

# Filtering System to Restrict Unwanted Message from OSN User Walls

Miss. Rashmi R. Atkare<sup>1</sup>, Pravin D. Soni<sup>2</sup>

<sup>1</sup>ME (CSE) Scholar, Department of CSE,  
P. R. Patil College of Engg. & Tech., Amravati-444602, India

<sup>2</sup>Assitant Professor, Department of CSE,  
P. R. Patil College of Engg. & Tech., Amravati -444602, India

## ABSTRACT

*The Online Social Networks (OSNs) have become an important part of daily life. The fundamental issue in today On-line Social Networks is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. Today OSNs provide very little support to prevent unwanted messages on user walls. Therefore a major task of today's online social network is information filtering. For that purpose, we proposed a new 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 (ML) based soft classifier automatically labelling messages in support of content-based filtering. The system exploits a ML soft classifier to enforce customizable content-dependent Filtering Rules. And the flexibility of the system in terms of filtering options is enhanced through the management of Blacklists. The proposed system gives security to the On-line Social Networks.*

**Keywords:** Online Social Network, Filtering Rules, Filtered Wall, Black List, Machine Learning.

## 1. Introduction

On-line Social Networks (OSNs) are today one of the most popular interactive medium to communicate, share and disseminate a considerable amount of human life information.

An OSN is a web-based service that allows individuals to:

1. Construct a public or semi-public profile within the service,
2. Articulate a list of other users with whom they share a connection,
3. View and traverse their list of connections and those made by others within the service.

Information and communication technology plays a significant role in today's networked society. It has affected the online interaction between users, who are aware of security applications and their implications on personal privacy. There is a need to develop more security mechanisms for different communication technologies, particularly online social networks. Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. OSNs provide very little support to prevent unwanted messages on user walls. With the lack of classification or filtering tools, the user receives all messages posted by the users he follows. In most cases, the user receive a noisy stream of updates. In this paper, an information Filtering system is introduced. The system focuses on one kind of feeds: Lists which are a manually selected group of users on OSN. List feeds tend to be focused on specific topics, however it is still noisy due to irrelevant messages. Therefore, we propose an online filtering system, which extracts the such topics in a list, filtering out irrelevant messages [1].

In OSNs, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private areas, called in general walls. In the proposed system Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. 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 OSN user walls. 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.

## 2. Related Works

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. With respect to this framework, our system is closer to content-based filtering systems, however we utilize other sources of information next

to the text of documents. Indeed, OSNs provide very little support to prevent undesired 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. 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. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences. 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. Therefore, in what follows, we survey the literature in both these fields.

### 2.1 Content-Based Filtering

In content-based filtering each user is assumed to operate independently[3]. 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 [4]. While electronic mail was the original domain of early work on information filtering, subsequent papers have addressed diversified domains including newswire articles, Internet “news” articles, and broader network resources [5], [6]. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The activity of filtering can be modeled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and non relevant categories [7]. More complex filtering systems include multi-label text categorization automatically labeling messages into partial thematic categories.

### 2.2 Policy-Based Personalization Of OSN Contents

There have been some proposals exploiting classification mechanisms for personalizing access in OSNs. For instance, in [8] a classification method has been proposed to categorize short text messages in order to avoid overwhelming users of microblogging services by raw data. The user can then view only certain types of tweets based on his/her interests. In contrast, Golbeck and Kuter [9] propose an application, called FilmTrust, that exploits OSN trust relationships and provenance information to personalize access to the website. However, such systems do not provide a filtering policy layer by which the user can exploit the result of the classification process to decide how and to which extent filtering out unwanted information. In contrast, our filtering policy language allows the setting of FRs according to a variety of criteria, that do not consider only the results of the classification process but also the relationships of the wall owner with other OSN users as well as information on the user profile. Moreover, our system is complemented by a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure.

