"A hybrid approach towards minimizing the semantic gap in image retrieval by combining classical and content based techniques"

Dissertation Report

Submitted in the partial fulfillment for the requirement for degree of,

Master of Engineering in Computer Science and Engineering

Submitted by

PALAK

801632033

M.E (CSE)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Under the guidance of

Rajkumar Tekchandani

Assistant Professor, CSED

Department of Computer Science and Engineering

Thapar Institute of Engineering and Technology

Patiala-147004

June 2018

Certificate

I hereby certify that the work, which is being presented in the thesis, entitled "***A hybrid approach towards minimizing the semantic gap in image retrieval by combining classical and content based techniques***", in partial fulfillment of the requirements for the award of the degree of **Master Of Engineering** in Computer Science and Engineering submitted in Computer Science & Engineering Department of **Thapar Institute of Engineering and Technology**, Patiala is an authentic record of my own work carried out under the supervision of ***Rajkumar Tekchandani.*** I have also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

The matter presented in this thesis has not been submitted elsewhere for the award of any other degree or diploma from any institution.

Signature: ______________

Palak

This is to guarantee that the above explanation made by the competitor is right to the best of our insight.

Signature: ______________

Rajkumar Tekchandani

Assistant Professor

CSED

Acknowledgement

This examination work would not be conceivable without affirmation of the general population who upheld and guided me for the fruitful fulfillment of this errand.

I want to thank Rajkumar Tekchandani, Assistant Professor, Computer Science & Engineering Department, Thapar Institute of Engineering and technology, Patiala for recognition my work and to give his important time and direction and furthermore to his constant consolation all through research work. It won't be conceivable without his scholarly direction which set off my insight right way and in measure up to extent.

I am appreciative to Dr. Maninder Singh, Hon'ble Head of Computer Science and Engineering Department, Thapar Institute of Engineering and innovation, Patiala for all the assistance and to keep up a sound domain which additionally assumes a noteworthy part in look into work.

I am additionally grateful to the Institution and all employees of Computer Engineering Department, Thapar Institute of Engineering and innovation, Patiala for giving right heading and their inspiration.

Palak

801632033

Abstract

Number of internet users has been increased drastically over past few years and so the need for secure, reliable and accurate information retrieval mechanisms. The increasing population over web creates loads of information and user data over web. It can include textual information, multimedia content, content specific URL's, personal information and other. Due to all this retrieval of information becomes difficult on fast and efficiency scale. There are search engines that are used to retrieve multimedia data called vertical search engines.

In spite of the upset in web indexes, there can be other issues which fails to meet client necessities and results in semantic gap between user search and the outcome. Other thing that can be often viewed is the content relevance. There are various search engines available that are classified into two categories known as *text based search engine (traditional)* and *content based search engines (content based).* Text-based Search Engines are simpler to implement but are sometimes contradictory in results, as they completely rely on the occurrence of query term in surrounding text. In case of absence of surrounding texts it fails to load the results or sometimes loads improper results. Also it fails to recognize the synonyms and fails to load those results. Whereas on the other hand content based retrieval is based on matching the user input query with that of the content present in the data pool. In this each feature of input query is matched against the available resources in database. Hence the results are generated with high quality and are more accurate. But it involves higher computation time and sometimes involves computation overhead.

Here, in following work, a technique is used which combines both the classical approach and content based. The overview of the task being done is extracting both textual features and visual features from images and its metadata. As both techniques are combined, result will have higher content relevance and much lower overhead time and response time.

Table of contents

List of figures

List of Tables

Chapter 1
Introduction

Since the evolution of internet has increase the count of its users which gave us the requirement of building up a system which provides security, accuracy, reliability with respect to retrieving of information. Ever since Web 2.0 world of online business, Ecommerce, Social media etc. had taken a huge part of our life. Huge amount of data is present on Web like Images, Multimedia, text formed, URL’s, Personal content etc. On top of this, the data is present on Servers which force us to have a secure and reliable mechanism for retrieval.

In spite of the time lag, there is another major area of concern i.e. relevance of the content. As there is increase in number of internet users, so the algorithms for retrieval of data must be efficient enough to search the correct information. Many a times user may search for something for which his understanding and the actual image is different. Another point of concern can be synonyms which may not match with the tags present in the images.

Due to changing needs of commerce and business, retrieval of data must be very efficient and accurate. We have quite a number of search engines for this. These are the utilities that are used to retrieve information from the web based environment based on the user input queries.

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Image search is a specific information look used to discover images. To scan for images, a user may give question terms, for example, catchphrase, image record/connection, or tap on some image, and the framework will return images "comparable" to the user query.

On the broader perspective we have text based search engines which are traditional and second are content based search engine. When we talk about text based engines, they always retrieve data based on the tags or surrounding text of the image. One advantage of these kinds of search engines is that they are relatively easy to implement. On the other hand, its result set is not accurate. The main functionality or basic idea of its working solely depend on occurrence of user input in surrounding text of the image. So usage of such search engines is less. Considering the case where tags or surrounding text is not properly embedded in the image, this fails. It also can't handle the case where matching words or synonyms occur. Content based search engines works when user input is matched with content of the image which is present in the database. Result set generated by content based is of high quality. Talking about terms of accuracy, it is better than text based search engine. The disadvantage of using this is basically the computation overhead time.

Whether it is text based or content based both occur in two stages- First links matched with content are loaded. Secondly going through any link to further search on content.

1.1 Horizontal search engine

The search method used in this is called Text based search. Horizontal search is basically searching through the entire web.

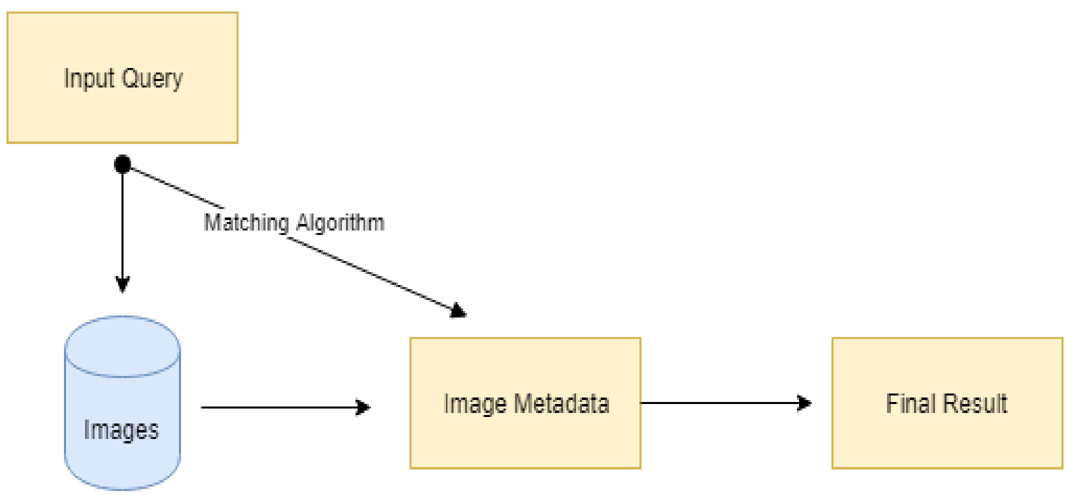

Figure 1.1 Basic flow of text based search

Here according to the user input query, links are loaded. Then navigation is done on basis of content matching to retrieve the actual information. We have such search engines named google and Bing. These basically search not in a categorical way but in general way. All the related links are retrieved first then links can be considered for detail information required. For images, it solely depends on presence of query term in image surrounding text or the metadata of an image.

1.2 Vertical search engine

The search method used in this is called Content based search. In vertical Search, it doesn't search through the entire web but only some specific parts like news, images etc. As in text based search no links are retrieved at first. This result into retrieving more relevant data as it gives us the direct information. Such engines are used in various online shopping sites like myntra.com, amazon.com etc.

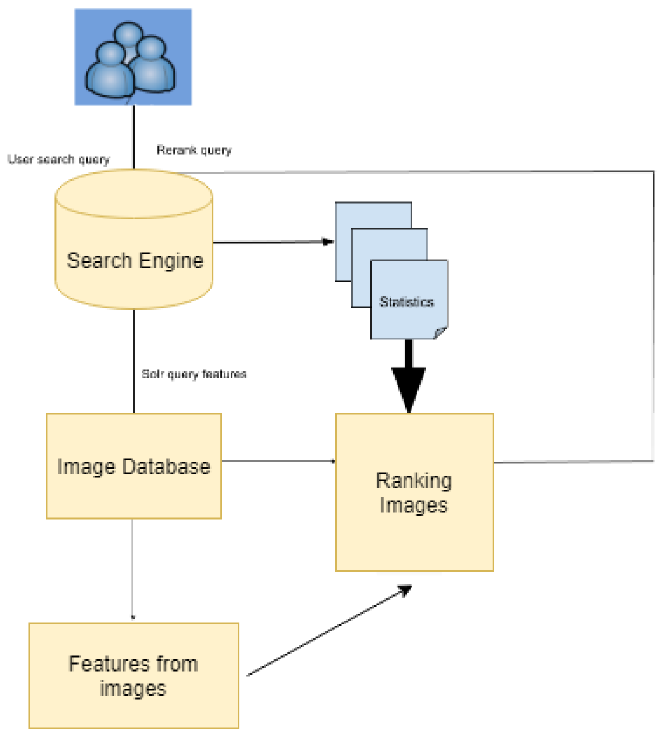

Figure 1.2 Basic flow of content based search

In spite of transformation in search approaches and different advances, in some cases users may not get the outcomes they are searching for. Main reasons for this could be:

- Interpretation of user input.
- Distinct meaning of user input and required output.
- Improper queries
- Absence of term used by user in database.
- Presence of synonyms instead of actual term.

Clearly there is a gap between what a user want and what it gets. This gap between user goal for search and the outcomes that are called as semantic gap and it needs to be reduced in order to retrieve better results.

1.3 Motivation

Text based search engines gives output based only on the surrounding text present in the images i.e. tags of the image. It has less computation time and reduced overhead. On the other hand, Context based search engine gives the output based on all image features with better accuracy than text based. It has high computation and overhead time. To achieve better results, an hybrid approach can be used which will be combination of text based and content based search engines giving lower computation time and higher accuracy.

1.4 Objective

The primary goal for this is to minimize the semantic gap between what user is trying to accomplish and the outcomes displayed. As text based search matches just the metadata of the images with input query term resulting in lower computation cost and lower overhead time. On the other hand content based search does not match input query with metadata, instead extract the features from images and compare it with query. In proposed work we combine these two approaches and matches image features and metadata with query term entered by user. A similarity calculator is designed where image from database is selected on basis of user input by calculating the similar features.

Chapter 2
Literature Survey

There are quite a number of search engines available in now days. Some of them are text based, content based and hybrid search engines.

2.1 Horizontal search engines

These types of search engines are classical and the first created engines which are used till now. They represent a category of horizontal search engines.

Considering metadata of images Hua et al. [1] proposed the tagging of images for text based engines. Here if the tag is not present along with the image, for this search engines, it is entirely not possible to proceed. Since query from user is matched with metadata part. There is a way of manually adding tags which is obviously difficult and time consuming. Now to remove the disadvantages of manual tagging, automated tagging system is made where ensemble classifier is trained and tags are created automatically after recognizing the image.

For efficient method of annotating image Sankar et al. [2] has implemented a trained tagging system which is made by hidden tagging procedure. Hashing techniques are used to improve the tags via Ant clustering algorithm.

After all tagging is done, Chen et al. [3] proposed a matching algorithm where query term from user is matched with the tags present in the images. Here an algorithm is given which states that best pair matching between the query by user and image features. It needs to check what user has given as input and what tag is present on the image and then use patch based word matching scenario.

Now, another point of consideration was clustering of similar tags of images. The calculation was for the most part worried about the bunching images in view of the literary portrayal related to the classification groups on basis of tag. For wrongly tagged images, this was certainly not possible to put similar tags in one cluster. Also there were chances that one image can be bunched into various groups. Thus the repetition of images was found in various clusters. Bo et al. [4] included an approach for image division and finding the similarity in view of the coordinating. These divisions parcels from the images were gone through the division coordinating calculation. To know which portion to choose was main point of concern. Additionally there was no semantic connection between portions chosen and tags of the images.

Portar [5], presented a hardware based conjunction for text based information retrieval in integrated environment. The system was able to perform text-search on large parallel connected systems. It demonstrated first application of classical text based search engine. The model was hardware based and was a primitive. It used limited set of keywords and specified set of instruction. Hence was not scalable. TIkeda et al. [6], in their work presents a DBMS with high performance text based search support. It was high speed full text-search engine including high precision ranking search. It made it easy to combine normal search attribute with full text search support. Krapac et al. [7], proposed two methods for improving web image search results using query-relative classifiers. Each of these methods included a query-ranking and classification algorithms.

Query Specific method: This method involved initial two step image retrieval mechanism. Firstly, images were retrieved using a classical text-based image search. It involved textual query acceptation from a generic user interface provided to end user. Based on the input text-query the initial result set is retrieved. Further a set of query-specific classifier is applied on the retrieved image set. The images are re-ranked based on the classifier along with textual mapping of the tagged images.

Query related approach: It involved training queries with relevance annotations. These annotations are used to get the initial dataset. Further a query-specific classifier is applied to results and images are re-ranked accordingly. In this mechanism the key challenge in this scenario was to form the classifier. It is difficult to form a generic classifier for the image set. A pre-specified set of classifier won't be sufficient to cover all the images in the image dataset and rank them accordingly.

2.2 Vertical Search engines

In vertical search engines, unlike text based tags are not considered for matching images with user input. Instead user input is matched with real time image features and then matched categorically. It actually takes a lot more computational time then classical text based, but primary used where accuracy is main point of concern. Zhang [8] proposed a method where it states that there is low accuracy of text based engines where they are unable to fetch correct output. In content based domain specific images are retrieved for better results than classical one.

For image features, Rasiwasia et al. [9] proposed a method for texture based image retrieval. In this method the images were mapped based on the annotations and are classified in to variety of predefined categories formed initially. As each image was classified in to initially mapped sample space there was a textual taxonomy involved. Based on the input queries the image texture was evaluated. Each category was identified based on the texture. Hence based on texture matching images of entire category were retrieved instead of the actual matching content.

