IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

Filtering Unwanted Contents from User's Wall in Online Social Networks

Miss. Ujwala S.Tambe[1], Prof. Archana. S. Vaidya[2]

Student, Computer Engg, G. E. S's R. H. Sapat College of Engg., Management Studies and Research, Nashik, India[1]

Asst.Prof. Computer Engg, G. E. S's R. H. Sapat College of Engg., Management Studies and Research, Nashik, India[2]

Abstract: Now a day's, people can communicate with each other by exchanging the multimedia information. It may include textual data, images, audio or video. Online social Network (OSN) provides different services to the user. In online social networking sites like facebook, twitter, etc., there is possibility of posting any kind of data on user wall. Such data may contain unwanted messages or images. Other user can view such data and can comment on such post. Such post may affect user social image. So security of such user's personal wall is an important issue. So, to avoid this, we propose a system which provides facility to user to control the unwanted messages and images posted on their wall and a system that allow users to restrict particular user on the basis of its social reputation, also user can extract labels from posted images and filter it, also take a decision whether to allow such content or not. This system also suggests malicious users from their behaviour and block frequency. Hence, this system, avoids nuisance of unwanted messages and images on the walls.

Keywords: online social networks, short text classifier, message filtering, labelled image, machine learning.

I. INTRODUCTION

Social networking sites are today's one of the most popular interactive mediums to communicate, split and spread around a lot of human life information. Daily communications are done by the exchange of several types of content, including textual data, image, audio and video data. As per facebook statistics [1], average user creates 90 pieces of content each month, whereas further 30 billion pieces of content are shared each month.

In Online social networking era, as per current scenario user has multiple choices and options to share their data privately or publicly. Such social networking system has an innovative idea that tempts user to take benefit of it. An online social network like facebook, twitter, etc. , there is posting facility using which user can directly post any kind of data like images, text messages, audio, video, etc. hence there may be possibility of posting any kind of data on user wall. Such data may contain unnecessary messages or images. For example: malicious political statement, vulgar data, personal teasing statement etc., which are publicly available to friends of wall owner. Also wall owner's friends can comment on it which is also publicly available.

Such post may affect user image in social networking systems and unnecessarily he will have to keep explicit watch on such own wall content which is not possible. Up to a certain extent some existing schemes like facebook allows users to define, who is allowed to put messages on their walls. However, no content-based preferences and filtering are supported and thusly, it is impossible to prevent posting of such undesired messages. To protect undesired message posting on user wall and to protect user social image is an important issue on social networking site.

The motivation behind this work is to avoid overwhelmed used of unnecessary data on user's wall. As we considered some existing system [1] like facebook, which permitted users to define who is allowed to insert messages on their walls .But content based filtering is not provided.

We considered another system called Film Trust [5] that deeds OSN trust relationships and original information to personalize access to the website. However, filtering policy is not provided by this system. Another social networking service is MyWOT. This service allows its subscribers the ability to the rate of resources in relation to four criteria: reliability, vendor reliability, privacy and child safe and set preferences that find out whether the browser should chunk access to a particular resource, or should send a warning message based on the specified value. So, this motivates us to work on content of data share on the user's wall in social networking.

Our system should filtered the unwanted text and images and enforces protection and productivity policies for business, schools and libraries to reduce legal and privacy risks while minimizing administration overhead. Filtering provide network administrator with greater control by automatically acceptable used policies.

In next section II, we are presenting the literature survey over the related work. In section III, the proposed system and its architecture are depicted. Section IV covers experimental result, conclusion of system.

II. LITERATURE SURVEY

The main contribution of this paper is, to design a system providing content-based message filtering for OSNs, based on Machine Learning techniques [2].

However, our work has relationships with content based filtering, as well as policy-based personalization for OSNs and, more in general web contents [1].

