Multi-dimensional Moderation in Online Communities: Experiences with Three Norwegian Sites

Gheorghita Ghinea[1], Bendik Bygstad[2], and Christoph Schmitz[1]

[1] School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge UB8 3PH, London, United Kingdom
[2] The Norwegian School of Information Technology, Schweigaardsgt. 14, 0185 Oslo, Norway
george.ghinea@brunel.ac.uk, bygben@nith.no, christoph.schmitz@gmail.com

Abstract. Online-communities and user contribution of content have become widespread over the last years. This has triggered new and innovative web concepts, and perhaps also changed the power balance in the society. Many large corporations have embraced this way of creating content to their sites, which has raised concerns regarding abusive content. Previous research has identified two main different types of moderation; one where the users have most of the control as in Wikipedia, and the other where the owners control everything. The media industry, in particular, are reluctant to loose the control of their content by using the member-maintained approach even if it has proven to cost less and be more efficient.

This research proposes to merge these two moderation types through a concept called multidimensional moderation. To test this concept, two prototype solutions have been implemented and tested in large-scale discussion groups. The results from this study show that a combination of owner and user moderation may enhance the moderation process.

Keywords: moderation, online communities, social media, prototype.

1 Introduction

Utilizing users to create both content and increase activity has become an important factor for many web sites. The concept has been implemented and used in e-commerce solutions, online newspaper editions and pure community concepts. While online communities have several merits, they also present challenges for the owners of these websites and the users themselves. These issues span from vandalism, harassment, false and potentially harmful advice to copyrighted material. The research shows evidence that these problems have already been present in the early days of online communication (Foley, 1989). A number of solutions have been proposed and deployed to deal with these issues. To this end, Preece (2004) argues that establishing a good etiquette and a positive environment is the ultimate solution for reducing these types of anti-social contributions from the users. In particular she recommends the use of moderators to achieve this. She also expresses a need to develop better processes that combine human-oriented approaches with good technical solutions.

A.A. Ozok and P. Zaphiris (Eds.): Online Communities, LNCS 5621, pp. 191–196, 2009.

Most of the research done so far, on how to create an efficient moderation mechanism, is based on sites that are of a member-maintained character. The aim of the study reported here is to investigate the balance when combining owner-maintained and member-maintained moderation. This will hopefully answer a need in the industry to continue ensuring the quality of content, yet still utilize the possibility of letting users do part of the moderation.

2 Related Work

Within online communities, issues with users misbehaving have also been reported several times in existing literature. Preece (2004), for instance, describes community users being tired of horrid comments, observing that it only takes one aggressive, insulting person to ruin a community for everyone else. Moreover, differences in the cultural origins of users strengthen the risk for one person's joke to be interpreted as an insult by another. Cosley et al. (2005) also state that not all contributions to a community are valuable. They identify the motivational component behind users contributing with low quality content, as being the response a user receives from a post. Predictions are made that reducing responses would take away some of this motivation. Analogies are made with the New York anti-graffiti campaign, which was based on quick removal of vandalism, and with Wikipedia, which takes an average of 3 minutes to correct a page deletion. This implies that taking quick action is somewhat logical and has a great effect on reducing the motivation behind these anti-social actions. From a different perspective, users being treated poorly online and the fear of aggressive responses is also one of the reasons identified by Preece et al. (2003) for users not contributing to online forums. Moreover, another concern identified in online communities is that of false information; however studies, in health groups for instance, have shown that the groups themselves correct the false information within a period of time (Jadad et al., 2006).

Inevitably, all of these issues point towards importance of sound moderation to a community as a whole, and this position is confirmed by Preece (2004), who argues that the ultimate solution for reducing such unwanted artefacts of user online behaviour is to establish a good online (n)etiquette and a positive environment. Good moderation is one of the elements mentioned as a mean to achieve this. She calls for a focus on etiquette, and online role models, and one type of role model mentioned is the moderator. However, her work also confirms that moderating can be demanding and time-consuming, as the busier community gets, the more effort from moderators is needed. However it is uncertain if this argument presumes owner-maintained moderation.