Considering indexing, Idrissi et al. [10] proposed a texture based image retrieval method that involved offline image indexing of images which are mapped with texture features with other images and kept in a database. In online retrieval of images input queries are mapped with the ones in the stored offline index. On matching selected index the relevant images are loaded and are presented on the user interface. Due to the initial cluster storage the need for space is increased thereby increasing storage cost. In this method also the cluster overlapping was unexplained. For CBIR Clustering, Patil et al. [11] proposed a method in which data mining approaches are used for setting a particular domain in vertical search. The approach is itself a particular domain based search for data retrieval. The major point is formation of clusters where similar images are grouped. Usage of k means clustering method is done to form basic clusters. This method of grouping is quick and gives more exact outcomes. For more suitable information recovery, this framework utilizes content element extraction calculation. After clustering, Ranking is important where Zhou et al. [12] proposed a method called geometric verification which helps to know user query text and re rank the images.

Images are to be stored in database where Yadav et al. [13] proposed an indexing technique of images present in database where images are in consideration with other images having similar features. When user query is matched with images, related all images are also received as a part of output. The only issue is at time of clustering where computational cost is high and space is needed. Main point in CBIR is to calculate the similarity between different images. In Blanco et al. [14] proposed a comparison method is proposed for calculating the similarity quotient between images. It has some regions which are labeled in an image and those portions are considered for comparison between images.

Smeulders et al. [15], proposed a Content Based Image Retrieval method for image retrieval. In this method the images were indexed by their visual content or visual features. This visual content related information includes color, shape, texture etc. which are used to define the image attribute. The attributes define image content based on partial or complete content matching. In some cases even a part of image can be used to define complete image content and its attributes. Torres et al. [16] suggested these features were automatically extracted from images at the time of feature extraction phase while processing images during image addition in to database. The method was effective but there was a semantic gap between low level image features with that of the high level features. Also Chathurani et al. [17] suggested the gap was persisted in low-level visual features and high level image semantics. Further it was challenging to format the content relevant query and the selection of query related features for content retrieval.

2.3 Hybrid search

These are exceptional kind of media recovery instruments that are utilized for data getting to and recovery. In this sort of web indexes a two stage media recovery calculation is connected. Content recovery depended on both of the content search or with respect to the classical based search. The example comes about included images with the labels and encompassing data. The images after recovery are gone through a weighing and positioning strategy. This included image content positioning. For doing as such a image weighting schematics is connected either in view of the semantic significance of literary questions that are utilized to characterize the image meaning. It incorporates the image particular substance.

In hybrid approach, Thomas et al. [18] proposed a text-based search algorithm combining the visual features semantic for performing a domain specific search. In this method the images were first retrieved by performing a tag based text search on the image surrounding text. Then the visual content was mapped and evaluated with that of the image content. The query relevant features were mapped with that of with the textual features. These features are combined with the image visual features to form a semantic pair of content and the tag. The semantic search enabled content retrieval on a higher scale making it versatile and efficient. Collection of data, Kurniawati et al. [19] proposed a method where we can collect the data from online web sites using a crawler. It uses the scrapping method to take all the HTML tags. Result of crawler would be metadata of images which are required to match the tags with user query. When we talk about combining these two approaches, a variable called visual synonym is used. Hua et al. [20] proposed the derivation of visual synonym from the analysis of visual content of an image. Based on visual content, similarity is measured between images. As a result of which images are re ranked and weight factor comes into play. For all this to happen, one of the major need is Visual lexicon, for which Amato et al. [21] proposed a method to build and use a visual lexicon. We need to basically perform the preprocessing of data from which we remove stops words, combine synonyms, retrieve unique terms and build lexicon at last which is to be used later while retrieving of images. While combining two approaches, Input for hybrid has to between text or an image. Here, Polsley et al. [22] proposed a method where images were given as input to retrieve similar images as output. The ranking was done in two levels i.e. low level with text based search followed by high level ranking system.

Feature extraction was highly essential part of the CBIR approach where Mahamuni et al. [23] proposed a method called Query by contents of image which describes main information of image like color, texture, patterns etc are used to define an image. These features are very low in scalability which is made linear to perform easy comparisons. With color, shape, texture, patterns and other piece of information from images, user query can be compared for retrieving results. Color and shape were two basic parameters for features where Mustaffa et al. [24] proposed two different approaches for color and shape. Color moments for two orders and i mean is used for color and shape respectively. This fusion method with CBIR results in better output. Also, Mustaffa et al. [25] proposed a method for determining the color and shape of the images where it experimented with Marine images using similar techniques as above. Along with color and shape, texture is also one of the key factors for feature extraction where Fakheri et al. [26] proposed a method where division in image is done into non-overlapping pieces. The highlights drawn from exchanged image with Gabor wavelets utilizing first and second minutes between the image tiles fill in as neighborhood descriptors of surface. This neighborhood data is caught for two resolutions and two matrix formats that give diverse points of interest of a similar image. Shape data is caught as far as edge images processed utilizing Gradient Vector Flow fields. In [27, 28, 29] Invariant minutes are used to record the shape highlights. The blend of the surface highlights amongst image and conjunction with the shape highlights gives a powerful list of capabilities to image recovery. Further, in Kumar [30] proposed a Semantic structure which is kept up for each preparation sentence where the relationship of each word in the preparation sentence is set up as cosine closeness weight and furthermore in Chen et al. [31] states an interface towards the conceivable terms of similar words are set up with weights. It required not just ascertains the likeness weight between the expressions of preparing sentence yet in addition to set up the semantic connection between sentences of a similar gathering.

Chapter 3
Architecture

This chapter clarifies about the essential requirements, for example, equipment and practical necessities. These also include the assumptions that are made during the period of development and the operational assumptions over the span of search activities.

3.1 Overview

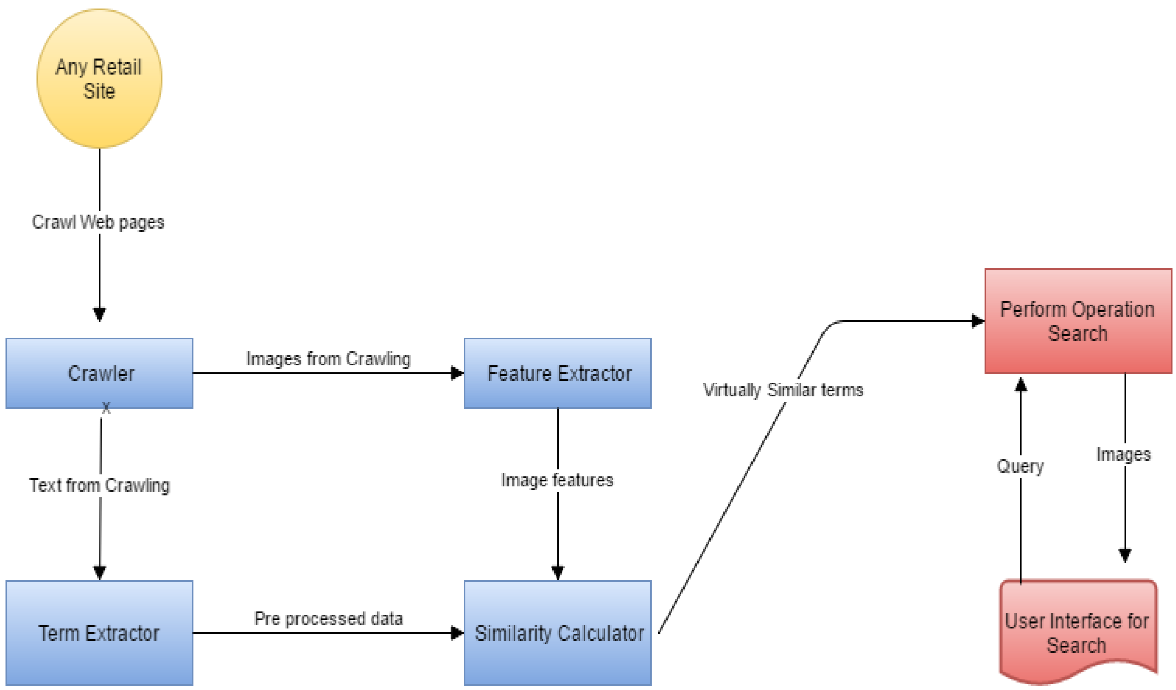

Figure 3.1: Architecture

3.1.1 Crawler

The process of extracting the content from Web pages is called crawling and term crawler is used to fulfill this. It comes under a Web based application which is used to gather Multimedia content online. There are number of Crawlers already available. They come with a pre-indicated algorithm utilized for crawling image content and have a specific format that can be consumed by applications to construct a meaningful conclusion. It can also be designed from existing one and used as a plugin in our browser.

3.1.2 Term Extractor

The crawled information contains part of contaminations in it. Thus it is difficult to isolate the online tagged image labels from the multimedia. To isolate them and concentrate totally impartial terms used to characterize visual importance of image it is must to isolate the significant words from the content. To do as such first the image information should be filtered, isolated and separated. To complete this process, a term separator and extractor parser is used for removing all the hindrance.

3.1.3 Feature Extractor

After removal of the hindrance and impurities, we have quite a number of crawled images. The next meaningful step should be utilizing the visual features of that. From visual features we mean defining the semantics of the image. For example if you consider a chair, it symbolizes four legs. Although there can be quite a number of patterns in them. The main way to define an image is by using its features. Now there can be number of ways in which its features can be extracted. Out of which we have image pixel values by which features are calculated.

3.1.4 Calculating Weight

When we extract the features from an image, the values retrieved are exponential which needs to be processed. Monotonous linear function is used in this case for scaling of feature value. It must take care of not affecting the actual meaning and value of the function along with proper scaling of values. Min Max normalization is used for normalizing these values after which these values are multiplied by a calculated threshold value experimentally for optimization. Then this value is being compared with inverse logarithmic function. These values are put back into database. In spite of textual factor, visual matching comes up with a more efficient relationship of semantic s of image. Result of this step is to chuck off all the mismatched tags and removes the irrelevancy.

3.1.5 Similarity Calculator

In this step, we need to compare the images. The basic flow of comparing is, consider that we have extracted features from bunch of images and we have them stored back in the database.

Now we have an image from Search query, we will need to match its features with existing features and compare and then give the output. Now in this step we need to work on synonyms. The calculated value in above step is passed onto a similarity algorithm. We need to find out Visual synonyms here which are represented as different in textual form but can be used to state a similar visual area. In this phase the weighted terms are passed through the cosine image similarity method to find out the image visual synonyms. As a result the textually different visual representations of image are generated and are combined together to form an offline image synonyms dictionary of terms. On input query from user interactive interface the synonyms for the given terms are loaded and expanded queries are formed. The results of these expanded queries are combined with the final results to generate the visually similar images from image database.

3.1.6 Assimilation server

The process did so far was a costly process, so these could not be performed online because this could increase computation and overhead time. There can be scenarios where number of new images could be added into the database periodically. So the main point of concern is to keep a track of online procedure going on along with offline database. As the content is offline calculated it is necessary to keep in sync with the online application deployed. To do this a continuous process needs to be executed. To achieve this Continuous Integration server is used with a scripting language.

3.1.7 Search User Interface

A semi-advanced, easy to understand UI has been given to users to play out a user driven inquiry in a multiuser domain. To do as such a platform independent user driven UI is used that is equipped for dynamic scripting and user preference assessment. Now as per requirement, User has to input a text which acts as a query according to him. With respect to this input provided by user, the appropriate images shall be retrieved. And for these images to show up, a UI is required which is built pretty simple and interactive with various edit features.

3.2 Pre processing

There are series of steps that are to be followed in preprocessing. As they are costly process, they are performed before online processing.

- Crawling of data from any website.
- Term extractor and working on stop words.
- Developing a parser which parses these terms to come up with various useful keywords for image recognition.
- After crawled text we have crawled images from which we have to extract features.
- Normalizing the terms for better comparison.
- Visual patterns extraction.
- Constructing offline image visual thesaurus

3.3 Online processing

Following are steps performed in this phase:

- On basis of user queries, perform conventional method of Image retrieval.
- Visual Query expansion to form expanded queries
- Expanded query based image retrieval
- Content relevance by Images ranking.
- Result display on UI.

3.4 Functional Requirements

This state all the necessities are the printed reported contract between what is to be developed and what is expected from the framework. It goes about as a detail report that involves about how framework must perform at whatever point it is asked. In our strategy following are the generic prerequisites that are expected from a non-specific search utility.

- It ought to give a natural language text processing and should give English as the essential mode for performing seeks.

- Framework ought to give an intuitive UI that can be gotten to by means of a few stages utilizing different gadgets, for example, PC, mobiles, tablets etc. It ought to be sufficiently adaptable to be conveyed and overseen over multi-user condition.
- Search should definitely be completed in less time along with relevant retrieval.
- It should worry about the safety efforts and should work as indicated by user inclinations set.
- If any new data comes in Database, it has to process accordingly. Finding out the keywords and visual pattern along with synonyms is one of the key requirements.
- This process shall not be manual whenever new image is added.

3.5 Non-Functional Requirements

These are the prerequisites that not at all like the practical necessities characterize the qualities of framework. Rather than characterizing what framework must do they characterize how framework must do it. These prerequisites don't add anything to usefulness yet goes about as quality accelerator agent for the framework. These occasionally are referred for quality factor of framework. Following are the non-useful prerequisites that are considered amid advancement of the proposed framework.

- User must not wait for infinite time to fetch the output. There must be a time frame in which results must be retrieved.
- We would expect the project to be up and ready to go.
- It ought to give stateless change from the sit out of gear state to the dynamic state without quite a bit of user execution overhead.
- It ought to likewise take in to thought different safety efforts. It shall give framework level, administrator level and user level parts each with the predefined parts get to levels and roles.

3.6 Requirements

3.6.1 Hardware Requirements

- RAM : 2GB or more

- Hard Disk : 32 GB
- Graphics : Intel
- Processor : 1.5Ghz
- Memory : 40GB

3.6.2 Software Requirements

- Integrated development environment : Eclipse/NetBeans
- Server : Apache Tomcat 5 onwards
- OS : Windows 2003 onwards/ Linux
- Storage (DB) : MySQL / Oracle 9g onward

3.7 Assumptions

It includes the pre-execution agenda that should be confirmed. There are set of pre-requisites that are required before really sending the application and making it open. They indicate the fundamental advances or set of steps that are should have been confirmed for a fruitful application arrangement over server. Following are some of the major points regarding it:

- Server must be configured with eclipse / NetBeans.
- Database must be configured and environment variables must be set. In eclipse, Run & Debug settings, under arguments section, localhost must be set and DB should be connected.

Chapter 4
Research Statement

In this chapter we have discussed the problems faced while retrieving images from Database one by one. Accordingly we will state methods to counter these problems and Implement in our project.

