ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 3, pp: (399-404), Month: July - September 2014, Available at: www.researchpublish.com

# A System to Filter Unwanted Messages from OSN User Walls

Author: K.R DEEPTI

***Abstract:*** **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, which 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.**

***Keywords:*** **Online social networks, information filtering, short text classification, policy based personalization.**

## I. INTRODUCTION

Most common interactive medium to communicate is online social network. several types of information or content will be shared between the users, the type of contents are audio, video, images etc. As the Amount of content will be very vast information filtering is used . OSN provide very less amount of security in posting unwanted messages. Information filtering is used for unrelated purpose. Ability of a user to automatically control the messages written on the user wall, by filtering additional communication will be termed as information filtering [1].

We exploit Machine Learning (ML) text categorization techniques [2] to automatically assign with each short text message a set of categories based on its content. The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminant features. Additionally, the system gives the support for user-defined Black Lists (BLs), that is, lists of users that are temporarily prevented to post any kind of messages on a user wall. OSNs provide 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). 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.

## II. LITERATURE REVIEW & RELATED WORK

Filtering is based on explanations of individual or group information preferences that typically represent long-term interests. Users get only the data that is extracted. Information filtering systems are intended to categorize a stream of dynamically generated information and present it to the user those information that are likely to satisfy user requirements. Feedback using previous related abstracts provided an efficient and simple way of demonstrating people's interests [4]. The main contribution of this is the design of a system providing customizable content-based message filtering for OSNs, based on ML techniques. Our work has relationships both with the state of the art in content-based filtering, as well as with the field of policy-based personalization for OSNs and, more in general, web contents.

A distinction is made between two types of text filtering systems: content-based and social filtering systems. In content-based systems, filtering is done by exploiting the information extracted from the text of documents. In social filtering systems, documents are filtered based on annotations made by prior readers of the documents. We use social features of the users to identify the ones who are more likely to post relevant content, however it is different from the social filtering systems where other users' feedbacks are used. In the OSN domain, interest in access control and privacy protection is quite recent.

### 2.1 Content-Based Filtering

Information filtering systems are designed to classify a stream of dynamically generated information dispatched synchronously by an information producer and present to the user those information that are likely to satisfy his/her requirements [3].In content-based filtering each user is assumed to operate independently. As a result, a content-based filtering system selects information items based on the correlation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences [5]. In content based filtering to check the user's interest and previous activity as well as item uses by users best match is found [6]. For example OSNs such as Facebook, Orkut used content based filtering policy.

### 2.2 Policy-Based Personalization Of OSN Contents

There have been some proposals exploiting classification mechanisms for personalizing access in OSNs. For instance, in [7] a classification method has been proposed to categorize short text messages in order to avoid overwhelming users of micro blogging services by raw data. The user can then view only certain types of tweets based on his/her interests. In policy based filtering system users filtering ability is represented to filter wall messages according to filtering criteria of the user. Twitter is the best example for policy based filtering.

Our work is also inspired by the many access control models and related policy languages and enforcement mechanisms that have been proposed so far for OSNs[8], since filtering shares several similarities with access control. Actually, content filtering can be considered as and extension of access control, since it can be used both to protect objects from unauthorized subjects, and subjects from inappropriate objects.

### Goal

Our goal is to design an online message filtering system that is deployed at the OSN service provider side. Once deployed, it inspects every message before rendering the message to the intended recipients and makes immediate decision on whether or not the message under inspection should be dropped.

### Limitation of Existing System

- However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages. No matter user who propose 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.

### Proposed Work

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 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. The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminate features.

### Advantages of Proposed System

- A system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationship and characteristics.
- The substantially extends for what concerns both the rule layer and the classification modules.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 3, pp: (399-404), Month: July - September 2014, Available at: www.researchpublish.com

## III. FILTERED WALL ARCHITECTURE

The architecture in support of OSN services is a three-tier structure (Fig.). The first layer, called Social Network Manager (SNM), commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management), where as the second layer provides the support for external Social Network Applications (SNAs).4 The supported SNA s may in turn require an additional layer for their needed Graphical User Interfaces (GUIs).According to this reference architecture, the proposed system is placed in the second and third layers. In particular, users interact with the system by means of a GUI to set up and manage their FRs/BLs. 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.

The core components of the proposed system are the Content-Based Messages Filtering (CBMF) and the Short Text Classifier modules. The latter element aims to categorize messages according to a set of categories. In compare, the first element exploits the message categorization offered by the STC module to implement the FRs specified by the user. 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.

