

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

*IJCSMC, Vol. 6, Issue. 4, April 2017, pg.195 – 200*

# Review Paper: A STUDY ON LITERATURE SURVEY ON AUTOMATED FILTERING OF UNWANTED MESSAGES ON OSN

**Ms. Shravani Krishna K<sup>1</sup>**

PG Scholar, Department of Computer Science and Engineering,  
New Horizon College of Engineering, Bangalore, Karnataka, India

<sup>1</sup> [kamatham.shravani@gmail.com](mailto:kamatham.shravani@gmail.com)

**Mr. Muralidhara S<sup>2</sup>**

Sr.Assistant Professor, Department of Computer Science and Engineering,  
New Horizon College of Engineering, Bangalore, Karnataka, India

<sup>2</sup> [muralidhars@newhorizonindia.edu](mailto:muralidhars@newhorizonindia.edu)

**Abstract:** *One fundamental issue in today's Online Social Networks (OSN's) 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, OSN's provide little support to this requirement. To fill the gap, in this paper, 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, which allows users to customize the filtering criteria to be applied to their walls, and a Machine Learning-based soft classifier automatically labelling messages in support of content-based filtering.*

**Keywords:** *Online social networks, Information filtering, short text classification, Trust.*

## INTRODUCTION

Social network is interactive medium to share and communicate some amount of data related to human life.

OSN is used to share some content type. It could be image, text, video and audio. It is platform to build relationship among people who are interested in sharing views, picture, real time connections and texts.

Social network provide various types of services such as profile, social links. It allows you to create a list of user with whom you want to communicate, to create a public profile and view the connections within system.

Given architecture is three-tier architecture

1) First layer is called as Social Network Manager (SNM)-It aims to provide the basic functionalities (i.e.) profile and relationship management.

- 2) Second layer provides support for external Social Network Applications (SNAs).
- 3) Third layer is Graphical user interface (GUI)-With the help of GUI user can interact with the system.
- 4) Filtering-It is used to filter the unwanted messages using blacklists.
- 5) Content Base filtering-It is used to select information based on the correlation between the content of the item and user performances.

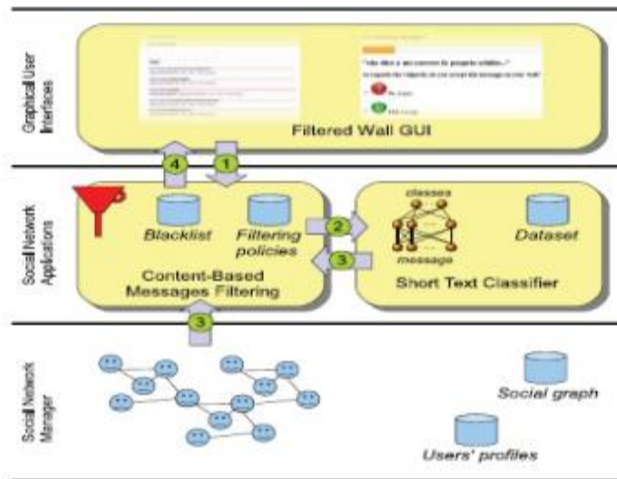


Fig.1: General Architecture of online Social Network

Some of the social network sites are Twitter, Facebook, and YouTube which are used worldwide. A social network can be defined as a set of factors and ties among them. Web mining is used to discover useful and related information from a large amount of data. Online social network information filtering can be used for a different purpose in OSN anyone can post on OSN wall as well as on others wall. With the help of information filtering user can able to control the messages written on their own wall by filtering out unwanted messages.

Currently OSN provides very less support to prevent not required messages on users wall for e.g. Facebook allows users to decide who is allowed to insert messages on their wall (i.e. Family members, friends, group of friends, friends of friends) However no-content based preferences are supported and hence it is not possible to prevent undesired messages such as vulgar or political one no matter who post them on the wall.

In our proposed system there are three methods message filtration by admin, message filtration by user and short text classifier. In message filtration by admin method the messages filtered by admin, Admin sets the word category. In message filtration by user messages are filtered by user, user sets the word category. In short text classifier short text word are set by admin in the data base.

**Message filtering** in this module, the unwanted messages are filtered. The other users that can send vulgar message to an OSN user is temporarily blocked by the OSN user. If the user sends vulgar messages that match the filtering pattern specified by the OSN user more than a specified threshold value than that user is unfriend permanently.