Web 2.0 and revolutions in web has acquired many changes in the method for recovering the data. Likewise the content over web has expanded to record level in recent years. The conventional data recovery frameworks are getting to be unessential as the size of content is evolving exponentially. A gradual scale defines the curve of information mechanisms against the exponential curve of content availability over web. It demonstrates a reasonable level of refinement between the image content and the strategies. As higher will be the rate of development lesser the proficiency of the data recovery techniques. Thus an expanded level of gap is seen in way data is recovered with that of the data produced. Therefore the demand of an efficient method for retrieving the data is required.

Next challenge comes to be relevant content. Now if a user enters a textual query like White shirt image. Mostly Content based retrieval is used where tags are manually added or at times we have trained tagging system for automating tagging process. While we are manually adding tags to image, it increases the overhead plus this is not the correct way to tag images. There might be a chance where some tags are unintentionally placed wrong or missed. So this can result in huge discrepancy and inconsistent result. This gap can affect performance of the system.

When a user has a query which is actually different from all the results retrieved by user is known to be a semantic gap. Our Main motive is to reduce this gap. Some of the reasons for existence of sematic gap are stated below:

- User queries according to his understanding

User may enter questions according to his or her comprehension and may neglect to frame appropriate query while operating the system. Also a user may expect a different form of understanding with respect to another user. Consequently it is important to keep the output of images according to user.

- Inappropriate tags present in image metadata

As we know tags are present in every image's metadata to define what is image about. Sometimes what happens is due to manual or even automated tagging, tags are assigned which doesn't match to visual meaning of that image. For example – Tagging shirt with skirt. Now after any search query related to skirt, this one image starts to appear which result in increasing gap. Such irrelevant images need to be improved for reducing semantic gap.

- Presence of similar terms in the query and tag

Rather than real term other similar words can be utilized to characterize the visual substance of image. Because of quality of equivalent words at times it ends up hard to pick the query related terms to recover the images. When we consider the most conventional method of retrieval, it solely depends upon presence of that word of query in image surrounding or metadata. Now if similar words would be present it may deliver different outcomes. In order to make systems efficient and better performing it is necessary to combine the results along with the visual synonyms.

- Linguistic gap between crawled text and crawled Image

In some cases the textual and visual importance of a term might be unique. Such terms are called as literarily similar yet visually unique terms. Keeping in mind the end goal to hold the expanded performance the framework must be capable of recognizing such terms and consolidate it with the visual representations there by coming about a higher review of execution.

- Comprehension about textual input

Different classifications of users get to the data accessible crosswise over web. Consequently variety of information queries can be framed. There can definitely be diverse understanding about the queries and henceforth can have distinctive search expectation for the similar queries.

4.1 Objective

Main & foremost objective for this system is to make a Vertical image search engine. There are basically two sorts of search engines – Horizontal and Vertical. Horizontal searches in the entire present Web and Vertical search in different domain require like image, news, maps etc. The purpose to achieve is reducing the linguistic gap between user text search and images we get. We will discuss counterparts of each issue discussed above one by one.

4.1.1. User queries according to his understanding

Selection of a query relevant image as the user understanding about the image helps to reduce the semantic gap as it makes clear what actually the user is looking for. Sometimes user has some similar terms to search but in the image the exact word is not present results into some major issues.

4.1.2. Inappropriate tags present in image metadata

To expel improper tags in the image surrounding data to make terms more significant to the image content. These helps to remove the undesirable and improper tags related to the image because of manual explanation mistakes that happen commonly. Expulsion of these unimportant tags to change the metadata related reduces the false positive outcomes.

4.1.3 Presence of similar terms in the query and tag

In order to increase the coverage and reduce the problem of synonyms, we need to form a dictionary where a term would contain all the similar terms to it. It would help easily to match the user query with visual meaning of that particular image. Considering the synonyms and visual synonyms the search operation must be performed else the efficiency of the system will be affected.

4.1.4 Linguistic gap between crawled text and crawled Image

The visual synonyms are calculated and are merged with the semantic significance of content. By doing as such the visual and textual meanings are consolidated with the goal that the outcome of linguistic differences can be reduced.

4.1.5 Comprehension about textual input

The recommendation of user relevant queries based on the popular inputs from data helps in retrieval of images. These recommendations are corporate while performing search operation thereby increasing the output result space.

4.2 Assumptions

There are some common points which are treated as fulfilled already before this project. Some of them are stated below:

4.2.1 Online Retrieval

Application would be showed locally as Server is already configured with one of the integrated development environment.

4.2.2 Search UI

A very simple and understandable UI is developed with minimal features related to filtering. According to any filer selected the output is retrieved.

4.2.3 Database

Now before storing the images to database, all the preprocessing is done. That means only appropriate image goes into Database.

4.2.4 User query

The form in which user is to enter any query must be text. Now from textual query entered features are to be matched from images and results are accordingly displayed.

4.2.5 Tags

Every image must contain a tag. When we talk about any image, it contains surrounding text and metadata. From that metadata we can actually retrieve tags. Presence of this is crucial as it affects the results majorly.

Chapter 5
Algorithms

This process is divided majorly into 3 parts which are as follow:

- Pre processing
- Offline processing
- Online retrieval of data

As we know there are majorly two image retrieval search engines- Text and content based. Here we attempt to construct a Retrieval mechanism where it utilizes both of the approaches as user has to enter a text query from UI.

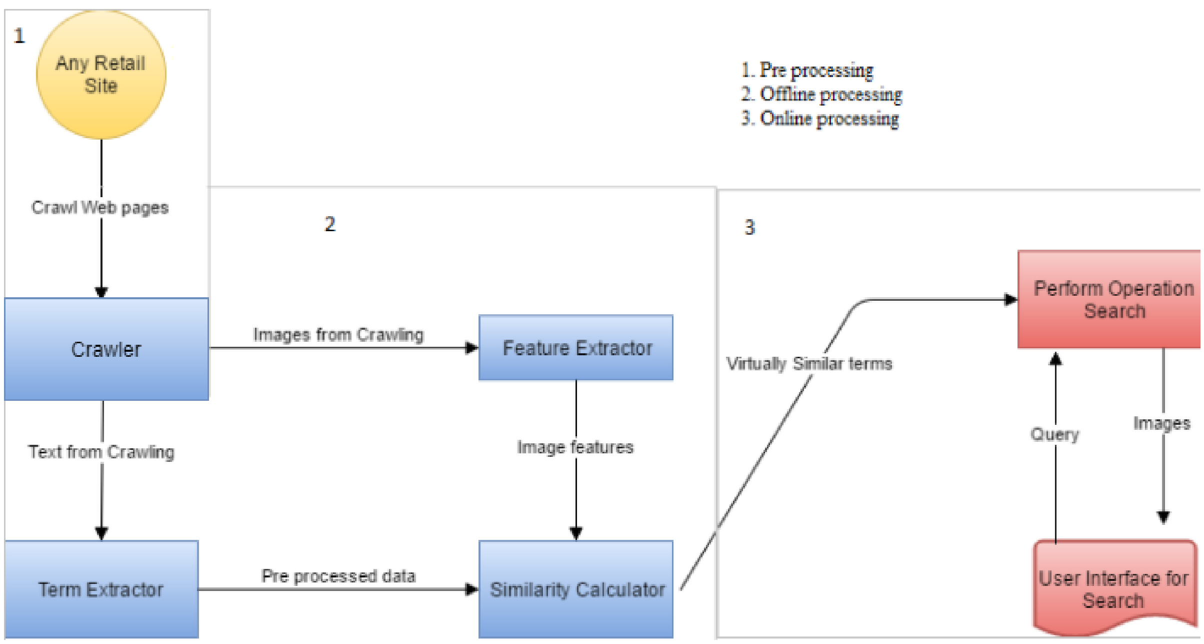

Figure 5.1: Components of project

5.1 Pre-Processing

The foremost thing we require to perform any kind of operation is data. We could use any Pre prepared data set which could contain certain set of images from various categories. Also crawler is used to retrieve all the images from any website. We have a Niche crawler which is used in order to crawl the data. From the well-known image web index the arrangement of pre-characterized queries are considered. These inquiries are utilized to perform look on the item pages.

5.2 Offline Pre-Processing

Now comes a huge part called offline processing where a number of operations are performed like handling synonyms, stop words, cleansing portions etc. If these operations were to be handled online and at the time of retrieval, it would have cost a lot. So, these operations are done before performing the actual operation.

This stage consists of Term Extractor, Feature Extractor and Similarity calculator.

5.2.1 Term Extractor

Serial Num	Image Metadata
1	{'category': 'footwear', 'thumbnail': u'http://myntra.myntassets.com/images/style/style_search_image/Fila-Men-White-Speed-Sports-Shoes_a680b6a824231fb1c3132aa8582aacde_images_180_240.jpg', 'subCategory': 'sports_shoes', 'image': u'http://assets.myntassets.com/v1/images/style/properties/Fila-Men-White-Speed-Sports-Shoes_a680b6a824231fb1c3132aa8582aacde_images.jpg', 'img_url_pattern': 'http://assets.myntassets.com/h_($height),q_($qualityPercentage),w_($width)/v1/images/style/properties/Fila-Men-White-Speed-Sports-Shoes_a680b6a824231fb1c3132aa8582aacde_images.jpg', 'brand': u'FILA', 'price': 2199.0, 'product_name': u'Fila Men White Speed Sports Shoes', 'sex': 'male'}
2	{'category': 'footwear', 'thumbnail': u'http://myntra.myntassets.com/images/style/style_search_image/FLX-Men-Sports-Shoes_a07fcd3c3c0b4cdf886151d475e6a088_images_180_240.jpg', 'subCategory': 'sports_shoes', 'image': u'http://assets.myntassets.com/v1/images/style/properties/FLX-Men-Sports-Shoes_a07fcd3c3c0b4cdf886151d475e6a088_images.jpg', 'img_url_pattern': 'http://assets.myntassets.com/h_($height),q_($qualityPercentage),w_($width)/v1/images/style/properties/FLX-Men-Sports-Shoes_a07fcd3c3c0b4cdf886151d475e6a088_images.jpg', 'brand': u'FLX', 'price': 1899.0, 'product_name': u'FLX Men

Table 5.1: Representation of metadata of an image

When we crawl any website that means we are making clone of all the multimedia in our local. Not like website mirroring but downloading the content of that website. After stage one, where crawling was done, we have data in form of images and text.

Now this image data contains metadata where we need to work. A term extractor is used which is basically a text parser that is used to filter some of the text required from metadata. For reference, consider the following image metadata which is a result of our crawler. It uses lexical syntactical rules which are used to work on ant text in metadata of the image. So when the images are being crawled, it contains a huge set of images which are not assigned with any tag or are inappropriate. So for better and smooth performance we need to remove any such existing impurities. Some of the methods for same are stated below:

5.2.1.1 Working on Stop words from image metadata

Presence of words like the, this, and, an, an etc. are stop words which are not at all useful in image retrieval. So these needs to be removed in order to proceed. There is an already defined set of these stop words, we just need to prepare a lookup table which contains all of these. This way we would match these words with words in image metadata.

5.2.1.2 Working on presence of Uniform resource indicator

As the images are crawled from any website pages, chances are they may also contain any URL and some other descriptions that are irrelevant for us. These must be expelled from metadata. For its removal we have a part which recognizes any URL presence in metadata. For reference, consider the following pattern

'Http://assets.myntassets.com/h_($height),q_($qualityPercentage),w_($width)/v1/images/style
/properties/Fila-Men-White-Speed-Sports-
Shoes_a680b6a824231fb1c3132aa8582aacde_images.jpg'
Such patterns are of no use to us, so can be removed and some important data can be filtered.

5.2.1.3 Working on explicit data and category of any product

The crawled images also contain the category and pricing related data which should be held yet in the different frame where its uniqueness can be kept up. The images are figured in to class and sub-classification i.e. on basis of such information we can divide them into its categories defined.

For reference, the data given about image in its metadata is in form ‘Category’: ‘Clothes’, ‘color’: ’black’ etc.

From these terms of explicit data of images, they can be categorized for offline processing. This can act as a first step of matching terms in metadata. Based on these category information first the images are separated to create offline indexed file hierarchy for information processing. Based on the input terms first the categorical information is matched to gain the image information followed by the text matching in surrounding text.

5.2.1.4 Working on distinct Keywords

After working on all the dependencies and isolating words from image, following thing is to distinguish the set of textually different unique keywords that are associated with images in its surrounding text. After successful identification of such terms, a lookup is made which has only one repetition of each term in it. Now this forms all the main terms required to describe an image. This lookup table is stored in form of a dictionary which is further used from matching query. Consider ‘Shirts’ stating a clothing category.

5.2.1.5 Equivalents and synonyms

This can be one of the scenarios where two different terms in lookup table means exactly the same. They will just add an overhead time to system performance because semantics are same. To solve this problem we have an Equivalent words algorithm in which these terms are passed. The word reference of terms is checked against the real dialect thesaurus to recognize any event of terms in the query. In the event that any match is discovered a passage in the literary equivalent word table is made against both the words where each term is an equivalent word of other. The same is relevant if in excess of one equivalent word is recognized.

5.2.2 Feature Extractor

In digital image processing each image is considered as a grid of pixel values where each pixel value is the combination of primary colors – Red, Green and Blue. Every one of these qualities characterizes a significant level of the given shading blend. Based on the mathematical derivation and formulations the numeric expressions of terms are formed these are known as image features. Features are the numeric expressions of pixel values which define images as matrices of image pixel values. There are various methods available to identify the pixel values and manipulate them. We use the following image features to identify image visual patterns:

5.2.2.1 Gray-level spatial dependence matrix

To proceed with any sort of image processing, first and foremost step is gray scaling. As a result of scaling, shading related data is changed with the grayscale extend from 0 – 255 which helps to know about intensity. This is also known as Gray level co-occurrence matrix where a method or function is used to evaluate texture of an image by considering its intensity values. Higher the intensity brighter the image content portions.

5.2.2.2 Hue, Saturations, Intensity

These three factors are responsible for distinguishing one color from the other. Hue is solely responsible for defining the purity of color. Saturation defines a quantity to measure amount of white in the original color. Intensity defines a subjective term called brightness. In view of the natural conditions the channels can be connected to evacuate impurities & other influence to pick up the good caliber of image.

For identification of any image, there are series of steps to be followed. File based storage is considered to store the images. To store them, we require a similar kind of structure to access the in easy manner. Image features are extracted using a designed algorithm for all images. Now these features could associate to image surrounding text data. These values are extremely small and exponential in form. We need to synchronize these in terms of a normalization factor which could diminish the surrounding values and give more focus on actual image. There we use min-max normalization as given in [31]:

$$norm\big(feat(x)\big) = \frac{feat(x) - \min(feat(x))}{\max(feat(x)) - \min(feat(x))}$$

Consider x as a particular feature and feat () as its value.