## 3. System Implementation and Working

First the pre-processing takes place on the posted message. And after that the pattern matching algorithm is applied for matching the text with the database contained word. Following Process is takes place for the Text filtering:

### 3.1 Pre-processing

The primary aim of the pre-processing phase is to remove from the input message all characters and terms that can possibly affect the quality of group descriptions.

#### 3.1.1 Text Filtering:

Text filtering step, all terms that are useless or would introduce noise in filtering process are removed from the input message. Among such terms are:

- HTML tags (e.g. <table>) and entities (e.g. &amp;) if any.
- non-letter characters such as "\$", "%" or "#" (except white spaces and sentence markers such as '.', '?' or '!')

#### 3.1.2 Stemming

Stemming algorithms are used to transform the words in texts into their grammatical root form, and are mainly used to improve the Information Retrieval System’s efficiency. To stem a word is to reduce it to a more general form, possibly its root. For example, stemming the term interesting may produce the term interest. Though the stem of a word might not be its root, we want all words that have the same stem to have the same root.

#### 3.1.3 Elimination of Stop Words

After stemming it is necessary to remove unwanted words. There are 400 to 500 types of stop words such as “of”, “and”, “the,” etc., that provide no useful information about the message. Stop-word removal is the process of removing these words. Stop-words account for about 20% of all words in a typical document. These techniques greatly reduce the size of the searching and matching each word in message. Stemming alone can reduce the size of an index by nearly 40%.



Figure1: Pre-processing of snippets

### 3.2 Modules

#### 3.2.1 Filtering Rules

The system provides a powerful rule layer exploiting a flexible language to specify Filtering Rules (FRs), by which users can state what contents should not be displayed on their walls.

#### 3.2.2 Online setup assistant for FRs thresholds

OSA presents the user with a set of messages selected from the dataset discussed. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are selected according to the following process.

#### 3.2.3 Blacklists

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs are directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when users retention in the BL is finished.

### 4. Conclusion:

In this paper, we describe our work to provide unwanted message filtering for social networks. we have presented a system to filter undesired messages from OSN walls. The system gives security to online social network. Our system is 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.

### References:

- [1] Measuring semantic similarity between words using web search engines. In WWW '07: Proceedings of the 16<sup>th</sup> international conference on World Wide Web, pages 757-766, New York, NY, USA, 2007. ACM.
- [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. J. Denning, "Electronic junk," Communications of the ACM, vol. 25, no. 3, pp. 163-165, 1982.
- [5] P. S. Jacobs and L. F. Rau, "Scisor: Extracting information from online news," Communications of the ACM, vol. 33, no. 11, pp. 88-97, 1990.
- [6] S. Pollock, "A rule-based message filtering system," ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.
- [7] P. J. Hayes, P. M. Andersen, I. B. Nirenburg, and L. M. Schmandt, "Tcs: a shell for content-based text categorization, " in Proceedings of 6th IEEE Conference on Artificial Intelligence Applications (CAIA- 90). IEEE Computer Society Press, Los Alamitos, US, 1990, pp. 320-326.
- [8] 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.
- [9] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in Provenance and Annotation of Data, ser. Lecture Notes in Computer Science, L. Moreau and I. Foster, Eds. Springer Berlin / Heidelberg, 2006, vol. 4145, pp. 101-108.
- [10] R. E. Schapire and Y. Singer, "Boostexter: a boosting-based system for text categorization," Machine Learning, vol. 39, no. 2/3, pp. 135-168, 2000.

### AUTHOR



**Miss. Rashmi R. Atkare** Received Bachelor's Degree in Information Technology from Amravati University in 2012 & Pursuing Master Degree In CSE from P.R. Patil College of Engg. Amravati-444602



**Prof. Mr. Pravin D. Soni** Received the Master Degree in Computer Science from VJTI, Mumbai in 2011. Working as Assistant professor in Department of Computer Science and Engineering at P. R. Patil College of Engg. Amravati-444602.