12-2017

Analysis and Implementation of Policy Inference of Uploaded Images on Networking Sites

Venkata Sai Sathvik Ponakala
St. Cloud State University, vsponakala@stcloudstate.edu

Follow this and additional works at: https://repository.stcloudstate.edu/msia_etds

Analysis and Implementation of Policy Inference of Uploaded Images on Networking Sites

by

Venkata Sai Sathvik Ponakala

A Starred Paper

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Information Assurance

December, 2017

Starred Paper Committee:
Susantha Herath, Chairperson
Dennis Guster
Balasubramanian Kasi

Abstract

Human or user increase in the usage of social networking through various platforms by sharing their images increased a need for developing more secure systems which help user to share information among groups or people that he or she authorizes without causing any security issues in future. People can then share their desired images by applying different settings on the image such as who can view the image, who can download the image etc., and least bother about the privacy. To solve the privacy problem of excessive and careless sharing of images and information by users, this paper analyzes and describes a privacy policy development system that helps users to create new privacy settings for the images they share and helps in reducing the security breach or fraud over internet. It also helps the user by providing individual settings to all the images they share individually. Depending on the user history on the sharing site, the system proposes a policy in which the user can approve or make changes to the privacy prediction. Polices are thus improved afterwards in time with all the user inputs. Along with the description, this paper also presents a part of implementation of the system.

Table of Contents

Chapter   Page

Chapter   Page

List of Figures

List of Tables

Chapter 1: Introduction

Introduction

Social Networking means connecting the people and expanding one's business or personal relationships through internet (what is social networking?, n.d.) Many individuals have become multi-millionaires in no time with the success of their networking sites. Facebook, Twitter, Instagram etc., come under such networking sites category. Billions of people all over the world have accounts in these sites and share their personal or public information with their friends.

People of all ages, from teens to old age are having social networking accounts on the internet. In the busy day-to-day world, where meeting every friend or family in person and spending some quality time has become a great deal of effort, people gathering together has reduced and sharing their information over the internet has increased.

It is pretty easy to share the images or the information in one's mind to their friends or family members and get their opinion on it quickly via social networking sites. There is no hassle of meeting everyone, discussing or showing the pics and getting their opinion. All one needs to do is to create an account, make friends and share their information. Moreover, the social networking sites over the internet are all ready to use without any charge if the individual has a proper internet connection at the time of use.

People get to know about each other and can make good friends over the internet using these sites. Anyone can know about a person's day to day life events and activities over the internet and there is no need for meeting and discussing at a certain place.

As a coin has two sides, using social networking sites has its own advantages as well as disadvantages. In physical point of view, now-a-days people are spending more time on social

networking sites and the involvement of kids or younger people in sports or other physical activities has been decreased.

There are many cases where youth are being targeted for money by stealing their personal information and using them against them resulting to increase in death rate via suicide due to constant threats and harassments. Such situations have increased enormously these days with the increase in usage of these social networking sites. Interaction between people is decreasing day by day which eventually decreases the human empathy.

User shares images over social networking sites to improve their connectivity with family, friends, and acquaintances and sometimes with people outside their social circle to discover new peers and identify new relations. In such cases, information shared by a user through pictures may contain key data of his/her peers or family members. This might result in exposure of an individual's personal information and also social environment which may lead to social abuse of the individual.

For example, consider a party where all the friends, family members and other acquaintances join and party with each other and click pictures. Later, the person hosting the party might want to share these images on his or her social networking site. In such scenario, the person might share a specific group of pictures to his friends and other specific group of pics strictly to his or her family in which only family members were present. In such cases, sharing the pictures on social networking sites need to have certain privacy and policies that are to be applied on the pictures to restrict misuse and also reduce the risk of sharing the private information to the wrong people in user's multiple groups.

Though many social sites enable privacy settings for user, it is a repetitive and boring task for a user to identify each person in his or her images and set according to their privacy

policies. For example, a user wants to share hundred pictures of his party. He or she needs to go through each and every picture and apply settings for each picture manually. In order to reduce this problem of hard and error prone privacy settings, this paper analyzes a more dynamic approach of privacy policy prediction which is adaptive to user's requirements. Having such policy prediction helps people to reduce the risk of losing their privacy (Squicciarini, Lin, Sundareswaran & Wede, 2015).

Criteria followed by the Adaptive Privacy Policy Prediction (A3P) system includes Social environment of the users: It can be referred to as the group of the people the user hangs out with, mostly interacts and also the social interests such as groups he or she included etc. User personal characteristics: User's personal characteristics may include personal interests such as hobbies or the user's passion such as sports, writing, photography etc. Image content: The content in the image that the user shared such as if a user shares a scenery, it might have the hills, or forest or lakes. Such things come under image content. Metadata: Meta data of the images like tags, location, and description of images. The user might want to specify the location where he or she is in the particular location and specify the people who are in the particular location or picture (Squicciarini et al.,2015).

Depending on these criteria, the system has been classified in to two main building blocks A3P-Social and A3P-Core. A3P core analyzes user's images and the metadata where as Social provides an environment perspective.

In order to develop the A3P, the system uses different data mining methods such as classification and policy prediction and also the wavelet transform in order to achieve the content based classification. The following sections are organized as literature review and background where other approaches for the problem and also the background information needed to

understand the terms in methodology are explained. Later the methodology is analyzed in methodology section and software environment chosen for the development of the system and then the implementation part where the system implemented is explained and also the system design which contains the UML diagrams of the system implemented and finally concluding with a brief conclusion and the future work that can be done on the system implemented.

Problem Statement

With the increasing social networking sites and lacking user privacy in data sharing, there is an increasing need for a shared image protection system which gives user a peace-of-mind by providing the user a set of privacy settings that can be used as-it-is or change them according to the user requirements. These settings are adaptive and adjust themselves to the users past decisions and future requirements. In this context comes the Advanced Privacy Policy Prediction system helping users stay safe and gives them a hassle free protection environment.

Chapter 2: Literature Review and Background

A3P systems is related to privacy settings of users in various social sites, hence privacy analysis of images is the key to the systems definition.

Details on Previous User Privacy Setting Systems

Concept of Privacy Settings proposed by Bonneau, Anderson and Church (2009), recommends a set of privacy settings which were earlier imposed by the user or by other groups or social circles of the user. It states a system of four building blocks such as abstraction, distribution, installation and maintenance. First a privacy suite will be created and distributed to all the person's groups. The System gathers the best selected suite and then imposed to the user's privacy policies. The UI should be dynamic to change the policies after the best suite is discovered. Lastly, the system needs to be maintained continuously by checking any changes in the person's or group's interests.

Bonneau, Anderson and Danezis (2009), proposes a system which automatically draws privacy setting recommendations from the context in which the data has been obtained by the user. It also collects profile and graph information from different users in popular social networking site Facebook and examine the profiles by various common attacks such as phishing, Data Aggregators, False profiles and malicious applications. Later they also describe the efficiency of above mentioned methods on crawled data.

One more concept of privacy settings based on the social circles of the user have been developed by Adu-Oppong, Gardiner, Kapadia and Tsang (2008), which questions the reliability of meaningfulness of user setting policies based upon the groups he or she made. The technique described analyses the social graph of user on groups of closely connected friends. It is also

based on the hypothesis that a user implies and connects with the same group of people he used to connect with. Later, if any changes occur, the system can be refined with the updated pattern.

According to Ravichandran, Benisch, Kelley and Sadeh (2009) has examined on the use of providing users with the default policies and to what extend does that help the users to better improve the privacy. The research is based on the context of location specific applications and the users need to specify under which they are happy to specify the locations and let the other people see it. They also learn a set of default depending on the locations of the data and user.

Mazzia, LeFevre and Adar (2012) introduces a system name PViz, which automatically creates groups and sub groups of friends with whom the user interacts with and shows it to the user for the acceptance and studied on feedback from huge number of people and found out the difficulties is making the user to use a new tool and get used to it.

Along with the above mentioned concepts, there are few more privacy setting systems proposed by many developers. But, all of them propose their rules basing on the social context of the user which are not solely enough for stating rules to an image. Because, images contain much more data than what a user sees in it.

Details on Previous Recommendations Systems

A System named SheepDog, proposed by Chen, Chang, Chang, Tien, Hsu and Wu (2008), which automatically categorizes the images into their respective groups and also adds the tags based on the people in Flickr. It involves concept detection and classification to generate the groups. They have also described two mechanisms to examine the performance of concept detection.

A framework developed by Choudhury, Sundaram, Lin, John and Seligmann (2009), connects the images among various social networks and the user get the right feedback from the

community the image fit in. They categorize image content in three sections namely visual features, user generated text tags and social interaction. And, Yu, Joshi and Luo (2009) proposed a system that automatically generates suggestions to the images uploaded categorizing them with the labels of the groups in which they fit.

On an overall note, there were many previous studies that have been conducted on user privacy and recommendation settings.

Classification

Classification can be defined as the method of classifying the things or objects into particular class according to their characteristics. In data mining terminology, classification can be defined as a function that separates an item in a group to its particular class. The main goal of classification is to correctly predict the class of a particular item in each instance (Data Mining Concepts, 2008).

A classifier classifies the item to its corresponding class. In order to do that, the individual need to train the classifier with huge amount of data and form separate classes for each object.

A training set is used to train the classifier to predict the class of an object and a test set is passed to the classifier to check the accuracy of the prediction. Hence, a good classifier is one which has the highest accuracy. There are different types of algorithms to develop a classifier. Some of them are:

1. Decision Trees
2. Neural networks
3. Bayesian network
4. K-Nearest Neighbor

Decision trees. The basic idea in decision tree algorithm is to recursively partition the given data set into multiple sub groups. It tries to get the solution using tree representation. The process involves placing the main attribute as root and split the training set into corresponding subsets (How Decision Tree Algorithm works, 2017). The record that needs to be classified is compared to each internal nodes of the tree until the leaf nodes are approached and thus predicts the class according to the leaf node. Leaf nodes are the final order in the tree structure.

Consider a scenario where one has to predict a car mileage. A car mileage can be based on different factors such as car's body weight, engine's horse power, acceleration and many others. Considering all the factors, one can judge whether the car gives a good mileage or not. In general, the person might consider the vehicle weight and if it is heavy, it can be stated as the vehicle gives less mileage. Later, if it is not a heavy vehicle, then the comparison goes to horse power, where, if the vehicle has high horse power, then it gives less mileage. Figure 2.1 gives an illustration of a graph of the example scenario explained above through which one can approach to calculate a car's mileage.
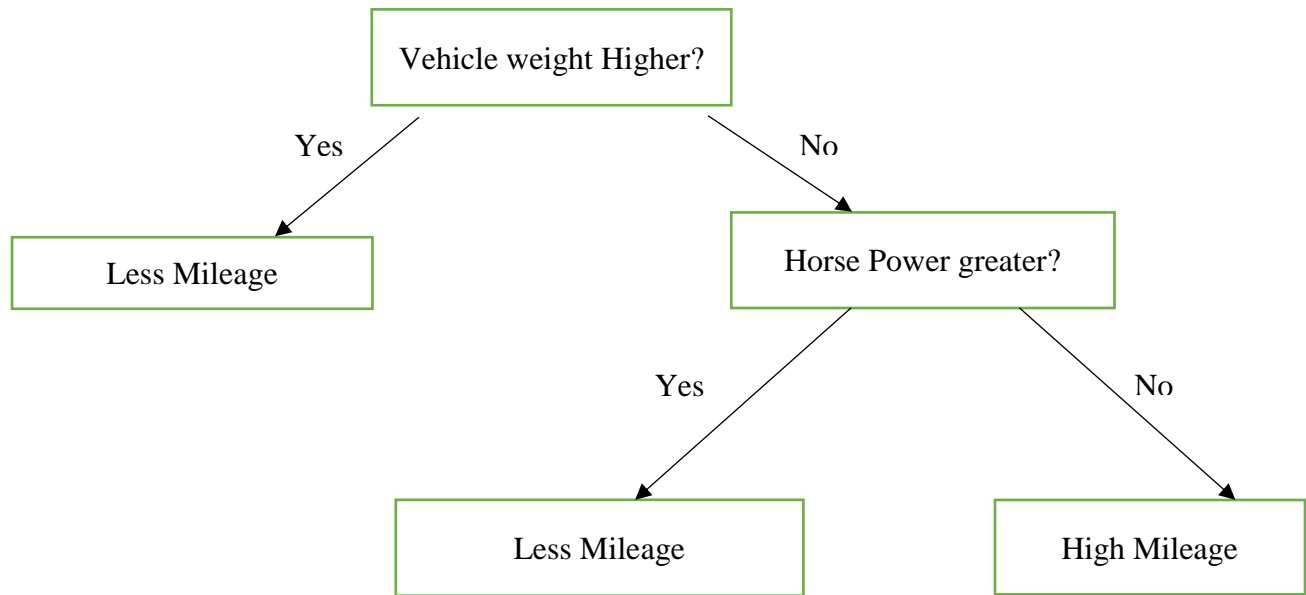
Figure 2.1: Decision Tree Graph

Neural networks. Neural Networks are based on the structure of the human brain and structured as network of neurons. All of the neurons are organized in different layers. The initial layer is known as input layer whereas the next n number of layers are hidden layers, n depending upon the individual. The last layer is known as output layer where different nodes exist with one node corresponding to one class (Han, Pei & Kamber,2006).

Like any other classifier algorithms, neural network also consists of training phase and testing phase in which the classifier is trained with a large data set to the initial nodes and the output layer is assigned with corresponding class nodes.

Each layer has different units and the inputs fed to the network are the attributes measured for each training set. These inputs are sent to the units making up the input layer simultaneously. They pass through each node inside the input layer and a corresponding weight is calculated and fed to the next layer which is known as hidden layer. The outcome of one

hidden layer might be the input of another layer or can correspond to the output layer directly. The number of input layers vary and are subjected to change for every neural network depending upon the user.

The units inside the input layer and the output layer can be called as input units and output units respectively. A neural network with two hidden layers can be called as three-layer neural network and at the same time an output layer can be one or two depending on the user choice. Each layer takes a weighted sum of outputs from the previous layer. The neural network is called feed-forward network, which means none of the weights or the weighted sum are sent back to the previous hidden layer or the previous output layer.

The network takes the testing set and compared with the desired outputs. If there are any errors, the error is then propagated back to the input layer which is often called as back propagation in neural networks. The data set is then undergone several number of repetitions until the desired output is achieved by changing the weights of the hidden layers. The figure 2.2 can be considered as an example for neural network which shows the different layers such as the input layer, a bunch of hidden layers and the output layer.
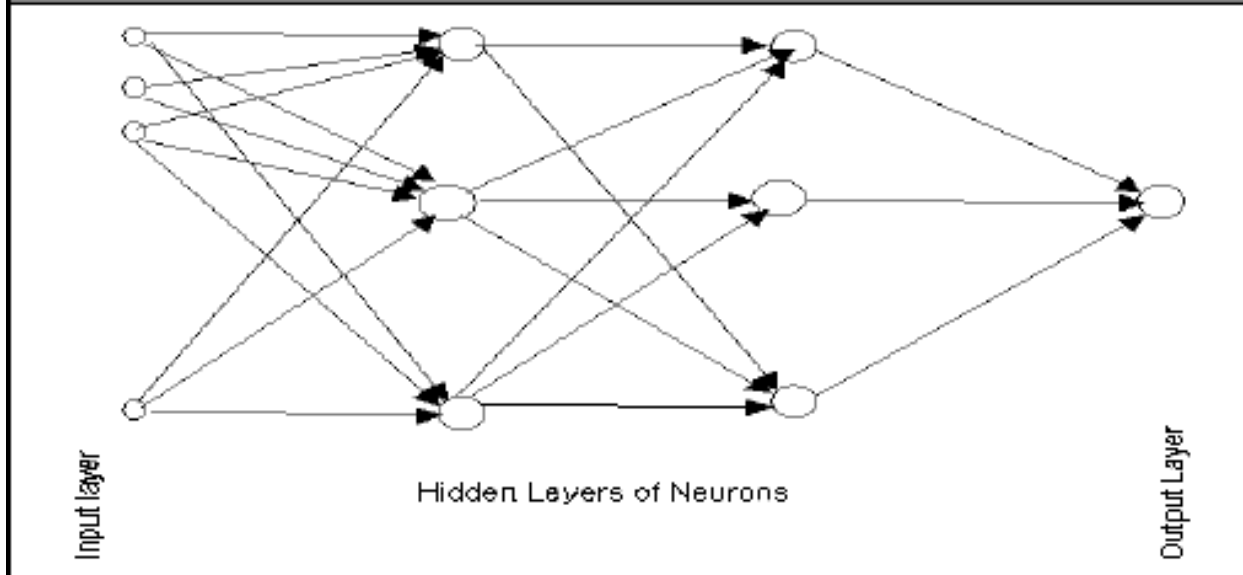
Figure 2.2: Neural Network

Bayesian network. A Bayesian network can be defined as a graph with probabilities containing a set of random variables (Bayesian network, n.d.). Consider a case where there are sprinklers for making lawn grass wet. In general, a rain can water the grass and at the same time, it fills the underground levels with water which later is pumped by sprinkler to water the grass. Figure 2.3 illustrates the scenario of sprinkler and watering the grass in which the variables are sprinkler, rain and grass and the figure shows the probability and the relation between rain, sprinkler and the grass getting wet.

Figure 2.3: Bayesian Network Example

Bayes theorem is used to develop the algorithm and the network consists of an acyclic

graph and a set of probability distributions. The number of possible structures in the network is

huge. Hence structure learning in the Bayesian network has very high computational complexity

(Khanteymoori, Homayounpour and Menhaj, 2008).

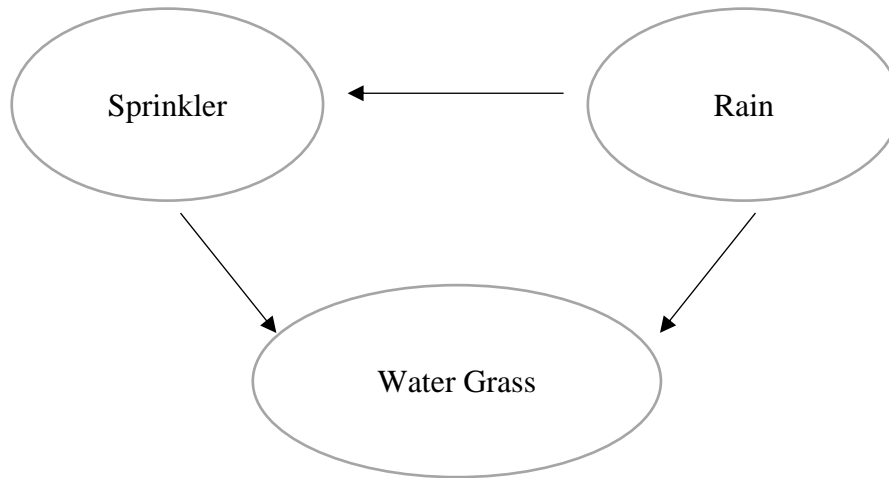**K-Nearest-Neighbor classification.** Classifying an object using single instance based

classifier means locating the closest neighbor and labelling the unknown object as the same as

the known neighbor. This approach can be named as nearest neighbor classification (Söder,

2008). But the disadvantage of the nearest neighbor classification is lacking of robustness in

resulting classifiers. The high value of sensitivity makes the nearest neighbor classifier

vulnerable to the noise in training data. Also if the training set increases in size, it is highly is

error prone.

