

# Content Based Filtering in Social Networking Sites using Web Application

M.Vamsi Krishna Reddy, Md Ali Hussain

**Abstract** — *In consideration to the today's globalized world everybody in the society are being addicted in using the Social Networks Sites.*

*The basic problem that we are gonna be seen in using these sites is "Lack Of Privacy". Till today, Social Networks Sites provide little support to this requirement. To sort out this problem, in this project we are proposing a system which will provide the indirect control to the users of these sites .This proposed model can be achieved through a modern rule-based system, that allows administrators to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically labeling messages in support of content-based filtering.*

**Index Terms:** *Social Networks Sites, Content-based filtering, Machine Learning, Rule-based system*

## I. INTRODUCTION

Social networking sites are one of the most efficient and effective medium of interaction. In today's society, we can't find a person who is not having a mobile phone neither with an account in social networking sites like Twitter, Facebook etc., Through the help of many existing surveys, we came to know that within the world we are having more number of mobile phone users rather than social sites users. And also even though a few people agree or not, neither mobile phones nor social networking sites are not providing "privacy" to the users. Only certain sites like [1]MyWot (<http://www.mywot.com>) is providing privacy. But it is limited to a certain area. Actually, it is a service which gives services to its users like rating the resources with respect to trustworthiness, child safety. It focuses on determining whether the browser should block to the desired resource or not. The system that we have been proposing is entirely different from the approach used by MyWot. That means only a few content is going to be filtered based on the rules that are being mentioned by administrators.

Web content mining strategies[3], are providing an active support in content filtering. The proposed system in mainly concentrates on content based filtering rather than collaborative based filtering or social filtering. The proposed content filtering system is designed for[6] unstructured or semistructured data, as opposed to database applications, which use very structured data.

Content filtering is a name used to describe a variety of processes involving the delivery of information to people who need it.

**Manuscript Received on May 2014.**

**M.Vamsi Krishna Reddy**, Student, Department of Electronics and computer Engineering, K L University, India.

**Dr. Md Ali Hussain**, Professor, Department of Electronics and computer Engineering, K L University, India.

Although this term is appearing quite often in popular and technical articles describing applications such as electronic mail, multimedia distributed systems, and electronic office documents, the distinction between filtering and related processes such as retrieval, routing, categorization, and extraction is often not clear. It is only by making that distinction, however, that the specific research issues associated with filtering can be identified and addressed. The aim of the majority of the proposals that are mentioned in the references are mainly to that provide users a classification mechanism to avoid the useless data in accordance to the administrators.

Filtering is based on descriptions of administrator. [6]Filtering is considered as removing of data from an incoming content or finding data in that content. In the first case, the users of the system see what is left after the data is removed like in an email filter designed to remove "junk" mail. Note that this means profiles may not only express what people want, but also what they do not want; in the later case, they see the data that is extracted with certain type of messages like the message contains bad or vulgar words. This list of features suggests that information filtering is a well-defined and unique process.

We are providing automatic filtering to the user through administrator. Here, Content based filtering in the sense that automatically controlling the unwanted messages in-between the users. We know that Facebook allows users to state who is allowed to insert messages in their walls. But it is not going to filter the unwanted messages i.e., the content may contain vulgar, offensive, sex, hate ones from the given content.

The main aim of our present work is proposing a system that will filter the content based on the filtering rules that are proposed by the administrator. Here, the entire control is under the administrator and the admin will propose "what are the content to be filtered and what not". This total work is done through the help of filtered wall. Here, we are gaining an advantage from Machine Learning text categorization techniques to automatically assign with each short text message a set of categories based on its content.

We also used the neural learning which is recognized as one of the efficient solutions in text classification. The proposed system is not going to block any users based on their content but it will block the content of the users. The major differences that are included in our proposed system are : 1) Content based filtering through filtered wall 2) Filtering rules/Filtered walls are proposed by administrator 3) Removing the blacklists concept i.e., list of users temporarily prevented to post any kind of messages on other users walls. 4) Administrator knowing everything about the chat between users, user details.

