www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 4 April 2015, Page No. 11680-11686

Unwanted Message Filtering From Osn User Walls And Implementation Of Blacklist (Implementation Paper)

Rakhi Bhardwaj[1], Vikram Kale[2], Prasad Morye[2], Manoj Dhaygude[2], Sagar Badhe[2]
[1]Department of Computer Engineering, KJEI's Trinity College of Engineering & Research, Pune, India
[2]Department of Computer Engineering, KJEI's Trinity College of Engineering & Research, Pune, India

ABSTRACT: Now-a-days there is a huge increase in usage of Online Social Networks (OSNs). But, these OSNs do not provide a security to the user walls against unwanted messages. Users can easily post any undesired contents on a user wall. This is a very important issue in OSN giving rise to misuse of these OSNs. Up to now OSN have provided little control regarding who can post on user's private wall. Here we have proposed a system that will prevent unwanted messages from being posted on user's wall. This can be done by scanning the messages for any undesired contents before they are being posted on any user wall. We have also proposed a Blacklisting mechanism which will block users frequently trying to post such messages and prevent them from further posting any messages on the user's wall.

Keywords: Text filtering, text categorization, online social network, blacklist, OSN security

Introduction

Online Social Networks (OSN) has proved to be an easy communication medium for a massive number of people. Statistics show that OSN's have become one of the most popular interactive medium. OSN have provided facilities for people such as to stay in touch, share their feelings, update their daily activities, travels, photos and political upraising. Daily communication results in the commencement of incalculable amount of data, text, images etc. The main section of OSN's is its public or private areas, called as walls. Today's OSN provides settings option as to whom the wall posts are visible i.e. only to the user's contacts or including other people (if the user allows to).But very less security is provided by OSN on the contents of the wall post. Up to date, a very little support has been provided by OSN's to prevent unwanted and undesired messages on the user's wall. Consider an example of the leading social networking site – Facebook, which allows user to determine who can post on their walls (i.e. defined groups of users such as friends, friends of friends or anyone) but there is no restriction on the contents to be posted. OSN does not scan the messages for vulgar, objectionable or political contents before posting it on user's wall. Thus, undesired and unwanted contents in the message are easily posted on user's wall, no matter who posts it. Thus, to overcome this problem, we can use unwanted message filtering. It will help to secure the OSN user walls against unwanted and undesired messages posting on his/her wall. Such

security requirement is a must in OSN service but is not provided in any OSN.

We aim to provide the user with a **Filtered Wall (FW)** mechanism in our proposed system. Filtered Wall is an automated system, which will provide filtering of unwanted messages from social network's user wall. Also, there will be a **Blacklisting** mechanism in our proposed system. User can block other users who repeatedly try to post unwanted content on former's wall because of this blacklisting mechanism. We can achieve it through text identification by scanning the message before posting it on a user's wall. Here first there will be classification of text using short text classifier technique for classifying unwanted contents, and then there will be text categorization, where the text will be subdivided into different domains. The next step is determining the rules for filtering and performing the blacklisting management. Filtering will be performed on the basis of unwanted or undesired, text or words in the message. Further on, there will be an automated system that will block users who repeatedly try to post such unwanted contents and such users will be kept in blacklist for a specific period of time. The contents of the messages posted by the user and the number of attempts made by him to post such messages will determine the time period for which the user is being blocked. Such technique will provide more security for OSN. This will help to ensure the prevention and circulation of undesired contents through online social networks.

Related Work

Elena Ferrari, Elisabetta Binaghi, Marco Vanetti, Moreno Carullo and Barbara Carminati, [1] [2] have proposed a system that possess a mechanism to avoid unwanted messages from any user on an OSN wall of other users. This paper aims at providing OSN users an ability to secure their walls through filtering the unwanted contents being posted. This system will block the undesired messages sent by the user which is achieved by an automated system called Filtered wall (FW). Content based message filtering and short text classifier support this system. But the drawback of this system is that the user posting unwanted messages will not be blocked; only the message posted by the user will be blocked. To overcome this problem of the system, the term Blacklist will be implemented as future enhancement.

L. Roy and R.J. Mooney [3] have proposed a content-based book recommending system using information extraction and a machine-learning based algorithm for text categorization.

M. Demirbas, B. Sriram, D. Fuhry, E. Demir and H. Ferhatosmanoglu[4]. In this paper technique to classify messages on micro-blogging sites such as Twitter is explained. Messages on twitter are short and hence lack sufficient word occurrences. Therefore there are limitations in traditional classification techniques like "Bag-Of-Words". Therefore, this paper proposes usage of small set of domain-specific features which are extracted from author's profile and text. This approach effectively classifies the text into sets of generic classes such as Private Messages, Opinions, Deals, Events and News.