5.2.2.3 Directions

This factor is responsible for knowing about image from another perspective of user. Basically helps to know about flow of image and its direction. Tangent moment is calculated in order to get its exact value.

5.2.2.4 Coarseness

This factor is responsible for calculating any patterns or textures of image. It acts as a sole way to know about value of image harshness and structural behavior.

5.2.2.5 Texture

In count we have nearly 48 features which are result of Gray level spatial dependence matrix. This factor is responsible for surface, edge and pattern of image. Hence a distinguish decision can be made on basis of texture features.

5.2.2.6 Haralick features

Properties like rigidness and softness of surface in any image are defined by 13 haralick features. These provide clarity on basis of surface thickness.

5.2.2.7 Tamura features

Angular moments act as a very necessary factor for image processing. This feature is related to that evaluation along with the flow. For reference consider a pattern going from left pocket to ankle and a pattern from right pocket to ankle. These patterns both might be different and here Tamura features come into image and distinguishes between such patterns.

5.2.2.8 Color correlogram

Color is also one of the important factors and one image could contain number of colors. Here a graphical state is required for all the data related to colors. This helps in forming a color histogram where a technique is used to evaluate color by calculating a function between adjacent pixels of image.

5.2.2.9 Color moments

Color moments are for the most part utilized for color ordering purposes as highlights in image recovery applications keeping in mind the end goal to look at how comparative two images depend on shading. Normally one image is contrasted with a database of computerized images with pre-processed highlights to discover and recover a comparable image. Every correlation between images brings about a comparability score, and the lower this score is the more indistinguishable the two images should be.

For better results, we need to perform three levels of features from which 18 features would be extracted.

5.2.3 Similarity calculator

Now after Text extraction and feature extraction from images, we need to define a parameter on basis of which we can say twsso images are similar. There are number of cases we need to consider.

5.2.3.1 Unique words

As we know, in text based image retrieval, if the query term is itself present in the image metadata, It turns out be to very beneficial for search. There are a lot of cases where user want something else from search and gives input slightly different. So solution to this is combining the query text along with visual patterns and features for better knowing the intensions of a user. If the text we have in metadata and text in query term matches, initial search result is formed. After that image features comes into image. We look for a common visual feature and name it as "Visual meaning" of the image. For each term T_i in term lookup table its done.

5.2.3.2 Base for each term

Consider a scenario where we have Y images in our lookup DB which consist of a specific term U in their metadata. Also consider that each term is represented by N-dim vector where N is count of features given in [31].

$$\underset{\mu}{\rightarrow} = \left(\frac{1}{N} \sum_{k=1}^{N} feat(x_{k_1}), \frac{1}{N} \sum_{k=1}^{N} feat(x_{k_1}), \ldots . \frac{1}{N} \sum_{k=1}^{N} feat(x_{k_1}) \right)^T$$

In the above formula we have u which is mean of all vectors for every term present in lookup Database.

5.2.3.3. Working on inverted index and p-values

A single image has a lot of values for features. In content based system, not every one of the feature can be utilized for examination as it builds the calculation time in this way corrupting framework execution. To maintain a strategic distance from that the features are joined to a single value that characterizes feature significance without changing the semantic meaning of the same i.e. now a single value will determine all such features. Now p-value is calculated on all image features for all values from lookup DB. To do so a two set Kolmogorov-Smirnov method of p-value calculation is applied to the set of image feature for each term T_i in term lookup table constructed during term extraction phase.

For example, a set of samples is there X(i) where each X is the sample value from lookup DB. These sets are removed independently for each term Ti in the term query table. Because of this we get an arrangement of two samples, one with the query term show in it and the other without in particular positive and negative sample space given in [31].

$$Seq_n(feat(x)) = \begin{cases} 0, & if\ feat(x) < X_{(1)} \\ \frac{k}{n}, & if\ X_{(k)} \le feat(x) < X_{(k+1)}, \\ & where\ k \in \{1,2,3 \ldots .., n-1\} \\ 1, & otherwise \end{cases}$$

Cumulative distribution is as follow:

$$K(feat(x)) = \frac{2\pi}{feat(x)} \sum_{i=1}^{\infty} e^{\frac{-(2i-1)^2\pi^2}{8x^2}}$$

5.2.3.4 Weight calculation

The figured p-value is very small in numeric form to perform any comparison. Consequently they should be scaled before really utilizing them. In order to work on this two stage normalization work is connected with the end goal that it is an entirely decreasing function and when it is set to a limit it should decrease exponentially till that and diminish slowly then after given in [31].

$$feat(x) = \frac{tan^{-1} - log(feat(x) - C) + \ tan^{-1}\, C}{\pi}$$

Here C act as a experimentally calculated constant value as following:

$$C = \left\{ \frac{max\big(feat(x)\big) - min(feat(x))}{2} \right\}$$

So this step basically scales the p-value calculated to easily work on it. Now these are assigned against each query word in table to make the terms significant both in terms of its query and text. These weights are used along with the semantic meanign of terms

5.2.3.5 Visual synonyms

Two terms that are textually unique yet are visually same are named as visual synonyms of each other. They show same visual properties when are thought about outwardly. So to work on such terms, visual patterns and histograms needs to be compared given in [31].

$$Similarity(T_1, T_2) = \frac{\sum_{i=1}^{M} (\mu_{Term_{1,i}}, \omega_{Term_{1,i}}) \cdot (\mu_{Term_{2,i}}, \omega_{Term_{2,i}})}{\sqrt{(\mu_{Term_{1,i}}, \omega_{Term_{1,i}})^2} \cdot \sqrt{(\mu_{Term_{2,i}}, \omega_{Term_{2,i}})^2}}$$

This may not be exact and to make it more precise and increasing the computation speed, we need to reduce it to linear scale. To calculate a similarity between values we have a formula above where M defines count of features. Values here refer to p-values calculated as a result of the features being extracted from the images which were normalized in order for comparing two p-values for two different images.

For Example: Consider some terms which can be a part of any user query along with their visual synonyms used.

Serial no.	Query term	Visual synonym
1	Unisex	Shoes, Adidas
2	Sports	Sport, Puma
3	Park	Avenue, Edition
4	Shirt	Semi-Formal, top

Table 5.2: User Input Query terms

Table 5.2 shows the example of visual thesaurus of the terms computed offline in this scenario.

5.2.3.6 Further optimization

There can be scenarios where image annotations in comparison with respective data have

- Irrelevant terms
- Visual synonyms
- Equivalent visual meaning

To exclude such terms from negative Result Set and let them be in positive Result Set, images need to remove false negatives. To do as such the expanded user queried are framed utilizing the term equivalent words calculated up until this point. Along these lines the queries are shaped with the equivalent words and are applied on the negative Result Set and the images are extracted. These extracted images are included in the positive Result Set and the p-values and altered records are recomputed. These recomputed values are considered as the image weights 'ω' and utilized for image similitude measures. The procedure is repeated until the point when every one of the equivalent words are calculated and the term lexicons are shaped. These terms alongside the equivalent words are included to the term lexicon framed called as visual synonyms.

5.3 Online retrieval of data

In this step we majorly have two sub steps for retrieving the data and rendering it on User interface

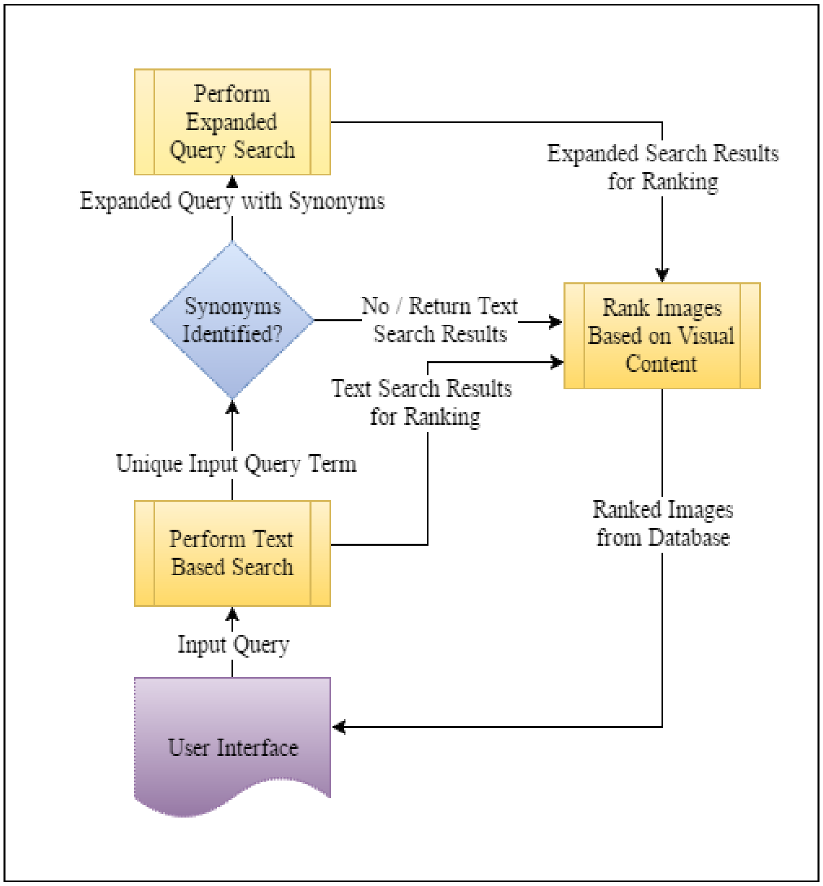

Figure 5.2 Search Algorithm

5.3.1 Search Algorithm

User has to enter a content contribution to the UI gave to pick up the data recovered from the space database. He can include the textual queries according to his or her comprehension. Internally it goes through all the steps for retrieving the content. Consider following points for in depth performance.

5.3.1.1. Search Algorithm for text

The most basic thing for retrieval of images we can think is retrieving images from offline database on basis of any query term occurring similar to one user has input. Think it like an agile process where in step one, query term is matched on basis of which data is retrieved. This offline dataset contains the data about the image alongside metadata, highlights, equivalent words etc.

In view of the queries the positive Result Set is retrieved from the information lexicon. Ranking needs to be done accordingly. Main fact for ranking should just be relevance of visual features. After this a buffer is used for the storage which is not permanent one.

5.3.1.2 Broader Search

After step one is completed, subsequent stage is to locate the relevant images exhibit in the negative Result Set. Keeping in mind the end goal to retrieve those, the query is broken into parts to get unique terms. These terms are cross checked with the term word lexicon for term presence. In the event that a match is discovered then the presence of equivalent words is checked. If a matching visual synonym is found in the term visual thesaurus the image expanded queries are formed using the visual synonyms. If no such equivalent terms exist then initially taken positive sample set is final result.

5.3.1.3 Image ranking

Now if we encounter any sort of equivalence and expanded query are shaped, these queries acts as a input for search being executed. The result given is the set of images that are visually similar yet do not have the query term in the image surrounding text. These images along with the initial sample result sets are combined with the latest results. The outcomes are as one weighed and are re-ranked in light of the visual feature significance. The last re-ranked comes about are surrendered as on the search UI.

Serial No.	Query for search
1	Orange causal shirt
2	Printed boots
3	Dark brown normal fit shirt
4	Red white printed t-shirt

Table 5.3: Examples of user inputs for results

5.3.2. User Interface

User interface is made having number of options and some filters with basic search tab. An established web crawler UI has been altered with the Bootstrap structure to make it significantly more clear and reasonable. A simple UI with help has been accommodated the user collaboration making it justifiable. Test queries are given to user to analyze the execution of framework in this way decreasing the effort in understanding. Likewise the established scan system is kept up for examination with the techniques to discover the distinction in the attempting to make it all the more engaging.

Chapter 6
Implementation

To execute any framework with a Proof of concept (POC) an assortment of ideas and information structures are utilized. These incorporate the class of information structures incorporate the fundamental usage points of interest and the design of elements in light of which the framework is based upon. A variety of data structures is utilized to execute the proposed framework. The MVC design is utilized for framework connection and execution displaying.

6.1 Overall system flow

The modules discussed are integrated in the manner shown in figure 6.1. The major modules depicted in this are falling under three categories- Preprocessing, Offline processing and online processing. Here we can see end to end flow from user query to all the calculations related to Weight calculation and similarity using cosine function under Assimilation server. Re-ranking is done after text based search operation is performed.

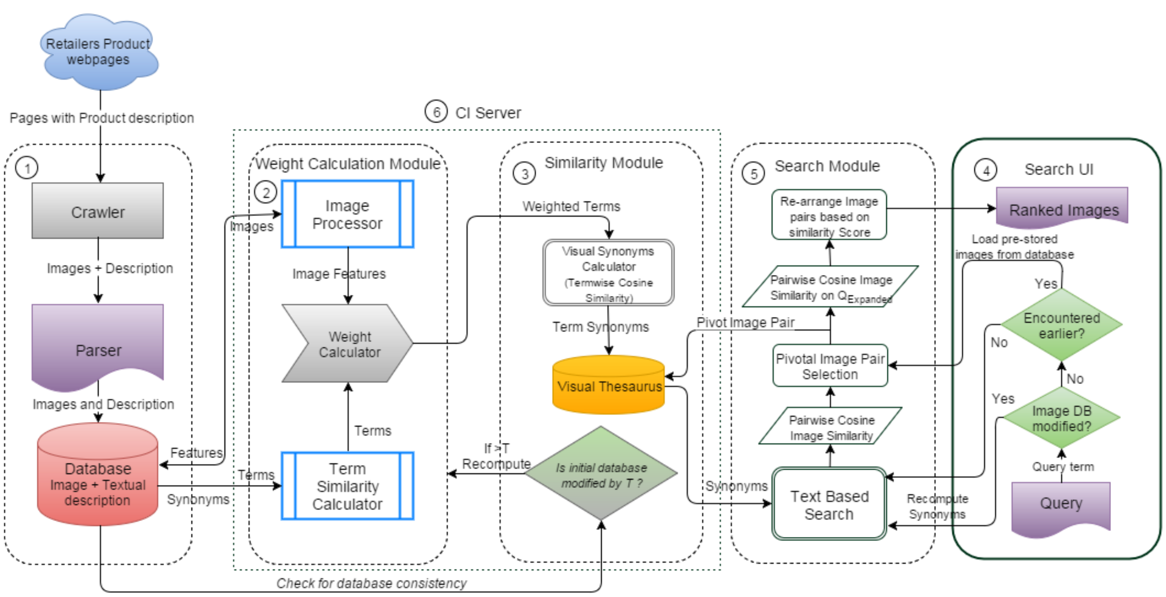

Figure 6.1: Overall system flow diagram

6.2 Class diagram

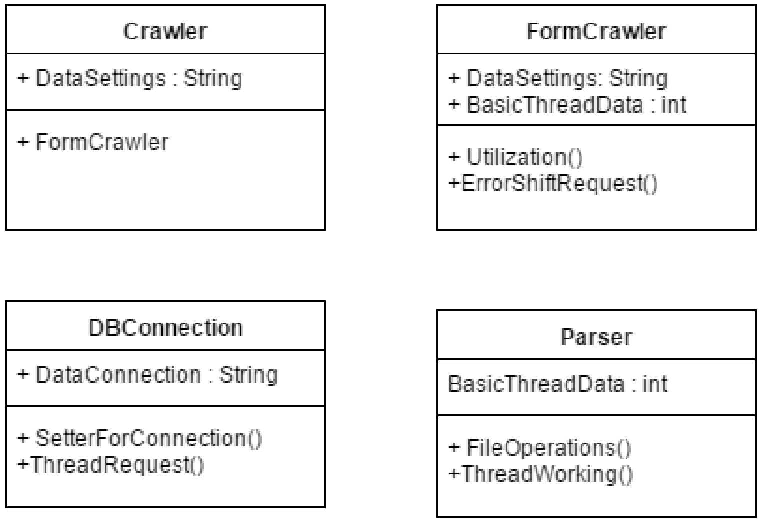

Figure 6.2: Crawler representation in a class diagram

Class is the dynamic portrayal of true idea. An object acts as an instance for a class. In the proposed work, different classes are recognized and are clarified as takes after. Each class contains two classifications of fields.