# Content Based Filtering in Social Networking Sites using Web Application

## II. RELATED WORK

In [1], the authors proposed 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 labeling messages in support of content-based filtering.

In [6], the authors proposed a system for information filtering and information retrieval. The systems proposed by them deals mainly with textual information, but they may also entail images, voice, video or other data types that are part of multimedia information systems..

In [5], the authors proposed a recommender systems improve access of retrieving user desired products based on user requirements and suggestions. The system here entirely uses collaborative filtering and uses machine learning algorithm for text categorization.

The work that is proposed by other authors like [2] M. Chau and H. Chen, [4]S. Pollock and [7] Brain Whitworth also useful in various in filtering the content based on certain analysis and some rule based systems.

Mainly the requirements(hardware/software) for technical systems as mentioned by the Brain plays an vital role in designing the real systems based on the proposed principles. Socio-technical systems use technology to connect people socially, for example through e-mail, electronic markets, social network systems, knowledge exchange systems, blogs, chat rooms, and so forth.

## III. SYSTEM DESIGN AND DESIGN GOALS

In the proposed system there are 3 major actors namely system, admin and user. admin can register,view user's profile,add filter words,view filter performance.

The user can update his profile image and status,can write comments on friends wall,can add new friends,can see other friends comments.

The system contains the databases of both admin and the user.It keeps track of records of each admin and user as new members are added every time.Admin logs in to the system and access the database of the users and performs filtering on the messages of each user.

design goals:

Here in this section we characterize certain design goals that are taken as pre requisites for designing the proposed work. These include aspects relating to access, privacy, storage and efficiency.

ACCESS:

Admin must register himself before his first login for obtaining the details of the user. Once he register admin can view user's profile,how many bad words a message contains and whenever admin filters any message then a pop up window is appeared to that particular user whose message is filtered.

PRIVACY:

When ever any user updates his status or his profile image immediately he wil logout of his account and he want to login again to view his status or image.it gives more security to the user.

STORAGE:

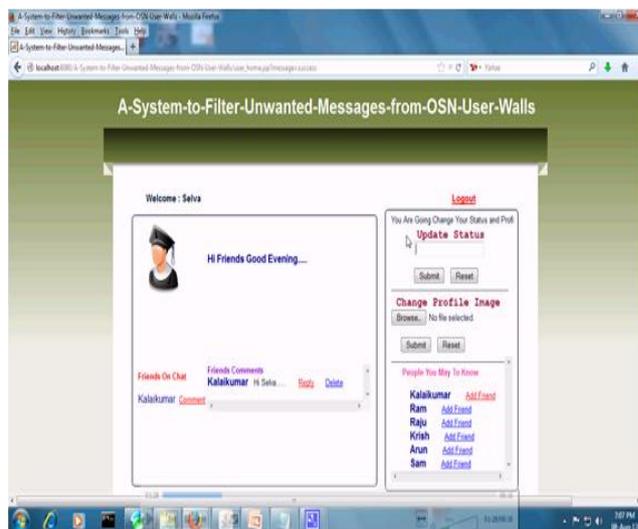
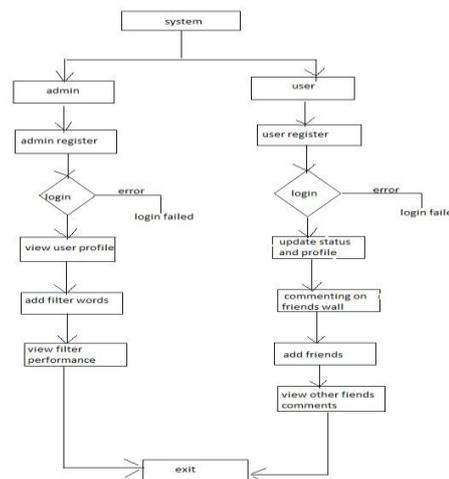
In the system we have many number of admins and users so the system provides a database which stores all these details.

When admin logs in he can directly access the users databases.

EFFICIENCY:

The efficiency of the proposed scheme works as follows: as we can have more number of admins the filter performance will be high. The proposed work is based on the content - based filtering which is in advance of the existing system.