**Filtering Rules** in defining the language for FRs specification, there are three main issues that, should affect a message filtering decision. First of all, in OSNs like in everyday life, the same message may have different meanings and relevance based on who write sit. As a consequence, FRs should allow users to state constraints on message creators. Creators on which a FR applies can be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on their profile's attributes. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth and trust values of the relationship creators should be involved in order to apply them the specified rules.

The same message on OSNs may have different meanings and relevance based on who write sit. It is necessary to apply constraints on messages. Constraints can be selected on several different criteria's. User can state what contents should be blocked or displayed on filtered wall by means of Filtering rules. Filtering rules are specified on the basis of user profile as well as user social relationship. FR is dependent on following factors,

1. Author
2. Creator Spec
3. Content Spec
4. Action

An authorize person who defines the rules. Creator Spec denotes the set of OSN user and Content Spec is a Boolean expression defined on content. Action denotes the action to be performed by the system on the messages matching content Spec and created by users identified by creator Spec.

**Social Network Manager (SNM)** The initial layer is Social Network Manager Layer provides the essential OSN functionalities (i.e., profile and relationship administration).It also maintains all the data regarding to the user profile. After maintaining and administrating all users data will provide for second layer for applying Filtering Patterns (FPs) and Black lists (BL).

**Social Network Application (SNA)** In second layer Content Based Message Filtering (CMBF) and Short Text Classifier is composed. Also we are detecting phishing links and filtering images posted on user walls in this layer. This is very important layer for the message, images and link categorization. Also Black list is maintained for the user who sends frequently bad words in message. Links are filtered and the user the alerted if phishing link detected. Images are scanned and if found hidden messages are displayed.

**Graphical User Interface (GUI)** Third layer provides Graphical User Interface to the user who wants to post his messages as a input and filtered wall is provided. In this layer Filtering Rules (FR) are used to filter the unwanted messages and provide Black list (BL) for the user who are temporally prevented to publish messages on user's wall.

**Blacklists** users are those users whose messages are banned from their contents. BL rules allow the wall owner to decide users to be blocked on the basis of their profiles and relationship with wall owner. This prevention can be done for a specified period or forever according wall owner's desire. BL is dependent on author, creator specification and creator behaviour.

## Related Research Works

[1] Ying Chen, Yilu Zho, "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety", ASE/IEEE International Conference on Social Computing, 2012. Since the textual contents on online social media are highly unstructured, informal, and often misspelled, existing research on message-level offensive language detection cannot accurately detect offensive content. Meanwhile, user-level offensiveness detection seems a more feasible approach but it is an under researched area. To bridge this gap, we propose the Lexical Syntactic Feature (LSF) architecture to detect offensive content and identify potential offensive users in social media. Results from experiments showed that our LSF framework performed significantly better than existing methods in offensive content detection. It achieves precision of 98.24% and recall of 94.34% in sentence offensive detection, as well as precision of 77.9% and recall of 77.8% in user offensive detection. Therefore we can refer n-gram method based on this paper.

[2] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls", IEEE Transactions on Knowledge and Data Engineering, February 2013. One fundamental issue in today's Online 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 fill the gap, in this paper, 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 applied to their walls, and a Machine Learning-based soft classifier automatically labelling messages in support of content-based filtering concept, therefore we can refer Machine learning based short text classifier.

[3] Ramnath Balasubramanyan, Aleksander Kolcz, "wOOT! Feeling great today!" Chatter in Twitter: Identification and Prevalence", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013. Microblogging services like Twitter are used for a wide variety of purposes and in different modes. Here, we focus on the usage of Twitter for "chatter" i.e., the production and consumption of tweets that are typically non-topical and contain personal status updates or conversational messages which are usually intended and are useful only to the immediate network of the producers of the tweets. We study the prevalence of chatter tweets in Twitter and present techniques to detect them using machine learning techniques that require minimal supervision. From this paper we have referred the following concept taking example of twitter as a "chatter", we are able to study or classify tweets as per their ranks and then filtering out the ones that are of high relevance.

