An Automated System to Filter Unwanted Message from OSN User Wall

Mr. Sagar A. Dhanake[1], Prof. V.S. Nandedkar[2]

PG Student, Comp. Engg, PVPIT, Pune, India [1]

Assistant Professor, Comp. Engg, PVPIT, Pune, India [2]

Abstract: One fundamental issue in today On-line Social Networks (OSNs) is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. Up to now OSNs provide little support to this requirement. To overcome this problem, we propose a system allowing OSN users to have a direct control on the messages posted on their walls. This is achieved through a flexible rule-based system, that allows users to customize the filtering criteria to be matter-of-fact to their walls, and a Machine Learning based soft classifier automatically labelling messages in content-based filtering.

Keywords: Facebook 4j API, Filtered walls, Machine Learning, Filtering Rules, Text Categorization.

I. INTRODUCTION

Today's modern life is completely based on Internet. Now a day's people cannot imagine life without Internet. From last few years people share their views, ideas, information with each other using social networking sites. Such interchanges might include diverse sorts of substance such as text, image, audio and video data. According to Facebook statistics average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month [1]. Information Filtering has been widely used and employed for the textual documents and web contents. However, the goal of this proposal is mainly to provide categorization techniques to give the security to user walls from useless and meaningless data [2]. This is especially for that in OSNs, the users can comment the post in public/private areas of another user walls [6].

These comments can be useless or meaningless or unwanted messages. So, here information Filtering plays a vital role to protect the user walls in OSNs from undesired messages and give the authority to user to automatically control the undesired data on their walls [3]. A System which will give ability to users to control the messages posted on their own private space to avoid unwanted messages displayed. Customizable Filtering Rules are used to filter the unwanted messages from OSNs users wall as well as Machine Learning approach, Short Text Classification and Black list techniques are applied on Users Wall [6]. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data [2]. This application is useful for common people who don't want to write any unwanted messages like vulgar, political, sexual messages on his / her wall by any third person [3]. OSNs provide very little support to prevent unwanted messages on user walls. For example, Facebook allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends) [1]. However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them [2]. However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design ad hoc classification strategies. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences [1] [2].

A. NEED
In today's OSN, there is a very high chance of posting unwanted content on particular public/private areas in order to control this type of activity and prevent the unwanted messages which are written on user's wall we need to develop an automated system.

B. BASIC CONCEPTS
User interface screen will ask for the registration or log in screen. After log in to the system, for accessing the system functions, user will need to authenticate with facebook login. So, will be directed to the facebook login page. If the user already authenticated with facebook, he will be directed to the system dashboard directly. A user friendly GUI will be provided. User will be able to access various features like Post Filtering, Auto Comment etc.

II. LITERATURE SURVEY AND BACKGROUND

N.Thilagavathi, R.Taarika "Content Based Filtering in Online Social Network using Inference Algorithm", 2014 International Conference on Circuit, Power and Computing Technologies [ICCPCT].

Information filtering is the process of providing appropriate information to the people who need it. It significantly searches for what actually concerns the textual document, specifically web contents, and offers a user with classification mechanism to avoid the unnecessary information. This information filtering process is used in the online social network for insightful objective. To facilitate the content based filtering, this article introduces the filtered wall architecture. It will filter the incoming post based on the content. The main goal of this system is to provide customizable content based message filtering for online social networks, based on machine learning techniques. Information Filtering Systems are designed to categorize the information which are generated dynamically and offer the information to the user fulfil their requirement. In the content Based Filtering system, each user is assumed to operate separately. So the filtering system selects the information based on the correlation between the content of the items and user preferences. In this system, Machine Learning based text categorization techniques are used to automatically allot each short text message with set of categories based on the content. Short Text Classifier is built to accurate extraction and set of discriminating feature in the message. The system also supports the user defined Blacklists (BL) that is, list of users that are temporarily blocked to post messages on the user wall.[1]

Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls", IEEE Vol:25, Year 2013.