The robustness of a classifier can be increased by leveling up the number of neighbors,

say k where k>1. The class of an object can be decided basing on the majority vote of the

 neighbors. Such classification can be termed as k-nearest neighbor classification. Thus nearest neighbor classification can be termed as a general instance of k-nearest neighbor classification.

Table 1 describes the differences between K-nearest neighbor classification and neural network classification with pros and cons through which a programmer can choose the best method he or she desires.

Few differences between K-nearest neighbor and neural network classification:

Table 1: Differences between K-Nearest Neighbor and Neural Network

Methods	pros	cons
K-nearest neighbor	Implementation is comparatively easier and training of classifier is faster	Quite sensitive to the noise and also requires huge storage. Testing can also be a bit slow
Neural Network	Ability to handle noisy data and also identification of relationships between independent and dependent variables	Requires high processing time and over-fitting

Ensemble classifier. All the classifiers such as k-nearest neighbor, neural network or Bayesian classifier does a good amount of classification provided good distribution of patterns in training set. If there is no good distribution, the classifiers learn poorly and the testing is difficult. That is, it might not be possible for every individual classifier to learn the patterns. In such

scenario, ensemble classifier which means including a set of classifiers in the initial level and then fusing the result vectors using another method or fusion classifier can relatively give a better accuracy in the classification (Polikar, 2006).

The initial level of classifiers is known as base classifier whose goal is to obtain a single class decision and the next stage of combining the class decisions from the base classifiers is done by the fusion classifier. The group of base classifiers can consist of different classifiers. The fusion classifier can also be any individual classifier of the programmer's interest. This entire method of classification is done by ensemble classifier. There can be different structures in ensemble classifiers. Such as

Parallel structure. In this type of structure, each base classifier is started individually with the same input set and the resulting set of class decision vector is fused using the fusion classifier. Majority of the ensemble classifiers use this structure as there is better opportunity of accuracy.

Hierarchical structure. It can also be referred to as cascading structure, where the classifiers are invoked one after the other in a serial manner. In order to attain accuracy, the inaccurate methods such as having small subsets are invoked first and then the accurate methods are invoked at the last. Figure 2.4 shows two ways of hierarchical structures which are feasible for developing an ensemble classifier, where one is a serial structure in which the output of one classifier is fed into next classifier and so on. The second one approaches a complex relation where output of one classifier is fed to two classifiers and their corresponding outputs are fed to a set of other classifiers and in the end all the class confidence vectors are fused by a fusion classifier to get single output.
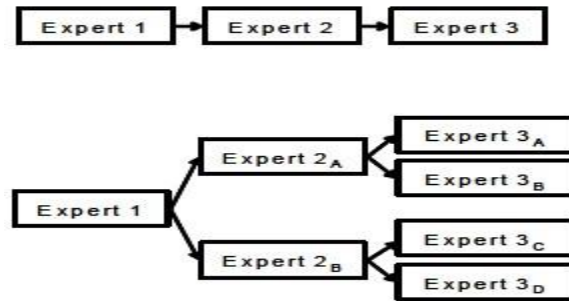
Figure 2.4: Serial Structure

Wavelets

Wavelets are a way to represent and analyze data in the form of mathematical functions. Wavelets are widely used in signal and image processing. Wavelets use in these areas dates to early 19th century with principles like Fourier transform (Wavelet, n.d.).

Scale and frequency play a major role in analyzing data. Scale in here is like Scale in maps where high scales represent a global view or a non-detail view where as low scales represent a detailed view. In-terms of frequency, low frequency (high scales) represent global information of a signal where as high frequency represent the detailed hidden patterns in the signal.

Scaling operation either dilates or compresses a signal. Dilated signals represent the larger scales and compressed signals represent small scales. Signals in the below image represent either dilated or compressed signals. The graph with s value 0.05 has the smallest scale and the s value with 1 has the largest scale. Figure 2.5 shows different scaling frequency representations.

Advantages of wavelets include:

1. Short time and localized waves with zero integral value.

2. Have a possibility of time shifting.
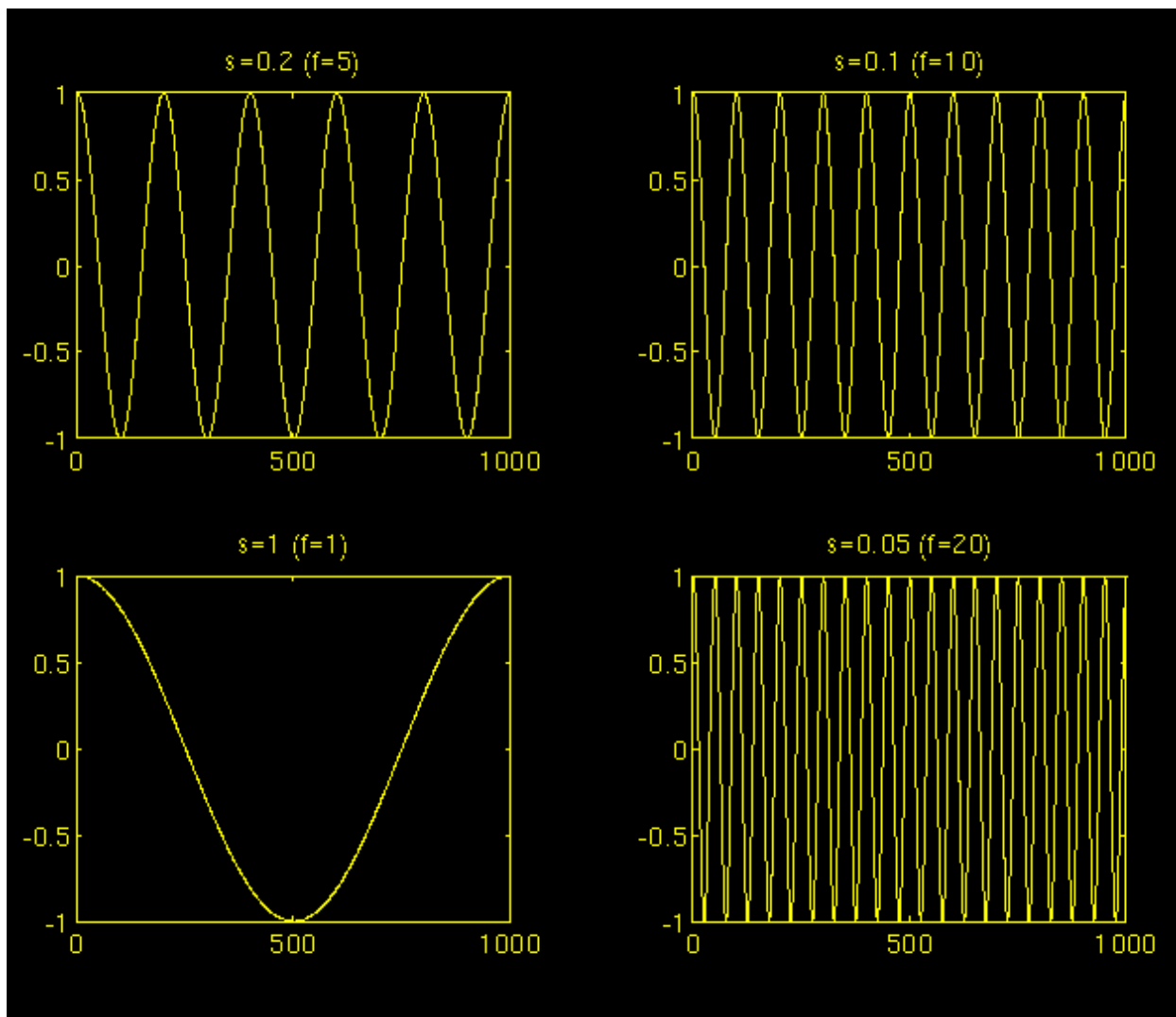
3. Have a greater flexibility.

Figure 2.5: Scale Frequency Representation of Wavelet

Figure 2.6 shows graphical representation of some of the popular wavelets included and explained in later in the section such as Haar wavelet transform, Daubechies, Daubechies 4, Gaussian, coiflet and others.
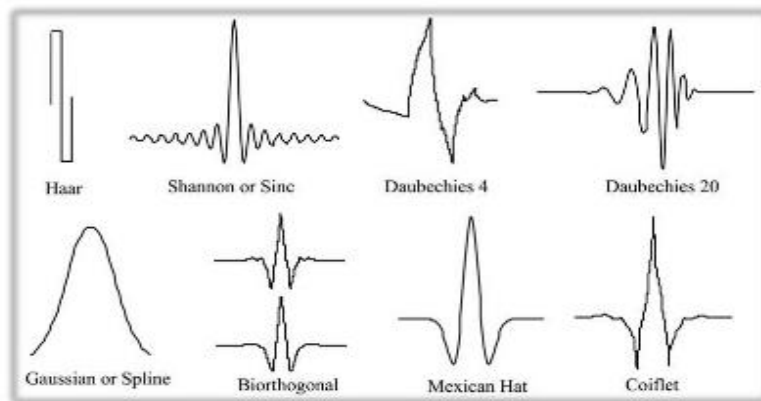

Figure 2.6: Different Wavelets

Wavelet transform. Wavelet transform is a mathematical representation of wavelets by a function. Wavelet transform has the scaling in its denominator which means the opposite of the above defined values of scale(s) work and that scales with s value greater than 1 dilates the signals and scales with s value less than 1 compresses the signals. This representation is used in the below explanations.

Discrete Wavelet Transform (DWT). Wavelets are used to refer to a set of orthonormal basis functions which are generated by dilation and translation of scaling function φ and mother wavelet ψ (Sahu & Choubey, 2013). In case wavelets are discretely sampled in a wavelet transform then it can be termed as Discrete Wavelet Transform or DWT. Frequency and location are both captured in DWT.

Discrete Wavelet Transform is a finite scale multiresolution representation of discrete function. DWT is invertible as well as orthogonal and is a fast-linear operation on a data vector

of length integer power 2. Inverse transform can be represented as a matrix and is a transpose of

the transform matrix (Difference between Fourier, n.d.).

Wavelet functions are more localized in space than the sine and cosine functions in

Fourier Transform but are localized in frequency.

Wavelet function can be given as
$$\psi_{(j,k)}(x) = 2^{j/2}\psi(2^j x - k)$$

And the scaling function is given as
$$\phi_{(j,k)}(x) = 2^{j/2}\phi(2^j x - k)$$

Where ψ is called as wavelet function and j and k variables are integers. These integers

scale and dilate the wavelet function. Factor 'j' represents the wavelet's width and can be called

scale index and 'k' represents the position. Dilation of wavelet is performed in the power of

number two and is translated by 'k' value.

Wavelet equation can be defined as below in terms of wavelet coefficients as

$$\psi(x) = \sum_{k}^{N-1} g_k\sqrt{2\phi(2x - k)}$$
where g0, g1, g2 are called the high pass wavelet

coefficients. Scaling function can also be defined as

$$\phi(x) = \sum_{k}^{N-1} h_k\sqrt{2\phi(2x - k)}$$
. Function $\varphi(x)$is defined as the scaling function

and h0, h1, h2.... hk are low pass scaling coefficients.

The wavelet and scaling coefficients have a relationship between them and is called quadrature mirror relationship and is shown as gn= (−1) h1-n+N. The factor N represents the number of vanishing moments.

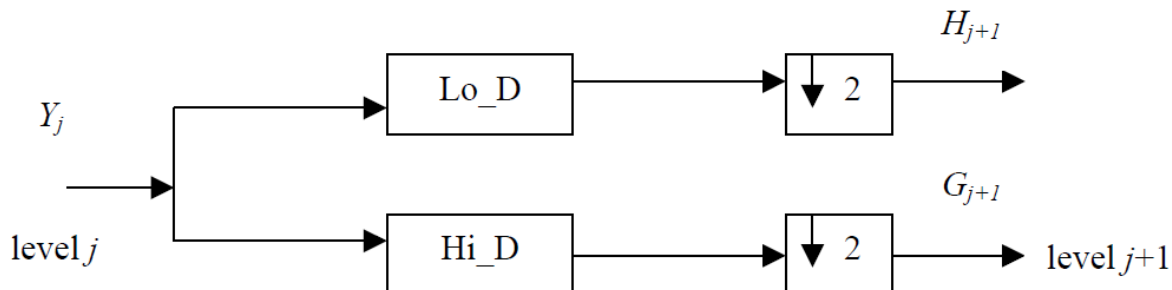Below figure 2.7 shows the graphical representation of DWT:

Figure 2.7: Decomposition of a One-Dimensional DWT

Many different wavelet families can be derived from wavelet equation such as Daubechies, Haar and so on. Number of vanishing moments N is used to classify wavelets into a wavelets family which in turn have wavelet subclasses which are distinguished by number of coefficients and level of iterations (Can anybody compare Haar, 2014). Table 2 describes the filter lengths and number of vanishing moments for some of the wavelet families which help in distinguishing the wavelets.

Table 2: Properties of Different Wavelet Families

Wavelet Family	Filter Length	Number of Vanishing Moments, N
Haar	2	1
Daubechies M	2M	M
Coiflets M	6M	2M-1
Symlets	2M	M

Wavelets multi-resolution processing. Resolution can be defined as number of bits per unit area or number of pixels used to represent an image (Wavelet, n.d.).

In general terms, resolution can refer to clarity and sharpness of an image. In day-to-day world, monitors, printers and many display goods like phone screens use the term resolution more often. In monitors, the number of dots or pixels on the entire screen represent the screen resolution for that monitor. For example, 1280 by 800-pixel screen implies the monitor screen can display 12800 dots on each of 800 lines.

Image resolution. As mentioned above, the number of pixels can be referred to resolution. Image resolution means the number of pixels inside an image and can be identified by the height and width of the image. If an image is said to be 2048 by 1536, then it contains 3,145,728 pixels or 3.1 Megapixel (Image Resolution, n.d.). In recent times, each mobile phone is distinguished by its camera resolution which means the number of pixels the camera can capture with in an image. The larger the number of pixels, the larger the quality.

An image is divided into four sub bands when a wavelet transform is applied on it. The four categories can be stated as LL, HL, LH, HH where LL means low-low, HL means High-Low, LH implies Low-High and HH implies High-High (Kaur & Kaur,2013).

LL implies Lowest frequencies which is an approximation and HL implies horizontal high frequencies which correspond to vertical edges and HH implies diagonal details which are high frequencies in all the corners. Finally, LH implies vertical high frequencies corresponding to the horizontal edges.

As Multi Resolution processing involves representing the signals of more than one resolution, features that might go undetected at on resolution can be easily spotted at another resolution. There are different methods to represent the image in more than one resolution. The following section describes various wavelet transforms where the representation of image in multi resolution can be done.

There are different wavelet transforms such as Daubechies wavelet transform, Haar Wavelet Transform, Image pyramids and so on. Some of them are

Daubechies wavelet. Based on the work of Ingrid Daubechies, this wavelet transform has been named as Daubechies Wavelet Transform. Daubechies wavelets are categorized into D1-D20 where the index number implies the number of coefficients. Each wavelet has the vanishing moments equal to the half of the number of its index which means D2 has one vanishing moments, D4 has two, D6 has three vanishing moments etc. Among all the twenty Daubechies wavelets, D1-D10 are most commonly used and let us see an example of D4.

The D4 wavelet transform has four scaling coefficients and in each step of the transform, the scaling function is applied to the data input. Smoothed values are calculated using the original data set values and the scaling functions are applied to each data set to get the smoothed

values half the number of original data set value. The scaling function coefficients can be

calculated as:

$$h_0 = \frac{1 + \sqrt{3}}{4\sqrt{2}}$$

$$h_1 = \frac{3 + \sqrt{3}}{4\sqrt{2}}$$

$$h_2 = \frac{3 - \sqrt{3}}{4\sqrt{2}}$$

$$h_3 = \frac{1 - \sqrt{3}}{4\sqrt{2}}$$

The wavelet function coefficients are assigned the scaling function coefficients in the the

decreasing order such as $g_0 = h_3$, $g_1 = -h_2$, $g_2 = h_1$, $g_3 = -h_0$ (The Daubechies Wavelet Transform, n.d.).

Calculating the scaling function:

It is calculated by taking the inner product of the coefficients and four data values.

Following figure shows the equation to calculate the D4 scaling function and D4 wavelet

function.

Daubechies D4 scaling function:

$$a_i = h_0 s_{2i} + h_1 s_{2i+1} + h_2 s_{2i+2} + h_3 s_{2i+3}$$
$$a[i] = h_0 s[2i] + h_1 s[2i+1] + h_2 s[2i+2] + h_3 s[2i+3];$$

Daubechies D4 wavelet function:

$$c_i = g_0 s_{2i} + g_1 s_{2i+1} + g_2 s_{2i+2} + g_3 s_{2i+3}$$
$$c[i] = g_0 s[2i] + g_1 s[2i+1] + g_2 s[2i+2] + g_3 s[2i+3];$$

Pyramid. In image processing the images are arranged in the shape of a pyramid which represents an image at different resolution. The base of the pyramid has the high-resolution image and the lower resolution images are arranged on the top in the form of a pyramid. Thus, it is named as pyramid resolution technique (Pyramid, 2017). It is developed by different techniques where image is smoothed repeatedly. It can also be referred to a type of multi-scale signal representation.

There are two types of pyramids which are lowpass and bandpass. Lowpass pyramid is made by smoothing and subsampling the smoothed image along each direction. The process is repeated multiple times until a smoothed image with smaller size is achieved. Highpass pyramid is achieved by calculating the difference between images and performing image interpolation between adjacent images in a pyramid.

Typically, for the pyramid generation in lowpass, the smoothing method used will be the binomial filter and applied typically twice and then subsample the image by the factor of two. There are different types of pyramids based on the generation such as Gaussian pyramid, Steerable Pyramid, Laplacian Pyramid etc.

