ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Filtering Noisy Contents in Online Social Network by using Rule Based Filtering System

Bala Kumari P[1], Bercelin Rose Mary W[2] and Devi Mareeswari M[3]

M.TECH / IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur-628215, Tamilnadu, India[1, 2, 3]

ABSTRACT: The aim of the present work is to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter noisy messages from OSN user space. In Online Social Network may has possibilities of posting some noisy contents so it need to filter without displaying to owner space. It has achieved by using Rule based System. The Rule Based System allows users customize to filter by applying Filtering Criteria. This work exploits Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content. Online social networks not only make it easier for users to share their opinions with each other, but also serve as a platform for developing new filter algorithms.

KEYWORDS: Rule Based Filtering System, Machine Learning Techniques, Online Social Network, Content Based Filtering.

I. INTRODUCTION

A social networking service is a platform to make interact among people, for example, shares ideas, posts, interests, activities in network. A social network service consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. These services are sometimes considered as a social network service, though in a broader sense. In social network may have possibilities of posting some unwanted messages on other user's space. So it has needed to avoid the displaying of unwanted words on user's space. This is achieved by using Rule Based Filtering System.

Sum up of our work as follows:

- We propose automatically filter post messages in OSN.
- We Propose a rule based Filtering Methods allows users to customize filter by applying filtering criteria.
- We propose and experimentally evaluate an automated system, called Filtered Wall (FW)
- We propose to filter unwanted messages from OSN user space.
- We propose to exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content.

The rest of the paper organized as follows: The model of our approach is in section 2.The text representation, short text classifier and their filtering rules of filter wall architecture in section 3. In Section 4 reviews several related works. Section 5 concludes the paper.

II. MODEL OF OUR APPROACH

Fig 1 shows the model of our work. The Filter Wall Architecture (Figure1) in support of OSN services is a three-tier structure. The first layer, called filter policy is used to add the filter policies. The second layer called Social Network Manager (SNM), commonly aims to provide the basic OSN functionalities, whereas the second layer provides the support for external Social Network Applications (SNAs). According to this reference architecture, the proposed system is placed in

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

the second and third layers. Moreover, the GUI provides users with a FW, that is, a wall where only messages that are authorized according to their FRs/BLs are published.

III. FILTER WALL ARCHITECTURE

The core components of the proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier modules. The latter component aims to classify messages according to a set of categories. In contrast, the first component exploits the message categorization provided by the STC module to enforce the FRs specified by the user. BLs can also be used to enhance the filtering process. As graphically depicted in Figure 1, the path followed by a message, from its writing to the possible final publication can be summarized as follows:

- After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.
- A ML-based text classifier extracts metadata from the content of the message.
- FW uses metadata provided by the classifier, together with data extracted from the social graph and users profiles, to enforce the filtering and BL rules.
- Depending on previous steps result, the message will be published or filtered by FW.