It was presented a system to filter undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent FR's. Moreover, the exibility of the system in terms of filtering options is enhanced through the management of BLs. The development of a GUI and a set of related tools to make easier BL and FR specification is also a direction we plan to investigate, since usability is a key requirement for such kind of applications.

A preliminary work in this direction has been done in the context of trust values used for OSN access control purposes. the developed Facebook application is to be meant as a proof of concepts of the system core functionalities, rather than a fully developed system. Moreover, it is aware that a usable GUI could not be enough, representing only the first step. In this context, many empirical studies have shown that average OSN users have difficulties in understanding also the simple privacy settings provided by today OSNs. To overcome this problem, a promising trend is to exploit data mining techniques to infer the best privacy preferences to suggest to OSN users, on the basis of the available social network data. [2]

K. Babu, P.Charles, "A System to Filter Unwanted Words Using Black List in Social Networks", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5(2), 2014.

A system to prevent the indecent messages from the Social Networking site walls has been presented. The Usage of Machine Learning has given higher results to the system to trace the messages and the users to distinguish between the good and bad messages and the authorized and unauthorized users in the Social Networking User Profiles automatically. Machine Learning Technique plays a vital role in this paper in order to generate the blacklist of the bad words and the unauthorized users. The user has to update his privacy setting in his account in order to add this method to prevent the obscenity in his public profile. Machine Learning is a system which can learn from the data and take decisions based on the learned data. A further component of our system is a Blacklist (BL) mechanism to avoid messages from undesired creators, independent from their contents. BL is directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when users retention in the BL is finished. A system automatically filters unwanted messages using the blacklists on the basis of both message content and the message creator relationships and characteristics. [3]

J.Anishva Rose, A. Pravin, "Machine Learning Text Categorization in OSN to Filter Unwanted Messages", International Journal of Computer Science and Information Technology Vol. 5 (1) , 2014.

Better privacy is given to the OSN wall using our system. System plan to implement the filtering rules with the aim of bypassing the filtering system, it can be used only for the purpose of overcome the filtering system. In this Blacklist mechanism is used, where the user's list will be avoided for the moment to post on user wall. All classification and filtering rules will be included, additionally BL rule is used. Based on the user wall and relationship, the owner of the wall can block the user. This prohibition can be approved for an uncertain period of time. Short text classifier is to recognize and eradicate the neutral sentences and categorize the non neutral sentences in step by step, not in single step. This classifier will be used in hierarchical strategy. Filtering rules will be applied, when a user profile does not hold value for attributes submitted by a FR. BL rule, owner can identify which user should be blocked based on the relationship in OSN and the user's profile. The user may have bad opinion about the users can be banned for an uncertain time period. [18]

III. PROBLEM STATEMENT

To develop a system that will filter unwanted messages from user walls using filtering rules which will overcome the problem of posting of undesirable messages on user's wall.

A. EXISTING SYSTEM
- Today OSN provides very little support to prevent unwanted messages on user walls.
- However, no content based preferences are supported and therefore it is not possible to prevent undesirable messages.

B. PROPOSED SYSTEM

The aim of the present work is to propose and experimentally evaluate an automated system called filtering wall that is able to filter unwanted messages from OSN user walls. We exploit machine learning text categorization techniques to automatically assign with each short text message a set of categories based on its content. Our contribution is that we are going to implement real time system using facebook app.

C. SYSTEM FEATURES

• Filter Unwanted Message (Before Posting on Wall)
This will not allow the user to post any message if it contains any abusive or vulgar word. The system will apply the filter on the posts and totally remove all the vulgar, unwanted, abusive words from that post.

