

On Reflecting Visitors' Opinions Fairly and Accurately on the Web

Won Kim, Hyungsuk Ji, and Hyunseung Choo

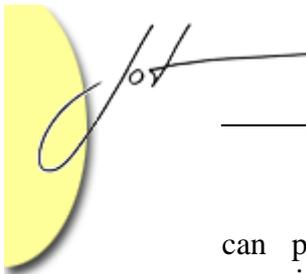
Abstract

Today very popular Web portal sites, social networking sites, online media sites, commerce sites, etc., provide platforms for millions of visitors to visit daily and express their opinions on a wide variety of subjects. The site operators strive to increase the number of visitors, and the visitors often make use of several means at their disposal to participate in the formation of collective opinions. In this article, we examine the various means becoming available to the Web site visitors to express their opinions, and the challenges that both the site operators and the general public face to ensure the visitors' opinions are fairly and accurately reflected in the collective opinions.

1 INTRODUCTION

Recently, social networking sites (such as YouTube, Digg, Flickr, MySpace, Facebook, LinkedIn, Cyworld (in Korea), etc.), Web portal sites (such as Yahoo, Baidu (in China), Naver (in Korea), etc.), media sites (such as New York Times, ESPN, CNN, FoxNews, Chosun (in Korea), etc.), commerce sites (such as Amazon, Hotel, Gmarket (in Korea), etc.), learning sites (such as Wikipedia, About, etc.) are drawing anywhere from hundreds of thousands to tens of millions of visitors daily. The site operators provide contents and/or platforms on which the visitors may upload and share user-generated contents (UGCs) and to express their opinions in any of several means. These means include posting comments, participating in discussions or forums, responding to polling questions, voting "like/dislike" (e.g., 'digg it/bury it' on Digg) on other people's postings, voting "thumbs up/down" on other visitors' comments, sharing contents with "friends", saving contents for future viewing, copying contents in their blogs, etc.

It is desirable that the opinions expressed by the Web site visitors be fairly and accurately reflected in the formation of collective opinions. The most important reason is that the collective opinions, although expressed by a minority of all Web site visitors, can influence the formation of the general public opinions, and in turn government's policies on a full range of momentous issues as the election of national leaders, waging a war, national security, education and welfare reforms, immigration policies, etc. Further, the collective opinions conveyed by the very popular Web sites



can provide important input to businesses, educational institutions, religious organizations, etc. that can influence the directions and decisions they make.

The seemingly modest goal of the Web sites' fair and accurate accounting of the visitors' opinions, however, presents significant challenges. The reason is that there are many ways that can lead to distortions of the collective opinions. Some of the ways are attributable to the visitors and others to the site operators. Most Web site operators need to pursue business goals and promote certain points of view. Further, a tiny activist minority may hijack some key policy decisions from the general public, and many Web site visitors do various things to distort the collective opinions.

In this article, we will examine various ways in which the collective opinions can be distorted, and offer approaches to mitigate the problems.

2 WAYS TO DISTORT THE COLLECTIVE OPINIONS

There are various reasons collective opinions cannot be entirely fairly and accurately formed when a very large number of Web site visitors express their opinions. They include the following:

1. Many visitors may not be knowledgeable enough to express correct opinions on certain subjects. In other words, their opinions may be based on wrong or partial information.
2. Many visitors may make mistakes when they express their opinions. For example, they may misread the polling question, or not understand the meaning of certain words or phrases, such as "digg it" or "bury it" (on Digg), and click them just to find out what would happen.
3. Some visitors may even hack into the Web sites and manipulate the database of the opinions and statistics on them.
4. Some visitors may actively try to manipulate the opinion counts or visibility of their postings. There are various ways they can do this. They may vote multiple times; they may post similar comments under multiple user IDs (called "sock puppeting"); they may time the posting of their opinions for higher visibility; they may mobilize their friends to support their opinions, etc.

The networking facilities provided by major social networking sites today have the potential to significantly facilitate the mobilization of activists. Sites such as Facebook, MySpace, YouTube, etc. allow the visitors to create or join groups, communicate with members of the groups through emails or instant messages, and share UGCs. Some sites allow visitors to copy UGCs into their blogs to share with other visitors or into their playlists for later viewing and sharing with registered "friends" or members of their groups.

The New York Times has reported sock puppeting of a chief executive of a company who assailed a competitor in an Internet forum [Stone07]. There are a number of similar ways to manipulate opinions using sock puppeting. A visitor may even intentionally post a weak argument to an opinion that he is against, so as to motivate like-minded other visitors to vigorously attack the opinion. One visitor may make use of multiple accounts in a similar way.



Some activists often mobilize others, or even hire people, to shape collective opinions along their beliefs. In Korea, a number of young voters used the Internet to mobilize support for their favored candidate in the very closely contested presidential election in 2002. It is widely acknowledged in Korea that the efforts of the Internet-savvy young people decided the outcome of the election. It was a sobering experience for those who supported the opposing candidate at the time, and it motivated them to learn to use the Internet. As a result, now it is said that, of the people who post comments to political news articles in online newspapers in Korea, those in their 40s and 50s outnumber those in the 20s and 30s.