Efficient moderation can only be undertaken if there is adequate tool-set support. Among several sets of common technically oriented tools used by the communities to aid this process are ones which enable moderators to identify, approve, reject, delete, edit or to request the sender to edit his or her messages. Search and visualisation tools are other tools that are presented. They include rating of users that contribute in message boards and ratings of the most valuable contributors. A fourth tool/mechanism is given by rating and reward schemes where the users themselves rate other users. In calling for better processes and tools for ensuring good online etiquette, Preece (2004) argues that the way forward is to develop processes that bring together the best human-oriented approaches supported by good, labour-saving, technical solutions.

3 Study Description

The study being described in this paper took place in Norway and three sites agreed to participate in it:

Ungdomsportalene.no is an online youth community based on the same social networking principles as Facebook and MySpace.com, with a general message board as a public add-on feature. The focus here is communication; users get their own profile, can send messages, and write in each other guestbooks, in addition to participating on message boards and commenting on news articles.

Kvinneguiden.no originally started as an online bridal magazine with a message board to discuss wedding topics. Over time it has transformed into a general online lifestyle magazine for (generally) female audiences. It has a news desk publishing articles for this segment and a message board.

Diskusjon.no is the second largest message board in Norway (Big-Boards.com, 2007). It started out as a website with tests and articles regarding hi-tech computer hardware, as well as a message board where the users could discuss different issues, tests and products. They have now grown larger in terms of target audience and types of categories covered, which now include digital cameras, computer games and mobile phones.

The consensus at all three sites participating in our study was that candidates for user-moderators should be found among their loyal members with a consistent level of posts, and that, moreover, such users should not have had any records in the abuse register over a certain period of time. This, as it was felt that this category of netizens is the most likely ambassador of online etiquette. Lastly, we used *Speed of Moderation* as the metric for our study.

Two prototypes incorporating user-moderation were implemented and ported to the three case sites of our study during March – June 2007. Both prototypes were based on the concept of multidimensional moderation, derived from the Slashdot and MovieLens experiences (Lampe and Johnston, 2005). The idea is that the first dimension is represented by the users in their separate roles, of owners and members. An extra dimension is then given by experienced users, who give feedback to other users in terms of a common understanding of what constitutes good or bad posts.

Whilst the underlying moderation principle was the same in both prototypes (experienced users reporting potential violating posts), what was different between the two versions was the number of negative reports before a post was suspended pending a final decision from the owner-moderators. Thus, whilst in the first prototype a post had to have had three such reports (consultation with the owners of the three sites involved in our study revealed worries that waiting for a higher number of reports would have in unacceptable negative impact on the moderation speed), in the second prototype this threshold was lowered to two.

4 Results

All results obtained using the two developed prototypes are compared to two baseline measurements obtained during the 28 days of February 2007. These consist of the average time, expressed in minutes, from an offending post being written to it being deleted, as well as the total number of such deleted posts on the three sites participating in our study.

4.1 First Prototype: Results

When investigating the results from the first implementation at Kvinneguiden.no (Figure1) one may see that the average time taken to delete an offending post decreased by 752 minutes compared to the baseline measurements. The same trend is also seen at Diskusjon.no, where there is a decrease in the average time such a post is online before it is deleted. However, the number of posts deleted at Diskusjon.no as a direct consequence of the prototype was only four. This hinted at the possibility that the number of negative flags before a post was deleted was possibly set too high for this particular site. On Ungdomsportalene.no, though, the results are mixed when compared to the baseline. Thus, the average time taken to delete such a post actually increased (from 2220 minutes to 3016 minutes). Nonetheless, those posts that were removed as a direct result of the implemented prototype, had been done so in the fastest time of all sites in our study (only 267 minutes compared to 1889 minutes on Kvinneguiden.no and 4821 minutes on Diskusjon.no).