• Analyse and Remove the Unwanted Post from User Wall (Real Time Posts)
This feature will read the posts from user's wall and analyse it. After analysis, the post will either be removed or deleted from the user's wall

• Auto Comment on a Particular Post
This feature will be able to make automatic comment on a particular post. First, the posts will be analysed and accordingly, the automatic comment will be given on that particular post like on the users Birthday, he will be having headache to check the posts and give comment "THANK YOU" for each post. So this will be automated.

IV. DETAILED DESIGN AND IMPLEMENTATION

In today's OSN, there is a very high chance of posting unwanted content on particular public/private areas, called in general walls. So, to control this type of activity and prevent the unwanted messages which are written on users wall we can implement filtering rules (FR) in our system. The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content.

We distribute our system into five modules, these are:

1. First of all the admin log on to the user.
2. After logging into server, admin see the friend's request, message send by other members on Online Social Network.
3. System First filter the message by using filtering rules.
4. System will check the type of the message, whether it will social, vulgar, sexual, non-neural.
5. If the unwanted message is occur then the message goes back to the user and it will not posted on user wall.
6. The person sending this type of message repeatedly then that person will directly blocked and goes to blacklist.

7. System will calculate trust value for each message it will give the trustworthiness of user.
8. If admin's friend post message "Happy Birthday" or "Congratulations" etc. types of messages then system can automatically send message thank you.
9. System can not only work English language but also work Hindi and Marathi language.
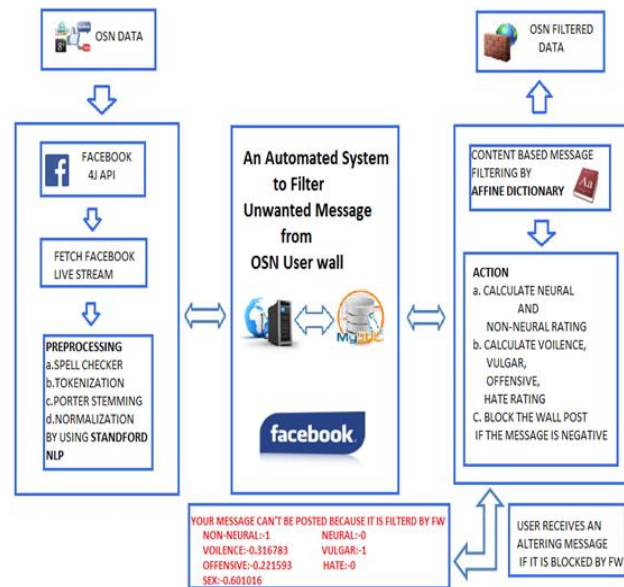
A. SYSTEM ARCHITECTURE

Fig1: System Architecture

The system aims to investigate the utility of linguistic features for detecting the sentiment of the posts done on person's time line wall. System will make use of FACEBOOK4J and FACEBOOK GRAPH API to integrate and communicate the system with facebook.

• Content-Based Messages Filtering (CBMF): For content-Based Massages filtering, we first filter out duplicate tweets and facebook comments, non-English tweets and non English facebook comments, and tweets that do not contain hashtags. From the remaining set (about 4 million), we investigate the distribution of hashtags and identify what we hope will be sets of frequent hashtags that are indicative of positive, negative and neutral messages. These hashtags are used to select the tweets that will be used for development and training.
• Short Text Classifier: Designing and evaluating various representation techniques in combination with a neural learning strategy to semantically categorize short texts.
• Integrate the System with facebook: The system will integrate with facebook and able to read the real time posts from users wall.
• Access Token Generation: As soon as the user logs in to facebook, the access token will be generated for that particular user.
• Post Reading from User Wall and Analysis:
a. With the help of that access token, the system will be able to read all the posts from user's timeline.
b. Pre-processing and NLP.

- Pre-processing:
a.   Tokenization: First of all we did the tokenization by which sentences are split into the words.
b.   Normalization: After that we used Stanford NLP to remove stop words from all the words.
c.   Part-of-speech (POS) tagging: Detects if the word token is noun, verb, and adjective.