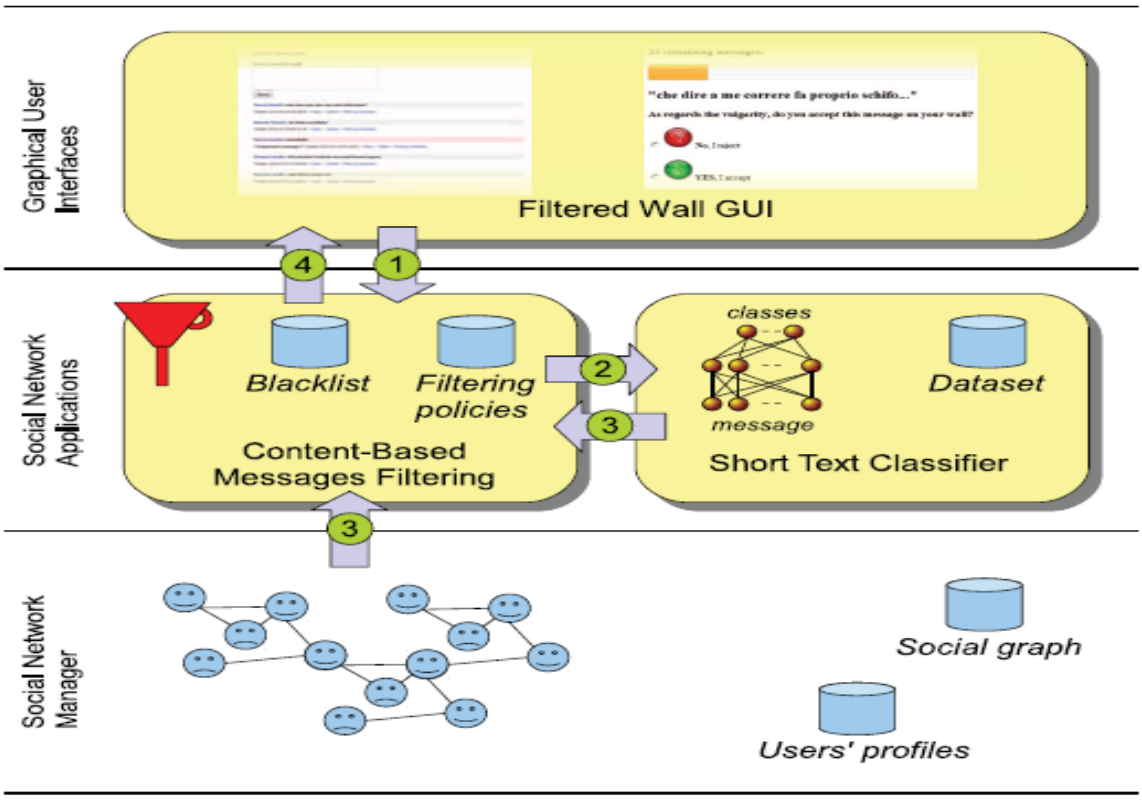


**Fig. 1: Filtered wall conceptual architecture and the flow messages follow, from writing to publication**

As graphically depicted in Fig., the path followed by a message, from its writing to the possible final publication can be summarized as follows:

1) After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.

2) A ML-based text classifier extracts metadata from the content of the message.

3) 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.

4) Depending on the result of the previous step, the message will be published or filtered by FW.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 3, pp: (399-404), Month: July - September 2014, Available at: www.researchpublish.com

## IV. SHORT TEXT CLASSIFIER

Established techniques used for text classifications work well on datasets with large documents such as newswires corpora [9] but suffer when the documents in the quantity are tiny. In this perspective critical features are the description of a set of characterizing and discriminant features allowing the representation of underlying concepts and the collection of a complete and consistent set of supervised examples. Our study is aimed at designing and evaluating various representation techniques in combination with a neural learning strategy to semantically categorize short texts. The first-level task is conceived as a hard classification in which short texts are labeled with crisp Neutral and Non neutral labels.[10] The second-level soft classifier acts on the crisp set of non-neutral short texts.

### 4.1 Text Representation

The extraction of an appropriate set of features by which representing the text of a given document is a crucial task strongly affecting the performance of the overall classification strategy. In the BoW representation, terms are identified with words. According to Vector Space Model (VSM) for text representation, a text document *dj* is represented as a vector of binary or real weights *dj* = *w1j*, . . . , *w*| *T*|*j*, where *T* indicates the set of terms that occur at least once in at least one document of the collection *Tr*, and *wkj* € [0; 1] denotes how much term *tk* contributes to the semantics of document *dj*. In the BoW representation, terms are identified with words. In the case of non binary weighting, the weight wkj of term tk in document dj is computed according to the standard term frequency—inverse document frequency (tf-idf) weighting function [11], defined as

$$tf - idf(t_k, d_j) = \#(t_k, d_j) \cdot \log \frac{|\mathcal{T}_r|}{\#\mathcal{T}_r(t_k)},$$

- ✓ **Correct words.** It expresses the amount of termstk 2 T \K, where tk is a term of the considered document dj and K is a set of known words for the domain language. This value is normalized byPjT jk=1 #(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 represents 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. For example, the value of this feature for the document "To be OR Not to BE" is 0:5 since the words "OR" "Not" and "BE" are considered as capitalized ("To" is not uppercase since the number of capital characters should be strictly greater than the characters count).
- ✓ **Punctuations characters.** It is calculated as the percentageof the punctuation characters over the totalnumber of characters in the message. For example,the value of the feature for the document "Hello!!!How're u doing?" is 5=24.
- ✓ **Exclamation marks.** It is calculated as the percentageof exclamation marks over the total number ofpunctuation characters in the message. Referring tothe aforementioned document, the value is 3=5.
- ✓ **Question marks.** It is calculated as the percentage ofquestion marks over the total number of punctuationscharacters in the message. Referring to theaforementioned document, the value is 1=5.