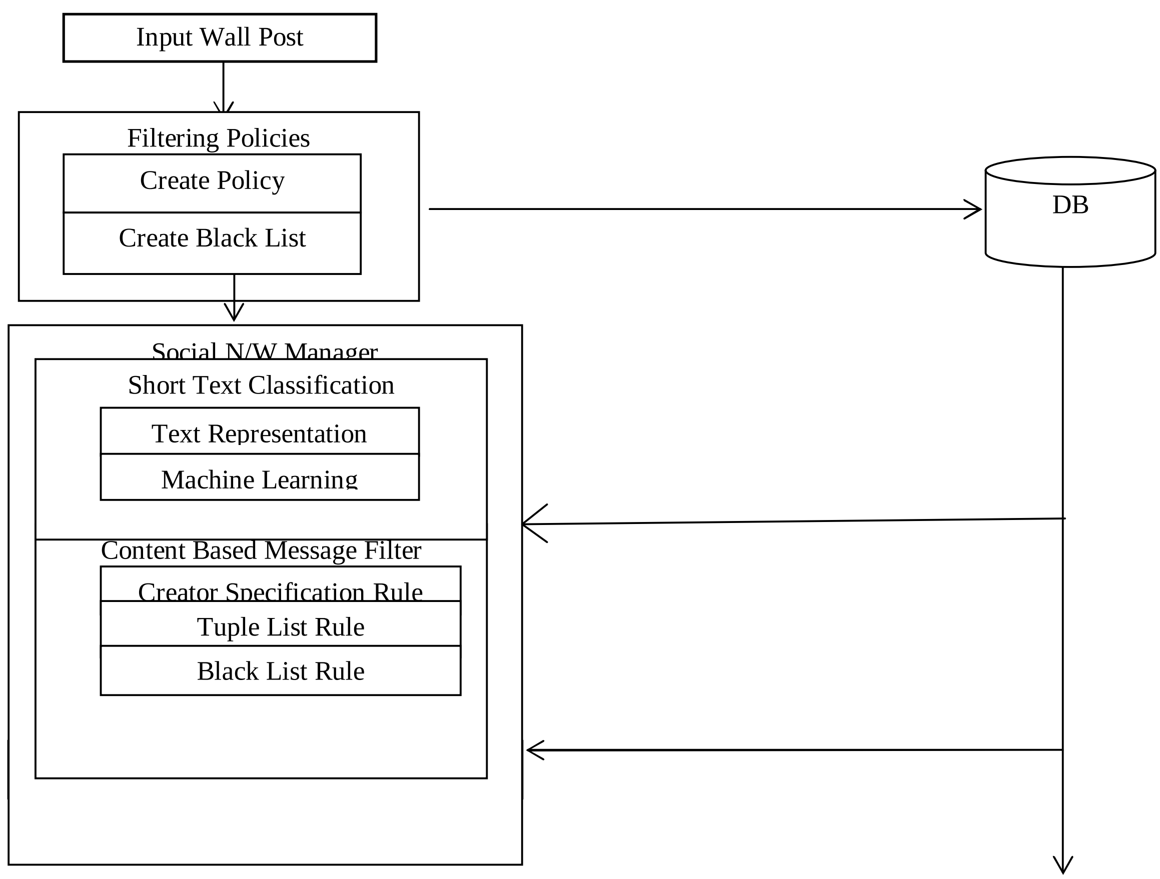

Fig1. Filter Wall Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

In Figure 1 has follows, this system explains the modules in more detail some of the above-mentioned architecture steps.

3.1 Input Wall Post

This module deals with the wall post details from the registered user. The Figure 2 shows user has to login and give the post to their friends.

3.2 Filtering Policy

These policies are used to filter the wall post by checking against the database policies which are already given by the user. In this section, this system introduces the rule layer adopted for filtering unwanted messages. This system starts by describing FRs, then this work illustrates the use of BLs.

3.2.1 Create Policy

In defining the language for FRs specification, we consider three main issues that, in our opinion, should affect a message filtering decision. As a consequence, FRs should allow users to state constraints on message creators.

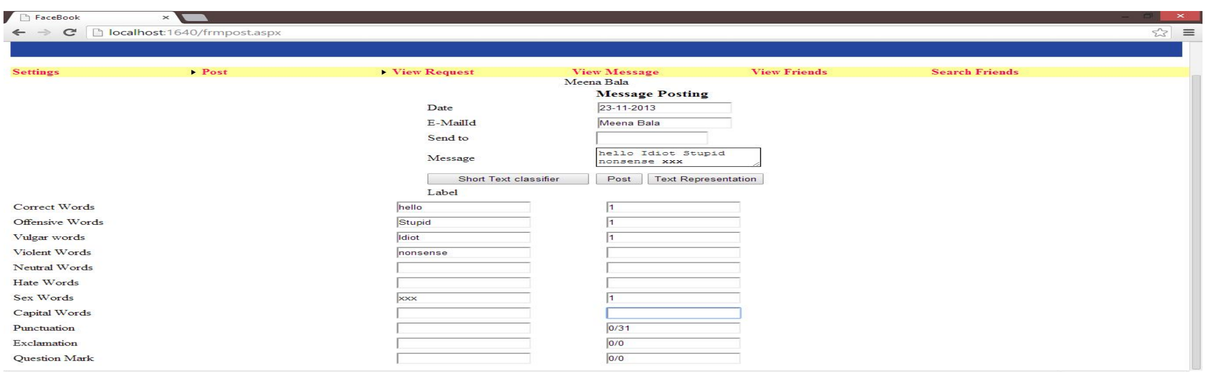

Fig 2. Input Wall Post

A) Rule 1 (Creator specification)

A creator specification creator Spec implicitly denotes a set of OSN users. It can have one of the following forms, possibly combined.