- NLP and Feature Extraction:
1. Apply Stanford NLP to separate part of speech from the sentence.
2. Porter Stemmer Algorithm will be applied for getting root of the word for adjectives.
3. After getting root of the word, we will compare weight / sense of each word with the affine dictionary.
4. Finding negative annotations in the sentence and reverse the weight.
5. Calculate overall weight using emotions approach.
6. Sum up both to draw final conclusion
7. Finally, positive, negative or neutral count for that particular post will be calculated.

- Action on Post: After the analysis, the action on the posts will be taken accordingly, whether to publish the post or not on the users wall. If found negative sense, the system won't allow the user to make the posts on his / her friends wall. In case of real time fetched posts, the system will either delete or hide the posts depending on the users choice.

- Maintenance: As said above, the access token will get expired after two months, the user will just needs to log in with facebook once in two months.

- Design and Implementation Constraints:
a. FB Login: User should login with his facebook account through the system for getting the access token needed by the system.
b. Access Token Renewal: As per FACEBOOKs constraints, User needs to login with the facebook once in two months two renew it.

V. ALGORITHMIC STRATEGY

Pre-processing
The primary aim of the pre-processing phase is to remove from the input message all characters and terms that can possibly affect the quality of group descriptions.

Pre-processing steps:-
/** Phase 1: Pre-processing*/.
Do text filtering;
Identity the document's language;
Apply stemming; mark stop words;

Algorithm 1 Pre-processing

1: d <- input message (Pre-processing).
2: for all d D do
3: perform text categorization
4: if d! = null then
Filter text for unwanted symbols
5: apply stemming and mark stop words in d;
6: end for

There are three steps to the pre-processing phase: Text filtering, Stemming and Stop words marking.

Stemming:
Stemming algorithms are used to transform the words in the texts into their grammatical root from, and are mainly used to improve the Information Retrieval System's efficiency. To stem a word is to reduce it to a more general form, possibly is root. For Example, stemming the term interesting may produce the term interest. Though the stem of a word might not be its root, we want all words that have the same stem to have the same root.

Elimination of Stop Words:
After stemming it is necessary to remove unwanted words. There are 400 to 500 types of stop words such as "of", "and", "the", etc., that provide no useful information about the message. Stop-words removal is the process of removing words. Stop-words account for about 20% of all words in a typical document. These techniques greatly reduce the size of the searching and matching each word in message. Stemming alone cam reduce the size of an index by nearly 40%.

Algorithm 2 Porter Stemming

Require: Words suffix stripping
1: Gets rid of plurals and -ed or -ing suffixes.
2: Turns terminal y to i when there is another vowel in the stem.
3: Maps double suffixes to single ones: -ization, -ational, etc.
4: Deals with suffixes, -full, -ness etc.
5: Takes o_ -ant, -ence, etc.
6: Removes a final –e

VI. MATHEMATICAL MODEL

A. For Filtering Rules:
1) Input
Filtering Rules are customizable by the user. User can have authority to decide what contents should be blocked or displayed on his wall by using Filtering rules. For specify a Filtering rules user profile as well as user social relationship will be considered. FR = Trustier, SOUs, Rule, TuV FR is dependent on following factors.
- Trustier
- Set of Users (SOUs)
- Rule
- Action
Trustier is a person who defines the rules.
SOUs denote the set of OSN user.
Rule is a Boolean expression defined on content.

2) Process
FM = SOUs, Rule == Category (Violence, Vulgar, Offensive, Hate, Sexual), TuV
- FM
- SOUs
- Rule
- TuV

Here,

- FM Block Message at basic level.
- SOUs Denotes set of users.
- Rule Category of specified contents in message.
- TuV is the trust value of sender.