A. Content-Based Filtering

Content Based filtering system approve a document by matching the document profile with the user profile, using information retrieval techniques such Term Frequency and Inverse Document frequency (TF-IDF). User characteristics are gathered over time and profiled automatically based upon a user's prior feedback and choices. The system uses item to item relationship in recommending the document to the user. The system initiates with the process of gathering the details about the content item, such as medicines, symptoms, etc. for disease related item and colour, model, wheels etc. for the car items. In the next step, the system asks the user to rate the items and system matches unrated item with the user profile item and assign scores to the unrated item and the user is presented with items ranked according to the scores assigned [3].

A new idea is discussed regarding text categorization in [6]. It explains that a content base application created with TCS comprises of the TCS run-time framework and a TCS rule base. Where, the rule base guideline characterizes what classes the application can allocate to messages and contains decides that the categorization decisions for particular texts. J. Belkin and W. B. Croft [7] proposed a technique to filter information is applied to unstructured or semi-structured data, as opposed to database applications, which use very structured data. Such as email. Where the format for sending email is a semi structured, but the body of the email may content unstructured data.

B. Collaborative Filtering

The Collaborative filtering is a method of building automatic predictions (filtering) about the interests of a user by collecting preferences from many other users (collaborating). It is widely used in many filtering systems, especially in ecommerce applications, like Amazon.com and e-Bay, where a user's past shopping history is used to make recommendations for new products [3].

C. Policy-Based Personalization of OSN Contents

A Recently, there have been some proposal operator classification mechanisms to customize the access in OSN. For example, in [4] a classification method was proposed to classify short text messages to avoid irresistible users of micro blogging services in the raw data. The system described in [4] and Twitter2 focuses on a set of categories associated with each describing updates its contents. The user can then see that certain types of tweets on the basis of his / her interests. However, Kuter and Golbeck [5] propose an application called Film Trust, which operates OSN relationships of trust and provenance information to customize access to the site. However, these systems do not provide a layer of the filtering policy by which the user can exploit the results of the classification process and decide how to filter unwanted information. Another social networking service is MyWOT. This service allows its subscribers the ability to the rate of resources in relation to four criteria: reliability, vendor reliability, privacy and child safe and set preferences that determine whether the browser should chunk access to a particular resource, or should send a warning message based on the specified value.

P. W. Foltz and S. T. Dumais [8] proposed a technique for filtration that analyzes the user's interest by analyzing the user feedback and keywords provided by the user. Keyword matching and Latent Semantic Indexing technique are used in it. By this user with similar preference are considered. Zelikovitz and Hirsh [9] proposed a system which enhance the classification of short text string building up a semi supervised learning based on a combination of labelled training data plus a secondary corpus of unlabelled but related longer documents.Bobicev and Sokolova [10] focus on circumvent the issue of error-prone feature development by receiving a statistical learning method that can perform sensibly well without highlight designing. On the other hand,, this system, called as Prediction by Partial Mapping, delivers a dialect model that is utilized as a part of probabilistic content classifiers, which are hard classifiers in nature and do not easily integrate soft, multi-membership paradigms.

III.PROPOSED SYSTEM

A. Filtered Wall Architecture

The main goal of the system is to filter the unwanted wall content posted by the other user on the particular user's wall. This post can be in text form or in an image form. The system should analyse the text / image content and allow desired content on the wall. In this system, when particular post is arrived to be published on his wall, all personal settings are considered and wall post is filtered accordingly.

While filtering the text message, it is first checked that whether it is from an authentic user or not. If it is from an authentic user then its content is analysed and properly categorizes using text classification techniques. Then system checks whether user preference is matching with derived post category. If it is matched, then particular post is published as it is on hold till user permits it. As shown in fig. 1 there are three modules

1) Message Filtering: For message filtering purpose, we have to extract textual data from user's wall. Whenever any user upload textual data on general wall, with the help of short text classification phase, the system can classify those data into different categories with the help of dataset.

2) Labelled Image Filtering: As part of the contribution we are filtering labelled image. With the help of OCR techniques, we can extract the text from an image and apply message filtering rules.