### 4.2 Machine Learning-Based Classification

Short text categorization is a hierarchical two-level classification process. The first-level classifier does a binary hard classification that labels messages as Neutral and Non-Neutral. The first-level filtering task enables the subsequent second-level task in which a finer-grained classification is performed. The second-level classifier carries out a soft-partition of Non-neutral messages assigning a given message a gradual membership to each of the non-neutral classes.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 3, pp: (399-404), Month: July - September 2014, Available at: www.researchpublish.com

## V. FILTERING RULES AND BLACKLIST MANAGEMENT

We introduce the rule layer adopted for filtering unwanted messages. We model a social network as a directed graph, where each node corresponds to a network user and edges denote relationships between two different users.

### 5.1 Filtering Rules

We consider three main issues in defining the language for FRs specification. First is related to the fact that, the same message may have different meanings and relevance based on who writes it. Message 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. Creator specification, defined as follows.

**Definition.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:

1. A set of attribute constraints of the form an OP av, where an is a user profile attribute name, av and OP are, respectively, a profile attribute value and a comparison operator, compatible with an's domain.

2. A set of relationship constraints of the form(m; r t; min Depth; max Trust), denoting all the OSN users participating with user m in a relationship of type rt, having a depth greater than or equal tom in Depth, and a trust value less than or equal to max Trust.

An FR is therefore formally defined as follows:

**Definition .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 content Spec and created by users identified by creator Spec.

### 5.2 Online Setup Assistant for FRs Thresholds

By conceiving and implementing within FW, an Online Setup Assistant (OSA) procedure, we address the problem of setting thresholds to filter rules. OSA presents the user with a set of messages selected from the dataset. For each message, the user expresses the system the decision to accept or reject the message.

### 5.3 Blacklists

BLs are directly managed by the system, and should be able to determine the users to be inserted in the BL and decide user's retention in the BL is finished. Such information are given to the system through a set of rules, called BL rules. We let the wall's owners to specify BL rules regulating who has to be banned from their walls and for how long. Therefore, a user might be banned from a wall, by, at the same time, being able to post in other walls.

A BL rule is therefore formally defined as follows:

**Definition. 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., thewall owner;
- ✧ creator Spec is a creator specification, specified according to Definition 1;
- ✧ creator Behavior consists of two components RFB locked and min Banned. RFB locked ¼ (RF, mode, window) is defined such that

--RF =#b Messages / #t Messages , where #t Messages is the total number of messages that each OSN user identified by creator Spec has tried to publish in the author wall (mode = my Wall) or in all the OSN walls(mode = SN); whereas #b Messages is the number of messages among those in #t Messages that have been blocked;

--window is the time interval of creation of those messages that have to be considered for RF computation; min Banned =(min, mode, window), where min is the minimum number of times in the time interval specified in window that OSN users identified by creator Spec have to be inserted into the BL due to BL rules specified by author wall (mode = my Wall) orall OSN users (mode =SN) in order to satisfy the constraint.

✧ T denotes the time period the users identified by creator Spec and creator Behavior have to be banned from author wall.

## VI. CONCLUSION

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 user walls. The future implication of this work is we exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content. Existing system is used to filter undesired messages from OSNs wall using customizable filtering rules (FR)enhancing through Black lists (BLs).

## REFERENCES

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, an Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls", 2013.

[2] F. Sebastiani, “Machine learning in automated text categorization,” ACM Computing Surveys, vol.34, no. 1, pp. 1–47, 2002.

[3] N. J. Belkin and W. B. Croft, “Information filtering and information retrieval: Two sides of the same coin?” Communications of the ACM, vol. 35, no.12, pp. 29–38, 1992.

[4] P. W. Foltz and S. T. Dumais, “Personalized information delivery: An analysis of information filtering methods,” Communications of the ACM, vol. 35, no. 12, pp.51–60, 1992.

[5] P. J. Denning, “Electronic junk,” Communications of the ACM, vol. 25, no.3, pp. 163– 165, 1982.

[6] P.S. Jacobs and L.F. Rau, “Scisor: Extracting Information from On Line News,” Comm. ACM, vol. 33, no. 11, pp. 88-97, 1990.

[7] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short text classification in twitter to improve information filtering,” in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, 2010, pp. 841–842.

[8] F. Bonchi and E. Ferrari, Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques. Chapman and Hall/CRC Press,2010.

[9] D.D. Lewis, Y. Yang, T.G. Rose, and F. Li, “Rcv1: A New Benchmark Collection for Text Categorization Research, ”J. Machine Learning Research, vol. 5, pp. 361-397, 2004.

[10] 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.

[11] G. Salton and C. Buckley, “Term-Weighting Approaches in Automatic Text Retrieval,” Information Processing and Management, vol. 24, no. 5, pp.513-523, 1988.