In processing, after giving input message, the system will compare the text with the different categories which are prevented. If message found in that prevented type of category then message will display to the user that can't send this type of messages, and still the user wants to send the message he/she can continue with sending the message. The Trustier, who gets the message, but the words which are defended in the rule are sent in **** format. After getting the message the Trustier will give the Feedback (FB) to the sender and the sender will gain the TuV accordingly. Process denotes the action to be performed by the system on the messages matching Rule and created by users identified by SOUs. E.g. FM == Friends, Rule == Category (Vulgar, Sexual), TuV = 5 i.e. Trustier will accept the message from friends but message should not contain vulgar or sexual words. Message containing such words will affect the TuV of sender. Now the question arises, calculation of TuV.

3) Trust Value Calculations

The trust value of any user in OSN is dependent on the feedback they gain by the user must also be trust worthy. That's why the FB can be categorized into followings:-

1. Positive with content (PC):- Good FB, message is acceptable with objectionable content. This will increase the TuV of sender.
2. Positive without content (PWC):- Good FB, message is acceptable as this message does not contain objectionable content. This will increase the TuV of sender.
3. Negative with content (NC):- Bad FB, such message must not be sent again, which are against the Rule. This will decrease the TuV of sender.
4. Negative without Content (NWC):- Bad FB, message doesn't contain any objectionable content but the Trustier is giving negative FB. Such type of FB from Trustier will affect the TuV of it's own, and the TuV of sender will remain same.

So, based on above categories the TuV will be calculated as follows: - FB as 1 and 2

$$TuV = TuV + abs \left[(PC + PWC) / (NC + NWC) \right]$$

FB as 3 $TuV = TuV - [1 + (NC + NWC) / (PC + PWC)]$ for $[(NC + NWC) / (PC + PWC)] < 1$.

Otherwise, send system generated message to sender, FB Negative with content exceeds limit of Threshold Value (ThV) and deduct 5 points from TuV, so $ThV = TuV - 5$.

FB as 4 $TuV = TuV$ of sender, but $TuV = TuV - [1 + (NC + NWC) / (PC + PWC)]$ for Trustier.

4) Output

PFM = Rule, M||Y

- PFM Percentages of filtered message in a year or month. In general, more than a filtering rule can apply to the same user. A message is therefore published only

if it is not blocked by any of the filtering rules that apply to the message creator.

B. Blacklists

BLs is directly managed by the system. This should be able to determine the users to be inserted in the BL and decide when to retain user back from the BL. To enhance exibility, such information is given to the system through a set of rules, here after called BL rules.

1) BL rules

INPUT = Sender, FB, TuV, ThV Where

- Sender is the OSN user who is sending the message;
- FB is the Feedback gain by the sender after sending the message.
- TuV is the new Trust Value calculated as formulas specified in A.3.
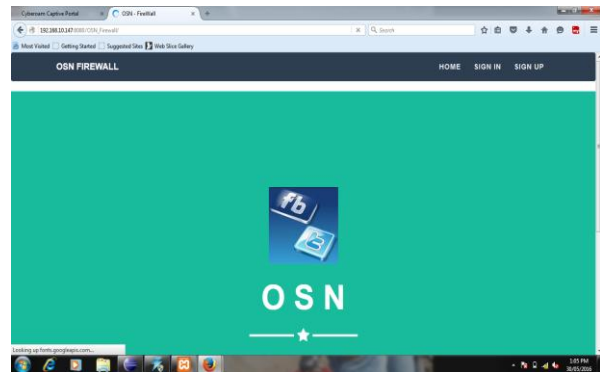- ThV is the Threshold Value.