Data flow diagram representing actions of the system as shown below:



User home page

Sno	Violence	Vulgar	Offensive	Hate	Sex	Total	Date & Time	Graph
9	0	1	0	0	0	1	25/03/2014 06:52:49	Graph

[View User](#)

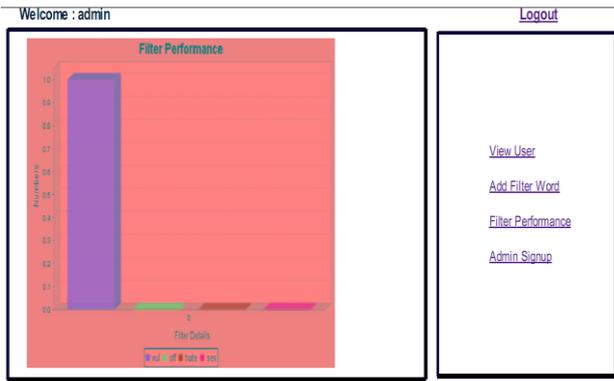
[Add Filter Word](#)

[Filter Performance](#)

[Admin Signup](#)

Message with vulgarity indication





Filter performance graph

#### IV. PROPOSED WORK

In the existing system everything is user controlled but our proposed system is about admin controlled .all the filtering analysis is done by the admin. when ever any user sends or chats with the other the information is stored in the database. The admin accesses this database and apply filtering techniques on each message content. 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, in support of content-based filtering.

If any user's message content is filtered immediately a pop up window will be displayed to that user saying that your message is filtered do you want to post the message .we have a filtering graph which represents how many bad words are used how many times by each user. The filtering technique we are using here is content based filtering which performs on the basis of content such as text. In existing system there is no content based filtering in our system we are implementing content based filtering.

Once you login to admin we can know how many users are there, he can add filter words like vulgar,violence,Sextual, Offensive, Hate type of messages and filter these messages. Administer can see monthly and yearly reports as well as Graphs like which category of messages are filtered (in percentage), who is message creator.

#### V. PERFORMANCE EVALUATION

This application is useful for common people who don't want to write any unwanted messages like vulgar, political, sexual messages on his/her own wall by any third person and the primary sOSN access control,we will provide only core set of functionalities which are available in current OSN networks services like Facebook, Orkut, Twitter, etc. some OSNs have difficulties in understanding,regarding privacy settings,but this problem will be overcome in the present OSN system.

#### VI. CONCLUSION

In this paper, we have presented a system to filter undesired content from user posting messages. Here the proposed system is limited to only certain extent.

This is because our developed application is applicable or implemented in only one system. Our work is done using facebook as a prototype. But the actual features that have

been offered by fb are not seen in our project. For example, two people can chat at a time using two different systems or through one system but ours is applicable to only one system. Even though this may be a difficulty problem in our proposed work our intention was providing privacy rather than simply implementing the existing features of social networking sites. we mainly focused on Privacy of users as mentioned above. The entire work i.e., specifying filter rules/filter words is going to be done by the administrator.The proposed procedure can be applied to any other social networking sites like twitter, gmail etc., The better privacy protection can be provided to users by using emerging data mining techniques.

#### FUTURE SCOPE

As specified from the above details, in future we can apply it to real applications so that we can work on it in a live environment. And also the basic principle of our application can be applied to facebook, twitter, whatsapp etc.,

#### REFERENCES

1. Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati and Moreno Carullo, "a system to filter unwanted messages from osn userwalls," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 2, february 2013.
2. M. Chau and H. Chen, "A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis," Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008.
3. M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, "Content-Based Filtering in On-Line Social Networks," Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.
4. S. Pollock, "A Rule-Based Message Filtering System," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.
5. R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in Proceedings of the Fifth ACM Conference on Digital Libraries. New York: ACM Press, 2000, pp. 195-204.
6. Nicholas J.Belkin and W.Bruce Croft " Information filtering and information retrieval: Two sides of the same coin?" in Communications of the ACM, DEC1992 V35 N12 P29(10).
7. BrainWhitworth notes on" The Social Requirements of Technical Systems"Massey University-Auckland,NZ.