

An Effective Approach for Post Supervision in Blogosphere

Kamanashis Biswas¹, Md. Liakat Ali², S.A.M. Harun³

¹Senior Lecturer, Daffodil International University, Department of CSE, Dhaka, Bangladesh.

²Lecture, International Islamic University Chittagong, Department of EEE, Chittagong, Bangladesh

³Project Manager, Right Brain Solution, Dhaka, Bangladesh.

ananda@daffodilvarsity.edu.bd, liakat3026@gmail.com, harun@rightbrainsolution.com

Abstract

A web log or blog in short is a trendy way to share personal entries with others through website. A typical blog may consist of texts, images, audios and videos etc. Most of the blogs function as personal online diaries while others may focus on specific interest such as photographs (photoblog), art (artblog), travel (tourblog) etc. Another type of blogging called microblogging is well known now-a-days which combines very short posts. Like the developed countries, the users of blogs are gradually increasing in the developing countries e.g. Bangladesh. Due to the nature of open access to all users, some people abuse it to spread fake news to achieve individual or political goals. Some of them also post vulgar materials that make an embarrass situation for other bloggers. The way to overcome this problem is to bring all the post under supervision of the site administrator. But it is totally contradictory to the blog concepts. In this paper, we have suggested an effective approach that would restrict the offensive entries from being posted. These entries would go through a supervision process to justify themselves as legal posts. Here, we have shown that this approach can eliminate the chaotic situations in blogosphere at a great extent.

Index Terms— Blogger, Blogosphere, Post Supervision, Moderator.

1. Introduction

The old fashion of writing diaries is now changed to a new form known as blogging – a modern way of keeping diaries online. It is a platform where bloggers can share their daily lives, thoughts, problems, suggestions to others and blog readers can also post their feedbacks through comments or personal messages. As the users of computer and internet have increased dramatically all over the world, blogging has become a habit to many users. There are more

than 189 million bloggers all over the world according to McCann. Besides the developed countries, the people of developing countries are becoming more addicted to blogging day by day. For example, Indian blogosphere comprises more than 2 million blogs today. More than 50,000 people blog regularly in Bangladesh.

Like any other entity, blogs, though increasing popular day by day, are also facing some serious problems which have great impact on overall society. One of the major characteristics of blogs is that it provides open access to all. Anyone can register to a blogsite and he/she can post his personal journal anytime using his/her account. As there is no process of gate keeping on blogging, there is no check on what kind of content is being published by the bloggers.

The world of User Generated Content is full of material which puts negative impact on an individual, place, organization etc. which leads to defamation of somebody [1]. “The blogosphere has created an open, unfiltered megaphone for anyone with a computer and a modem, a striking characteristic of contemporary times is an unseemly incivility.” Says Robert Dilenschneider in his book ‘power and influence’. This incivility includes making an untrue statement to a person or organization that damages the subject’s reputation. Libel and Slander are two subcategories of defamation where first one is committed in printing media such as articles of a magazine or newspaper and the second one is in spoken form such as person-to-person or broadcast over a radio or television channel. Blogosphere very often suffers from defamation in a printed forum. A number of researches are done which include protection against spam blogs (Splog) [3, 7], instant blog updating and retrieving [2], providing fast news alert to the RSS feeds subscribers [4]. But no mechanism is proposed to overcome the above prob-

lem. In this paper, we have suggested a new approach which will work as a gate-keeper and will be able to minimize the number of offensive posts and comments.

2 Blogging Communication Model

2.1 Current Blogging System

The following figure illustrates the current blogging system where bloggers directly write their entries using user interface module and submit to the storage database. Immediate, it is published to the desired page. There is no gate-keeping mechanism followed in this process.

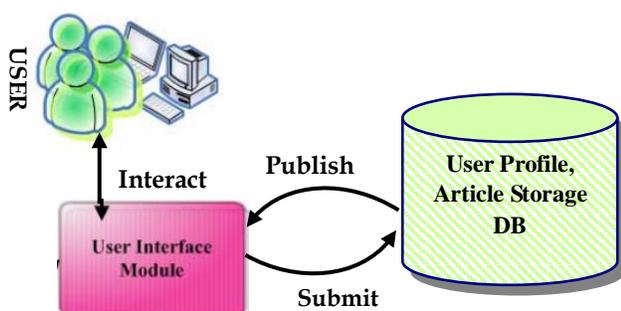


Fig. 1: Traditional Blogging System

2.2 Blogging System under Supervision

This proposed system is improved from the earlier model as each of the post must pass the verification phase. The following figure illustrates the process. In this system, every entry is assessed using an algorithm and if it achieves certain level of scores then will be published directly or directly with a notification. Otherwise it will be pending or rejected on the basis of the score.

