Knowledge-base Systems, Soft Computing (Fuzzy Systems, Neural Networks, and Genetic Algorithms), Optimization and Graph Theoretic techniques, Web Technology, Internet of Things, Deep Learning.

Hemavati C. Purad received the Bachelor's Degree in Information Science and Engineering and the Master's Degree in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, and submitted her Ph.D. thesis in the Department of Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, Karnataka, India. Her areas of interest include multimedia processing and pattern recognition, soft computing, and deep learning.