- A set of relationship constraints of the form (m, rt, minDepth, maxTrust), denoting all the

In Fig 3. shows the creator specification rule

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

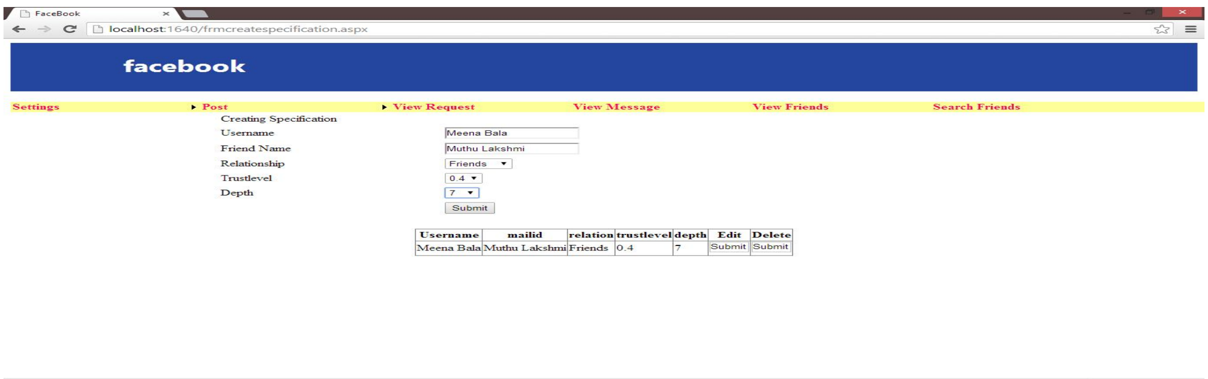

Fig 3. Creator Specification

An FR is therefore formally defined as follows:

B) Rule 2 (Filtering rule)

A filtering rule FR is a tuple (author, creator Spec, content Spec, action), where

- author is the user who specifies the rule;
- Creator Spec is a creator specification, specified according to Definition 1;
- Content Spec is a Boolean expression defined on content constraints of the form (C,ml), where C is a class of the first or second level and ml is the minimum membership level threshold required for class C to make the constraint satisfied;
- action ∈ {block; notify} denotes the action to be performed by the system on the messages matching contentSpec and created by users identified by creator Spec.

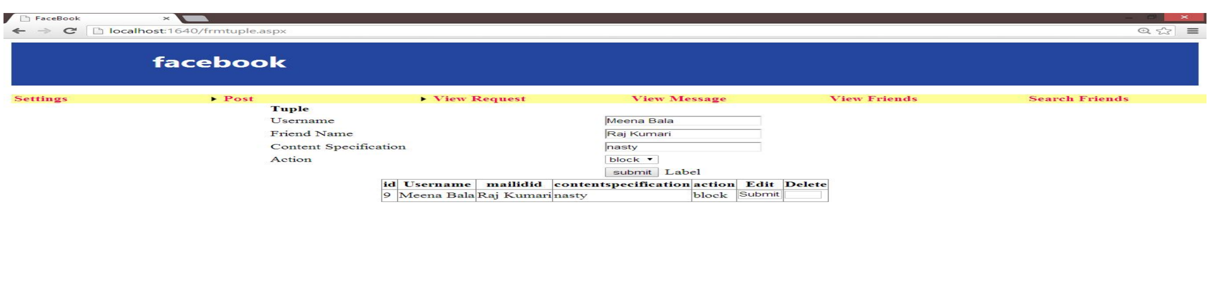

Fig 4. Tuple List Rule

In Fig 4, Shows the Tuple list rule to be specified

C) Rule 3 (BL rule) :
A BL rule is a tuple (author, creator Spec, creator Behavior, T), where

- author is the OSN user who specifies the rule, i.e., the wall owner;
- creator Spec is a creator specification, specified according to Definition 1;
- creator Behavior consists of two components RF Blocked and min Banned. RF Blocked = (RF, mode, window) is defined such that

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) **Vol.2, Special Issue 1, March 2014**