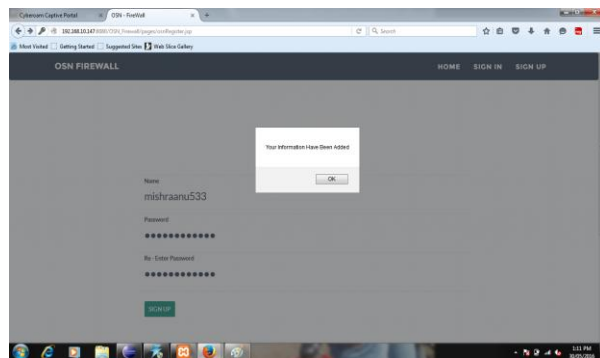## VII. RESULT AND EVALUATION

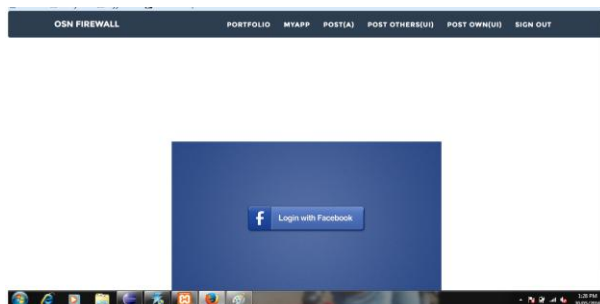Fig2: Login Page

Fig3: Registration Page

Fig4: Login into Facebook

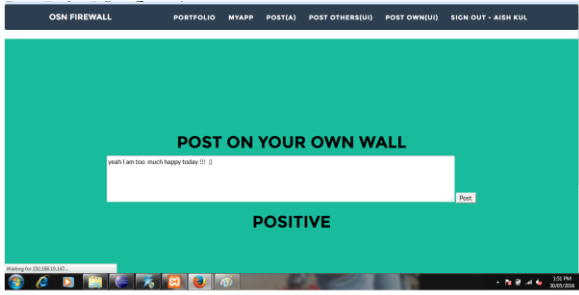**DOI 10.17148/IARJSET.2016.3611**

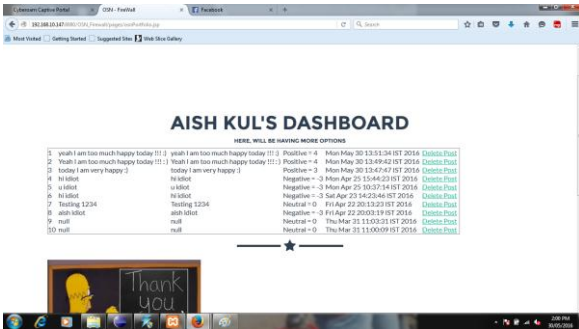Fig5: Posting on own wall and message Categorized

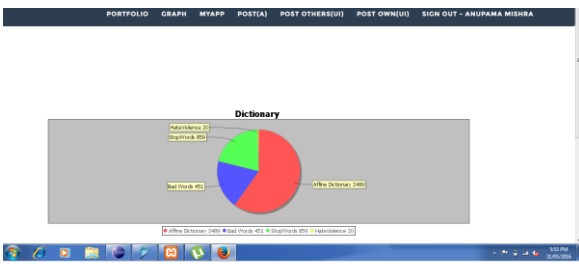Fig6: Message Categorized Dashboard

Fig7: Graph

Fig8: Posting Happy Birthday reply Automatically

Fig9: Post others

VIII.   CONCLUSION

We have presented a system to filter out unwanted messages from OSN user walls. The system exploits a Machine Learning soft classifier to enforce customizable content depended filtering rules. The exibility of the system in terms of filtering options is enhanced trough the management of BLs. we plan to enhance our filtering rule system, with a more sophisticated approach to manage those messages caught just for the tolerance and to decide when a user should be inserted into a BL. we proposed a system with the exible rules to filter the unwanted messages posted on user wall.

As the future work and our contribution we are going to implement a system to prevent undesirable messages from OSN walls. It makes OSN more reliable, secure, trustworthy and comfortable for the user.

ACKNOWLEDGMENT

REFERENCES

[1]   N.Thilagavathi, R.Taarika "Content Based Filtering in Online Social Network using Inference Algorithm", 2014 International Conference on Circuit, Power and Computing Technologies [ICCPCT].