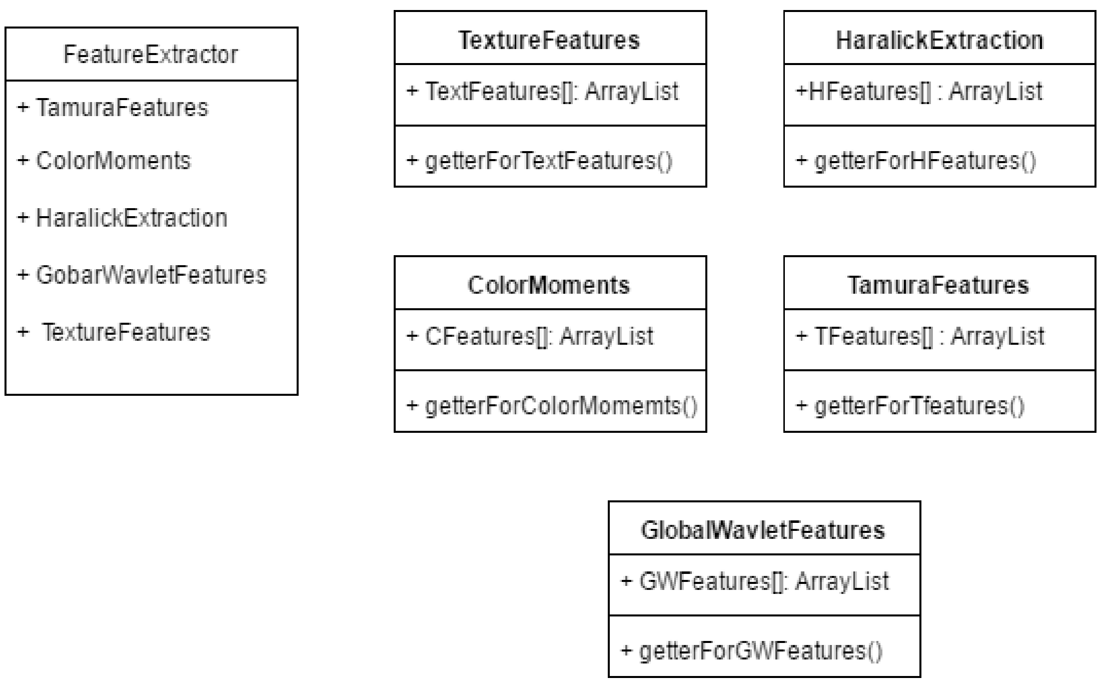

Figure 6.3: Feature Extraction class diagram

Attributes represents the class related data. For example, company, salary, position, and so on for the class employee and Behavior defines what an attribute is and the activities are characterized in the behavior segment.

After the crawler's job is done of getting the images from any website, a Comma separated file is formed which contains all the related features and properties along with metadata of the image. The information extracted so will be a crude content field including various URL's introduce in the image surrounding data. These are stored in file system as images in both of these PNG, JPG or JPEF organize. Now a file system is used to maintain these images and sole criteria of distinguishing is category. So images are stored on basis of category to which image belongs. Hence indexing is done after they are categorized. These are kept up in a hierarchal order in view of the categorical descriptors accessible in the image surrounding text. These are additionally utilized as the contributions for the offline pre-processing and numeric calculation.

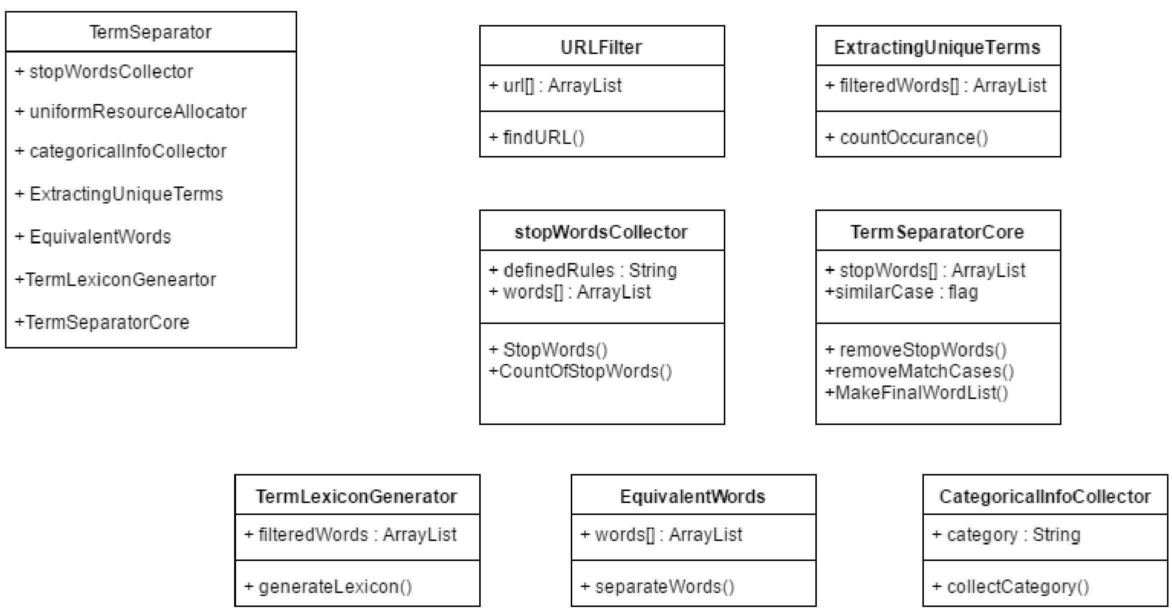

Figure 6.4: Term Separator Class diagram

In figure 6.4, Term Separator is designed where all the pre-processing part is done. It include removal of all the stop words, removing any URL present, Collection of all the synonyms and equivalent terms, defining the category as required by each image. The result of Term Separator is Lexicon which is later used.

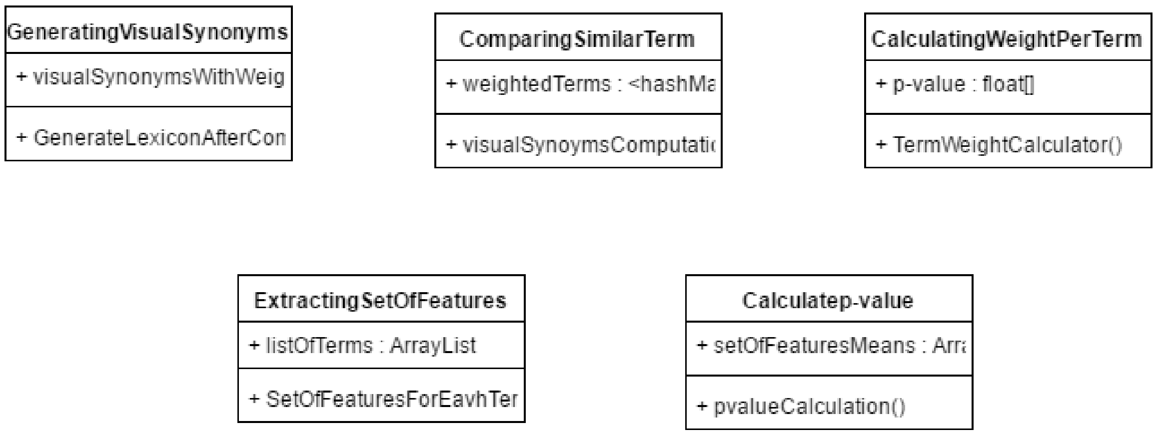

Figure 6.5: Similarity Calculator class diagram

In figure 6.5, Weight calculation and similarity module class diagram is shown. Here p-value is calculated for terms which need to be normalized since they are very low in scale. After which computation is done for calculation of weight.

6.3 Classes Interaction diagrams

Following are the diagrams which state the dependencies and interaction amongst classes.

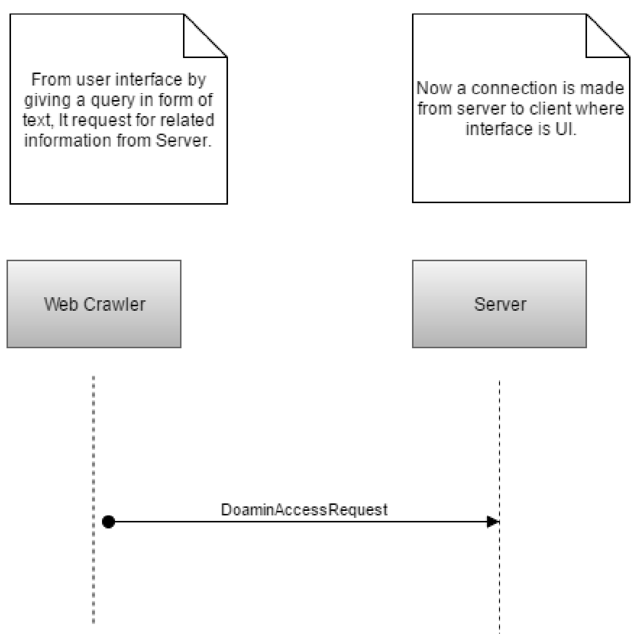

Figure 6.6: User Server Model where Crawler acts as a user

The main point of contact between various classes can be through number of methods like implementing same interface or some variables. Here we have shared variables. In figure 6.6, Integration of crawler is shown with the server. Since crawler is used as a plug in with browser which needs a connection.

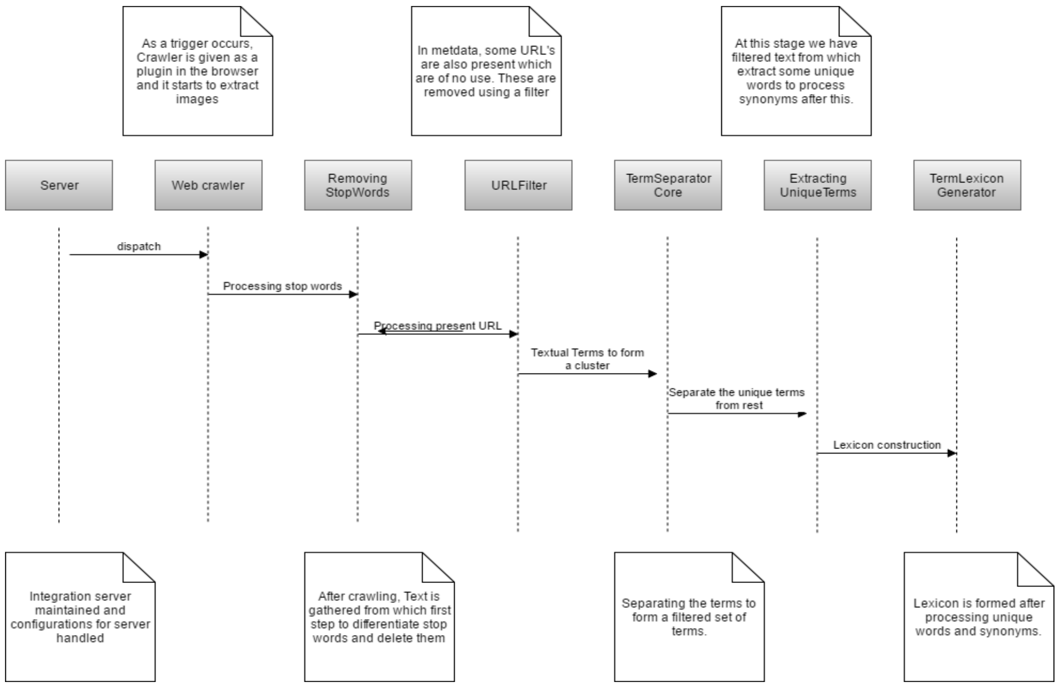

Figure 6.7: Term separator interaction with Server while offline processing

In figure 6.7, offline processing is depicted where term separator works on stop words, remove all the URL's, forming the filtered set of terms, extracting unique terms and results is formation of term lexicon to be further used while calculating the similarity between two different images to find visual synonyms.

Here Presence of words like the, this, and, an, an etc. are stop words which are not at all useful in image retrieval. As the images are crawled from any website pages, chances are they may also contain any URL and some other descriptions that are irrelevant for us. These must be expelled from metadata.

Distinguish the set of textually different unique keywords that are associated with images in its surrounding text. After successful identification of such terms, a lookup is made which has only one repetition of each term in it. Now this forms all the main terms required to describe an image. This lookup table is stored in form of a dictionary which is further used from matching query.