Existing System

Today's existing OSNs such as Facebook, MySpace, etc. provides a facility to user to allow the latter to choose a group of other users who can post the messages on latter's wall (i.e. defined groups of users such as friends, friends of friends or anyone). But, this provides little security to user's wall because the allowed people can post any kind of messages on the wall possessing unwanted contents like vulgar, objectionable or political words.

Disadvantages of existing System

- The existing system does not scan the messages for unwanted contents before posting it on wall, no matter who posts it.
- It does not filter the messages possessing unwanted contents which the users don't want to be on their wall.
- It doesn't automatically block the people who keep posting unwanted messages on a user's private wall.

Proposed System

In this paper we propose a technique known as filtered wall (FW), which is used for filtering unwanted messages. The Filtered Wall scans each message before being posted on wall. Filtering rules are used to determine which contents should be allowed on user's wall and which messages should be blocked. Further it will also provide a Blacklisting mechanism. Blacklist will be an automated mechanism which will block users posting undesired messages on the user walls. The prohibition can be approved for uncertain period of time.

System Implementation

In this paper we propose a technique known as filtered wall (FW), which is used for filtering unwanted messages. The Filtered Wall scans each message before being posted on wall. Filtering rules are used to determine which contents should be allowed on user's wall and which messages should be blocked. Further it will also provide a Blacklisting mechanism. Blacklist will be an automated mechanism which will block users posting undesired messages on the user walls. The prohibition can be approved for uncertain period of time.

The Filtered Wall mechanism works on Conflation algorithm. **Conflation Algorithm** works in three phases as follows:

1. Removal of high frequency words.
2. Suffix stripping.
3. Detecting equivalent stems

1. **Removal of high frequency words**

   The removal of high frequency words (stop words) is done by comparing each word from the document to the list of high frequency words. High frequency words are those words which occur more number of times in the text. This words does not contain semantic of the text**.** Non-significant words are removed and also the size of document is reduced.

2. **Suffix Stripping(Stemming)**

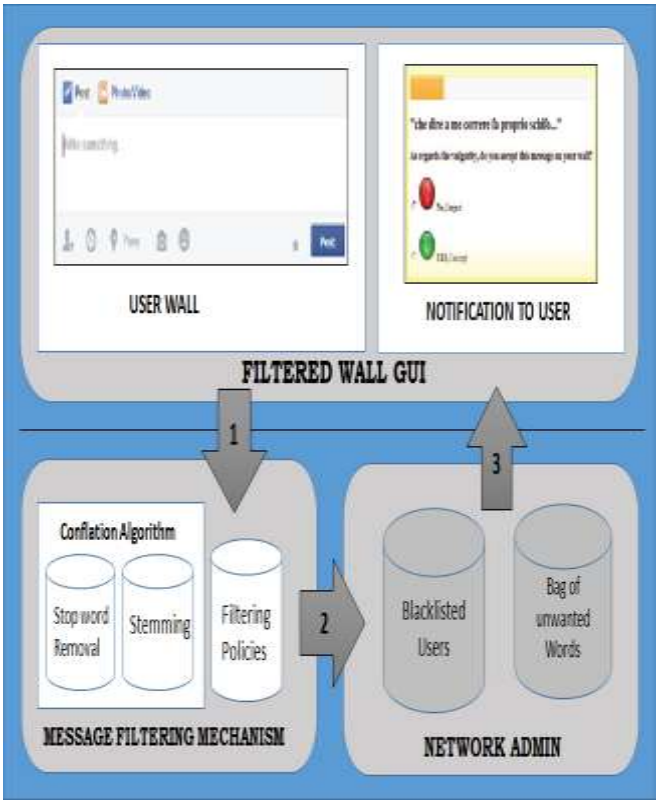
   Each word is handled from output of first step. If any word is having suffix then, the suffix get removed and the word is converted in its original form. This process is called as stemming.

3. **Detecting equivalent stems**

After stemming we have list of words from which only one instance of word is kept in the list. Each word is called as 'stem'.

The output of Conflation algorithm is set of stem which is used for further processing.