[2]   K. Babu , P.Charles "A System to Filter Unwanted Words Using Blacklists in Social Networks", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014.

[3]   Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno "A System to Filter Unwanted Messages from OSN User Walls" IEEE Transaction on knowledge and data engineering, vol. 25, no. 2, February 2013.

[4]   Facebook development platform. http://developers.facebook.com

[5]   Ms. Shruti C. Belsare, Prof. R.R. Keole, "OSN user filtered walls for unwanted messages using content mining" march-2014.

[6]   Dipali D. Vidhate, Ajay P. Thakare "To avoid unwanted messages from osn user wall: Content Based Filtering Approach", IJCSMC,vol.3, Issue.4, April 2014.

[7]   M. Vanetti, E. Binaghi, B. Carminati, M. Carullo and E. Ferrari "Content-based Filtering in On-line Social Networks" 2010.

[8]   Dactchayani Sukumar1, Maniraj .S.P "Context Aware Message Filtering in OSN user Walls" IJIRCCE, Vol. 2, Issue 3, March 2014.

[9]   Rashmi R. Atkare, P.D.Soni " Survey of Filtering System For OSN (Online Social Networks)", IJCTA, Vol 4 (6),Dec 2013.

[10] Aarti Supe , R.N. Phursule "A System to Clarify Unwanted Messages from OSN User Surface", International Journal of Scientific Engineering and Research(IJSER), Vol. 3, Issue 5, May 2015.

[11] Bhushan Jagtap, Rahul Rathod, Vitthal Anpat, Utsav "A System to Filter Unwanted Messages from OSn User Walls", International

Journal of Advance Foundation and Reasearch in Computer(IJAFRC), vol. 2, Special Issue, January 2015.

[12] Vikrant Sanghavi, Amol Nanaware, Divya Nadar, Chitra Bhole "A System to Filter Unwanted Messages from OSN User Wall", International Journal of Research in Advent Technology, Vol. 1, Issue 5, December 2013.

[13] Mayuri Uttarwar, Yogesh Bhute "A Review on Customizable Content-Based Message Filtering from OSN User Wall", International Journal of Computer Science and Mobile Computing(IJCSMC), Vol. 2, Issue 10, October 2013.

[14] Aishwarya Kashyup, Himanshu Kothari, Ravina Kothari, Geetika Narang "Filtering Mechanism for Unwanted Messages on OSN User Wall Using CBMF(Content Based Message Filtering)", International Journal of Advanced Research in Computer and Communication Engineering(IJARCCE), Vol. 4, Issue 2,February 2015.

[15] Amruta Kachole, S.D.Jondhale "Unwanted Messages Filtering System from OSN User's Wall Using Customizable Filtering Rules and Black list Techniques", International Journal of Emerging Technology and Advanced Engineering(IJETAE), Vol. 4, Issue 2, February 2014.

[16] Sruthi. T, Greeshma T.R. "Filtered Wall-An Online Social Network Filter", International Journal of Science, Engineering and Technology Research (IJSETR), Vol. 3, Issue 4, April 2014.

[17] K. Arulmurugan, P. Ranjitkumar "An Analysis of Unwanted Messages Filtering Methods from OSN User Walls", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 2, Issue 1, March 2014.

[18] J.Anishya Rose, A.Pravin "Machine Learning Text Categorization in OSN to Filter Unwanted Messages", International Journal of Computer Science and Information Technologies (IJCSIT), Vol 5(1), 2014.

BIOGRAPHY

Mr. Sagar A. Dhanake Completed his B.E. from Savitribai Phule Pune University, Pune. Interested in Data Mining and Computer Networking Area, M.E. pursuing from Dept. of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune, Maharashtra, India.

Prof. V.S. Nandedkar Assistant Professor, Dept. of Computer Engineering, Padmabhooshan Vasantdada Patil Institute of Technology, Bavdhan, Pune, Maharashtra, India.