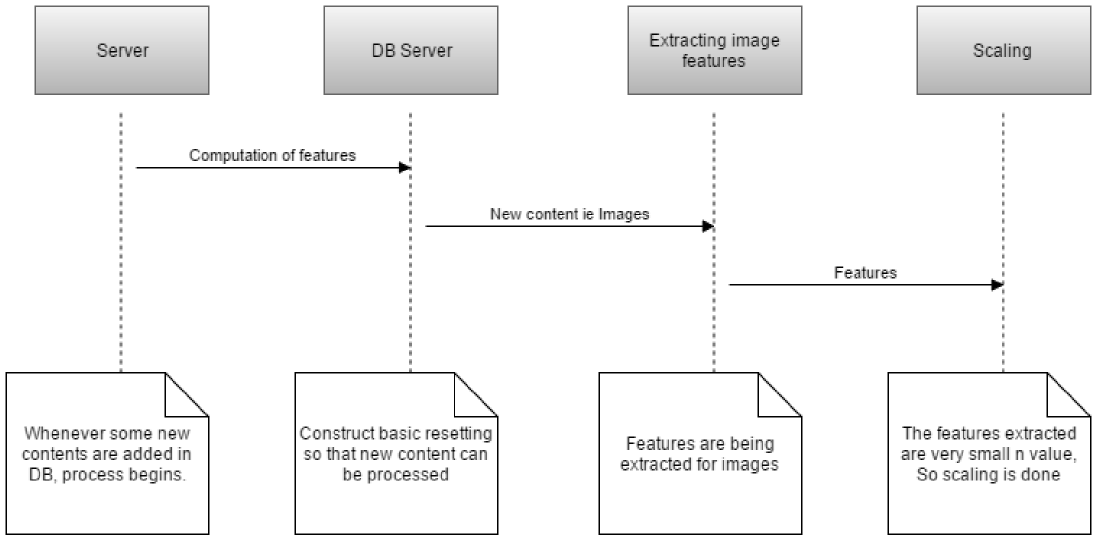

Figure 6.8: Feature extraction interaction with Server while online processing

In figure 6.8, p-values calculated are scaled since they are very small in value. It is difficult to compare these values. So normalizing these values are necessary.

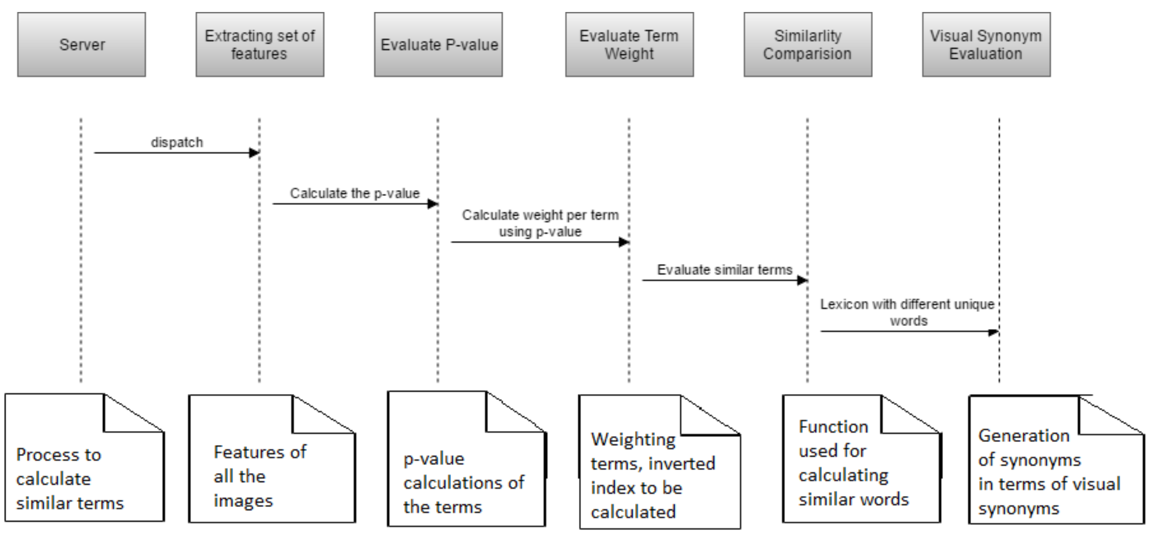

Figure 6.9: Similarity calculator interaction with Server

Here, after calculating and normalizing p-value, Inverted index is to be calculated. Calculation of similar terms is to done and dictionary of terms is generated.

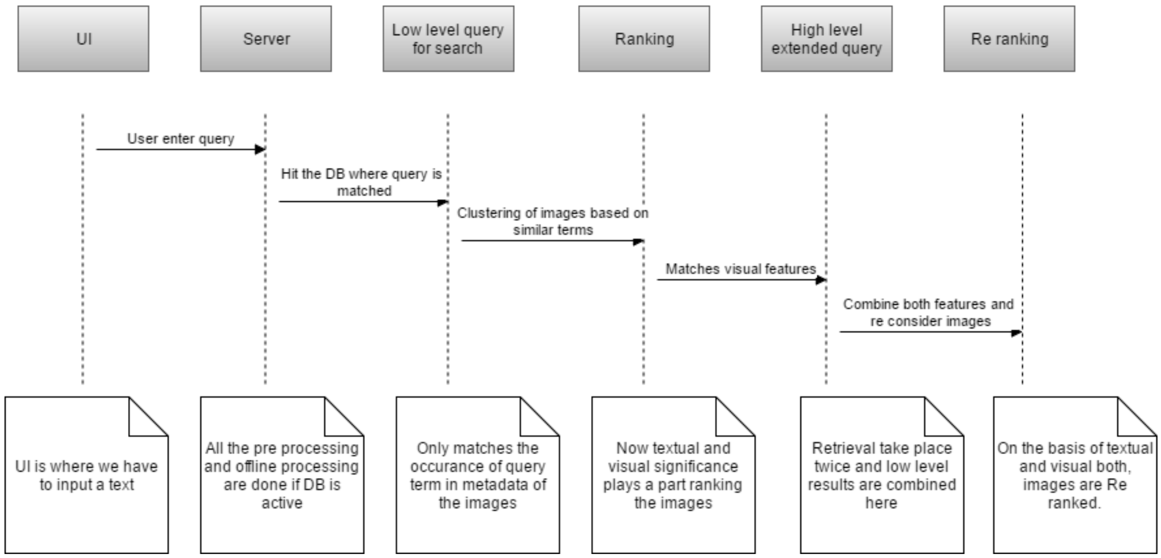

Figure 6.10: Retrieval dependency with server diagram

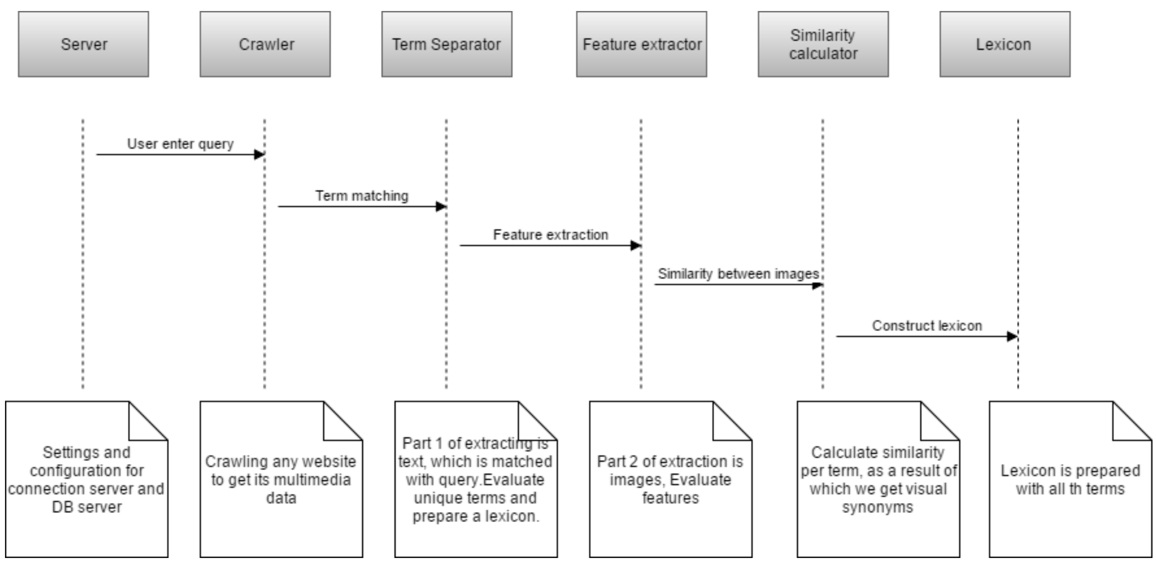

Figure 6.11: High level system interaction diagram

By methods for Interaction charts one can comprehend the correspondence between different framework segments and procedure of framework at usage level. Following outlines demonstrates the communication graph of planned framework.

6.4 Sequence diagrams

These diagrams are mainly used to understand the flow of operations taking place. It gives a detail structure about the input and output. It states about all the steps to cover before reaching the output. Hence with the help of these diagrams, it becomes quite simpler to understand system functionality.

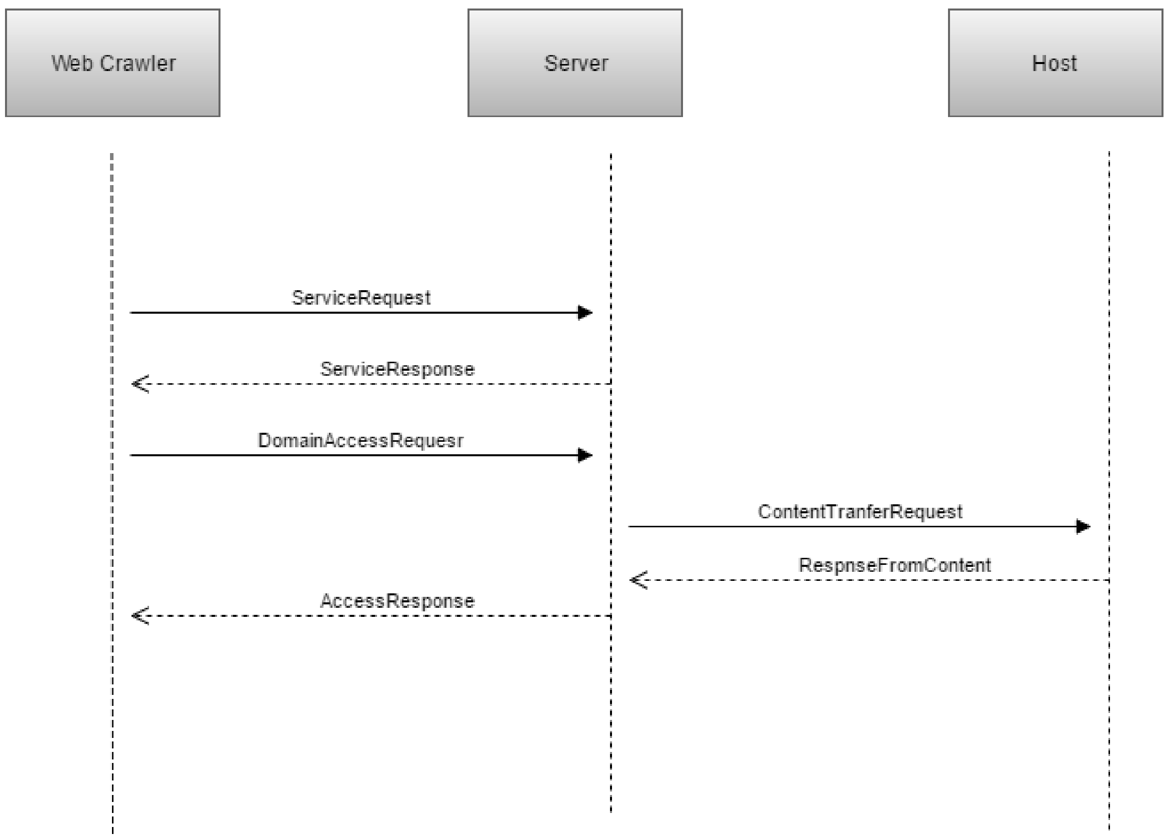

Figure 6.12: Sequence diagram for Crawler

It clarifies the crawler arrangement chart where user first demands for the server particular administrations for data getting to. A text query is passed for getting the information from domain. Already set of queries is made and fetch it from file present on server. Now these act as main point of request to server. It responds back with appropriate response based on request to be accepted or not. Output of this contains images with literary explanations along with some part of metadata. The data is extracted and given to the crawler interface where the surrounding text is removed and put away in the content document for each info inquiries. The data is crude and should be handled before utilizing it that is done in later stages.

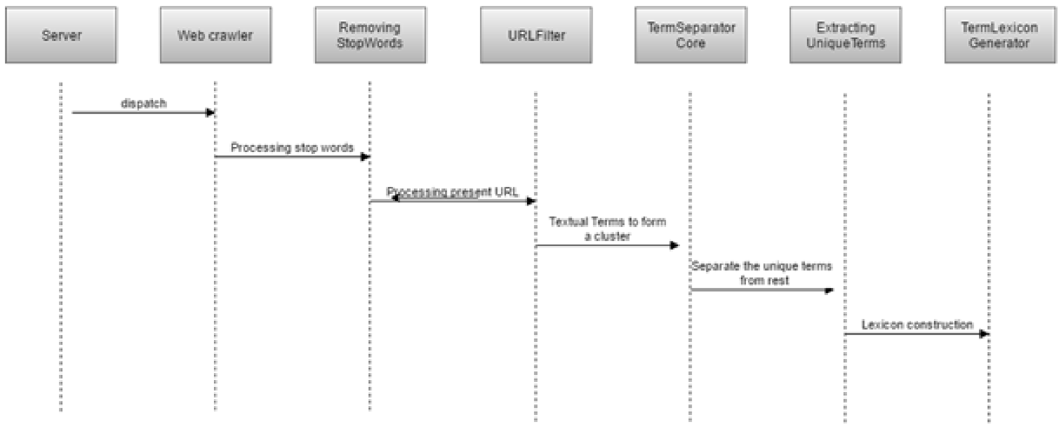

Figure 6.13: Sequence diagram of Term separator

Above diagram shows all the operations taking place while offline processing of data. This gives more vision and clarity of how offline processing is operated. The activities shown here are on high level.