[4] A Review on Filter Undesired Text from Social Networks International Journal of Computer Applications (0975 – 8887) Volume 107 – No 14, December 2014 this paper explains techniques to filter undesired messages in online social network. Additionally, different machine learning techniques for classification purpose. In this article, proposed system represents the core set of functionalities needed to provide a proper tool for OSN message filtering. This system approach when the user decides to be inserted into a blacklist. The system developed GUI and a set of tools that make BLS easier and simpler specifications. Moreover, the flexibility of the system in terms of filtering options is enhanced by the direction of BLS. In future work, we can filter unwanted images from user wall. We refer content based learning and different machine learning technique

[5] Multidisciplinary Journal of Research in Engineering and Technology, Volume 2, Issue 2, Pg.398- 402, 2015 Filtration of unwanted messages from online social website user's wall in existing system user can block unwanted person but cannot block unwanted messages. Through proposed system user can block unwanted messages and the person depend on his trust value. Advantage of proposed system is that there is no need to block a person directly but you can block his/her bad messages and without blocking that person you can evaluate the person who is good or bad depend on trust value analysis.

[6] Filtering unwanted messages from osn walls 2016 1st International Conference on Innovation and Challenges in Cyber Security (ICICCS 2016) today various social networking sites are available which Make people remain in constant touch with each other. Sharing any type of data has become easy. There are great advantages of such social networking sites excepting a few minor drawbacks like poor security which create huge problems to people when they were active on such sites. As we have seen Facebook allows users to post comment on another user's wall even when they were unknown to each other. But if that comment is a vulgar one then it may cause serious problem to user reputation. To avoid such a problem Information filtering is used to filter the content of the message. So we have analysed various Information filtering methods like content based filtering, policy-based filtering, and collaborative filtering in this paper. Content-based filtering method is best filtering method than any other methods, because it has filtered out bad or non-neural words from the input message and allows posting only pleasant comment to be posted on a user's wall. This will help us to avoid unwanted messages from ever spoiling reputation which carries the utmost importance in the world of socialization.

## CONCLUSION

In this paper, we have presented a system to filter undesired messages from OSN walls. The system exploits a Machine Learning soft classifier to enforce customizable content-dependent Filtering Rules. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of Black Lists. The early encouraging results we have obtained on the classification procedure prompt us to continue with other work that will aim to improve the quality of classification. Since the underlying domain is dynamically changing, the collection of preclassified data may not be representative in the longer term. The present batch learning strategy, based on the preliminary collection of the entire set of labelled data from experts, allowed an accurate experimental evaluation but needs to be evolved to include new operational requirements.

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. In particular, we aim at investigating a tool able to automatically recommend trust values for those contacts user does not personally known. However, the design of these audit-based tools is complicated by several issues, like the implications an audit system might have on user's privacy and/or the limitations on what it is possible to audit in current OSNs. A preliminary work in this direction has been done in the context of trust values used for OSN access control purposes. Even if we have complemented our system with an online assistant to set FR thresholds, the development of a complete system easily usable by average OSN users. As such, the developed Facebook application is to be meant as a proof-of-concepts of the system core functionalities, rather than a fully developed system.

We explain techniques to filter undesired messages in online social network. Additionally, we studied different machine learning techniques for classification purpose. In this article, proposed system represents the core set of functionalities needed to provide a proper tool for OSN message filtering. This system approach when the user decides to be inserted into a blacklist. The system developed GUI and a set of tools that make BLS easier and

simpler specifications. Moreover, the flexibility of the system in terms of filtering options is enhanced by the direction of BLS. In future work, we can filter unwanted images from user wall.

## REFERENCES

- [1] Ying Chen, Yilu Zho, “Detecting Offensive Language in Social Media to Protect Adolescent Online Safety”, ASE/IEEE International Conference on Social Computing, 2012.
- [2] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo, ”A System to Filter Unwanted Messages from OSN User Walls”, IEEE Transactions on Knowledge and Data Engineer, February2013
- [3] A Review on Filter Undesired Text from Social Networks International Journal of Computer Applications (0975 – 8887) Volume 107 – No 14, December 2014
- [4] Filtration of unwanted messages from online social website „Multidisciplinary Journal of Research in Engineering and Technology, Volume 2, Issue 2, Pg.398- 402, 2015
- [5] Filtering unwanted messages from osn walls 2016 1st International Conference on Innovation and Challenges in Cyber Security (ICICCS 2016)