The site operators can also cause, or even lead, the formation of distorted collective opinions. They may actively do certain things, or they may be careless in the procedures they use to collect visitors' opinions, or they may turn a blind eye to some of the irregular things the visitors do.

They may actively delete opinions and ratings that are contrary to ones that they want. There are indications that some news media sites do this, judging from some of the visitors (in postings that obviously were not deleted) who complain that "all my postings get deleted".

They may even have people post opinions they dictate. The people they use may be employees, family and friends of employees, or part-time free-lancers.

They may display vastly inflated visitor counts or ratings. Some of the Web sites may assign a large number when starting to count visitors to a UGC or votes to a poll.

They may manipulate the ordering or placement of the postings; they would place those posting that match their views on more visible pages and more visible locations of a page. They may also purposely highlight certain postings. For example, they may include certain postings in the "Top 10" list or "Featured" list, even if those postings may not be very popular or rated highly.

For example, Daum, the second largest Web portal in Korea has a system that displays the titles/snapshots of UGCs on the main page. It displayed more than 20 opinions highly critical of the Korean church that had sent missionaries to Afghanistan who were kidnapped by the Taliban. These postings generated some 70,000 page views on the subject in Daum, while only 700 page views on the subject were generated in Naver, the largest Web portal in Korea.

They do not attach a warning to certain types of postings. For example, authors of a book may post a few favorable reviews on their book, and those may be the only reviews on the book. Similarly, travel sites may have a tiny number of customers who had bad experience in certain hotels post harsh comments on the hotels. Such statistically insignificant, and overly biased, postings can mislead other visitors. However, the site operators mostly do not provide a statement on the potentially misleading nature of the postings to advise the visitors to check additional sources to reach a balanced and accurate conclusion.

3 ISSUES WITH THE WAYS WEB SITE OPERATORS CONVEY THE VISITORS' OPINIONS

The Web site operators have various means to gather, process, and display opinions expressed by the visitors. Using these means, they are in a position to influence the formation of the collective opinions. In this section, we examine some of these means, and problems with them.

A basic thing the site operators do is display visitors' opinions or aggregated counts of the opinions. This, however, is not as simple as it may appear. The site operators must first block or delete opinions expressed in "unacceptable" ways – profane language, libelous statements, threats, etc. They also need to ensure that the visitors not vote multiple times or use multiple user IDs. Moreover, they need to protect the visitors from spammers and launchers of malware (virus, worm, spyware); many social networking sites require user validation when the visitors try to share UGCs via email.

Once the site operators have validated the postings, they would display them. To display a posting, the position and ordering must be determined. Let us examine the ordering of the postings.

The postings may be ordered either in an ascending order and a descending order of the timestamp associated with them. In the descending order, the most recent posting is placed at the top, and the earliest one at the end. It encourages new visitors to post their opinions. The online USA Today uses this order for users' comments on articles. This is also used by YouTube, MySpace, Friendster, Fotolog, Skyrock, Flickr, Digg, etc. One problem with this order is that visitors intent on distorting the collective opinions may post multiple times the same or similarly worded opinions to ensure that their opinions be placed on or near the top of the list. The ascending order is used by online Washington Post for users' comments on newspaper articles. It discourages late comers from posting their opinions, for their postings would tend to be invisible, since they will tend to discourage visitors by requiring additional clicks to access them.

Once a large number of valid postings have been collected, the site operators can aggregate them, and report various types of aggregated counts. (YouTube, Hi5, Skyrock, etc. use the aggregated counts to also determine the display ordering.) A few types of aggregated counts have been in use for a while: the number of votes cast in response to polling questions, the number of page views, and the number of comments and responses. The online New York Times uses counts of emails received about articles in order to select distinguished articles. The significance of the response count can be questionable, however, since the count may simply reflect the fact that the issue or UGC is very controversial or provocative. Some visitors may manipulate the response counts by posting unnecessarily provocative or polemic opinions or UGCs.

Recently, social networking sites have introduced additional types of aggregated counts. For example, Digg measures the number of "digg"s (likes), the number of "share"s (the visitors want to share a particular URL with friends), and the number of "blog"s (the visitors want to use the URL in their blogs). YouTube measures the



number of “favorite”s (the visitors save the URL in their playlists for viewing or sharing later), and the number of “share”s. The significance of the differences among these measures need to be clearly understood when looking at the counts. In both online forums and UGC-sharing sites, some visitors manipulate the view counts by resorting to the use of sensational title/snapshot to attract attention.

4 WAYS TO IMPROVE FAIRNESS AND ACCURACY

In this section we will outline approaches that may be taken so that the site operators may reflect the visitors’ opinions more fairly and accurately. Basically, the site operators should curtail what visitors may do that can distort the collective opinions, and government agencies and/or non-profit civic groups should monitor and curtail what the site operators do that can distort the collective opinions. We discuss what site operators should do.