3) Image Filtering. As part of contribution, we are also implementing the vulgar image filtering. For vulgar image filtering we used two algorithms. Skin detection algorithm and pornography detection algorithm.

We also manage black list of users for a particular member so that the sender is automatically blocked till user allow him. Our system provides auto blocking facility to the

DOI 10.17148/IJARCCE.2015.41114

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

user. If a particular user wants to block someone, then he/she can manually block that user.

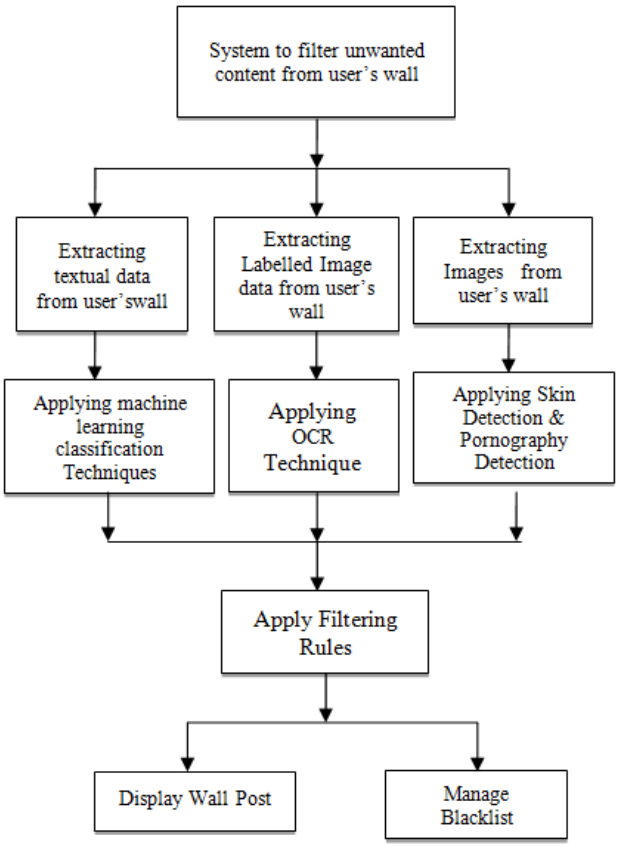

Fig1.Filtered Wall Architecture

B. Algorithm

1) OCR Algorithm.

The OCR (Optical Character Recognition) algorithm basically depends on a set of learned characters. It compares all characters in the image file with the characters in the learning set. Learned set requires an image file with the desired characters in the desired font be created, and a text file depicts the characters in this image file. Testing image is given to the user as input. In pre-processing phase, the text gets extracted from the image. Features extraction method searching thematch between the features extracted from the given character's image and the library of image models. If match found, it display classified a letter of the image. [15]

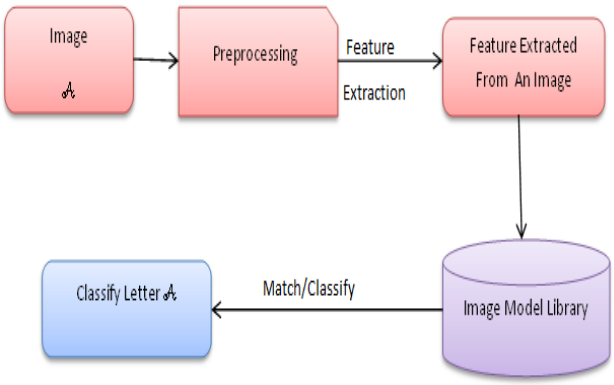

Fig.2. OCR Processing

2) Lemmatization Algorithm.

Lemmatization is the process of stemming the words. Using this process words is traced are converted to their basic forms, e.g. "playing" is word converted to "play" after lemmatization.

Lemmatization actually has two steps

Step 1: Word forms are paradigmatically associates with basic forms (lexemes), which serve as lemmas. This process is called lemmatization.

