# INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

**A PATH FOR HORIZING YOUR INNOVATIVE WORK**

## IMPLEMENTATION OF FILTERED WALLS FOR UNWANTED MESSAGES USING CONTENT MINING

### MS. SHRUTI C. BELSARE[1], PROF. R. R. KEOLE[2]

1. Second year student M.E, Department of, Information of Technology, H.V.P.M's college of Engineering & Technology Amravati, India.
2. Assi. Professor, Department of, Information of Technology, H.V.P.M's college of Engineering & Technology Amravati, India.

**Abstract:** In today's On-line 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, 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. To study strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of bypassing the filtering system by creating a instance randomly notifying a message system that should instead be blocked, or detecting modifications to profile attributes that have been made for the only purpose of defeating the filtering system. Automatically user will get a mail notification.

**Keywords:** Online Social Networks, Short Text Classification, Information Filtering, Policy-Based Personalization, Filtering Rules and Blacklist Management.

**Corresponding Author: MS. SHRUTI C. BELSARE**

**Access Online On:**

www.ijpret.com

**How to Cite This Article:**

*PAPER-QR CODE*

## INTRODUCTION

Online Social Networks (OSNs) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. According to Facebook statistics1 average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (weblinks, news stories, blog posts, notes, photo albums, etc.) are shared each month.

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. Indeed, today OSNs provide very little 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. 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 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 [4] 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 (STC) are concentrated in the extraction and selection of a set of characterizing and discriminate features. The solutions investigated in this paper are an extension of those adopted in a previous work by us [5] from which we inherit the learning model and the elicitation procedure for generating reclassified data. The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to the context from which the messages originate. As far as the learning model is concerned, we confirm in the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification [4]. In particular, we base the overall short text classification strategy on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers, in managing noisy data and intrinsically vague classes. Moreover, the speed in performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks. We insert the neural model within a hierarchical two level classification strategy. In the first level,

the RBFN categorizes short messages as Neutral and Nonneutral; in the second stage, Nonneutral messages are classified producing gradual estimates of appropriateness to each of the considered category. Besides classification facilities, the system provides 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. More precisely, FRs exploit user profiles, user relationships as well as the output of the ML categorization process to state the filtering criteria to be enforced.

## 2. LITERATURE REVIEW

M.Chau and H.chen [3] proposed a machine-learning-based approach that combines Web content analysis and Web structure analysis. We represent each Web page by a set of content-based and link-based features, which can be used as the input for various machine learning algorithms. The proposed approach was implemented using both a feedforward/backpropagation neural network and a support vector machine. Two experiments were designed and conducted to compare the proposed Web-feature approach with two existing Web page filtering methods - a keyword-based approach and a lexicon-based approach. The experimental results showed that the proposed approach in general performed better than the benchmark approaches, especially when the number of training documents was small. The proposed approaches can be applied in topic-specific search engine development and other Web applications such as Web content management.

Marco Vanetti [1] presented a key OSN service that has not been provided so far. Indeed, today OSNs provide very little 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). Limitations are, 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.

Elisabetta Binaghi[6] states that, the main contribution of this paper is the design of a system providing customizable content-based message filtering for OSNs, based on ML techniques. Content-Based Filtering Information filtering systems are designed to classify a stream of dynamically generated information dispatched asynchronously by an information producer and present to the user those information that are likely to satisfy the requirements.

P.W. Foltz[7] and P.S. Jacobs[8] proposed 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.
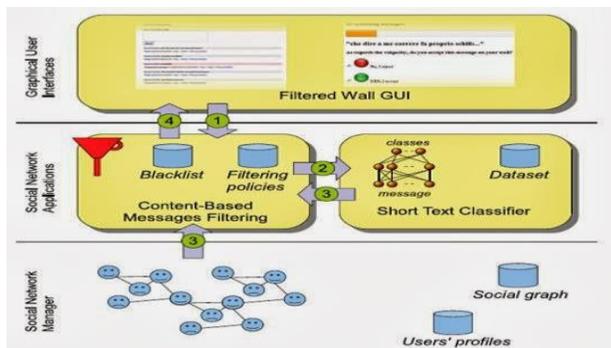
## 3. SYSTEM ARCHITECHURE



**Figure1. System architecture**

The architecture in support of OSN services is a three-tier structure (Fig. 1). The first layer, called Social Network Manager (SNM), commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management), whereas the second layer provides the support for external Social Network Applications (SNAs).The supported SNAs 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.