They should make the polling questions clear and at least as well-formed as offline opinion research firms have learned to do for the past several decades. The responses to ambiguous questions are not useful and the aggregated results may be misleading.

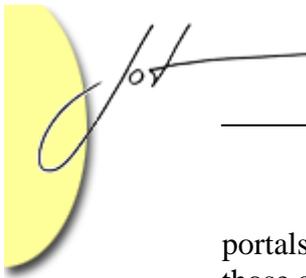
They should make the meanings of the vote-gathering functions clear, so that fewer visitors will make mistakes.

They should try to reduce chances of unintentionally misleading the visitors with statistically insignificant aggregated counts. There should be some threshold that makes counts statistically significant, and the site operators may post a warning on the statistical insignificance if the count is well below the threshold.

They should be able to identify multiple identities belonging to the same person in order to combat sock puppeters. This is not easy to realize, especially when the sock puppeteer simultaneously uses different user IDs and different IP addresses. Of course, even if it is technical feasible to identify sock puppeters, it is expensive to actually detect them and take corrective actions.

Page views are certainly one of the most important measures that site operators seek to maximize. As such, site operators may find it difficult to diligently block all forms of visitors’ efforts to distort the collective opinions, if those efforts would substantially increase page views. Examples of such efforts include, as mentioned earlier, the use of sensational topics and provocative discussion subjects, posting pornographic and provocative UGCs, posting sensational rumors and even fabricated stories, etc. The site operators may also be tempted to actively do things to increase page views at the expense of distorting collective opinions. Therefore, it may be necessary to require the site operators to maintain a log of all the opinions and votes the visitors leave, and show the log to some organizations, such as a government agency or non-profit organizations empowered to demand to see it.

It may also be necessary for some organizations to monitor how the Web portals display UGCs on the main page. In Korea, some NGOs do this, especially during critical periods like a presidential election. It is important for countries where dominant Web portals, which should maintain neutrality, are encroaching the news media’s territory. In Korea, the majority of the Web visitors read online news on Web



portals that have either copies of news articles from online news media or links to those articles.

5 CONCLUSION

As tens of millions of Web site visitors now record on the Web site their opinions (and UGCs), the opportunities are becoming greater to influence or create public opinions on a wide range of issues by using or manipulating the collective opinions on the Web sites. Therefore, it is becoming important that both the Web site visitors and the site operators recognize the impact of the fair and accurate accounting of the collective opinions expressed by the individual Web site visitors. The public opinions formed on certain momentous issues, manipulated by a tiny activist minority or some of the site operators, may sometimes be good for the general public, but sometimes may not be what the general public would have desired, if they had gone through the normal process of learning and deliberating on the issues.

The site operators should adopt better techniques and carefully designed procedures to block the visitors' actions that may distort the collective opinions, and to fairly and accurately convey the aggregated opinions and counts. In the near future, government agencies and non-profit civic groups may also need to be drawn into the monitoring and regulating of some aspects of the operating practices of the dominant Web sites.

At the end of the day, people should remind themselves that, despite the paradigm-shifting nature of the Internet and the Web, most of the manifestations of the negative and frail aspects of the human spirit in the offline world have counterparts in the online world; and as such, in a sense, it is only natural that the collective opinions conveyed by the Web sites are often distorted.

ACKNOWLEDGMENTS

This research was supported by MIC, Korea under ITRC IITA-2007-(C1090-0701-0046).

REFERENCES

- [Stone07] Brad Stone and Matt Richtel: "The Hand That Controls the Sock Puppet Could Get Slapped", *The New York Times*, July 16, 2007.



About the authors



Won Kim, Professor and Univeristy Fellow with the School of Information and Communication Engineering at Sungkyunkwan University, Suwon, S. Korea. He is Editor-in-Chief of ACM Transactions on Internet Technology (www.acm.org/toit). He is Global General Chair of the [Human.Society@Internet](http://www.human-society-internet.org) International Conference. He is the recipient of the ACM 2001 Distinguished Services Award, and is an ACM Fellow. He can be reached at wonkim@skku.edu



Hyungsuk Ji is a research professor in the School of Information and Communicatin Engineering at Sungkyunkwan University, Korea. He received a Ph.D. in cognitive science from the Institut National Polytechnique de Grenoble, France. His research interests include corpus linguistics, cognitive linguistics, semantic representation with a computational model and other areas in natural language processing (computational linguistics). He participated in developing the Atlas Project <http://dico.isc.cnrs.fr>.



Hyunseung Choo received a Ph.D. degree in computer science from the University of Texas at Arlington, USA in 1996. After a stint as a patent examiner at Korean Industrial Property Office. Since 1998, he joined the faulty of the School of Information and Communication Engineering at Sungkyunkwan University. He is currently an associate professor and Director of the Intelligent HCI Convergence Research Center with financial support from Korea's Ministry of Information and Communication Korea. His research interests include wired/wireless/optical networking, mobile computing, and grid computing. He can be reached at choo@ece.skku.ac.kr