System Architecture

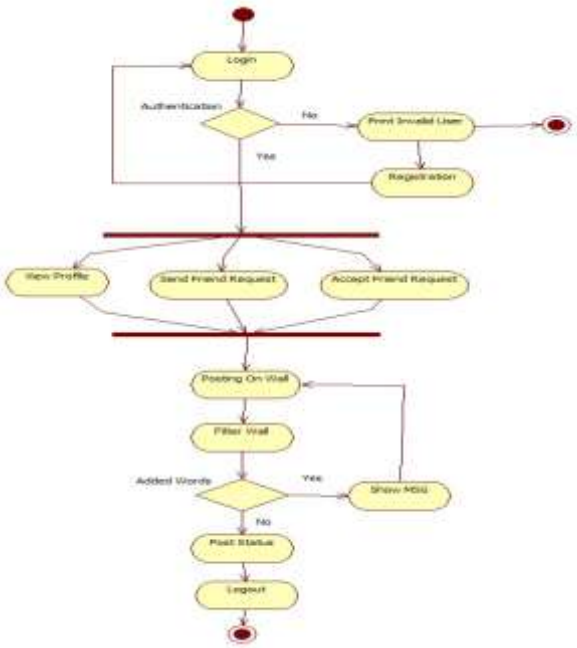


The above diagram represents the system architecture of our proposed system. Filtered wall provide the basic functionality of posting messages on OSN. It is generally supported by the filtering mechanism.

The filtering mechanism provides the facility to filter each and every message posted on wall. After filtering the contents an acknowledgement is sent to network admin. The network admin is an automated system that performs the function of notifying the user for unwanted contents as well as performing the task of blacklisting. The contents of message are classified with the predefined set of unwanted words stored in data dictionary. If any unwanted contents are found then the user trying to post the message is notified. If the user repeatedly tries to post such unwanted messages on OSN wall then he/she is blacklisted from the OSN. The time period for which the user is been blocked is determined by an automated mechanism. The network admin maintains the list of blocked users separately.

Working

The system flow graph is as follows.

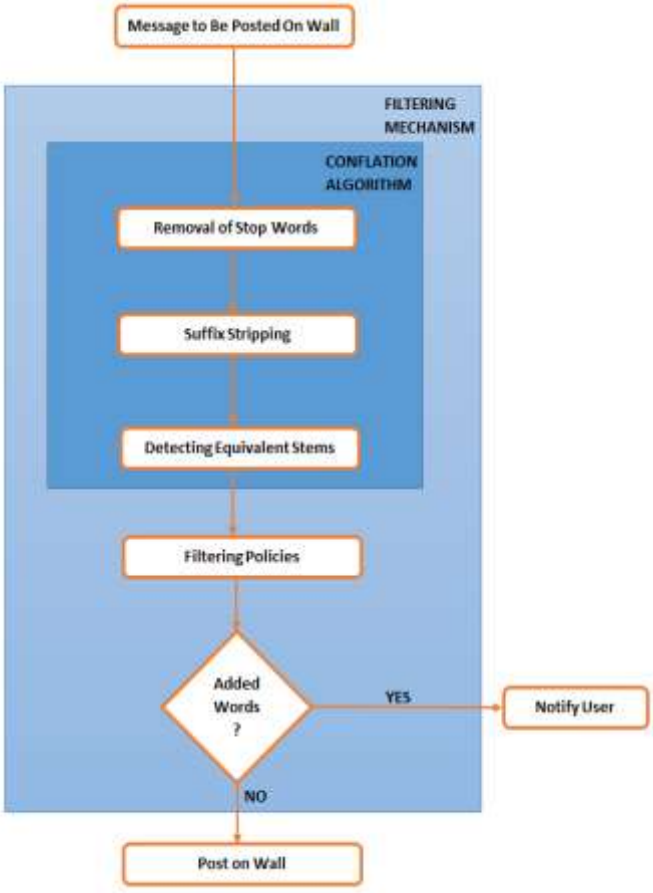


User has to login in order to get access to his profile, only valid users are authenticated. If the user is invalid he/she has to register first. Once the user is successfully logged in, he can view his profile, send friend request to other users, accept request, chat with friends, view their profile and also post on wall.

When the user tries to post a message on wall the message goes through filtering stage before getting posted. The message is filtered for any kind of unwanted words specified by either user or network admin. If the message is neutral then it gets posted easily on the OSN user wall, but if it

contains any non-neutral words then the user trying to post such message is either notified or blocked.