## 4. PROPOSED WORK

User Friendliness is provided in the application with various controls provided by system Rich User Interface.

The system makes the overall project management much easier and flexible.

It can be accessed over the Intranet.

The user information can be stored in centralized database which can be maintained by the system.

This can give the good security for user information because data is not in client machine.

Authentication is provided for this application only registered Users can access.

There is no risk of data management at any level while the project development is under process.

Categorization of computers in the database make easy to identify the various computer peripherals.

## 5. IMPLEMENTATION

According to limitations and literature reviews of this paper will experimentally evaluated and implemented an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN 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. Additionally, we plan to investigate the use of online learning paradigms able to include label feedbacks from users and we plan to enhance our system with more sophisticated to decide when a user should be inserted into a BL. 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. The solutions investigated in this paper are an extension of those adopted in a previous work by us from which we inherit the learning model and the elicitation procedure for generating pre-classified data. The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to the context from which the messages originate. As far as the learning model is concerned, we confirm in the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification. In particular, we base the overall short text classification strategy on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers, in managing noisy data and intrinsically vague classes. The goal of the paper is therefore to propose and experimentally evaluate an automated system, called Filtered Wall. System proposed Machine Learning (ML) text categorization techniques. Create neural model within a hierarchical two level classification strategy. Generate user-defined Blacklists (BLs) module.

## 6. MODULES

Filtering Rules and Blacklist Management:

In what follows, 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. In particular, each edge is labeled by the type of the established relationship (e.g., friend of, colleague of, parent of) and, possibly, the corresponding trust level, which represents how much a given user considers trustworthy with respect to that specific kind of relationship the user with whom he/she is establishing the relationship. In this paper, we do not address the problem of trust computation for indirect relationships, since many algorithms have been proposed in the literature that can be used in our scenario as well. Such algorithms mainly differ on the criteria to select the paths on which trust computation should be based, when many paths of the same type exist between two users.

Filtering Rules:

In defining the language for FRs specification, we consider three main issues that, in our opinion, should affect a message filtering decision. First of all, in OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, FRs should allow users to state constraints on message creators. 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. In such a way it is, for instance, possible to define rules applying only to young creators or to creators with a given religious/political view. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth, and trust values of the relationship creators should be involved in order to apply them the specified rules. All these options are formalized by the notion of creator specification

## 7. CONCLUSION

In this paper, we have presented a system to filter unwanted messages from OSN walls. On-line 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, we enhance the system by creating a instance randomly notifying a message system that should instead be blocked, or detecting modifications to profile attributes that have been made for the only purpose of defeating the filtering system. Automatically user will get a mail notification. Additionally, we plan to

investigate the use of online learning paradigms able to include label feedbacks from users and we plan to enhance our system with more sophisticated to decide when a user should be inserted into a BL.

## 8. REFERENCES

1. Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo," A System to Filter Unwanted Messages from OSN User Walls", IEEE Trans. Knowledge and Data Eng., vol. 25, no. 2, pp. 1041-4347 February 2013.

2. A. 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.

3. 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.

4. 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.

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

6. 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.

7. P.W. Foltz and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information Filtering Methods," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992.

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

9. S. Pollock, "A Rule-Based Message Filtering System," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.

10. G. Amati and F. Crestani, "Probabilistic Learning for Selective Dissemination of Information," Information Processing and Management, vol. 35, no. 5, pp. 633-654, 1999.