The figure 2.8 describes an image pyramid where the base image has the highest resolution and the lower resolution images which are smoother are placed on the top of the base image and also describes the algorithm for generalization for smoothing of the images so that the images can be placed in the shape of a pyramid.
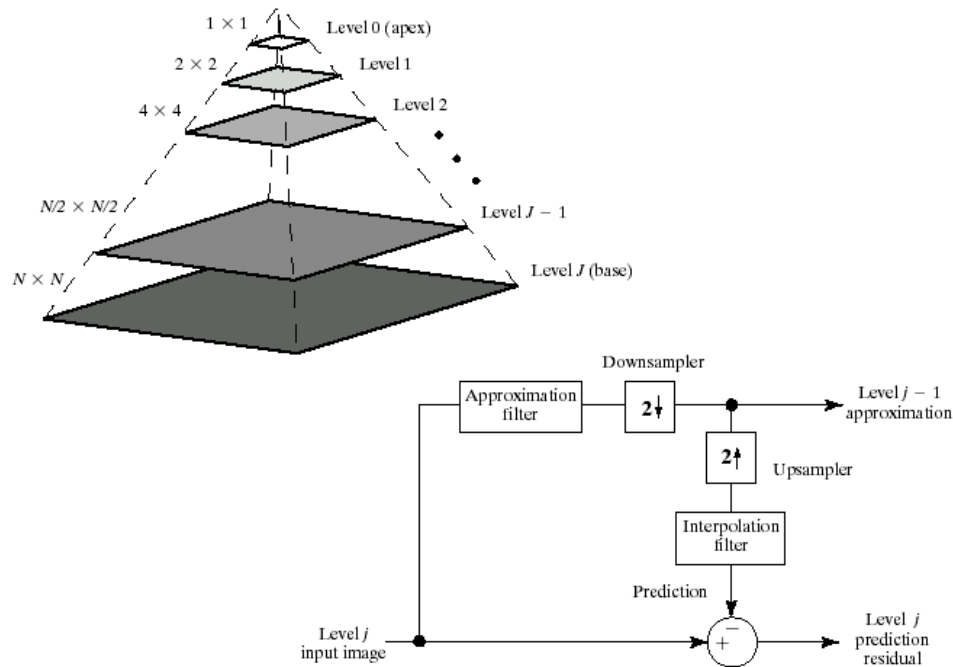
Figure 2.8: Image Pyramid and the Generalization for Smoothing

Existing System

Social Networking sites has become a part and parcel of the human being. Almost all the people have social media accounts and share the images of everything happening in their life.

Almost all the social networking sites offer to set the privacy settings for the profile but still there are many cases where the information is misused. Moreover, in general people find the process of setting the privacy settings for each and every image as tedious and there is a risk of some images sliding out of box.

Proposed System

In the world where time is as precious as life, setting up privacy settings for each and every image might not be the ultimate goal of a user, thus having the possibility of information misuse. In such case, there is a necessity for the development of privacy policy prediction system in

which the system proposes a certain policy that can be used as-is or change them according to the user requirements.

This system also takes the social context of the user in order to propose the privacy setting in order to achieve higher accuracy for the prediction. Thus, the primary advantage of the proposed system is to reduce the time consuming process of applying privacy settings for each and every image the user has uploaded and also reducing the misuse of the information share on these social networking sites.

Chapter 3: Methodology

Overview of A3P System

A3P is used to predict the settings for an image uploaded by the user basing upon the user's social context and also the image content.A3P system can be categorized into two main components name: A3P-Core and A3P-Social (Squicciarini et al.,2015).

A3P-Core classifies the images using the content of the image, image's metadata using data mining approach named as classification and predicts policies depending on the user behavior using policy mining.

A3P-Social defines settings by analyzing the type of the image. And if the current image is similar to an existing or already uploaded image by the user, it tries to impose the same existing rules. Else it generates rules/settings depending on the image data and location data.
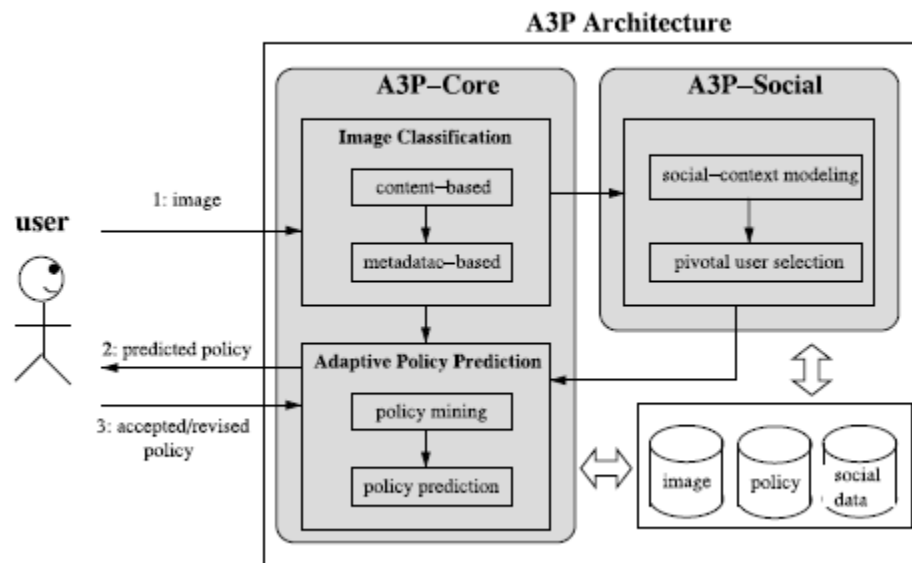

Figure 3.1: A3P System Architecture

Figure 3.1 shows the architecture of the A3P system, where a user interacts with the system which has two modules in order to predict the policy. The first module is named as A3P core which has the methods such as image classification and policy prediction and the second module is named as A3P social where methods such as user's social context modelling and selection of the user's groups are done.

A3P Core Analysis

A3P-Core can in turn be classified into two major segments. Namely, Meta data based and content based image classification segment and Adaptive Policy prediction segment.

Content based classification is carried out by using wavelet transform and by deriving the signature from the transform, it is compared to the existing signatures and included in the most suitable sub category.

There are many wavelet transforms that can be used in order to get the signatures. One of them would be Haar Wavelet transform which is most suitable.

Haar Wavelet Transform

Haar wavelet transform is the first and simplest discrete wavelet transform which divides the wavelet on orthonormal wavelet basis. It was proposed by Alfred Haar in 1909 (Haar wavelet, n.d.). Haar Wavelet transform is conceptually simple and reversible without any edge effects of other wavelets. It is also memory efficient and computationally cheap. It reflects the changes among adjacent pixel pairs and does not have overlapping windows.

The mother Haar Wavelet transform can be represented as:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Haar wavelet function $\psi_{n,k}$ can defined on the real line by the formula:

$$\psi_{n,k}(t) = 2^{n/2}\psi(2^n t - k), \quad t \in \mathbf{R}.$$

Haar wavelet can be described as a sequence of rescaled "square-shaped" function which form a wavelet family. The wavelet allows target function in orthonormal basis over an interval. The Haar sequence is the most recognized and extensively used over the world.

Haar matrix. The 2 by 2 Haar matrix which is used in Haar Wavelet is

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

When a given sequence has length equals to multiples of four, 4 element blocks can be built and transformed similar to a 2X2 matrix using the below specified 4X4 matrix and in this, two stages of Haar wavelet transform are combined.

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

In a similar manner, below specified is an unnormalized eight point Haar Matrix and this has to be normalized before being used in a transformation process.

$$H_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Haar wavelet transform properties. The main properties of Haar Wavelet are:

1. Real Continuous functions can be uniformly approximated by linear combinations of $\varphi(t)$, $\varphi(2t)$, $\varphi(4t)$…, $\varphi(2^n t)$ and also includes their shifted functions too.

2. And also any real continuous functions on [0, 1] can be uniformly approximated on [0, 1] using linear combinations of functions 1, $\varphi(t)$, $\varphi(2t)$, $\varphi(4t)$ …, $\varphi(2^n t)$ including their shifted functions.

3. Unlike Fourier transform, Haar matrix only has real elements and is asymmetric.

4. Haar wavelet transform works using additions and not multiplications. Since many values in the Haar matrix are zeros, the transform computational times are really short.

5. Input and output will be of same length and only in the powers of 2.

6. Haar function can be used to analyze the frequency components of the input signals with the help of its orthogonal property (Haar wavelet, n.d.).

Using the Haar wavelet transform, anyone can transform sequence of even length into a sequence of two component vectors. In order to get the original sequence, one has to right-

multiply each vector with the matrix. Haar wavelet transform can be applied over all dimension images in order to compress, derive a signature etc.

Differences of DWT and CWT. Continuous wavelet transform differs from a discrete wavelet transform in the below stated aspects:

1. When a Signal is cross correlated with your wavelet, the CWT and DWT differs in the number of samples skipped.

2. When a wavelet is dilated, CWT and DWT differs in the number of samples skipped.

3. DWT uses both a Wavelet and a Scale-let where as a CWT uses only the wavelets.

4. Scale is discretized more finely and it uses scales which are less than 2 in value, whereas Scales are not as discretized and are in the base powers of 2.

Use of DWT. Image quality after all the processes without any significant data loss is what is required as a result from any transformation processes performed. In subject based corrective solutions as in this case, humans are the image uploaders and the end users. Any data loss from the image would result in wrong information to be read and a wrong set of rules set for the images uploaded. To avoid this, use discrete wavelet transformation which uses both scale and frequency to determine the image information.

Metadata based classification segment generates/groups the image data basing on the keywords, tags, and comments of the image. And this process helps in finding the sub-category in which the user image fits in.

Adaptive policy prediction deals with the user privacy concerns by practicing Policy Mining and Policy prediction. Let us take a look into policy mining.

Policy Mining

Data mining is the process of discovering the patterns between the data objects (Data mining, n.d.). There are many mining techniques which help in getting the desired pattern. Some of them are:

1. Association rule mining

2. Anomaly or outlier detection

3. Clustering analysis

4. Regression Analysis

Let us take a deeper dive into each of the techniques which help to identify the matching group or policy.

Association rule mining. It is a data mining technique which is a process of finding the associations, relations or correlations between different kinds of data basing upon the relations between particular kind of data that is stored inside the databases such as relational databases, repositories (What is Association Rule Mining, n.d.).

In a given instance, it helps to find the occurrence of the data based on the occurrence of other data in the same instance. For example, consider a shopping mall where a customer bought both milk and eggs, then from the association rule mining, one can say that, he will buy bread too. In general terminology, association rule mining can be termed as set of 'if' and 'then' to discover the relationships among the objects.

Clustering analysis. Organizing of objects who have similar characteristics into same groups can be referred to as clustering. So, a cluster can be stated as collection of objects that are similar to objects inside but different to objects in other clusters (A Tutorial on Clustering Algorithms, n.d.).

In the figure 3.2, there are four different sets of unlabeled data which can be any type on the left side. Each data object is grouped into similar cluster and the right side of the figure shows the clusters of similar objects. The similarity criterion can be distance between the objects or any conceptual criteria among the objects such as shape or color etc.
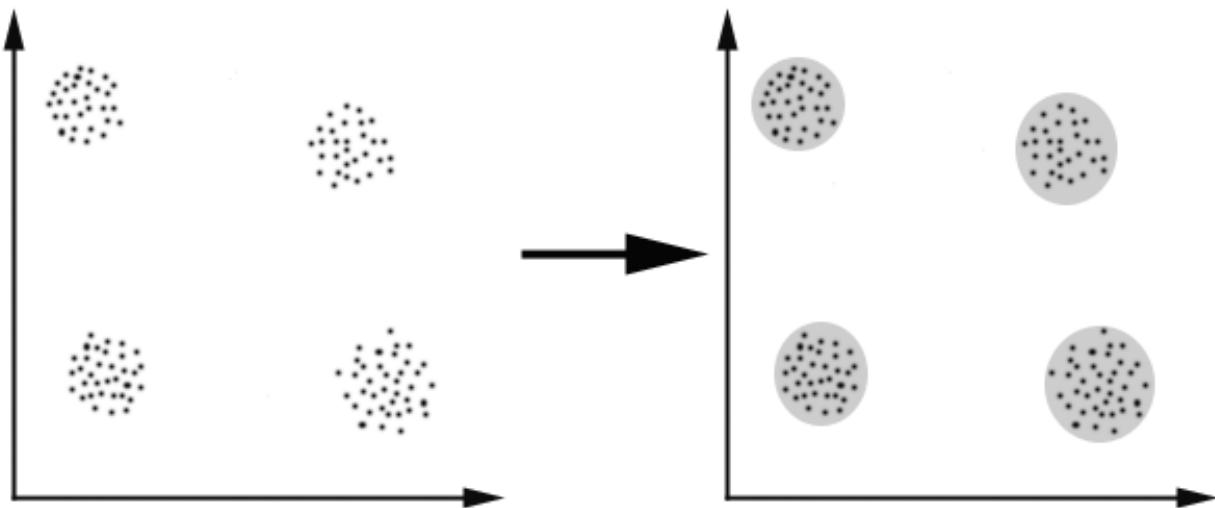
Figure 3.2 Clustering Analysis

Main goal of clustering is to find the similarities between the unlabeled data and group them into different clusters. The criteria to group the objects into clusters depend on the user's perspective in such a way that the resulted groups should meet his or her needs. The objects which doesn't belong any cluster are stated as outliers. There are many real world applications for clustering such as finding the groups of similar books in a library which makes the organizing of books a way lot easier. In marketing, where finding customers with similar interests which helps the companies to promote the products to the specific group of people. Even in biology, to organize the plants and animals of similar kind and differentiate them.

There are many types of algorithms used for clustering such as Exclusive clustering, overlapping clustering, hierarchical clustering, probabilistic clustering. One of the major difference in each technique is the type of data used as initial data for clustering.

Anomaly detection. Any object which cannot be compared with certain behavior or which doesn't include in any of the grouped objects which possess the expected behavior can be defined as an anomaly or an outlier [30] (Choudhary, P., 2017) The technique used to identify the anomalies can be referred to as anomaly detection or outlier detection. Anomalies or outliers can be grouped as:

Point Anomalies: Any single instance of object or data which is far away from the grouped objects or expected behavior can be stated as point anomalies.

Contextual Anomalies: It is the difference in any context such as day to day activities.

Collective Anomalies: Group of instances that help in detecting the anomalies can be referred to collective anomalies.

A3P Social Analysis

A3P-Social considers different criteria to deal with privacy policy prediction. It tries to deliver policies by considering mostly the user's social environment as well as user's attitude towards privacy of his/her images.

A3P-Social's modelling of social context states that user's with similar background make similar decisions on their data privacy. So depending on this statement, algorithms with many users social context are developed. Data for the study of user social context has been gathered considering user's age, ethnicity, location, likes and dislikes, past settings of the user, groups which they share, social events they participate (Squicciarini et al.,2015).

Identifying the social group of the user in which the user image can be shared follows a recommendation process. If a user has uploaded an image, the A3P-Core invokes A3P-Social in-order to find the best social group in which the image can be shared without any doubt depending on the user's previous settings and the data that the image is carrying. The keywords thus derived can be of help for the system to generate the social groups for the user.

Chapter 4: Software Environment

Java

Java is a high level programming language developed by sun micro systems in the year 1995 (Java: A Beginner's Guide, n.d.). There are many versions since its release and currently every company/application is tending to move to the newest version java 8 from 7. Java 7 is most predominantly used in all over the world.

Java is object oriented and everything is wrapped inside classes which gives more security. Main Object-oriented concepts of java are inheritance, polymorphism, abstraction and encapsulation. Implementation in this paper uses Java because of its easy to use and also easier for debugging.

Some of the advantages of java are:

1. Java is object oriented, which gives the programmer the advantage of reusing the code and code modularity.

2. Java is platform independent. Which indicates that the code can be written in any platform and then used in any other platform.

3. Java is easy to learn. It is designed to use easily and compile and debug.

4. Java is robust. Robust means reliability. The errors are shown up while compiling rather than at the time of execution.

5. The other main feature of java is multithreading. Several tasks can be performed simultaneously which is lack in many other programming languages.

6. One of the other feature is its security. All the compiler, interpreter and each integrated part are designed with security as main motto.

7. Java is distributed. It supports TCP, UDP and basic socket communications and pretty flexible to access objects around the internet.

These can be specified as java's major advantages over other programming languages. As a coin has two sides, Java also has its disadvantages. Object oriented as being very helpful, it can be tedious and time consuming in real time. Being it as the main reason, many developers are tending towards new technologies which are not object oriented such as JavaScript. But, its advantages overshadow the disadvantages and most of the people are still willing to use and upgrade for newer versions.

MySQL

SQL stands for structured query language (Margaret, n.d.). It is an open source database management system copyrighted by Oracle corporation. Most of the web based applications use it because of its reliability, performance and ease of use. In order to support the industries with low cost and feasibility, it is available with Oracle MySQL cloud service. The current application has MySQL server as the database and MySQL query browser which has a GUI for easy execution and result of the queries carried out while creating the data. It also helps database administrators to visually create, model, design the databases.

Apache Tomcat

It is also an open source web server developed by Apache group. Under the collaboration with sun microsystems, java server pages use the apache servlet container as reference for implementation. It is a web server and supports the web applications designed by the programmers in the real world. One has to download the tomcat software from the official site and later need to configure the class path and then specify the port on which the web applications

run. General ports are 8080, 8888 and others. It contains web container named Catalina and bin

directory. It helps in creating the response methods for POST, GET, PUT, DELETE requests.

Responses can be checked through a tool named postman through which only the

response objects can be seen rather than the GUI in web pages. The current version of tomcat is

version 8.

JDBC

JDBC stands for java database connectivity (Java JDBC API, n.d.). It is the SQL database

access system created by sun microsystems through which programmer can connect the

application to database. The JDBC objectives are critical and are of high importance.

Chapter 5: Implementation

Installing Java

Firstly, Java Development Kit (JDK) package has to be downloaded from Oracle which is an open source and can be downloaded for free. Then, install the downloaded JDK package on the system the programmer intends to develop the applications.

Java Runtime Environment is needed to run any program written in Java programming language. One can download JRE by the following commands.

Sudo get-update

Sudo apt-get install openjdk-7-jre

One can check the java version through command line by 'java-version' command. The latest is java 8 (How Do I Install Java?, 2017).

Installing Apache Tomcat

There are many different ways to install Apache tomcat in your corresponding operating system. A detailed description is given in the apache tomcat documentation. First, download the corresponding tomcat software from the apache tomcat site which is also an open source. Later unzip the folder and move the folder to appropriate place. Next, add home directory to the path and configure the tomcat user roles by editing the users.xml file inside the tomcat bin folder.