The message filtering process goes through following phases:



When the message is been posted the **Filtering Wall** mechanism comes into picture it works in following manner. First the input string is passed through **Conflation Algorithm**. Conflation Algorithm performs following operations on the string:

1. **Removal of high frequency words.**
   All the stop words are removed from the input string and the string is passed to next step for further operations.
2. **Suffix Stripping**
   The word having suffix are searched and the suffix part is removed from it. Only the root word is maintained for further operation.
3. **Detecting equivalent Stem**
   Stems are detected and from that only one instance of word is kept.

Now the **filtering policies** are applied to the stemmed words. In this step the stemmed words are scanned to detect the presence of any unwanted words. If any undesired contents are found then the message is not posted on wall. The user is well notified for the undesired contents in his/her message. If the user continues to do so he/she is been blacklisted from another user's wall.

Blacklisting: This feature is an automated system which performs the function of blocking the user who repeatedly tries to post undesired messages on user walls. The user will be blocked depending on the impact of his messages and also the number of attempts the user makes to post such messages on the wall. The impact of these undesired messages is calculated considering the following factors:

- The total number of non-neutral words with respect to total words in message.
- The priority of the domains of non-neutral words.

The system works on following Algorithm

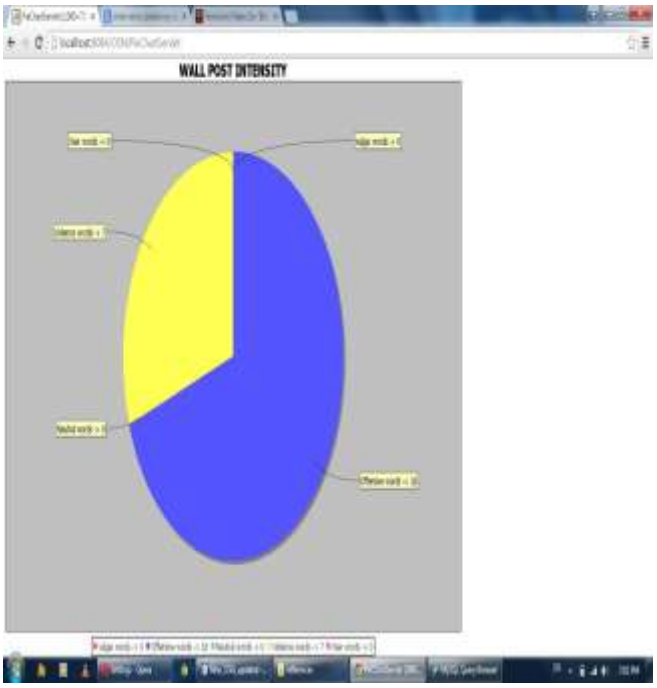
Algorithm

Require: String S Inserted by User; Data Dictionary D of Stop Words.

1: **for** all words w Є S **do**
2: **if** w belongs D **then**
3: skip w
4: **else**
5: copy w to S1
6: **end if**
7: **end for**
8: **for** all words w Є S1 **do**
9: suffix stripping of words w
10: **end for**

The output from obtained from above algorithm is used for further filtering operations.

Expected Result



The result will be displayed in form of pie chart. It will represent the intensity of word from each domain. It represents the intensity of unwanted contains in the message to be posted.

Conclusion

Inspecting the messages posted on OSN user walls is important issue in today's world. Our proposed system uses an automated mechanism to scan the messages before being posted on the user's wall and further filters those messages from OSN user walls which are unwanted and undesired. We have also proposed an automated Blacklisting mechanism which blocks the users who repeatedly try to post such undesired messages ignoring the given warnings. Hence, our proposed system provides more security to OSN user walls and therefore no objectionable or undesired contents can be circulated through our proposed mechanism for OSN user walls.

References

[1] Elena Ferrari, Elisabetta Binaghi, Marco Vanetti, Moreno Carullo and Barbara Carminati, "A System to Filter Unwanted Messages from OSN User Walls", IEEE Transaction, Vol. 25,No. 2,Feb 2013.

[2] Elena Ferrari, Elisabetta Binaghi, Marco Vanetti, Moreno Carullo and Barbara Carminati, "Content-based Filtering in On-line Social Networks", IEEE Transaction, Vol. 25,No. 2,Feb 2013.

[3] Raymond J. Mooney, Loriene Roy, "Content-Based Book Recommending Using Learning for Text Categorization".

[4] M. Demirbas, B. Sriram, D. Fuhry, E. Demir and H. Ferhatosmanoglu, "Short Text Classification in Twitter to Improve Information Filtering".

[5] Christian Bizer, Richard Cyganiak, "Quality-driven information filtering using the WIQA policy framework", Web Semantics: Science, Services and Agents on the World Wide Web 7 (2009) 1–10.