Step 2: Lemma selection is a process in which basic forms fetched from step first is selected for adding it into the dictionary.

3) KNN Algorithm.

This algorithm is used for text classification. Text classification is the process of assigning various short texts provided by user to one or more target categories based on its content. KNN algorithm is used to classify instances based on nearest training examples in the frame space. The object is classified into the particular class which has maximum number of nearest instances. Following are the steps of KNN algorithm execution [14].

- To Store the output values of the M nearest neighbours to query scenario Q in the vector r=(r1….rm) by repeating the following loop M times.
  a) Go to the next scenario Si in the data set, where I is the current iteration within the domain {1……P}
  b) If Q is not set or q < d (q, Si): q ←d (q, Si), t ←Oi
  c) Loop until we reach the end of the data set.
  d) Store q into vector c and t into vector r.
- Calculate the arithmetic mean output across r as follows:

$$r = 1/M \sum_{i=1}^{M} n$$

- Return r as the output value for the query scenario q

4) Skin Detection Algorithm.

Skin Detection Algorithm is to extract all skin areas of an image. We group and label all pixels independently and mark them in a binary image. We label pixels that, we characterize as skin in grey, and pixels we order as non-skin pixels in white. Consequently, we separate binary or grey levelled image with labels for skin and non-skin pixels [16].

Skin detection algorithm steps:

1. Scale the image to a width of less than 100 pixels.
2. Execute an auto contrast function to maximize the image's contrast.
3. skinmapNewImage(imgwidth,imgheight,white)
4. for all pixels in img do

R, G, B ←pixel

H, S, V ← ConvertRGBtoHSV(R, G, B)

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

If IsSkin (R, G, B, H, S, V) then

skinmap [pixel x, pixel y] ← grey

else

skinmap [pixelx,pixely] ←white

end if

end for

grey closing(skinmap,size ← (6,6))

return skinmap

5) Pornography Detection Algorithm.

This algorithm calculates a bounding polygon around the big three ski areas of an image. The percentage of skin pixels inside the polygon area are then used for classification. Other important features are the total amount of skin in the image, the number of skin areas and the sizes of the three largest ski areas. The idea behind this set of heuristics is that skin maps from pornographic content are large, connected areas, while ordinary pictures show these areas disconnected by clothes [16].

Pornography Detection algorithm steps:

1. function ClassifySkinmap(skinmap)
2. skinareasFloodFill(skinmap)
3. EliminateSmallAreas(skin areas)
4. features ← []
5. for skin area in skin areas do
   Features ← Extract Features (skinners)
6. end for
7. skin areas ←Shape Elimination(skin areas)
8. if CheckSpatial(skin areas) then
   Return false
9. end if
10. if CheckFace(skin areas) then
    Return false
11. end if
12. ConvexHull(max(skinareas,3))
13. ConvexHullFillRate
14. Scale Features(features)
15. is Nude ←SVMClassification(features)
16. Return is Nude
17. end function

6) Data Set.

We are using message dataset downloaded from [11] to categorize the message, we get word list. We have 2 types of word category. One is violence-vocabulary [11] and the other is banned word list [12]. The category list is in update format. User can add our own words in the list. For Testing, we are using spam message dataset [13] to classify messages. User can write his own messages that can also be getting filtered.

C. Filtering process

In defining the language for FRs specification, we consider three main issues that influence a message filtering decision. First, in OSN, the same message may have distinctive implications and pertinent in light of whom composes it. As an outcome, FRs should permit users to state limitations on message makers. Makers on which a FR applies can be chosen on the premise of a few distinctive criteria; a standout amongst the most significant is by forcing conditions on their profile's attribute. For example, conceivable to characterize rules applying just to youthful makers or to inventors with a given religious/political perspective.