Sample users.xml file:

```
<tomcat-users>
<role rolename= "manager"/>
<role rolename= "admin"/>
<username= "<username>" password= "<password>"
roles= "manager, admin"/>
<tomcat-users>
```

User's name and password and his corresponding roles are assigned under <user> tag. Role tag

specifies the various list of roles.

By Default, the tomcat server runs on port 8080, but can be configured on other ports by changing the port number in server.xml file in .conf file.

There are two ways of starting up the server. One can directly go through the $catalina_home/bin and double click the startup.sh file or can manually start the tomcat from the services in windows. The manager role and admin role in the roles help to access the manager web application and host manager web application.

Access to Apache Tomcat. The localhost path that has been set is used to access the tomcat homepage where the application is deployed. Figure 5.1 shows the apache tomcat home page when the user hits the localhost, where he or she have the options to select different modules such as tomcat manager, documentation and so on.
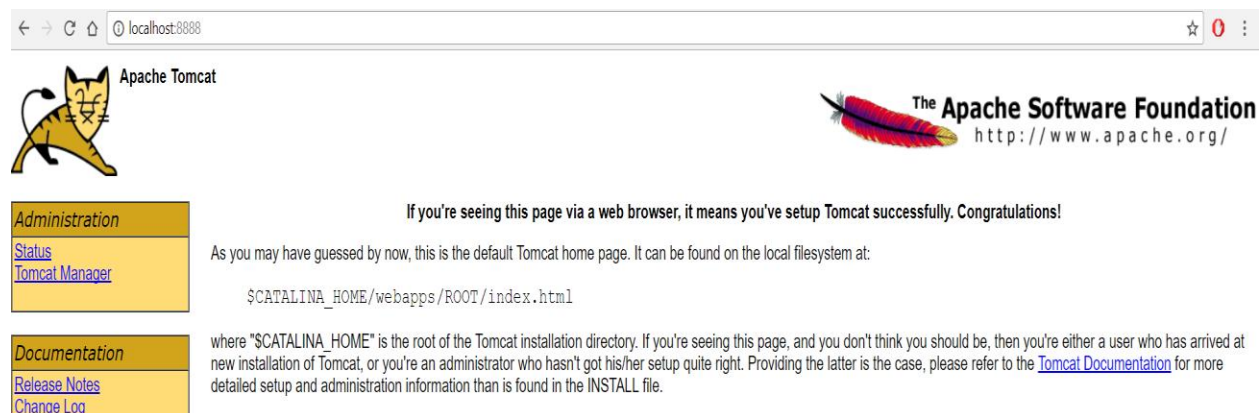
Figure 5.1: Tomcat Home

The option Manager App allows the programmer to deploy the application and installing. Login credentials for the Manager App are set when configuring the tomcat-users.xml and the same are

to be used in order to access the server as an admin. Fig 5.2 shows the user logging in to the

manager app to access the application by the credentials he or she has set up while configuring

the users.xml file.
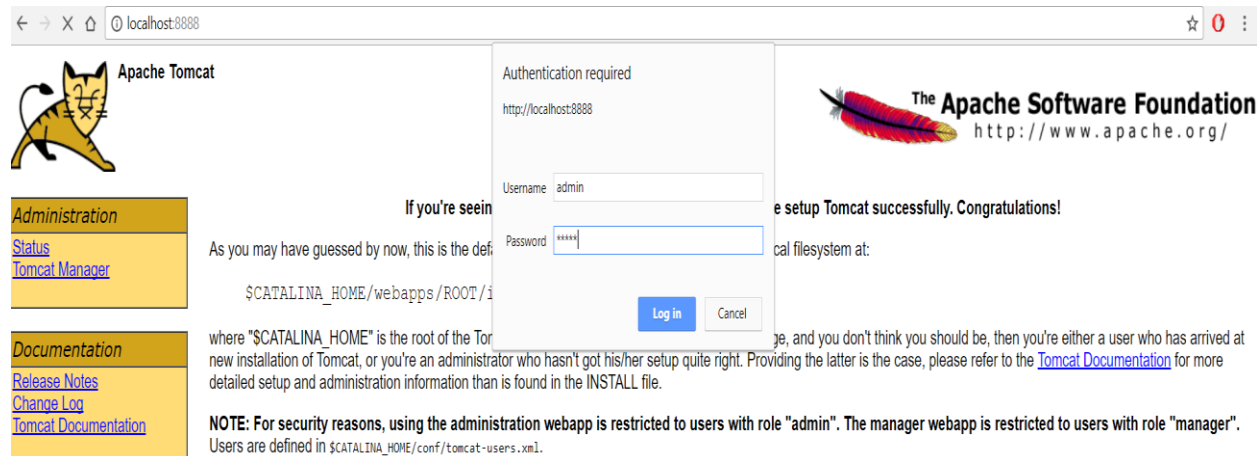
Figure 5.2: Tomcat Manager Login

The source code of the application can be deployed using a WAR file which can be

chosen from the tomcat application by choosing the .war file. After successful deployment, the

application can be started by clicking on the /folder-name, folder-name being the name given to

our application folder while deploying. Figure 5.3 shows the user accessing the application in

manager app and starting the application in a web browser.

Figure 5.3: Tomcat Manager Home

Installing MySQL Server

MySQL server can be easily installed through downloading the installer provided by the MySQL community. It is also an open source and can be customized while installing. The user needs to set a username and password while installation so that the server can be accessed only through the credentials. The server will be started automatically after installation. User can connect to the server through tools like MySQL workbench and perform operations. Tools like workbench provides a rich GUI to perform the queries to the database.

Application and User Interface

The implemented application has a rich GUI connected to the database at the back end where users created can connect with each other like many social networking sites like Facebook, Instagram, Twitter etc. Users connected can share the images of their interest to other connected friends along with settings. There are mainly two modules in the implementation. They are:

1. Admin

2. User

After starting the apache tomcat server, the localhost address directs the user to the home page. Figures 5.4 and Figure 5.5 show the home page and general about page which describes the application.
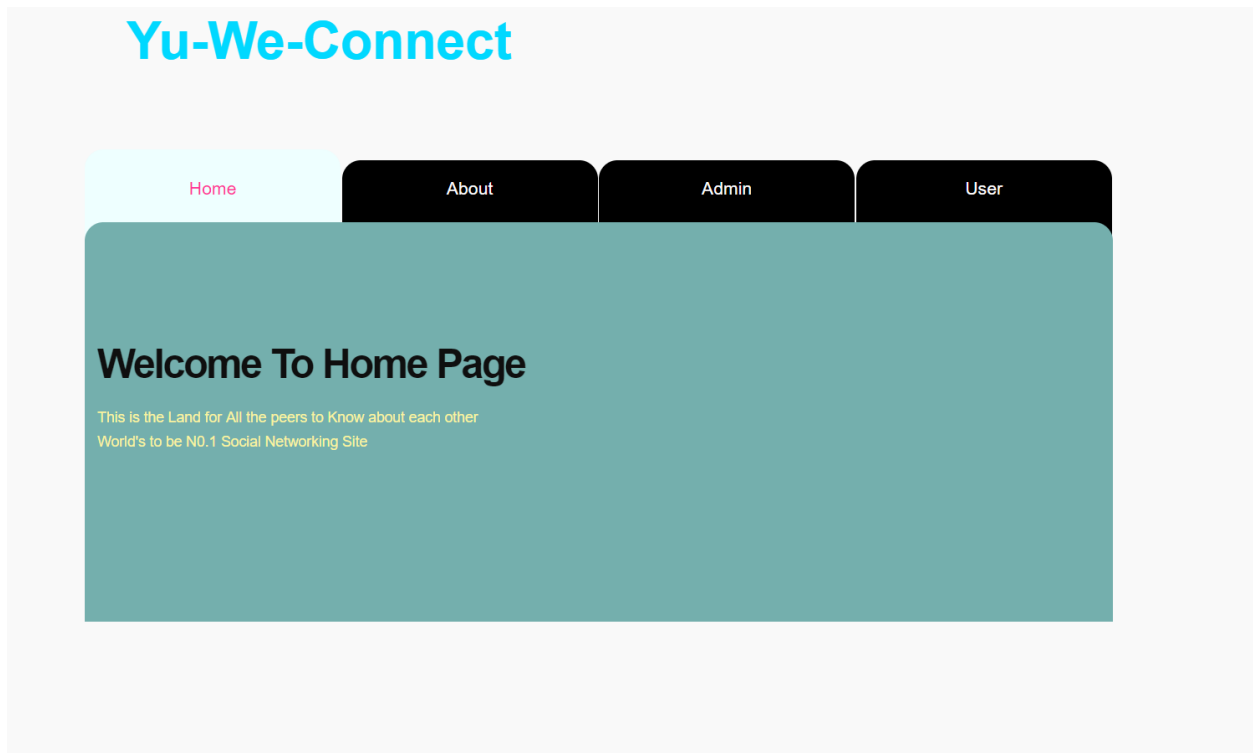
Figure 5.4: Application Home Page
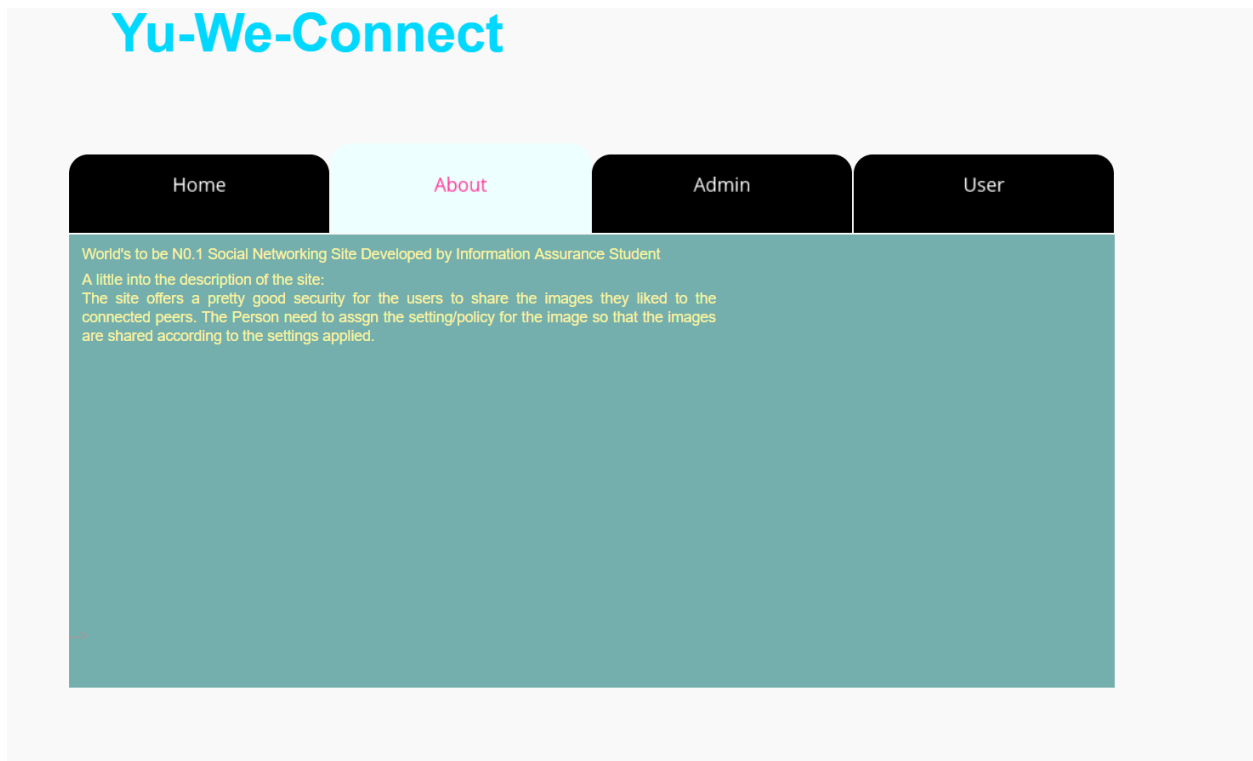
Figure 5.5: About Page

Admin role. Admin is generally a person who monitors the issue happening inside the system. Admin has all the rights to modify or delete or change any information or image from the system if it is found to be non-compliant or if the image or information is disliked by the people. Figures 5.6 and 5.7 shows the login page for the admin where he or she needs to provide the correct credentials in order to access as admin. If one enters wrong credentials for consecutively five times or forget the credentials, the credentials needed to be changed from database which involves the database access.
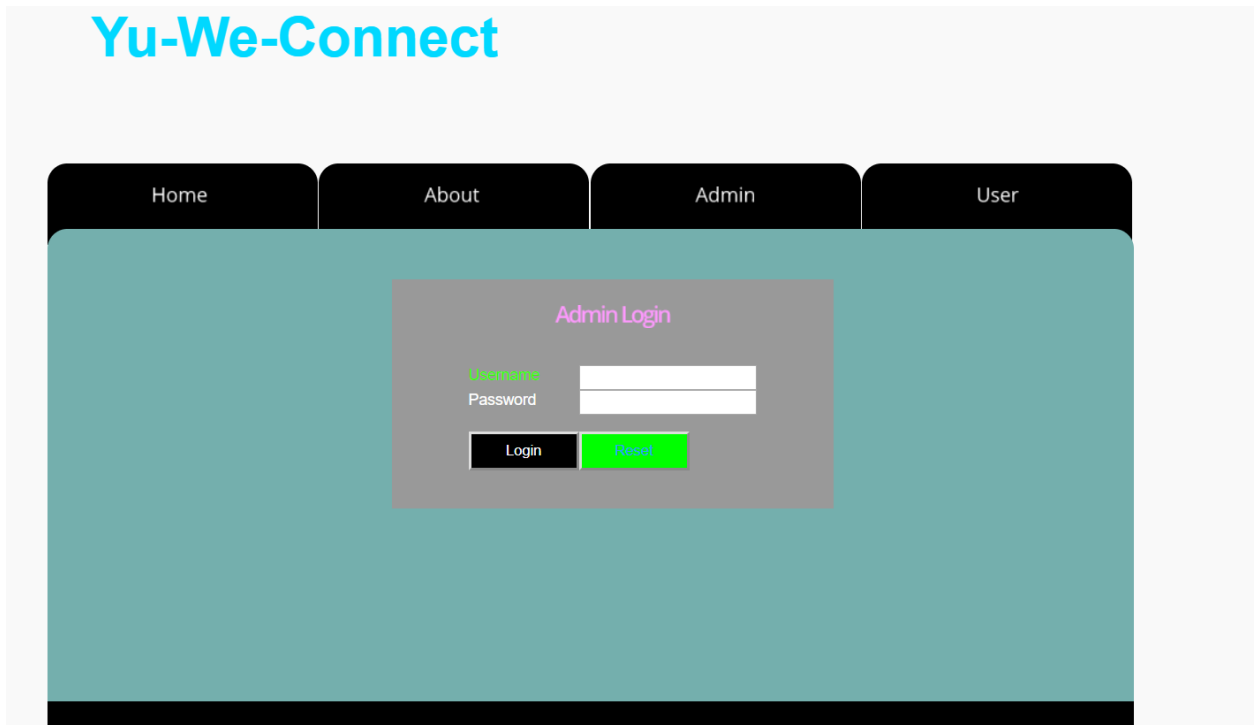
Figure 5.6: Admin Login Page
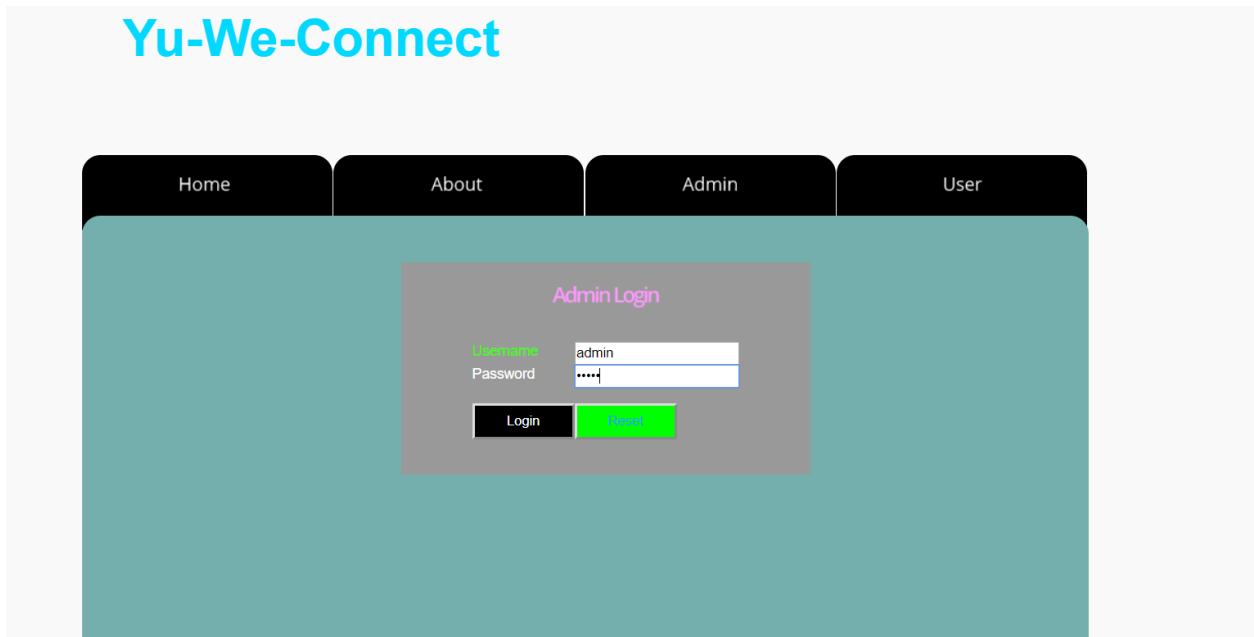
Figure 5.7: Admin Logging with Credentials

After the admin has successfully logged in with the credentials, the home page will be shown for

the admin where he or she have all the options. Figure 5.8 shows the admin home page.
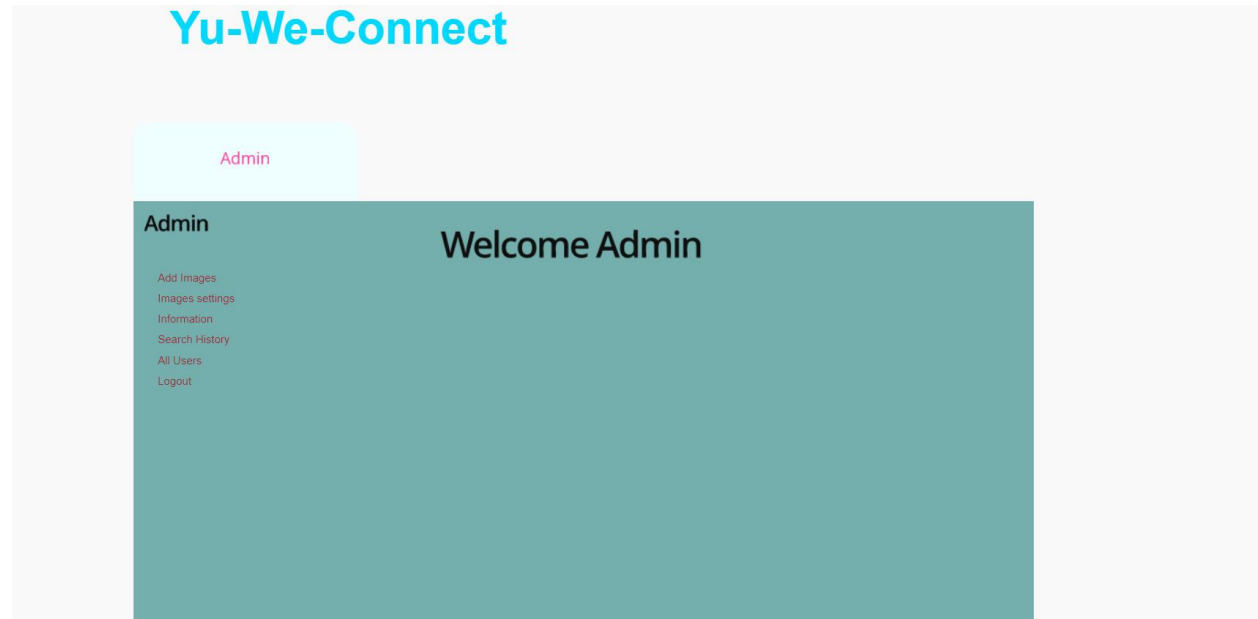
Figure 5.8: Admin Home Page

The main role of the admin is to permit the user to create his/her profile and then the user can

access other user's profiles and send request to the peers. Figure 5.9 shows the view of admin

where he or she can view all the users who are requesting for profile creation inside the system.