4.2 Second Prototype: Results

When evaluating the results from the first prototype, it became clear that the number of notices from the users was too low on two of the sites, with few posts receiving

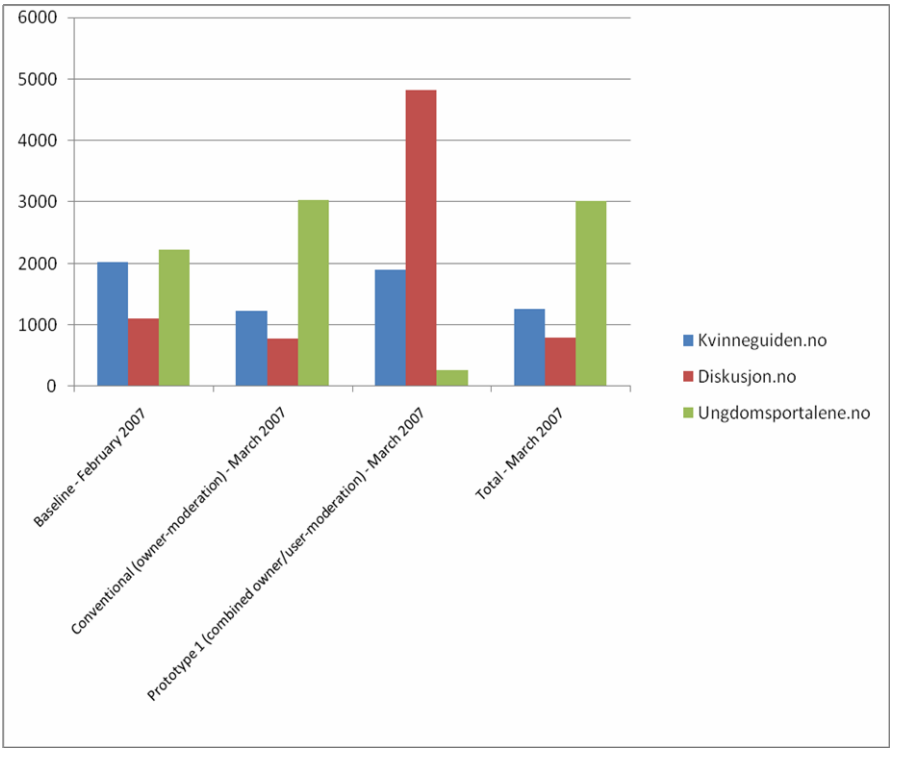

Fig. 1. Average time (mins.) to delete posts as a result of prototype 1

more than two notices from users. Indeed, it became apparent that there was a rather large possibility that a potentially offending post would not have had enough time to receive a sufficient number of notices to suspend it under the new regime before being filtered by the old (owner-led) reporting system. Thus, a second prototype was implemented, where the number of notices needed before any action was taken on a potentially violating post was lowered to two.

Looking at the effect of the second prototype at Kvinneguiden.no (Figure 2), one may see that the average time taken to delete a post actually went down (by exactly 500 minutes). Moreover, while combined owner-user moderation affected nine posts in March, it only affected five in May, notwithstanding the reduction of the number of notices needed for a post to be suspended under the second prototype. This may be explained by the fact that the time taken to delete a post using traditional owner-led moderation had actually witnessed a dramatic reduction between March and May (from 1226 minutes to 767 minutes). Thus, it may well be that 767 minutes was not a long enough time-span for a post to gather the minimum number of two notices required for it to be suspended using combined owner-user moderation.

A different pattern is seen at Ungdomsportalene.no where the average time taken to delete a post decreased down to 2530 minutes, and at Diskusjon.no, where the average time taken to delete a post also decreased by 106 minutes when compared with the first prototype.