D. Blacklisting Process

A further segment of our system is a Blacklist (BL) system to avoid messages from undesired makers, free from their substance. BL is is clearly administered by the by the system, which should have the ability to figure out, who are the users to be added in the BL and choose when user's maintenance in the BL is done. To upgrade adaptability, such data are given to the system through an arrangement of guidelines, for this Point forward called BL rules. Such principles are not characterized by the Social Network Management; consequently they are not implied as broad abnormal state mandates to be connected to the entirety group. Maybe, we choose to let the clients themselves, i.e., the wall proprietors to indicate BL principles directing who must be banned from their wall and for to what extent. Consequently, a user may be banned from a wall, and at the same time, he won't have the capacity to post in the wall.

IV. EXPERIMENTAL EVALUATION

An experiment is conducted with the short text classifier and filtering rules, by categorizing messages posted on the Social Networking Site. The graph shows the different categories of messages, which are accomplished by the proposed system.

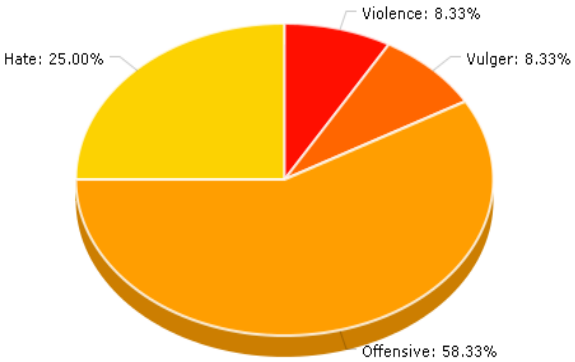

Fig.3 Graphical Representation

A. .Results

As shown in fig. 4 it displays the page where users can select friends wall, put a message and able to post the comment.

Fig.5 shows the number of block messages which are banned by the system.

DOI 10.17148/IJARCCE.2015.41114

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 11, November 2015

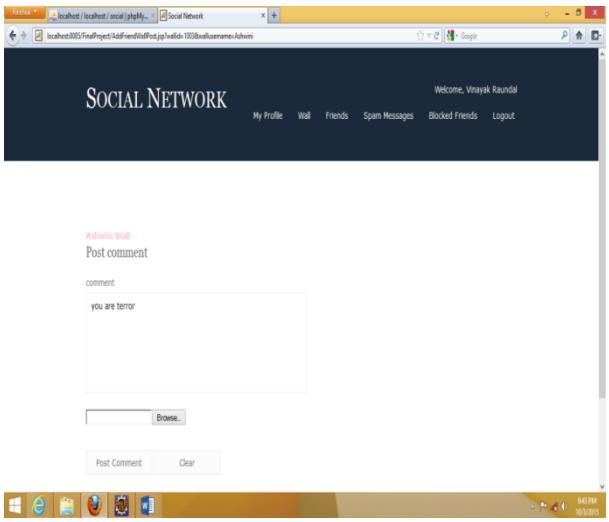

Fig 4.Post Comment Page

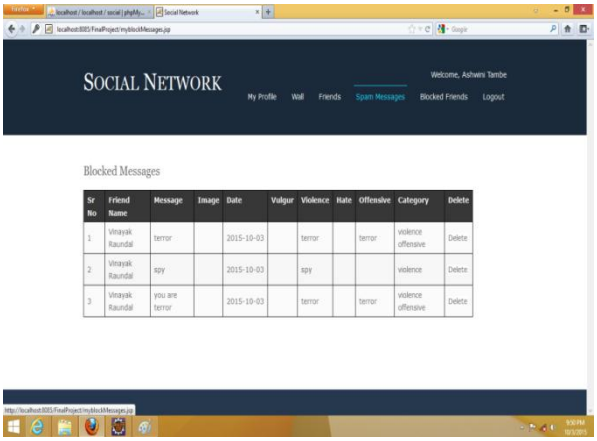

Fig.5.Block Messages

Fig.6 shows the unwanted labelled images blacklisted by the system. It will also depict the image which is posted by the user and extracted text from the image.