The admin might accept or reject the user basing upon his or her interest.

Figure 5.9: Admin Accepting the User

In addition to that, an admin can add a specific image along with his/her own settings and later

share the images to the users with his/her interest. Figure 5.10 shows the view of the admin to

add an image to the system where after he or she can share the image to other users of the

system.

Figure 5.10: Admin Add Image

The admin has the facility of viewing all the images that are shared by the users and can later delete if any of the pics are irrelevant or inappropriate and also share those with other people he wants. Thus, the admin can have an eye towards what was happening among the users and can control any explicit content to make the site better for people of all the ages.
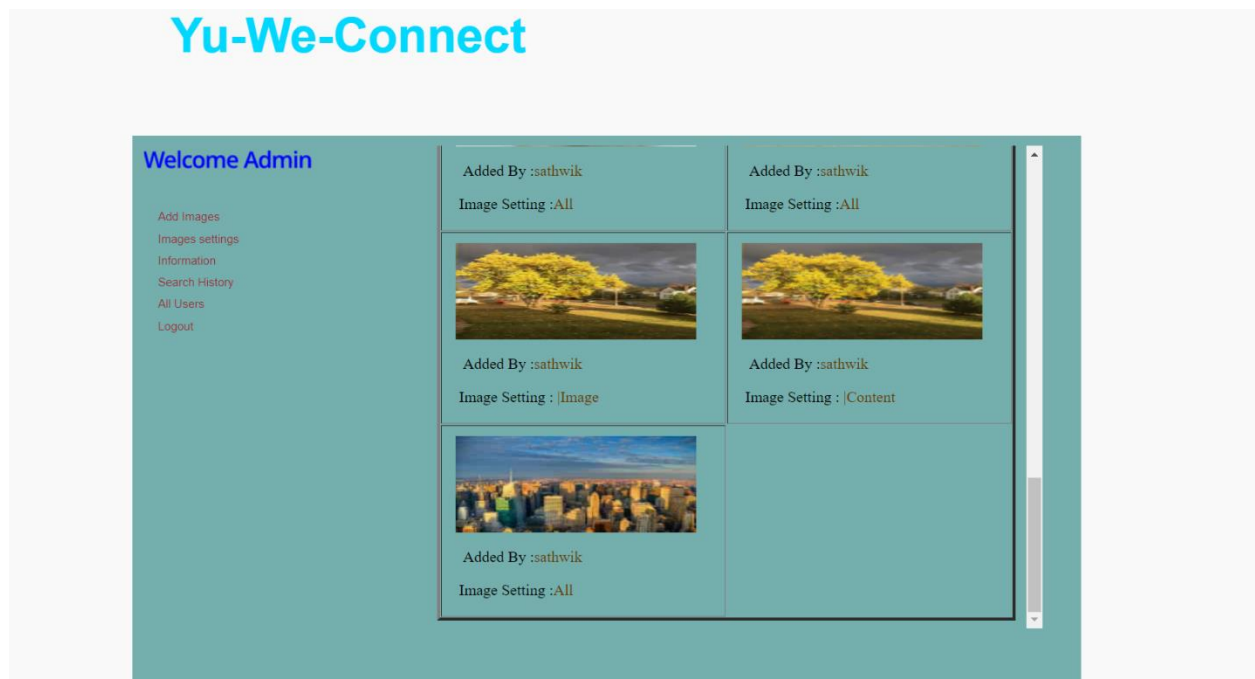

Figure 5.11: Admin View All Images with Settings

Admin can also see the images details shared by the user and can make changes to the content if he/she finds it inappropriate or irrelevant. Thus, creating a better environment for the users. Figure 5.12 shows the view of the admin where he or she can see the images with all the description and also the details of who added the picture.
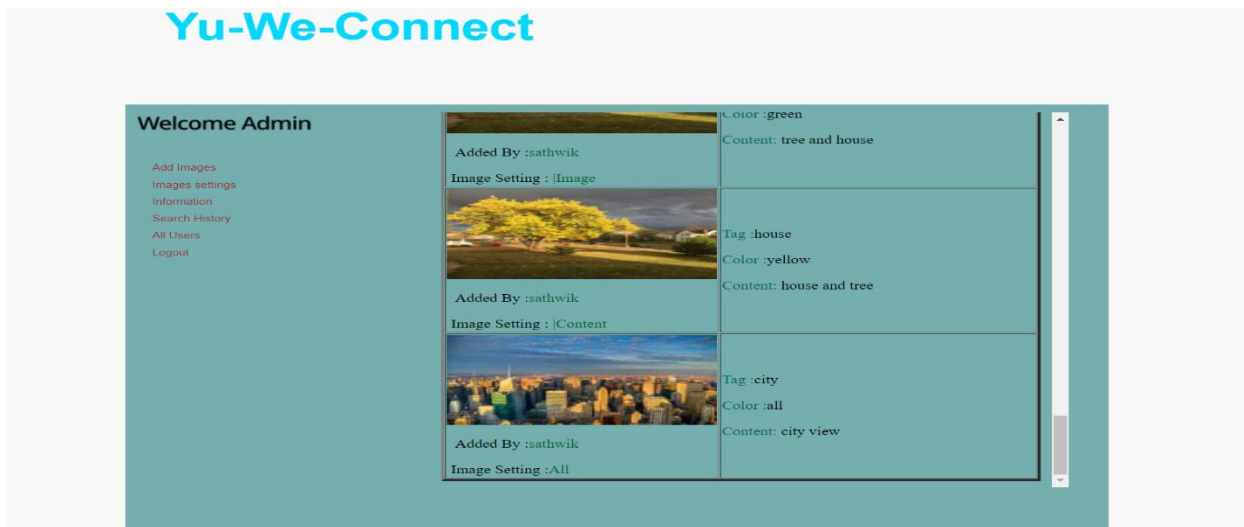
Figure 5.12: Admin View All Image Details

In addition to those, admin can view the search history of all the users and store the information for any later purposes like enhancing the GUI of the site, making available many options that are being searched. Admin has the facility to view all the users information to grant access to the site in the beginning and revoke the access when he/she believe the user has shared any illegal or inappropriate images or information. Figure 5.13 show the view of the admin where he or she can view all the users.
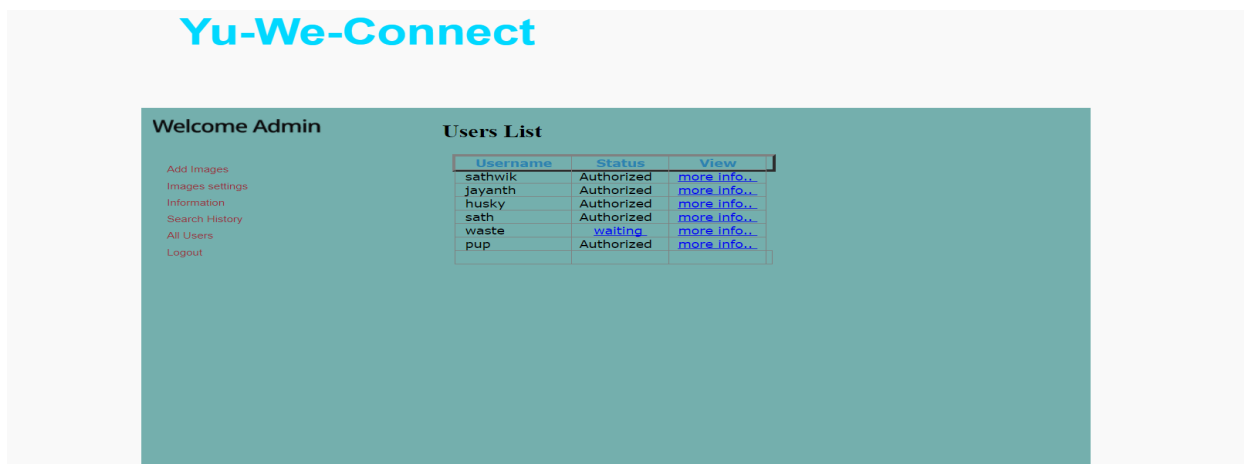

Figure 5.13: Admin View All Users

User Module

As admin, user can login to the system using his or her credentials through the user home page.

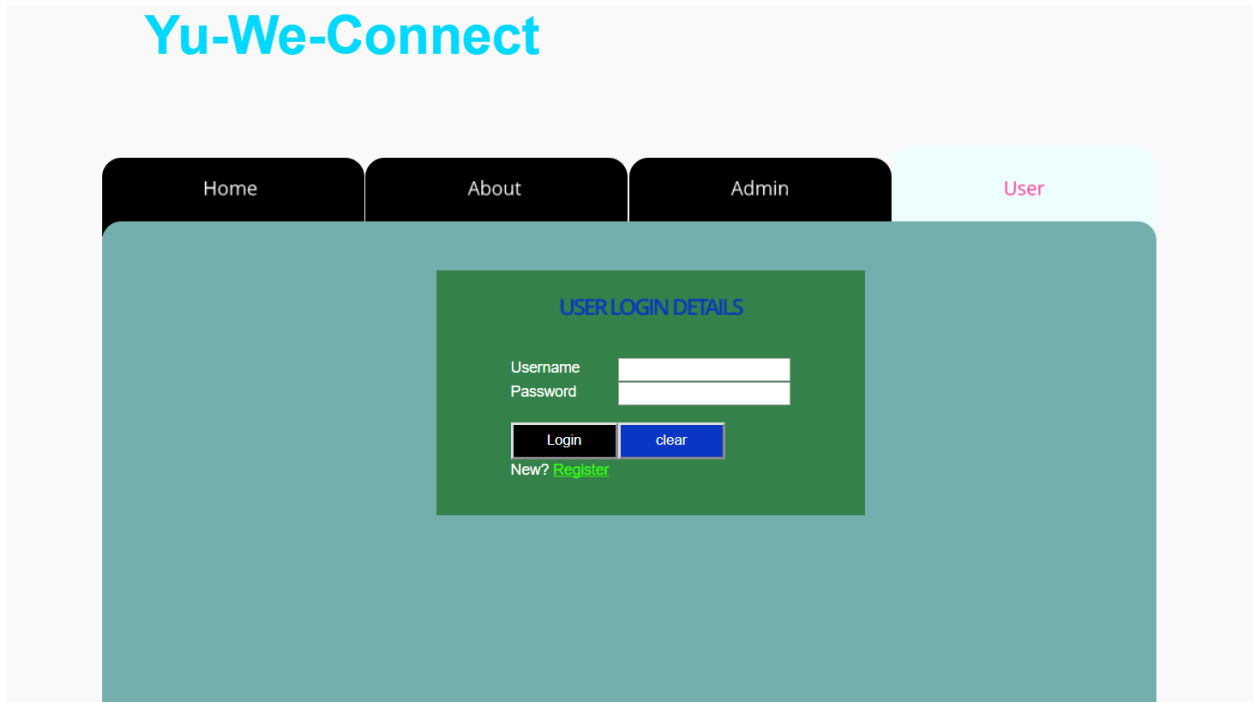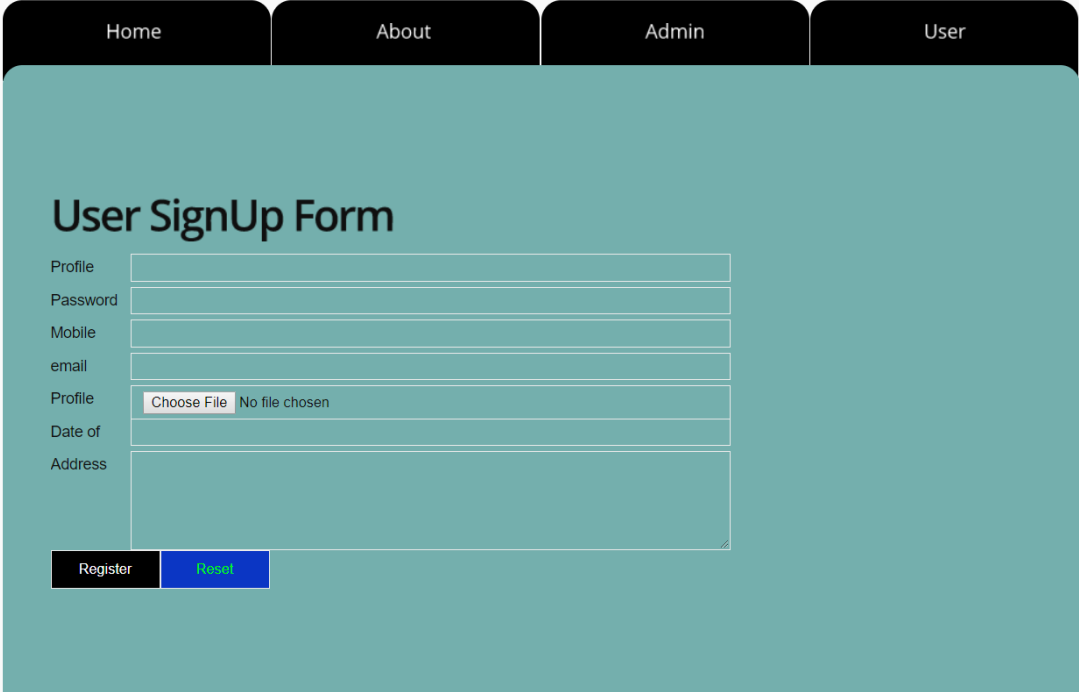Figure 5.14 shows the user home page.

Figure 5.14: User Login Page

A user initially registers in the signup module made available for specifically signing up the people to the site. Figure 5.15 shows the user registration form in order to register into the system.

Figure 5.15: Signup Registration Form

He/she give the information such as name, Date of Birth, mobile number, email, profile picture and sets up a password which he/she can remember and clicks the register button. Note that none of the details should be null in order to create a user profile. The site later redirects to the user profile page which shows the user information. Figure 5.16 shows the user entering the details without failing to fill all the blocks.

Figure 5.16: User Entering Details

At first, the information wouldn't be available since the admin has to accept the user.

Figure 5.17: New User without Authorization

Later, once the admin accepts the user, he/she can view the information that is shown in the profile by logging in to the system using user login page. Figure 5.18 shows the user logging in to the system with his or her own credentials.
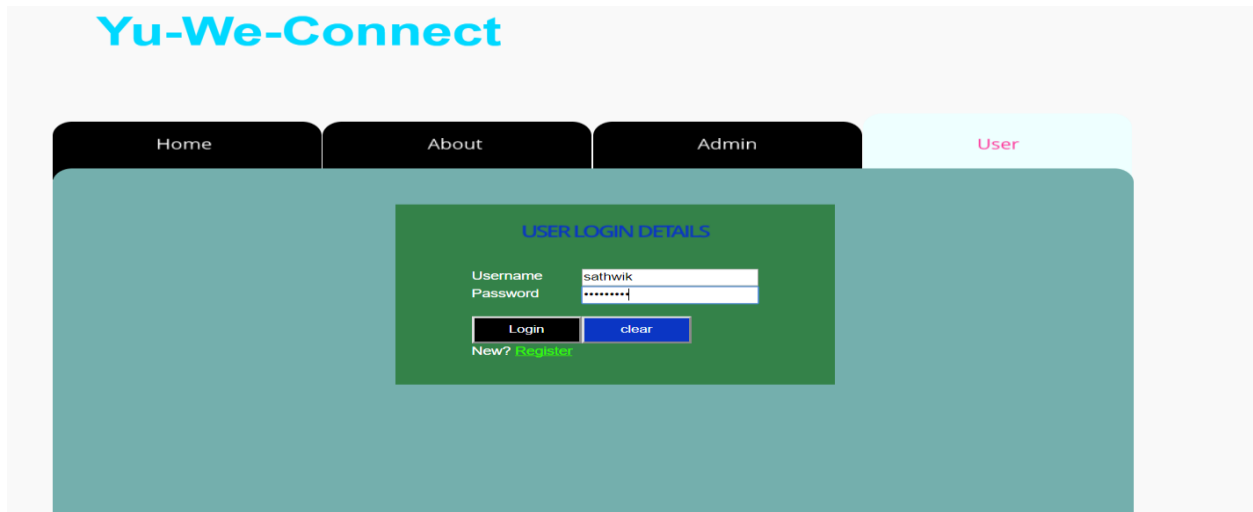

Figure 5.18: User Logging In
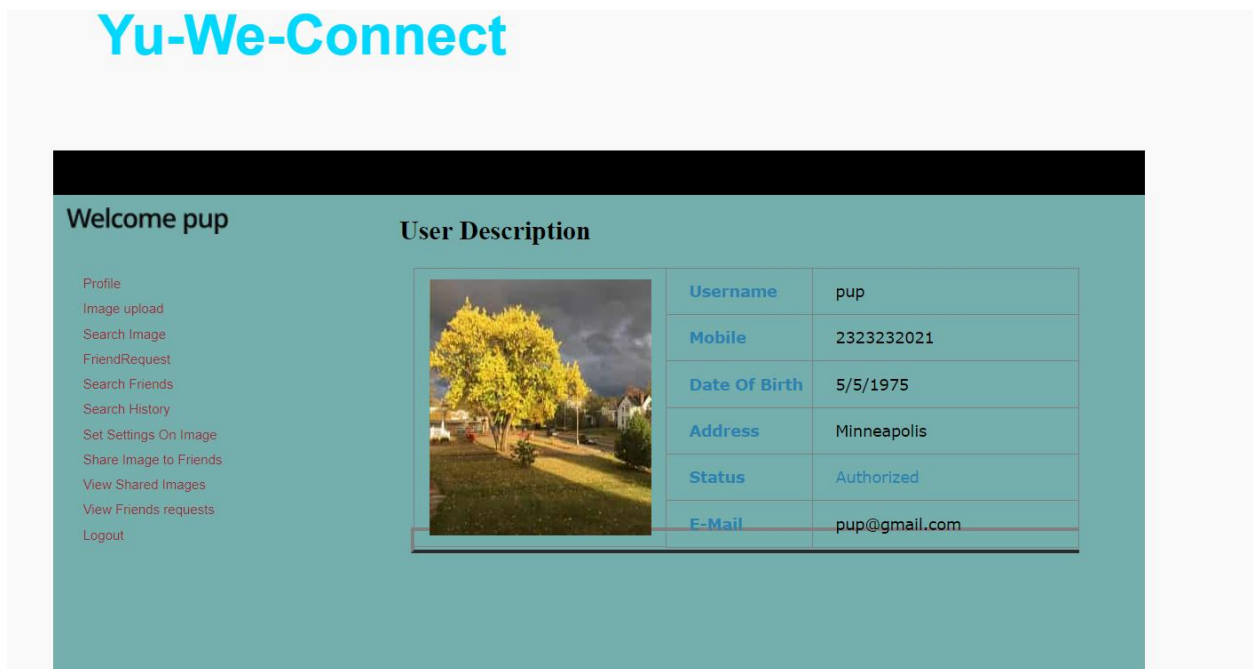
Figure 5.19: User Home Page

Figure 5.16 shows the home page of the user where one is redirected to after logging in to the system.