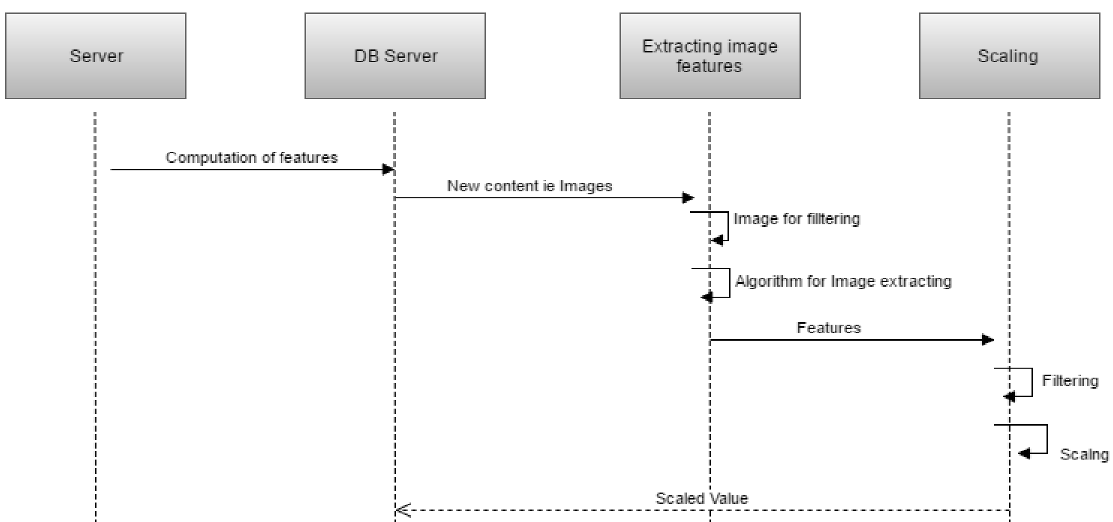

Figure 6.14: Sequence diagram in Feature extraction

Above diagram shows about all operations taking place which extracting features from images. This basically deals with addition of new data into the database and processing that with extraction of features. These newly added contents when passed through the feature extraction algorithm image visual features are extracted and updated in to database and further processes are carried out on these newly added feature values.

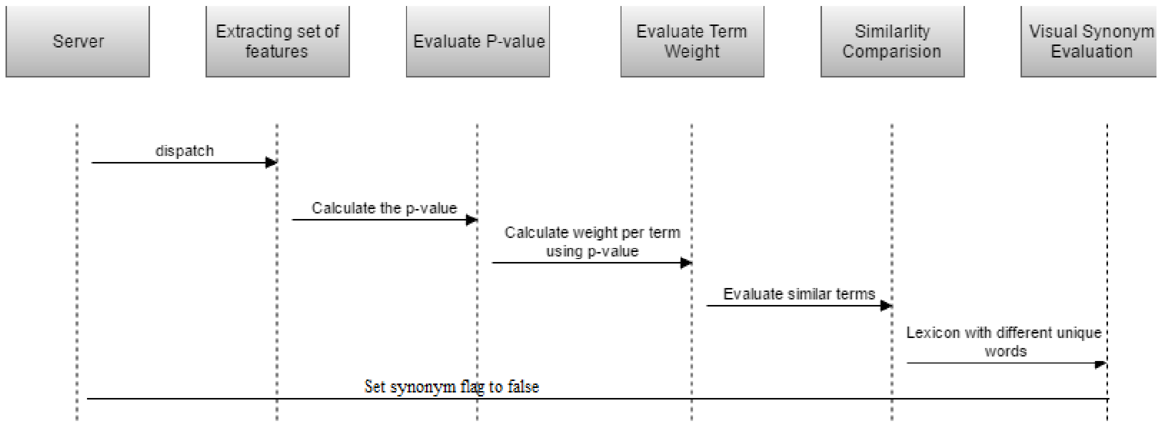

Figure 6.15: Sequence diagram in Similarity calculator

As we know, Text based search engine solely depends on presence of the user query term in the image surrounding text. If images are not tagged properly, this method fails. But this should retrieve results related to visually similar terms. In order to make this possible, we calculate term similarity which is necessary to make two level search possible. This is used in extended query generation. This term similarity is to be evaluated using p-value and Cosine similarity function shown in figure 6.16 and 6.17.

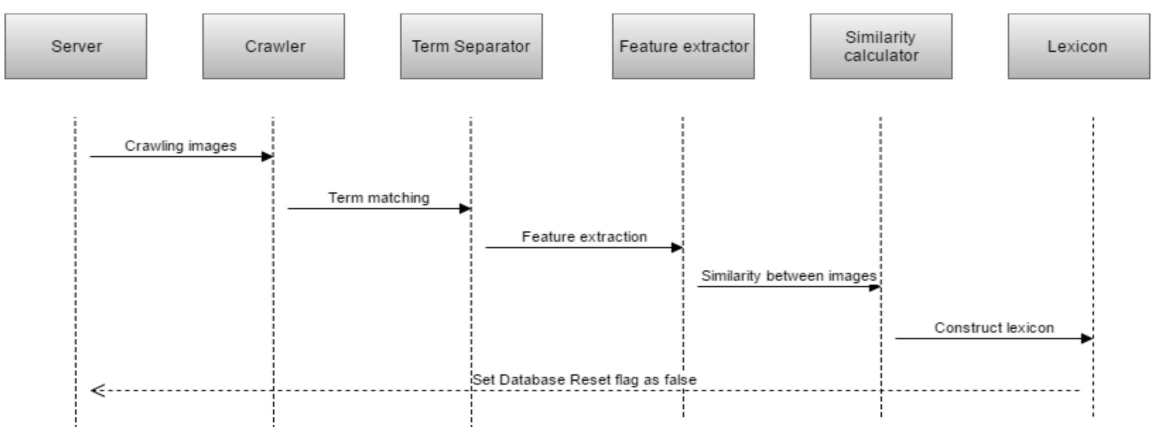

Figure 6.16: System Collaboration offline system sequence

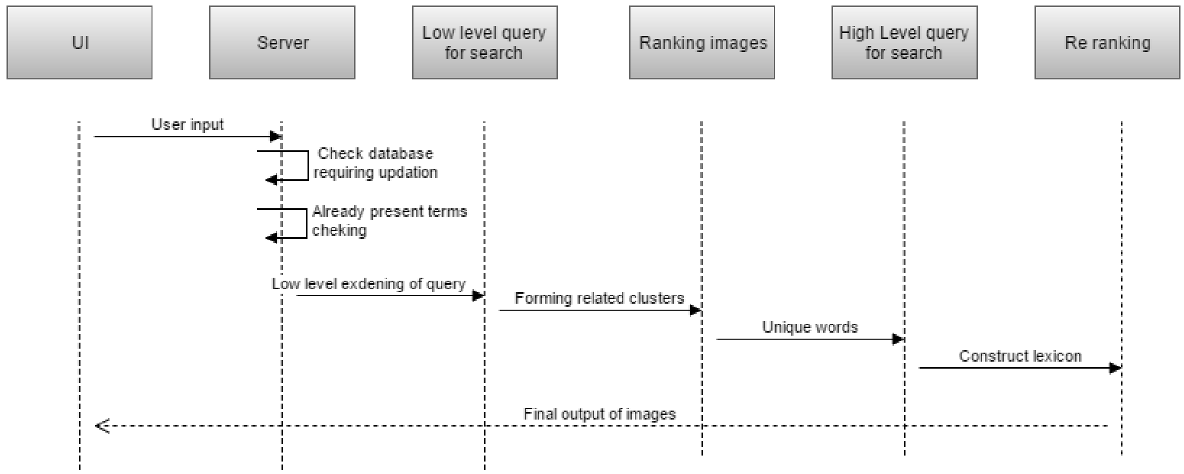

Figure 6.17: System Collaboration online sequence diagram

6.5 Use Case Diagram

These diagrams basically represent user perspective towards the system. User does an essential part while interacting with the system. These cases characterize the usefulness of framework from users' point of view. It is essentially an agreement between end users and the framework. It empowers developer to comprehend the extent of utilization and the users' desire from the framework. It outlines reenact a typical comprehension between the improvement angles with the customer point of view.

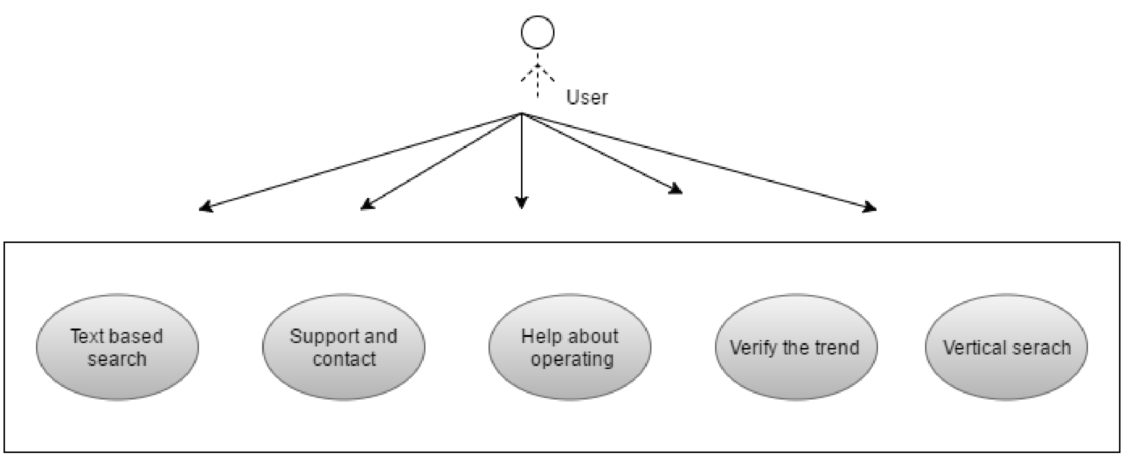

Figure 6.18: User view of system

It crosses over any barrier between the semantic aspects of system as an implementer with the market needs. It characterizes the rundown of on-screen characters that perform activities on the framework under development. In our strategy we order users in to two classifications manager and end user.

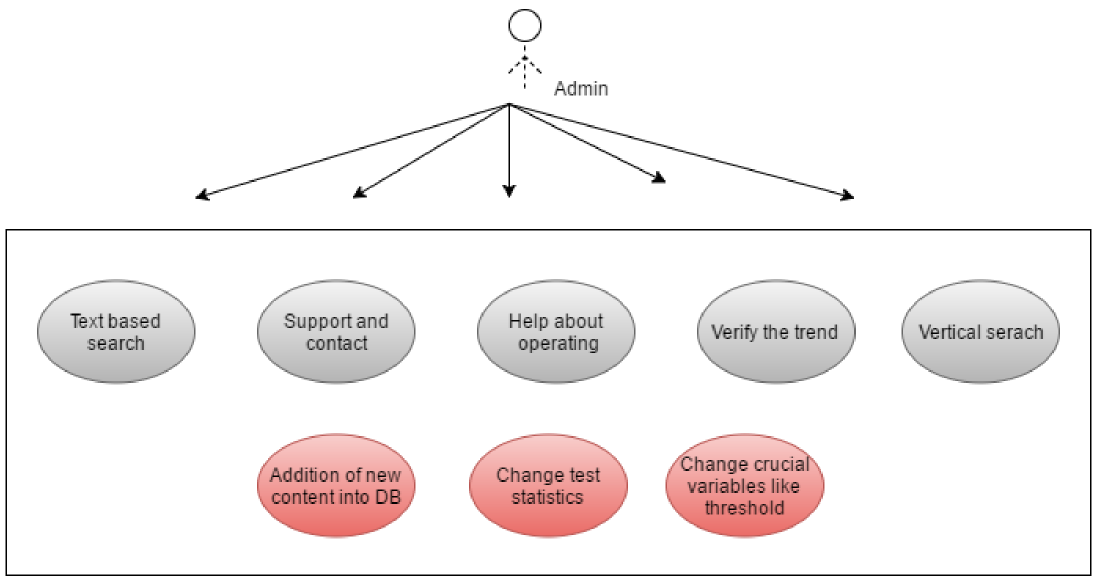

Figure 6.19: Admin view of the system

Additionally characterizes the rundown of permitted activities took into consideration every user part. Consider two major roles– Admin and user. Following are respective diagrams for both of them. On the basis of username and password, users are distinguished into respective database where per user is marked with its role. Accordingly various options from UI are given or removed. So, Admin may have lot more roles and permissions than a normal user.

Chapter 7
Results and Performance testing

This framework contains mix and match of both the search techniques i.e. text based and content based. Using advantages of these and removing limitations, there are number of techniques used. To increase the performance, following are some of used set of measures:

7.1 Lexicon generation

The content being crawled consists of clothes related data. In order to gather unique keywords that are related to images, data needs to go through a parser. As a result of this a lexicon is formed which contains keywords. These unique terms are one of the most important modules in this work.

7.2 Gathering content

The principal segment is to assess the framework in view of the choice of a decent selection of dataset. To analyze the system robustness a universal dataset was considered. To evaluate the same we have crawled randomly from the domain specific product pages. To do so a crawler was used. A set of hundred sample queries were taken from the random search engines. These queries were given to a custom web crawler. The system was capable of crawling 4393 image description with content. After filtration of images only 698 images were taken out with proper images and tags.

7.3 Visual synonyms

Visual synonyms are gathered after calculating all the p-values and inverted index. The similarity calculator was used to compare two p-values and comparing it with one threshold value. Those values which were higher than the threshold were marked as visual synonyms. They are basically all terms that are different in text but have equivalent meaning. The observed p-values were in the range of 10^{-12} to 10^{-49}. These values were normalized using Kolmogorov-Smirnov Distribution.

7.4 Testing

Utilizing these watchwords fifty test inquiries were framed. The resultant inquiries were connected on calculation given to perform content based pursuit so as to assess the execution. The execution of both the techniques was equivalent.

7.5 Performance and business Analysis

For testing purpose, we require some of reference queries mentioned in Table 7.1. The performance was evaluated by retrieving all the images and respective time to gain the output mentioned in table 7.2. We evaluate trend and recall values mentioned in below figure 7.1.

7.6 Scale Testing

In order to test it on larger amount of images, some more queries were considered. Also along with increasing the number of images, some more keywords were added. Count of total keywords goes up to 564. Accordingly visual synonyms were updated.

7.7 Feedback

Serial no.	User Input
1	Plain colored causal t-shirt
2	Round neck red t-shirt
3	Light brown shoes
4	Dark blue Shirt
5	White causal shoes
6	Peace colored formal Shirt
7	White Slim fit Shirt
8	Orange Black T-shirt
9	Black Pattern Formal shoes
10	Purple colored causal Shirt

Table 7.1: Some user inputs

Serial no.	User input	Classical method	Our Method
1	Plain colored causal t-shirt	21	24
2	Round neck red t-shirt	25	28
3	Light brown shoes	97	107
4	Dark blue Shirt	101	134
5	White causal shoes	7	19
6	Peace colored formal Shirt	25	35
7	White Slim fit Shirt	48	50
8	Orange Black T-shirt	83	89
9	Black Pattern Formal shoes	65	64
10	Purple colored causal Shirt	211	231

Table 7.2: Results of search retrieval number

Following are some of the graph representations regarding User rate, Recall and time taken for query to execute vs trend.