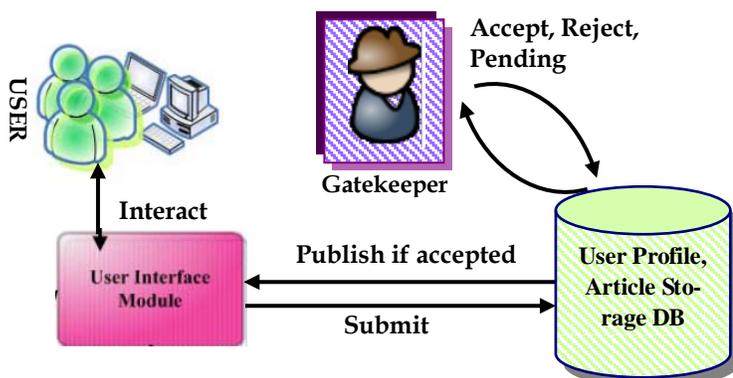


Fig. 2: Blogging under Supervision

3 Case Study

There are about 10 blogs in Bangladesh and most of them are developed for blogging in Bangla platform. The very first Bangla blog is somewhereinblog [5] which has around 40,000 registered accounts. About 10,000 bloggers visit, post, and interact on this site daily in average. At the beginning, this blog was open for everyone, i.e. anyone could register as a user and then could post his entries activating his/her account. Unfortunately, some bloggers abused this opportunity by posting offensive materials. Some of these posts were to spread ideology of some forbidden groups. On the other hand, some posts and comments were made to spread fake news intentionally. During mid November, 2009, this misuse had been occurred so frequently that the site administrator imposed some restrictions. Now, the user's entry does not immediately appear in the home page after registration. It is only shown at his personal page. The blog administrator will observe new bloggers' activities and he/she will get access to the home page after getting safe blogger notification from administrator. There is also a rating system for the bloggers which indicates whether the blogger is in safe state or not. However, this chaos in somewhereinblog compelled many users to leave the site and at the same time a new blogging platform developed named as 'sachalayatan' [6] which is known as 'online writers' community'. This forum does not provide post directly to the home page. Users have to write their entries as guests and the entries have to wait for the approval of the moderator. Each user has to post as guest for long time before getting access to write as a regular member. Both the two blogs are following a supervision process which requires human interaction. Sometimes this takes long time and users get bored in blogging to these sites. This finally leads to decreasing users' interest.

4 Blog-Post Supervision Process

We proposed an algorithm which will check the blog objects that are submitted to be published. Here, we used a dictionary database which has two types of words and also a list of links of restricted sites. First type is slang word list which will be used to measure frequency of these words in a blog-post. Another type is demand-based word list. Anytime, new words can be added to or deleted from this list.

Sometimes it is necessary to add some words which are not slang but required to control unusual events. For example, when a disaster (such as earthquake, firing etc.) or political violence (such as revolt in PilKhana) happens, then a number of blogs are posted immediately but most of them include wrong information about the incidents. These posts quickly spread rumor which may be harmful for the whole society. Demand base word list will help to protect the blogosphere from this type of chaotic situation.

4.1 Proposed Algorithm for Supervision

Procedure POST_SV (Blog_Object)

```
//Blog_Object may be Title, Post or Comments
// Post or Comment includes body which may contain
    Texts, links of audio, video, documents or websites
//Organize Blog_Object data for search, i.e. links will be searched first
]

Do for each part of Blog_Object
While (True)
{
    //Continue till to end of selected module
    Integer frequency_level, check
    //Initialize frequency_level and check to zero

    Select an Item from Blog_Object data
    IF (Item is link and found in list)
        Reject the post, send a notification and set check to 1
    ELSE IF (Item is found in demand-based list)
        Keep it pending for the administrator's approval and set
        check to 1
    ELSE
        {
            For each Item in Database
            {
                IF (Found)
                    Calculate the overall frequency of the Items
            }// End_For
        } //End_ELSE

    IF (check is not 1)
        IF (frequency_level > 60%)
            Reject the Post and send a notification
        ELSE IF (frequency_level is between 20-59%)
            Place it for moderator's approval
        ELSE IF (frequency_level is between 1-20%)
            Publish it and send a notification
```

```
ELSE
    Just publish it
} //End_While
```