Firstly, the user can add an image of his/her interest by clicking on image upload button. The site redirects to another page where the user has the options to describe the image, add the tags for it and describe the color and uses of the image and upload it. Figure 5.20 shows the view of the user to add an image with the details he or she might want to share and apply settings one might interested in.
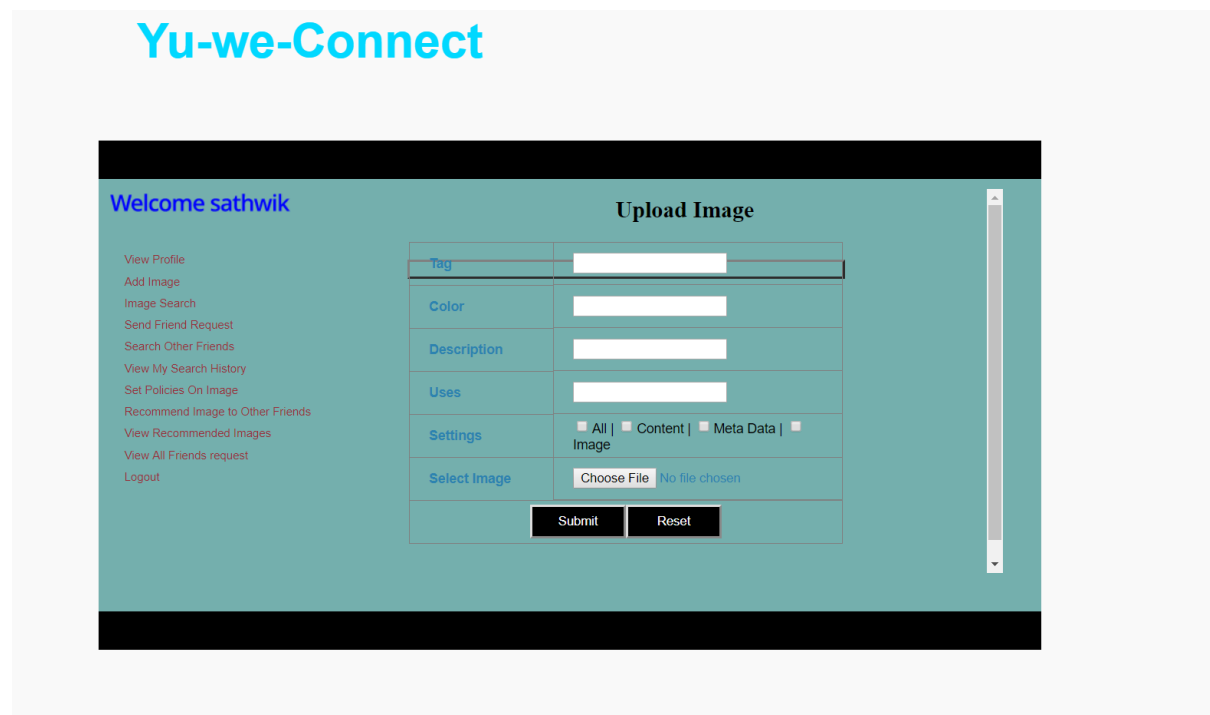
Figure 5.20: User Add Image

Once the user has uploaded, later he or she can share it to the people they like.

Figure 5.21: User Adding Image Details

The user can also set the settings for the image like all, metadata, content and image which makes the fellow users to access the image with the corresponding settings.

The user has the option to search the image by clicking on search image button. It redirects to the options to search image based on content, meta data and friends. Figure 5.22 shows the view of the user where he can search an image by his or her interest.

Figure 5.22: User Image Search

Depending on the words he/she anticipates to search, the user needs to select the option of searching by content or metadata or basing on friends. If the user gives the input correct, he/she will get the image based upon the settings applied by the image uploader, who is the user's connected friend.

User has the friend request button through which he/she can send a request to connect with other users whom he/she might know or to get to know. Figure 5.23 shows the list of users in the system and the user can send a friend request to anyone he might be interested to.

Figure 5.23: User Sending Friend Request

He can also search the friend by using the search friends button available in the left tab.
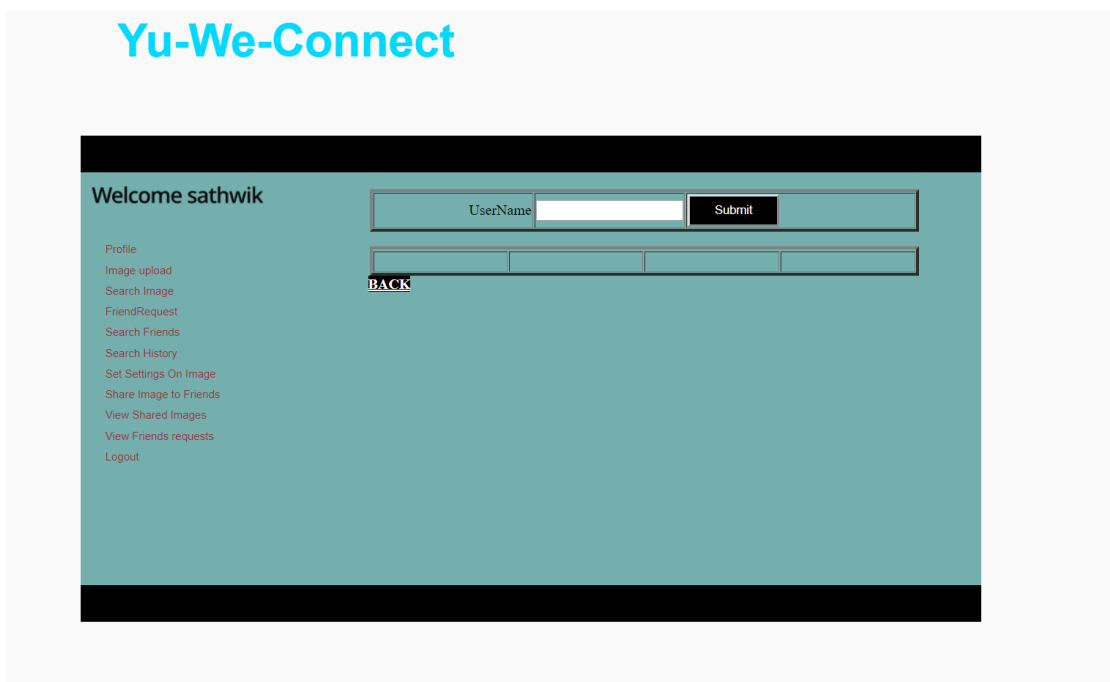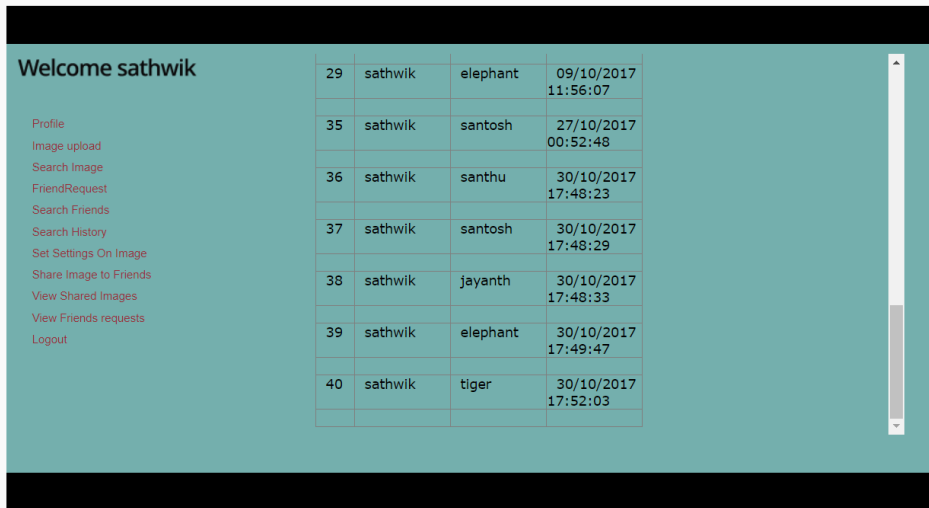
Figure 5.24: Results of Friend Search

In addition to those, the user has the options to search the history in which they have searched for the images. Figure 5.25 shows the list of searches by the user.

Figure 5.25: User Search History

User can share the images to other friends by clicking on share images to other friends. User can also change the settings that are applied on the image at the time of uploading by clicking on set settings on image. Figure 5.26 shows the view of the user where he or she can apply settings even after uploading the image, which gives the flexibility.
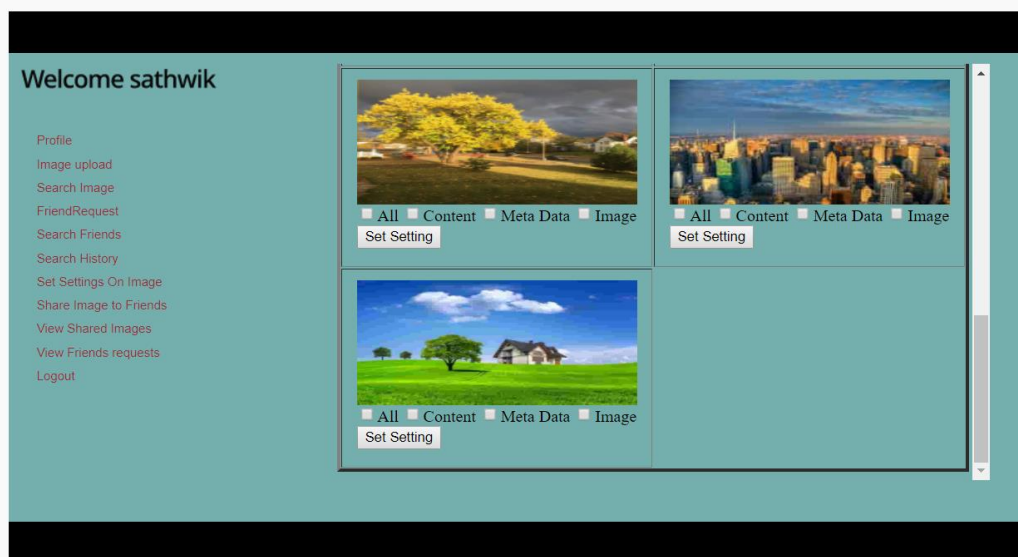
Figure 5.26: Set Policies on Image

He or she can view the shared images or information shared by friends basing upon the settings applied by the user that shared the original image.
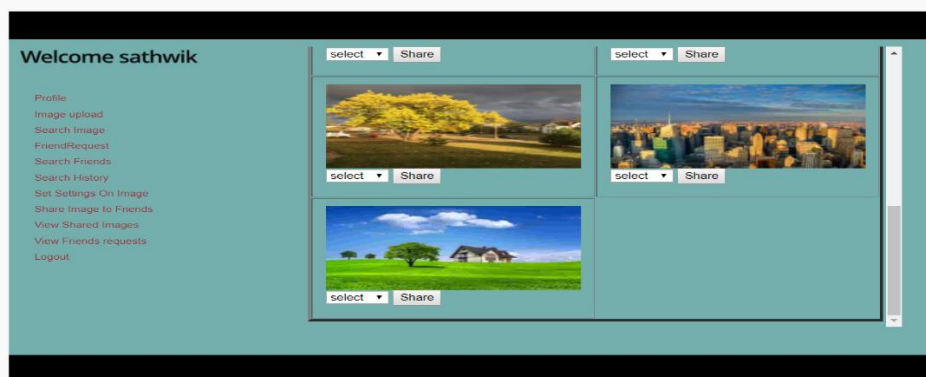
Figure 5.27: Share Image to Friends

The user has the option to check his or her old or pending friend requests so that he can have a clear idea of who accepted and who are still in decision making, through which the user can get in contact with friends or come to know of new peers. Figure 5.28 shows the view of the user where he or she can see all the user's and can send the requests if interested.
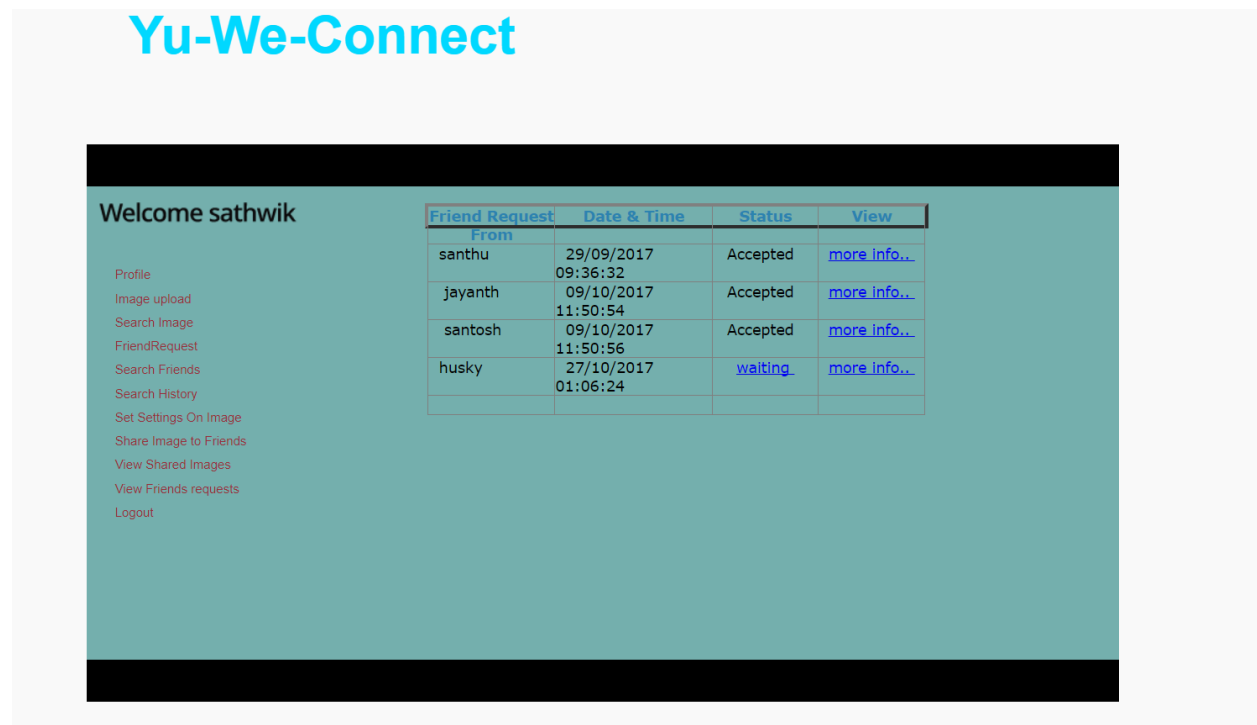
Figure 5.28: Checking Friend Request

Later, he/she can logout by clicking on logout button available at the end of the panel.

Hardware Requirements

Any computer system with processor not before Pentium IV and having a hard disk space of minimum of 30GB and good fast RAM for example 1GB RAM can be used to deploy or develop the application.

Software Requirements

1. Windows, mac OS, Linux

2. MySQL Server 5.6

3. Eclipse IDE for Java EE Developers

4. Java 1.7

5. Apache Tomcat7

Chapter 6: System Design

UML Diagrams

In order to visualize, document the flow or artifacts of systems, a language called UML is used. UML stands for Unified Modelling Language (Tutorials Point, n.d.). As many people say, a picture is worth of a thousand words, and the meaning can be easily made understandable to the viewers than a hard explanation, UML fits in perfectly in such cases. In order to organize the much ancient developed concepts such as object oriented concepts, UML has been developed. The main aim of UML is to develop a common modelling language which can be understood by all the levels of programmers and thus it needs to be made simple to use. UML is used to model both simple and complex systems in the hassle world of computer.

UML diagrams can be categorized into two types:

1. Static representations such as class diagrams

2. Dynamic representations such as activity diagrams, use case diagrams and sequence diagrams.

The section is organized as architecture, data flow, class diagram, activity diagram, use case and sequence diagrams. The architecture shows the major components of the system that are web server, web database, admin and a couple of users, it shows the actions performed by the system.

Attributes. Uploading attributes are image id, image tag, image color, image uses and the user registration has username, password, address, email, mob no, date of birth, gender, zip code.