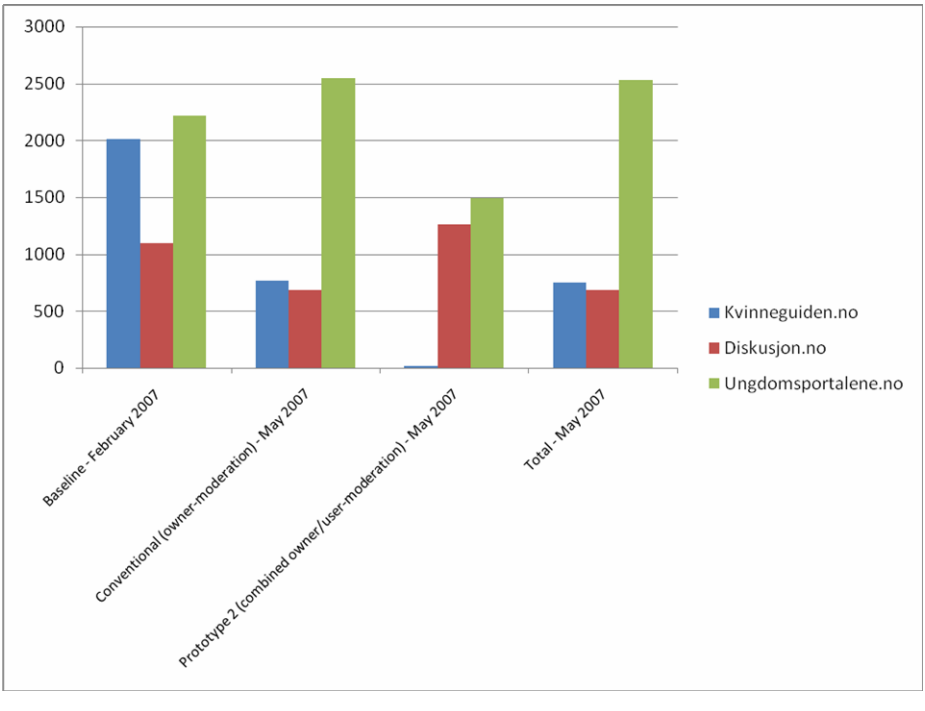

Fig. 2. Average time (mins.) to delete posts as a result of prototype 2

5 Conclusions

This paper reports the results of a study which examined the impact of multi-dimensional (i.e. combined owner-user) moderation on three Norwegian case sites. Whilst our work is inevitably limited by the context in which it took place, the participation of three leading bulletin boards of an IT-intensive country such as Norway in the study has offered some interesting and valuable insights into this particular type of moderation.

Our results seem to suggest that, generally speaking, in the case of a post needing two reports for it to be deleted, user involvement does speed up the moderation process – either through posts being deleted more quickly as a direct result of user-reporting, or, indeed, user involvement seemingly acting as a catalyst for owner-moderations to do a faster moderation of posts. Nonetheless, there are exceptions to these observations within our results and one would need to examine the impact of user involvement over a longer period of time in order to obtain a better picture of the trends concerned.

References

1. Cosley, D., Frankowski, D., Kiesler, S., Terveen, L., Riedl, J.: How oversight improves member-maintained communities. In: Proceedings of the SIGCHI conference on Human factors in computing systems, Portland, Oregon, USA, pp. 11–20 (2005)
2. Foley, T.J.: Managing Campus-Wide Information Systems: Issues and Problems. In: Proceedings of the 17th annual ACM SIGUCCS conference on User Services, Bethesda, Maryland, USA, pp. 169–174 (1989)
3. Jadad, A.R., Enkin, M.W., Glouberman, S., Groff, P., Stern, A.: Are virtual communities good for our health? British Medical Journal 332(7547), 925–926 (2006)
4. Lampe, C., Johnston, E.: Follow the (Slash) dot: Effects of feedback on New Members in an Online Community. In: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA, pp. 11–20 (2005)
5. Preece, J.: Etiquette online: From nice to necessary. Communications of the ACM 47(4), 56–61 (2004)
6. Preece, J., Nonnecke, B., Andrews, D.: The top five reasons for lurking: improving community experiences for everyone. Computers in Human Behavior 20(2), 201–223 (2003)