5 Conclusion

It is too hard to keep the blogosphere free from misuse. Only one way that can be done is check each and every part of the user generated contents. But it is totally contradictory of blog characteristics and also time consuming. Our proposed algorithm will play very effective role in this case. It will be very easy to find out the offensive posts and prohibit them from being post immediately. Though it is not possible to eliminate the current problem completely but still it will help to improve the chaotic situation of blogging world. We hope that it is possible to reduce about 60% of offensive posts from being post. However, one of the big problems is that in some cases some vulgar words may be used in a post for different purpose (for example, tutorials). And there is a chance that this algorithm will treat the post as suspected one. Another possibility is that a post with very few vulgar words may be published due to low frequency level. However, except these problems, we can say that the algorithm is efficient enough to find out unpleasant entries very quickly and to handle the anomalous situations in the blogosphere.

5.1 Futre Work

This algorithm can not check some of the posted or attached contents such as image or pdf files. It also cannot check the audio file contents for selecting a victim. There is scope to work in this area to improve the checking capability. Another property of the algorithm is that it is very simple. This can be improved using decision tree mechanism which will make a decision on the basis of the correlation of previous behavior of the user from the database and the frequency level. Finally, we did not implement the algorithm in real scenario. Hence, all the concepts are hypothetical. A good conclusion can be drawn after implementing the algorithm in practical field and analyzing the outcomes.

6. References

- 1 [Gore, S.], *I Blog Therefore I Am*, by Sneha Gore, Dissertation Report, Master in Communication Studies, Depart-

ment of Communication Studies, 2009, University of Pune

- 2 [Leu, J. Chi, Y. Chang, S. and Shih, W.], *BRAINS: Blog Rendering and Accessing INSTantly System*, by Jenq-Shiou Leu, Yuan-Po Chi, Shou-Chuan Chang and Wei-Kuan Shih, IEEE International Conference on Wireless And Mobile Computing, Networking and Communications, 2005. (WiMob'2005), 22-24 Aug. 2005, Volume-4, P.1-4, Print ISBN: 0-7803-9181-0
- 3 [LIN, Y. SUNDARAM, H. CHI, Y. TATEMURA, J. TSENG B.L.], *Detecting Splogs via Temporal Dynamics Using Self-Similarity Analysis*, by Yu-ru Lin and Hari Sundaram, Yun Chi, Junichi Tatemura, and Belle L. Tseng, ACM Transactions on the Web, Volume 2 , Issue 1 (February 2008), ISSN:1559-1131.
- 4 [Sia, K. C. Cho, J. and Cho, H.], *Efficient Monitoring Algorithm for Fast News Alerts* by Ka Cheung Sia, Junghoo Cho, and Hyun-Kyu Cho, IEEE Transactions on Knowledge and Data Engineering, Volume 19 , Issue 7 (July 2007), P.950-961, ISSN:1041-4347
- 5 Somewhereinblog, <http://www.somewhereinblog.net/>, Access date: May 14, 2010
- 6 Sachalayatan, <http://sachalayatan.com/>, Access Date: May 14, 2010.
- 7 [Takeda, T. Takasu, A.], *A Splog Filtering Method Based on String Copy Detection*, by Takaharu Takeda and Atsuhiko Takasu, Applications of Digital Information and Web Technologies, IEEE, 4-6 August, 2008, P-543-548, ISBN: 978-1-4244-2623-2



Kamanashis Biswas, born in 1982, post graduated from Blekinge Institute of Technology, Sweden in 2007. His field of specialization is on Security Engineering. At present, he is working as a Senior Lecturer in Daffodil International University, Dhaka, Bangladesh.



Md. Liakat Ali, born in 1981, post graduated from Blekinge Institute of Technology, Sweden 2007 in Security Engineering and 2008 in Telecommunication. Now he is working as a Lecturer in International Islamic University Chittagong, Bangladesh.



S. A. M. Harun is graduated from International Islamic University Chittagong. He is a programmer and ACM problem setter. Now he is working as a project manager in a software company. His major area of interest is developing efficient algorithm.