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

3.3 Social Network Manager

The supported SNAs may in turn require an additional layer for their needed Graphical User Interfaces (GUIs).It has contains Short Text Classifier and Text Representation are as follows in below

3.3.1 Short Text Classifier

Our study is aimed at designing and evaluating various representation techniques in combination with a neural learning strategy to semantically categorize short texts. It has 2 level task. The first-level task is conceived as a hard classification in which short texts are labeled with crisp Neutral and Nonneutral labels. The second-level soft classifier acts on the crisp set of nonneutral short texts and, for each of them, it "simply" produces estimated appropriateness or "gradual membership" for each of the conceived classes, without taking any "hard" decision on any of them

3.3.2 Text Representation

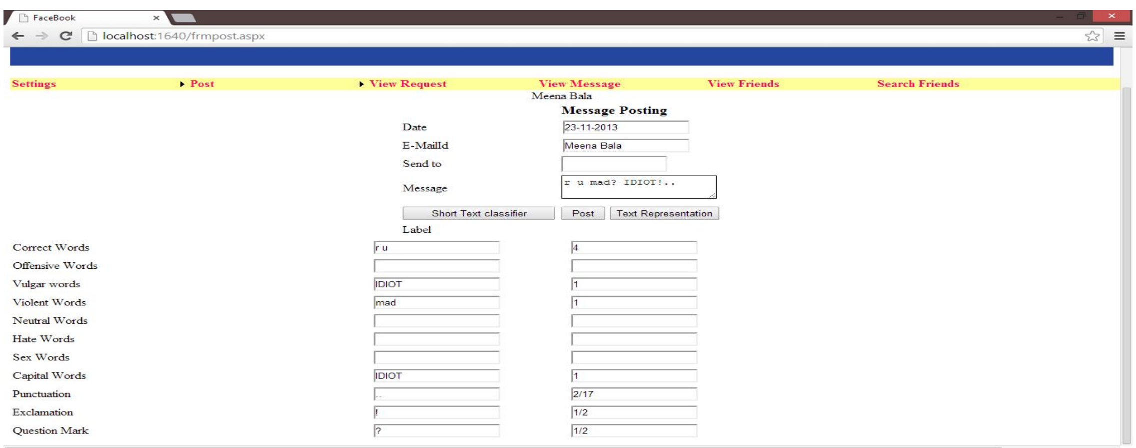

Fig 5. Text Representation

Required trial-and-error procedures. In more detail,

- Correct words. It expresses the amount of terms $t_k \in T \cap K$, where tk is a term of the considered document d_j and K is a set of known words for the domain language. This value is normalized by $\sum_{k=1}^{|T|} \#(tk, dj)$
- Bad words. They are computed similarly to the correct words feature, where the set K is a collection of "dirty words" for the domain language.
- Capital words. It expresses the amount of words mostly written with capital letters, calculated as the percentage of words within the message, having more than half of the characters in capital case.
- Punctuations characters. It is calculated as the percentage of the punctuation characters over the total number of characters in the message.
- Exclamation marks. It is calculated as the percentage of exclamation marks over the total number of punctuation characters in the message.
- Question marks. It is calculated as the percentage of question marks over the total number of punctuations characters in the message.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

IV. RESULTS AND DISCUSSION

In this section, this work illustrates the performance evaluation study this work has carried out the classification and filtering modules. This work starts by describing the data set. It have been selected and extracted by means of an automated procedure that removes undesired spam messages and, for each message, stores the message body and the name of the group from which it originates. The messages come from the group's webpage section, where any registered user can post a new message or reply to messages already posted by other users. The set of classes considered in our experiments is omega = {Neutral, Violence, Vulgar, Offensive, Hate, Sex}, where omega – {neutral} are the second-level classes.This project hopes that our data set will pave the way for a quantitative and more precise analysis of OSN short text classification methods.

V. CONCLUSION

In this Paper has concluded a system to filter undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent FRs . Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs.

REFERENCES

[1] Adomavicius and G. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[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] R.J. Mooney and L. Roy, "Content-Based Book Recommending Using Learning for Text Categorization," Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000.

[4] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[5] 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.

[6] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, "Inductive Learning Algorithms and Representations for Text Categorization," Proc. Seventh Int'l Conf. Information and Knowledge Management (CIKM '98), pp. 148-155, 1998.