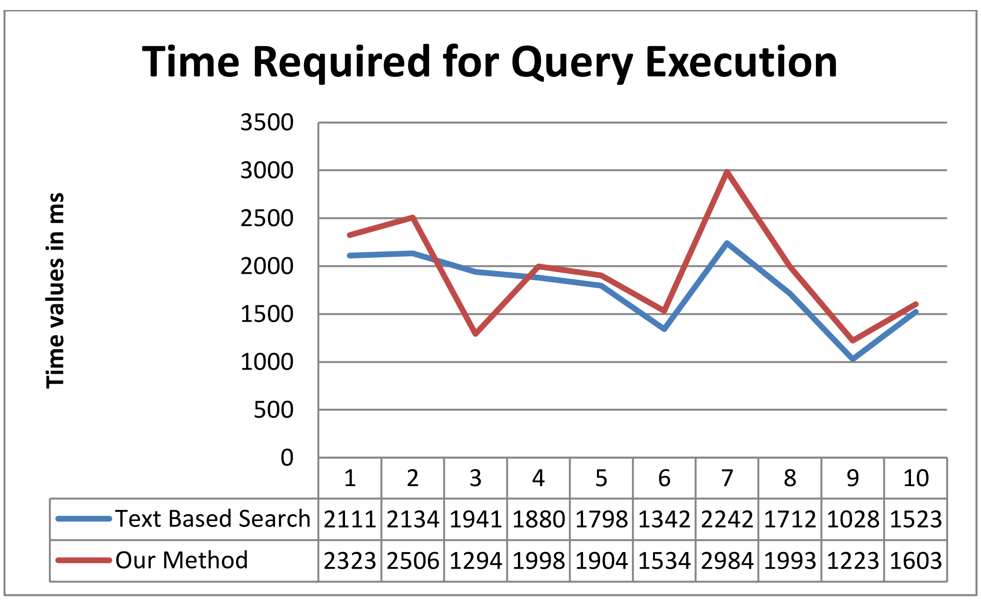

Figure 7.1: Graph representing Time vs trend

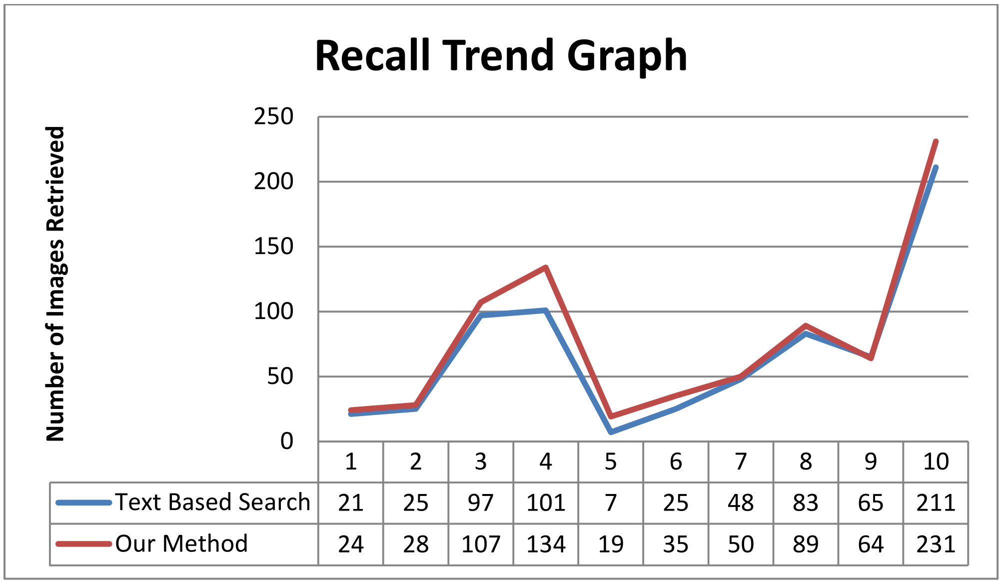

Figure 7.2: Graph representing Recall vs trend

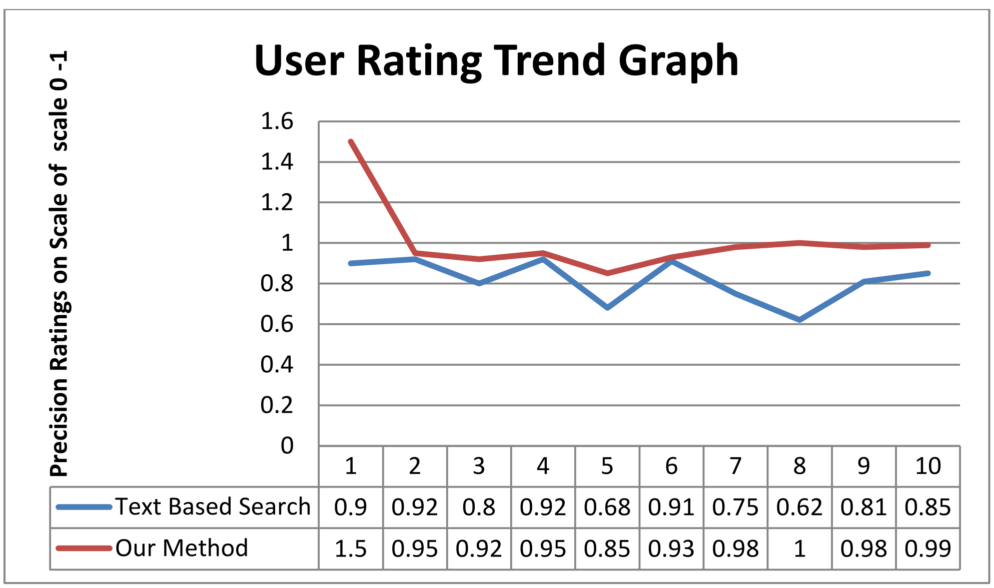

Figure 7.3: Graph representing user rate vs trend

Chapter 8
Conclusion

A proof of idea for a hybrid search system is proposed in this work. This framework combines features of both classical text based and content based image retrieval approaches and techniques to deliver an adaptable area particular vertical image internet searcher. The system is utilizing a web based UI thus can be made available anywhere. The classical approach strategy made us remove the semantic dissimilarities amongst images. Giving more importance to image surrounding text, good results are retrieved. It conquers the constraint of both the techniques with lower computation time and high performance time. Considering text based, the computation time is comparatively low but not as accurate as content based whose computation time is high. So texts based were mainly used in shopping sites and content based where accuracy is more important than time constraint like thumb impression. Combining the features and working on limitations of text based and content based, proposed hybrid approach gives better result.

The intelligent UI and FAQ's assistance one to comprehend the usefulness and make it effectively available with a little learning of pursuit. Testing is done for this framework against the user criticism. The testing has been finished with an aggregate of 989 images and an arrangement of 78 queries. The user input has been assessed by 90 users on a variable one to ten, where one is the most reduced significance score and ten is for most elevated. An arrangement of 436 watchwords is utilized to make the execution more solid. It has passed against various systems with failure assuring results and it has performed in a better manner. The computed importance has demonstrated the framework proficiency is superior to anything other with 87% achievement rate.

The framework is capable for retrieving images in view of the textual input inquiries. The model can be upgraded and tried against greater count of content and queries to make the execution testing of the same. Failure free tasks with a lower response time builds adequacy of the framework with 89% exactness score. Image similitude measures can be consolidated to build accuracy. This is not restricted to just Retail websites but can be extended to any sort of platform.

Chapter 9
References

[1] G. Hua, Z. Liu, "Meta-tag propagation by co-training an ensemble classifier for improving image search relevance", *in proceeding of Computer Vision and Pattern Recognition Workshops, IEEE Computer Society Conference,* July 2008.

[2] K. Sankar, G. Babu, "Implementation of Web Image Search Portal with automated concealed annotation", in *Communication and Electronics Systems*, March 2017.

[3] T. Chen, Chenliang, J. Luo, "Improving Text-Based Person Search by Spatial Matching and Adaptive Threshold", in *Applications of Computer Vision,* May 2016.

[4] Q. Bo, J. Pen, "Mixture Feature Based Junk Image Filtering", in *Computer Distributed Control and Intelligent Environmental Monitoring,* April 2012.

[5] K J Portar, "Textract-a hardware based parallel text search system", in *Parallel Techniques for Information Retrieval, IEE Colloquium on, IET publication, London,* April 1989.

[6] T. Ikeda, H. Mano, H. Itoh, H. Takegawa, T. Hiraoka , S. Horibe , Y. Ogawa, "TRMeister: a DBMS with high-performance full-text search functions", *in proceeding of 21st International Conference on Data Engineering,* IEEE, pp. 958 – 967, April 2005.

[7] J. Krapac, M. Allan, J. Verbeek, F. Jurie, "Improving Web Image Search Results using Query-Relative Classifiers", in *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, July 2009.

[8] D. Zhang, "A generic content-based image search model and its key factors", *in proceeding of Applications of Digital Information and Web Technologies, Second International Conference* , October 2009.

[9] N. Rasiawala, J. Pareira, E. Coviello, G. Doyle, G. Lanckriet, R. Levy and N. Vasconcelos, "A New Approach to Cross-Model Multimedia Retrieval," in *MM'10,* October 2010.

[10] N. Idrissi, J. Martinez and D. Aboutajdine, "Bridging the Semantic Gap for Texture-based Image Retrieval and Navigation" in *Journal Of Multimedia,* vol. 4, no. 5, pp. 277-283, October 2009.

[11] A.R. Patil, A. Manjrekar, "An innovative approach to classify and retrieve text documents using feature extraction and Hierarchical clustering based on ontology", *in proceeding of Computing, Analytics and Security Trends International Conference,* May 2017.

[12] Y. Zhou, K. Chen, X. Yang, "Google image search refinement: Finding text in images using local features", *in proceeding of Computer Science and Automation Engineering, IEEE,* August 2012.

[13] S. Yadav, P. Bhanushalli, S. Jain, T. Kaur, "Word matching and retrieval from images and indexing", *in proceeding of Electronics, Communication and Aerospace Technology, International conference,* December 2017.

[14] G. Blanco, M.V.N. Bedo, M. T. Cazzolato, L. F.D. Santos, A. E. S. Jorge, C. Traina, P.M. Azevedo-Marques, A. Triana, "A Label-Scaled Similarity Measure for Content-Based Image Retrieval", in *Multimedia IEEE International Symposium* , January 2017.

[15] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years", in *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, no. 12, June 2000.

[16] R.S. Torres, C.B. Medeiros, E.A. Fox, "An OAI compliant content-based image search component*", in Proceedings of the 2004 Joint ACM/IEEE Conference on Digital Libraries*, June 2004.

[17] N. Chathurani , S. Geva , V. Chandran "An effective Content Based Image Retrieval system based on global representation and multi-level searching*", in proceeding of 10th International Conference on Industrial and Information Systems IEEE*, Feburary 2016.

[18] Thomas, D. Keysers and H. Ney, "Features for Image Retrieval: An Experimental Comparison", in *German Research Center for Artificial Intelligence, Kaiserslautern, Germany,* November 2007.

[19] D. Kurniawati, D. Triawan, "Increased information retrieval capabilities on e-commerce websites using scraping techniques", *in proceeding of Sustainable Information Engineering and Technology International Conference* , March 2018.

[20] G. Hua, Q. Tan, "What can visual content analysis do for text based image search?", *in proceeding of Multimedia and Expo, 2009 IEEE International Conference on*, August 2009.

[21] G. Amato, V. Magionami, P. Savino, "Region Based Image Indexing and Retrieval Inspired by Text Search", in *Image Analysis and Processing Workshops, 2007,* January 2008.

[22] S. Polsley, J. Ray, T. Hammond, "SketchSeeker: Finding Similar Sketches", in *IEEE Transactions on Human-Machine Systems* Volume: 47 Issue: 2, April 2017.

[23] C. Mahamuni, N. B. Wagh "Study of CBIR methods for retrieval of digital images based on colour and texture extraction", *in proceeding of Computer Communication and Informatics (ICCCI), IEEE International Conference* , November 2017.

[24] M. R. Mustaffa, W. S. San, "Content-based fauna image retrieval system", *in proceeding of Signal and Image Processing Applications, IEEE International Conference* , December 2017.

[25] M. R. Mustaffa, N. M. Norowi, S .M.Yee, "Content-based image retrieval system for marine invertebrates", *in proceeding of Information Retrieval and Knowledge Management, IEEE International Conference* , January 2017.

[26] M. Fakheri, M. C. Amirani, T. Sedghi, "Gabor Wavelets and GVF Functions for Feature Extraction in Efficient Content Based Colour and Texture Images Retrieval", in *Machine Vision and Image Processing,* January 2012.

[27] P. Bhatt, P. Rusiya, V. Birchha, "WAGBIR: Wavelet and Gabor Based Image Retrieval Technique for the Spatial-Color and Texture Feature Extraction Using BPN in Multimedia Database", in *Computational Intelligence and Communication Networks,* March 2015.

[28] M. Unser. "Fast Gabor-like windowed Fourier and continuous wavelet transforms", in *IEEE Signal Processing Letters*, vol. 1, no. 5, May 1994.

[29] P. Moreno, A. Bernardino, and J. Victor, "Gabor Parameter Selection for Local Feature Detection", in *IBPRIA - 2nd Iberian Conference on Pattern Recognition and Image Analysis, Estoril, Portugal*, June 2005.

[30] C. S. S. Kumar, M. Mohanapriya, C. kalaiarasan, "A new approach for information retrieval in semantic web mining involving weighted relationship", in *Innovations in Information, Embedded and Communication Systems* , February 2018.

[31] Y. Chen, H. Sampathkumar, B. Luo and X. Chen, "iLike: Bridging the Semantic Gap in Vertical Image Search by Integrating Text and Visual Features,", in *IEEE Transactions on Knowledge and Data Engineering,* vol. 25, no. 10, pp. 2257-2270, October 2013.

List of publications

1. P. Arora, R. Tekchandani, "A hybrid approach towards minimizing the semantic gap in image retrieval by combining classical and content based techniques" , in *CONFLUENCE-2019, 9th International Conference on Cloud Computing, Data Science & Engineering*, August 2018, Amity School of Engineering and Technology, Noida, U.P., India (Communicated).

Palak thesis

ORIGINALITY REPORT

6%	4%	5%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Hardeep Singh, Dheeraj Agrawal. "A meta-analysis on content based image retrieval system", 2016 International Conference on Emerging Technological Trends (ICETT), 2016 Publication	<1%
2	ir.canterbury.ac.nz Internet Source	<1%
3	Venkatakrishna, D., and B. Ankayarkanni. "A Description of Content Based Image Retrieval using from Block Truncation Coding and Image Content Description", Indian Journal of Science and Technology, 2016. Publication	<1%
4	ittc.ku.edu Internet Source	<1%
5	dblp.dagstuhl.de Internet Source	<1%
6	C S Saravana Kumar, M. Mohanapriya, C Kalaiarasan. "A new approach for information retrieval in semantic web mining involving	<1%