Architecture
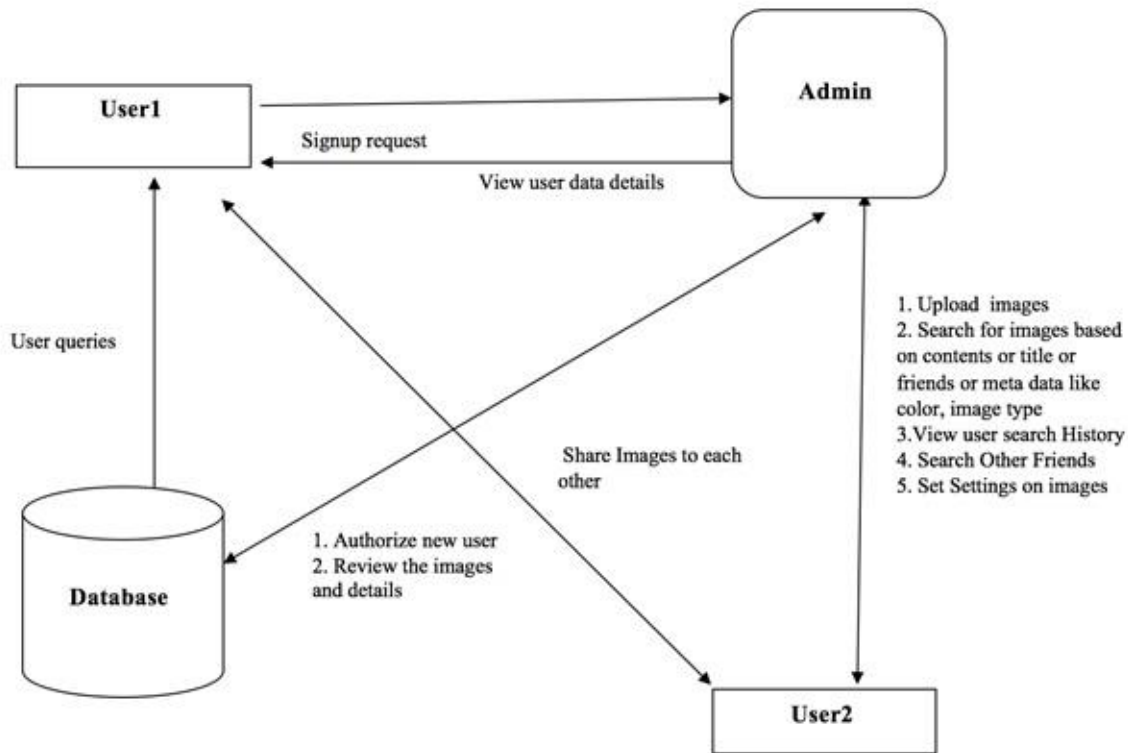
Figure 6.1: Architecture

Data Flow

Data flow diagrams are used to represent the flow of the input and the steps carried out inside the system and the output generated by the system (Terry, 2003). It represents how the information is being carried inside the system and also how the output is being generated. Figure 6.2 shows the representation of the system flow in the application that is discussed where the admin can add image and review the image and the user can add image, apply settings and search for the users.
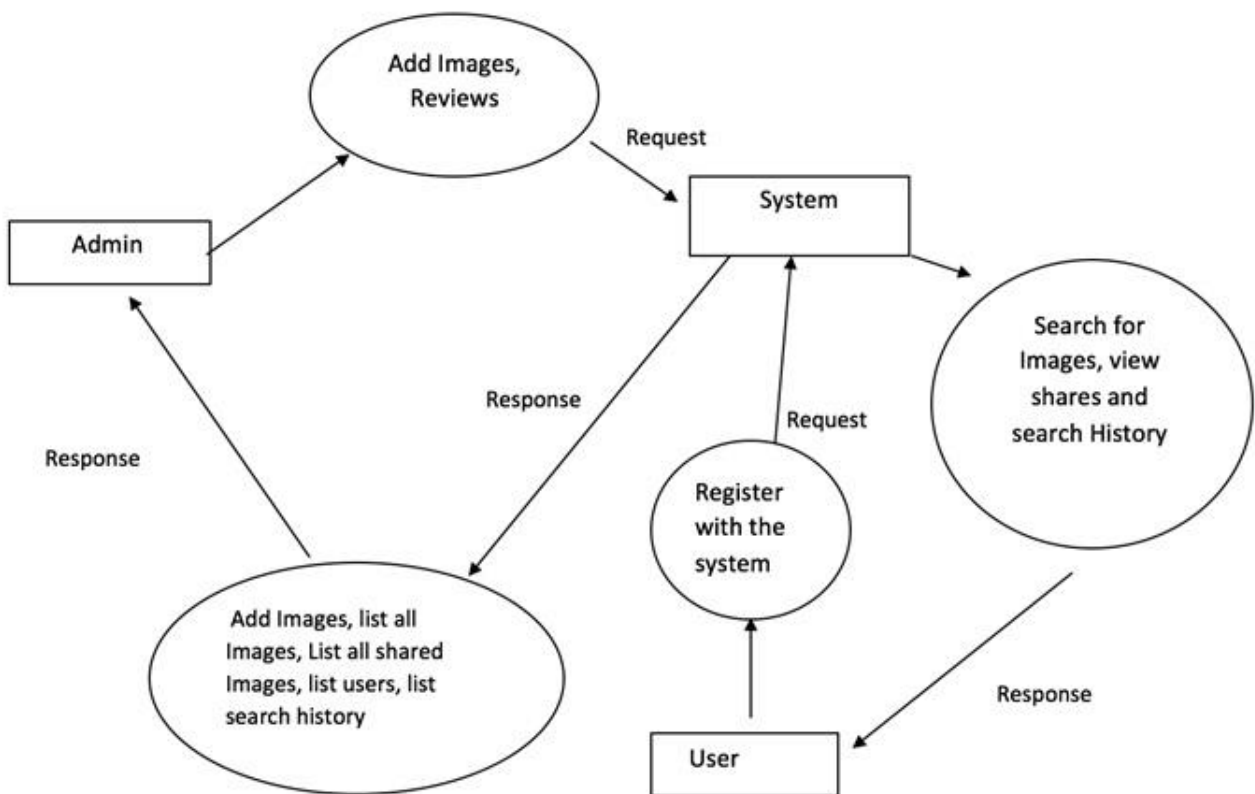
Figure 6.2: Data Flow

Class Diagram

The figures which provide the pictorial representation of classes in a system can be referred to as class diagrams (Terry, 2003). Class diagram in the logical view can be referred to as the pictorial representation of packages in the system. Each diagram mainly comprises of major classes in the system and has its own main class.

Class diagrams picture the generic descriptions of the systems. Collaboration diagrams and class diagrams are both helpful to represent the object models in system. A class can have icons which represent classes, packages, operations and relationships.

A class is the set of objects which have similar packages, attributes and relationships. Packages and attributes can be any object or place or anything that is applicable to the system. Operations may contain methods which perform specific function inside the class with the specific attributes. Relationships can be further grouped to Association, aggregation and inheritance.

A relationship that specifies connections among two or more classes is known as an association. A connection that makes the relationship between a whole and its parts can be grouped as aggregation. Inheritance is the ability to get the methods or the attributes from the whole.

Below is the class diagram for the system which has main class as the web server and all are the other counter parts. The containers have methods on the top half and the attributes on the bottom half. Figure 6.3 illustrates the classes in the application and the methods and attributes inside the classes.
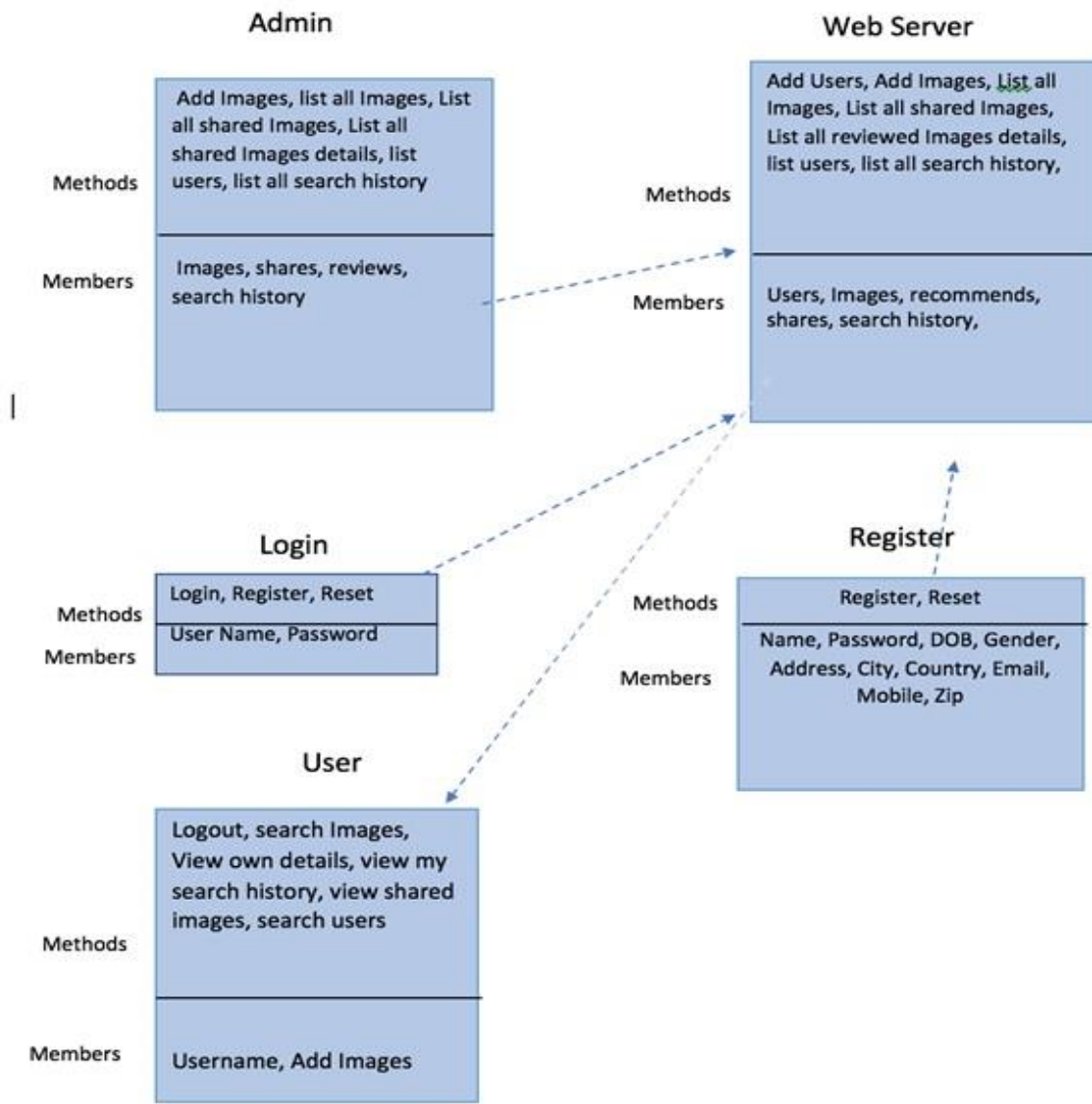
Figure 6.3: Class Diagram

Activity Diagram

An Activity diagram can be stated as change of a special case of a state machine (Terry, 2003). The states in the activity diagram are activities that represent the performance and the

transitions of the operations. They represent the operational and business workflow of the system. They can simply be also considered as a flow chart where the flow goes from one operation to the other operation which are interlinked or interdependent on one another.

The main purpose of activity diagrams is to provide a complete view of the operations and the flow between them and also what exactly happens after one operation and the other. It can also be used to specify the work flow inside a use case and also for describing the code specific information.

Activity diagrams represent the dynamics of the system like showing the workflow of the system from activity to activity where things can be done either in parallel or in a series. They may be created to represent the flow across use cases or with in a use case.

Below, you can find the two activity diagrams where the figure 6.4 describes the activity flow for a user who logs in and creates account. The user registers his account and then login with the credentials and share his or her posts to the other friends

Figure 6.5 describes the activity diagram for the admin who accepts the user registration and also share the images or see the user's work to make sure he or she is acting within the norms.
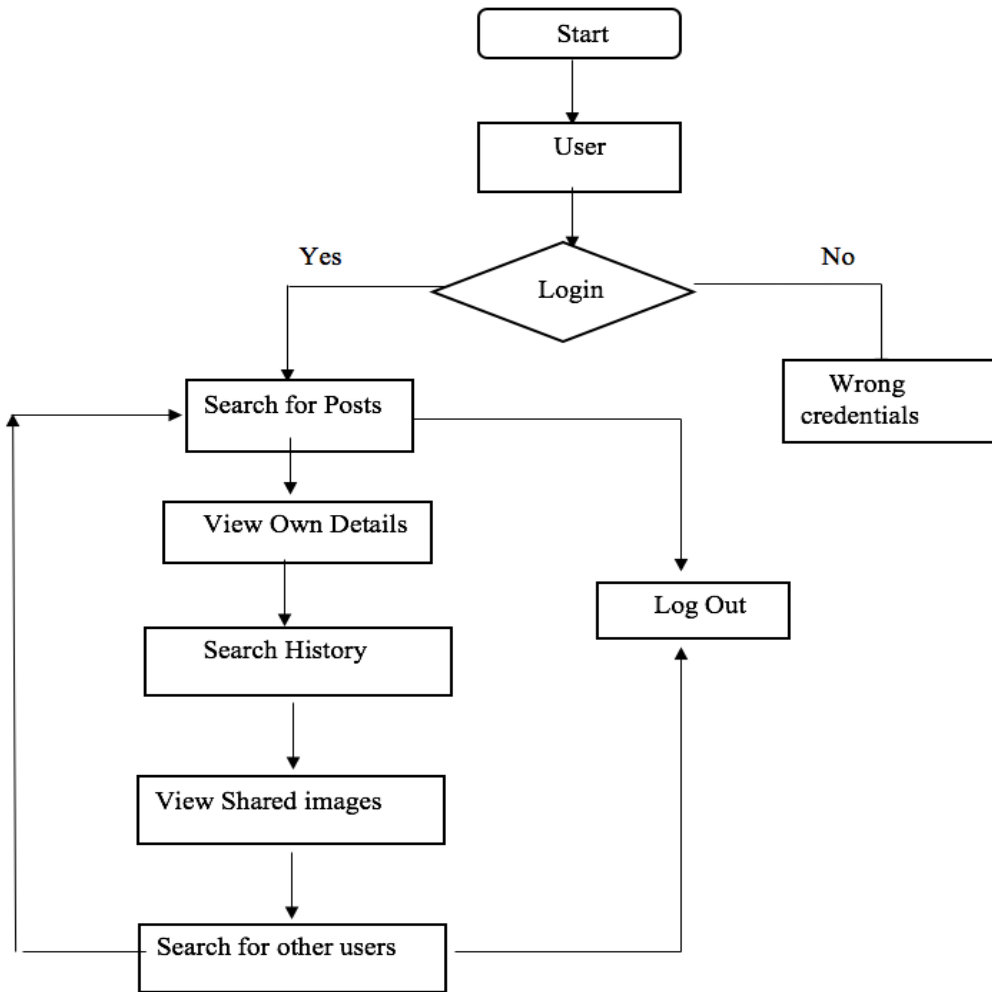
Figure 6.4: Flow Chart 1
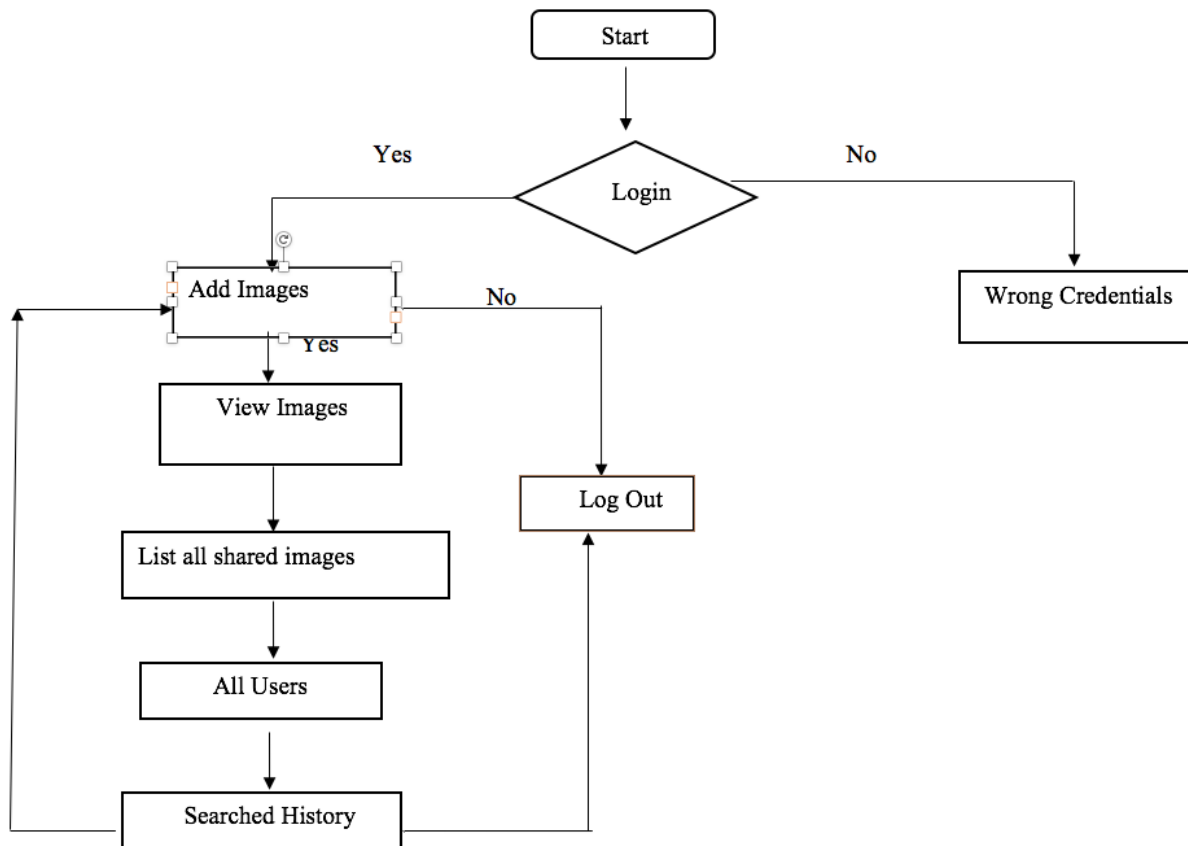
Figure 6.5: Flow Chart 2

Use Case Diagram

Use case diagrams can also be called as the functions that are provided by the system to its users (Terry, 2003). They model the system using use cases and actors.

Use case. It represents the actions that are initiated by the actors. A use case can be stated as the bunch of scenarios that describe the interaction between the system and the users.