Fig.7 shows the vulgar image post by the user. As per skin detection algorithm, first convert the image into grey and white pixel, and then used Pornography detection algorithm.

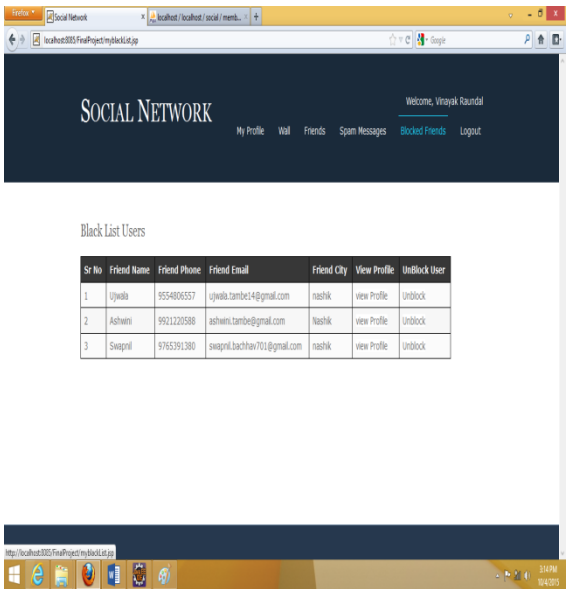

Fig.7 Blacklist Users

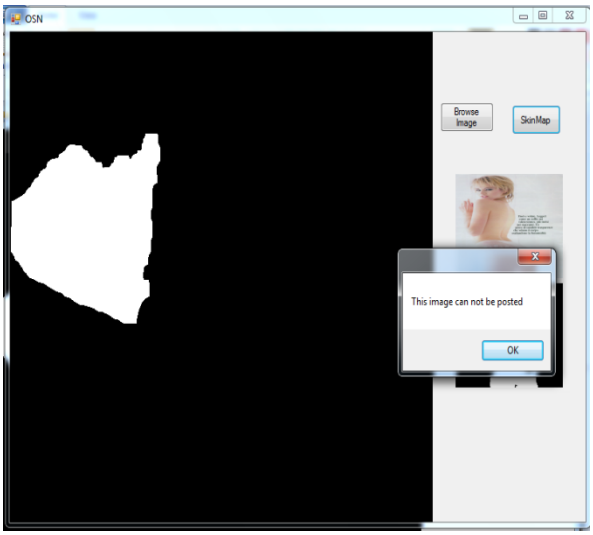

Fig.8.Image Filtering

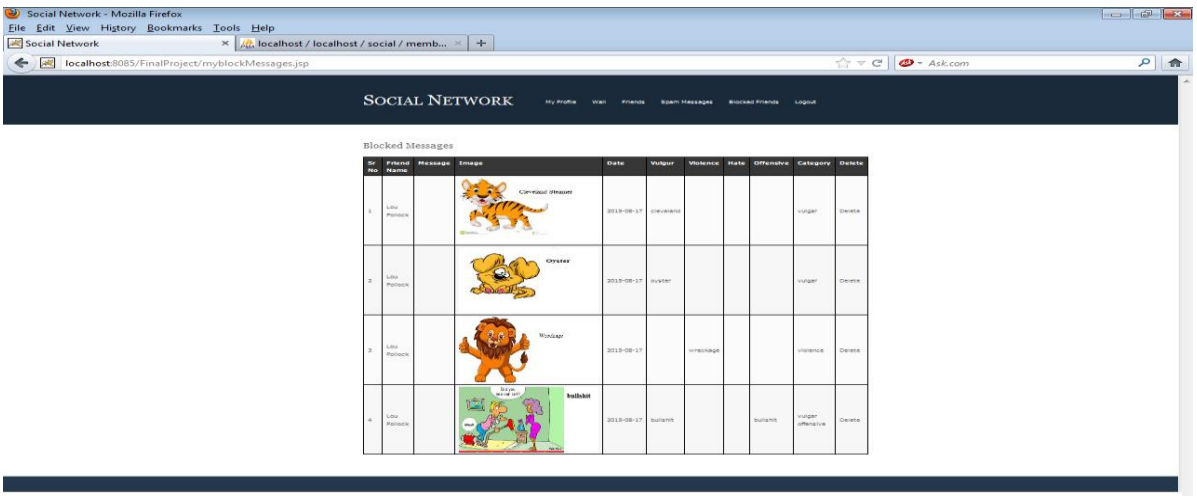

Fig.6.Blacklisted labelled Image Messages

DOI 10.17148/IJARCCE.2015.41114

We also provide an auto blocking facility by using threshold value. If the particular user post more than three undesired content on the wall, then it automatically block by admin.

V. CONCLUSION

With the expanding utilization of online informal communities, there is an increasing interest for evacuation of undesired posts; a fundamental usefulness gave by all social network applications, which spam a user's wall. There is a need for an automated filtering system that filters undesirable messages with great precision. In our system, we provide text as well as image filtration from user wall.

We have used KNN algorithm for short text classification and SVM technique is used to filter images. We have provided a facility to the user to add a custom filtration rule. This provides personalization in wall filtration. We provide customized Black list facility to the user. User specific setting can adaptively manage the user's personal black list. Currently filtering rules are defined by the user.

REFERENCES

[1] Vanetti, ElisabettaBinaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo Department of Computer and Communication, University of Insubria "a system to filter unwanted messages walls OSN user" IEEE Transactions on Knowledge And Engineering Flight Data: 25 Year 2013.

[2] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002. J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, "Predicting positive and negative links in online social networks," inProc. 19th Int. Conf. World Wide Web, 2010, pp. 641-650.

[3] Vinaitheerthan Renganathan1, Ajit N Babu2 and SuptendraNath Sarbadhikari3 " A Tutorial on Information Filtering Concepts and Methods for Bio-medical Searching".

[4] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas,"Short text classification in twitter to improve information filtering," in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, 2010, pp. 841–842.

[5] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in Provenance and Annotation of Data, ser. Lecture Notes in Computer Science, L. Moreau and I. Foster, Eds. Springer Berlin / Heidelberg, 2006, vol. 4145, pp.101–108.

[6] P. J. Hayes, P. M. Andersen, I. B. Nirenburg, and L. M. Schmandt,"Tcs: a shell for content-based text categorization," in Proceedings of 6th IEEE Conference on Artificial Intelligence Applications (CAIA-90). IEEE Computer Society Press, Los Alamitos, US, 1990, pp.320–326.

[7] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin?" Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992.

[8] P. W. Foltz and S. T. Dumais, "Personalized information delivery: An analysis of information filtering methods," Communications of the ACM, vol. 35, no. 12, pp. 51–60, 1992.

[9] S. Zelikovitz and H. Hirsh, "Improving short text classification using unlabeled background knowledge," in Proceedings of 17th International Conference on Machine Learning (ICML-00), P. Langley, Ed. Stanford, US: Morgan Kaufmann Publishers, San Francisco, US,2000, pp. 1183– 1190.

[10] V. Bobicev and M. Sokolova, "An effective and robust method for short text classification," in AAAI, D. Fox and C. P. Gomes, Eds. AAAI Press, 2008, pp. 1444–1445.

[11] http://www.bannedwordlist.com/lists/swearWords.txt

[12] http://www.myvocabulary.com/word-list/violencevocabulary

[13]https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection

[14] Priti Gaur, International Journal of Computer Science and Mobile Computing "An Efficient Routing Implementation Using Genetic Algorithm"Vol.2 Issue. 7, July- 2013, pg. 250-257

[15] Qing Chen"Evaluation of OCR Algorithms for Images with Different Spatial Resolutions and Noises"

[16] Christian Platzer, Martin Stuetz, MartinaLindorfer"Skin Sheriff: A Machine Learning Solution for DetectingExplicit Images"