Finding actors. Any person or system who interacts with the current system can be called as an actor [27]. Actors cannot be called as part of the system.

An actor can be something or someone who give the input to the system and receives the output from the system or who is totally outside of the current system and does not possess any control over it who uses or interacts with the system or who directly uses and responsible for the system or which can also be the external software used by the current system

The direction of the arrow from the actor to use case or the use case to the actor defines the initiation of interaction. Figure 6.6 shows the use case diagram of the application where the admin and user are the actors and different use cases the actors might involve in.
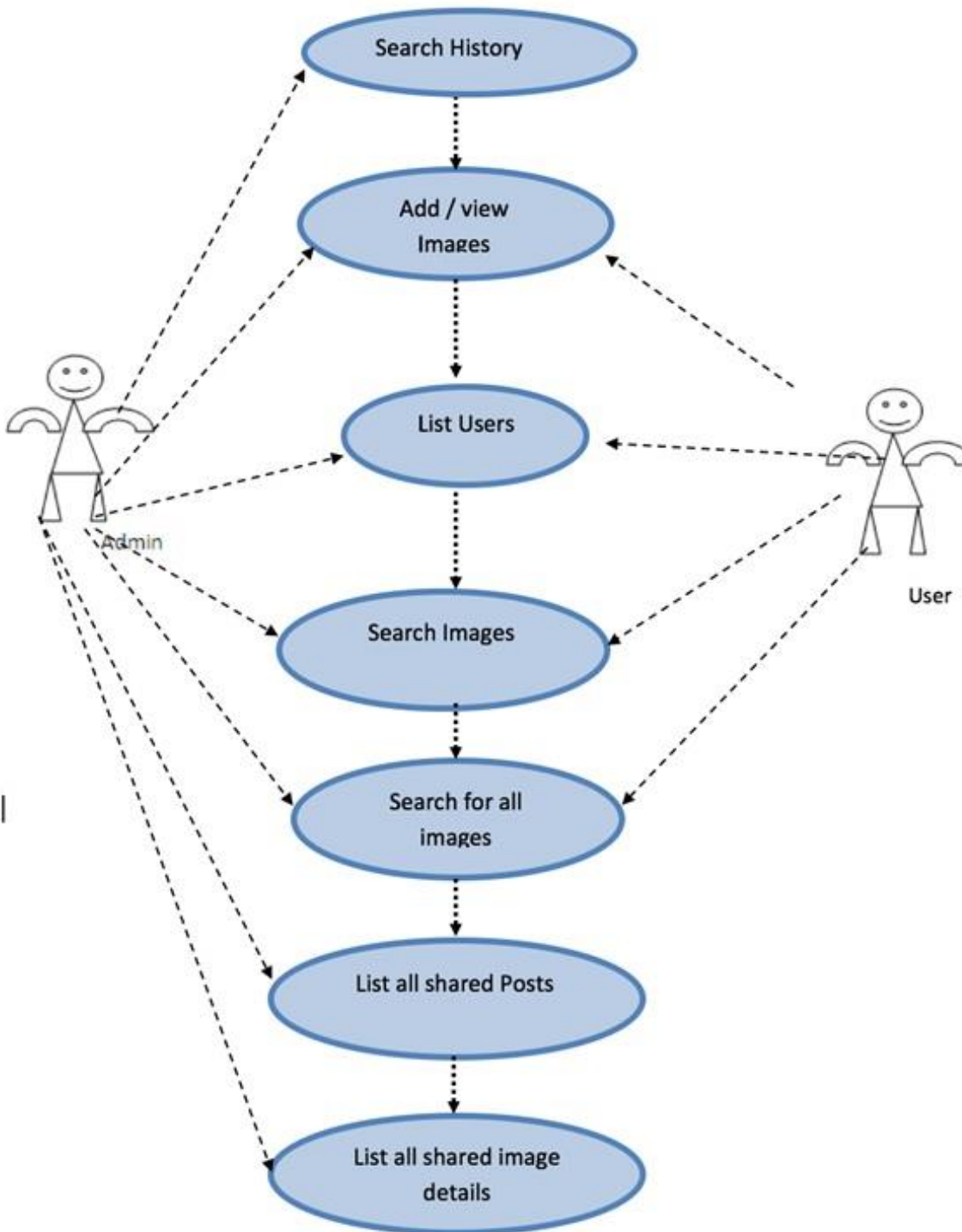
Figure 6.6: Use Case Diagram

Sequence Diagram

It is a graphical representation of interactions between the objects in a system with in a time sequence. Sequence diagram show the representation of what action or sequence happens first and then what happens next within a certain period of time. They help in providing the necessary information to determine the responsibilities of the classes and interfaces. There are different parts which on a whole make a sequence diagram (Terry, 2003). They are:

Actor: Actors are generally the person or the thing which interacts with the objects inside a system. A stick figure is used to model the actor.

Objects: Instances of the class which are arranged horizontally are stated as objects. A rectangle is used to show the object with a name and a semi colon.

Lifeline: Lifeline shows the time in which the objects exists in the system for the use of the persons or the systems which interact with the system.

Activation: The solid rectangle in between the line is called as activation line and it represents the time which the object operates.

Message: The arrow lines between the objects can be referred to as message where the objects interact between each other. They are modeled in between the activations. The solid arrows indicate that the caller waits in an operation for receiving the object to finish and also may include response.

Creating a sequence diagram includes identifying all the above mentioned states or parts and organizing them in a meaningful manner. Figure 6.7 illustrates the sequence inside the application where the actors, admin and the user, interact with the system, web server.

Figure 6.7: Sequence Diagram

Chapter 7: Testing

All the applications that are developed need to be tested to ensure the accuracy and the suitability of that application in real world. Thus, testing pl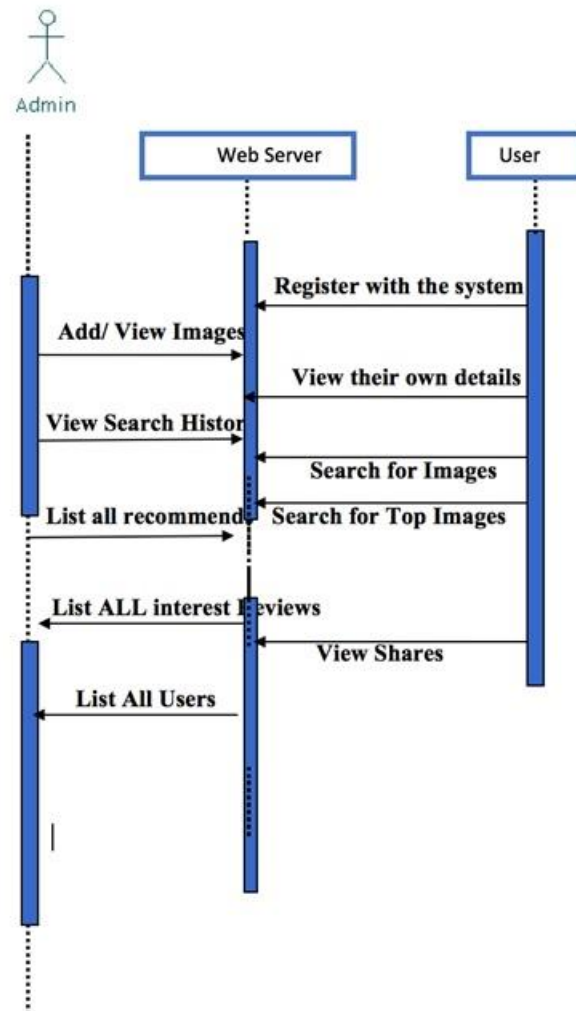ays a major role as any other mechanism in the software development. The real-time errors or the bugs which cannot be identified by developers come in to the light at the stage of testing. Thus, helping the programmers to correct the mistakes. Testing can be classified into many different types in which integration testing, unit testing, white box testing, black box testing and functional testing are a few. Each of the above-mentioned types have their own advantages and helps in discovering the bugs in the entire application. Let us go through some of the testing methods.

Unit testing. Unit testing is performed at the beginning of the development stage in which the developers themselves write the testing code in order to check whether the particular function is performing the specific role, whether the output of the function as a whole is correct or not. One can stub the functions and variables in order to validate the result without the requirement of the original function. As, it needs the logic of the code within the function, this type of testing is generally performed by the programmer. One can achieve a 100% code coverage with the help of unit testing. It can also be specified as a part of white box testing.

Integration testing. Integration testing is a form of both white box and black box testing and is also generally carried out by the programmers itself. Integration testing tests the entire module of the application after grouping all the pieces of the function together and checks the output and the flow of the module. Usually programmers use the entire code including the functions and variables without stubbing any thus giving more reliable output for the code. This form of testing is also helpful to test the modules that are grouped over different networks.

Performance testing. Performance testing verifies the credibility and effectiveness of the application and gives out the performance of the system under different circumstances though which the marketing people can show up the capability of the application. It is a form of black box testing.

Black box testing. Black box testing is a form in testing where the tester ignores the internal design of the application and entirely relies on the output of the system/application. All the test case scenarios are validated in this form of testing.

White box testing. In contrast to the black box testing, white box testing validates the internal design/code of the application and checks whether each and every function inside the application works as expected. It can be stated as white box testing is used for verification of the application whereas the black box testing is used for validation.

Stress testing. Stress testing is a form of testing that helps the developers to know how the application behaves under sheer pressure such as system overload, less memory, system crash and very huge commands. Thus, testing the efficiency of the application.

System testing. This form of testing deals with the testing of the entire project when loaded on different operating systems such as windows, Mac OS, Linux and others. It is a form of black box testing where the entire working efficiency is only tested.

Regression testing. After several changes made to the application due to the errors popped up with the different forms of testing, the testing made to check the capability of the testing is called as regression testing. It mainly focusses on the output of the system after application of many changes. Several automation tools are used to perform this type of testing.

Acceptance testing. This can also be called as user acceptance testing. After the delivery of the project to the client, the acceptance of the end users will be specified under this testing.

Thus, neither developers nor testers go through this process. If any of the end user come up with dissatisfaction or any errors that showed up after many days are bought back to the developers and dealt accordingly.

      Testing performed for the implemented project. The testing performed in this project has main objectives. They are:

1. Checking all the UI fields

2. The work flow of all the pages and the functioning of all the pages

3. User functions such as sign-in, sign-out, image sharing, making other friends

4. Admin function such as accepting the user, viewing all the images with the policies

In testing all the objective, the following tests are performed.

1. Tested the UI login screens for successful Login

2. Tested for registering of the user successfully uploading the profile picture.

3. Tested the user uploading the images and sharing them with the desired settings for his/her friends.

4. Tested for the user changing the policies of the image and sharing them.

5. Tested for the admin to login and log out

6. Also Tested for the admin to authorize the user and check all the images with settings/policies along with the images.

7. Also tested the specific functions which perform the sharing and setting up the policies.

Chapter 8: Conclusion

With the increase of social networking sites and population over the world, security has become a great interest and the need for more and more secured systems has been increasing day-by-day irrespective of the evolving technologies. In light of these situations, the proposed paper analyzed an adaptive policy prediction system used for proposing privacy policies to the users in order to make the social networking a better place for the current and also for future generations. Along with the analysis, the paper also discusses the implementation of a social networking site with which the registered users can create their profiles, make friends and share their images with specific settings.

The users who share images over internet have different opinions and the opinion for a person changes day-by-day. Hence, the security models need to be updated thoroughly to make the user feel safe while using any social networking site at any given time. Application of such systems in each and every networking site improves the security enormously and the amount of frauds over the internet through the social networking site decrease rapidly.

Even the people need to be aware of the incidents happening due to the misuse of the social networking sites and need to check the settings they apply or accept a set of policies proposed by the security system over the images they intend to share to avoid any misuse with the shared content.

Limitations

The system described in the paper takes user's social context, personal characteristics and image content of the information shared by user and generated settings. These settings are proposed to the user for individual images or group images shared on the networking site and the user reviewing the proposed settings might accept or modify them according to their interest.

Any repetitive image sets added in future also needs user consent before applying the same settings.

Image content classification needs maximum accuracy in classifying the image in order to achieve higher acceptance rate of the settings or policies that are being proposed by the system. Implementation of highly accurate content based classification is the main challenge in generating perfect settings for the users to approve. There are many image classification systems available but a deeper analysis in to the outputs generated by the classification system needs to be performed before finalizing an algorithm to propose user settings for a given social networking site. This can be done differently for different social networking sites which are built on different platforms. Hence a different approach needs to be incorporated for different sites depending on their requirement.

Future Work

There is a wider scope to further extend the work implemented in this paper and some of them include; proposing a policy while the user uploads an image basing upon the image content description and other factors like social context. Various content based classifications and policy mining concepts can be implemented in the application in order to improve the image classification techniques and train the system to propose more accurate settings for the user depending on the networking platform and there by achieve a more accurate and dependable policy proposing system. Robust technologies like various JavaScript frameworks which are not object oriented with simpler implementation formats and fast response times can be used to improve the performance of networking site's or networking applications developed in future and also provide more advanced user interaction.

References

A Tutorial on Clustering Algorithms. (n.d.). Retrieved March, 2017, from

https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/

Adu-Oppong, F., Gardiner, C. K., Kapadia, A., & Tsang, P. P. (2008). Social circles: Tackling

privacy in social networks. Proceedings of the 5th Symposium on Usable Privacy and

Security.

Bayesian network. (n.d.). Retrieved October, 2017, from

https://en.wikipedia.org/wiki/Bayesian_network

Bonneau, J., Anderson, J., & Church, L. (2009). Privacy suites. Proceedings of the 5th

Symposium on Usable Privacy and Security - SOUPS 09.

Bonneau, J., Anderson, J., & Danezis, G. (2009). Prying Data out of a Social Network. *2009*

International Conference on Advances in Social Network Analysis and Mining.

Can anybody compare Haar, Daubechies, coiflet, and symlet wavelets?.(2014). Retrieved May,

2017, from

https://www.researchgate.net/post/Can_anybody_compare_Haar_Daubechies_coiflet_and

_symlet_wavelets

Chen, H., Chang, M., Chang, P., Tien, M., Hsu, W. H., & Wu, J. (2008). SheepDog. Proceeding

of the 16th ACM international conference on Multimedia - MM 08.

Choudhary, P. (2017). Introduction to Anomaly Detection. Retrieved October, 2017, from

https://www.datascience.com/blog/python-anomaly-detection

Choudhury, M. D., Sundaram, H., Lin, Y., John, A., & Seligmann, D. D. (2009). Connecting

content to community in social media via image content, user tags and user

communication. 2009 IEEE International Conference on Multimedia and Expo.

Data Mining Concepts. (2008, July 01). Retrieved June, 2017, from

https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/classify.htm - DMCON004

Data mining. (n.d.). Retrieved November, 2017, from https://en.wikipedia.org/wiki/Data_mining

Difference between Fourier transform and Wavelets. (n.d.). Retrieved November, 2017, from

https://math.stackexchange.com/questions/279980/difference-between-fourier-transform-
and-wavelets

Haar wavelet. (n.d.). Retrieved November, 2017, from

https://en.wikipedia.org/wiki/Haar_wavelet

Han, J., Pei, J., & Kamber, M. (2006). Data Mining. CA: Southeast Asia Edition.

How Decision Tree Algorithm works. (2017, April 21). Retrieved October 2017, from

http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/

How Do I Install Java?. (2017, February 21). Retrieved June, 2017, from

http://www.java.com/en/download/help/download_options.xml

Image Resolution. (n.d.). Retrieved October 2017, from http://www.microscope-
microscope.org/imaging/image-resolution.htm

Java: A Beginner's Guide - Oracle. (n.d.)., retrieved from

http://www.oracle.com/events/global/en/java-outreach/resources/java-a-beginners-guide-
1720064.pdf

Java™ JDBC API. (n.d.). Retrieved October, 2017, from

https://docs.oracle.com/javase/7/docs/technotes/guides/jdbc/

Kaur, R., & Kaur, R. (2013). Remove the Noise from Medical Images Using Discrete Wavelet

Transform. International Journal of Engineering Research and Applications, 3(2).

Khanteymoori, A. R., Homayounpour, M. M., & Menhaj, M. B. (2008). A Bayesian Network Based Approach for Data Classification Using Structural Learning. Communications in Computer and Information Science Advances in Computer Science and Engineering, 25-32.

Margaret, R. (n.d.). What is SQL (Structured Query Language)? - Definition from WhatIs.com. Retrieved December 13, 2017, from http://searchsqlserver.techtarget.com/definition/SQL

Mazzia, A., Lefevre, K., & Adar, E. (2012). The PViz comprehension tool for social network privacy settings. Proceedings of the Eighth Symposium on Usable Privacy and Security - SOUPS 12.

Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21-45.

Pyramid (image processing). (n.d.). Retrieved October, 2017, from https://en.wikipedia.org/wiki/Pyramid_(image_processing)

Terry, Q. (2003). Visual Modeling with Rational Rose 2002 and UML. Addison-Wesley

Ravichandran, R., Benisch, M., Kelley, P. G., & Sadeh, N. (2009). Capturing social networking privacy preferences. Proceedings of the 5th Symposium on Usable Privacy and Security - SOUPS 09.

Sahu, J., & Choubey, A. (2013). Study and Analysis of Multiwavelet Transform with Threshold in Image Denoising: A Survey. International Journal of Science and Research, 2(8).

Söder, O. (2008, May 29). Retrieved June, 2017, from http://www.fon.hum.uva.nl/praat/manual/kNN_classifiers_1__What_is_a_kNN_classifier_.html

Squicciarini, A. C., Lin, D., Sundareswaran, S., & Wede, J. (2015). Privacy Policy Inference of

       User-Uploaded Images on Content Sharing Sites. IEEE Transactions on Knowledge and

       Data Engineering, 27(1).

The Daubechies D4 Wavelet Transform. (n.d.). Retrieved May, 2017, from

       http://www.bearcave.com/misl/misl_tech/wavelets/daubechies/index.html

Tutorials Point. (n.d.). UML Tutorial. Retrieved May, 2017, from

       http://www.tutorialspoint.com/uml/

Wavelet. (n.d.). Retrieved December 13, 2017, from http://www.wikidoc.org/index.php/Wavelet

What is Association Rule Mining? - Definition from Techopedia. (n.d.). Retrieved April, 2017,

       from https://www.techopedia.com/definition/30306/association-rule-mining

What is social networking? - Definition from WhatIs.com. (n.d.). Retrieved 2016, from

       http://whatis.techtarget.com/definition/social-networking

Yu, J., Joshi, D., & Luo, J. (2009). Connecting people in photo-sharing sites by photo content

       and user annotations. 2009 IEEE International Conference on Multimedia and Expo.