The Bots That Got Some Votes Home

Updates on the Digital Natives video contest voting results that was declared in May 2012

It was a hint of suspicion raised by one of our colleagues at The Centre for Internet & Society that spurred our Web Analytics team to check into the voting activity of the contest that was all about the 'Everyday Digital Native'. And while we acknowledged and celebrated the 'digital' in the native (users of technology), we forgot the human part that the digital has to engage with. Following weeks of deliberations, we now have conclusive evidence that points to irregularities in voting numbers of the Top 10 contestants. We are now staring at the White Elephant in the room: those innocuous little automated scripts we have sweetly nicknamed 'bots'.

Internet bots, also known as web robots or simply bots, are software applications that run automated tasks over the Internet. Typically, bots perform tasks that are both simple and structurally repetitive, at a much higher rate than would be possible for a human alone. The largest use of bots is in web spidering, in which an automated script fetches, analyzes and files information from web servers at many times the speed of a human. Each server can have a file called robots.txt, containing rules for the spidering of that server that the bot is supposed to obey. In addition to their uses outlined above, bots may also be implemented where a response speed faster than that of humans is required (e.g., gaming bots and auction-site robots) or less commonly in situations where the emulation of human activity is required, for example chat bots (Source: Wikipedia).

What irregularities?

You would see how a script or bot would have played a role in 'automating' the votes for a video. The Top 10 videos received a combined voting number of 20,000+. The discrepancy occurs at the juncture where the votes polled on the front end (the webpage where the contestant video was visible to the public) did not match with the number of hits the page received on the backend (this is the analytics part). For instance, the top polled video has some few thousand votes more than the number of people who actually visited our CIS website in the same duration. This prompted a review of the logs and the possible "hand" of a *non-human agent* acting on its human creator's command to drive up the votes.

How was this done? The Technicalities

The following graph shows the extremely high level of voting requests just before the closing date (March 31, 2012). This would not be extraordinary except for the fact that two or three entries had an exceptionally higher vote count relative to their page views as per the analytics statistics.

~~GRAPH IMAGE.~~

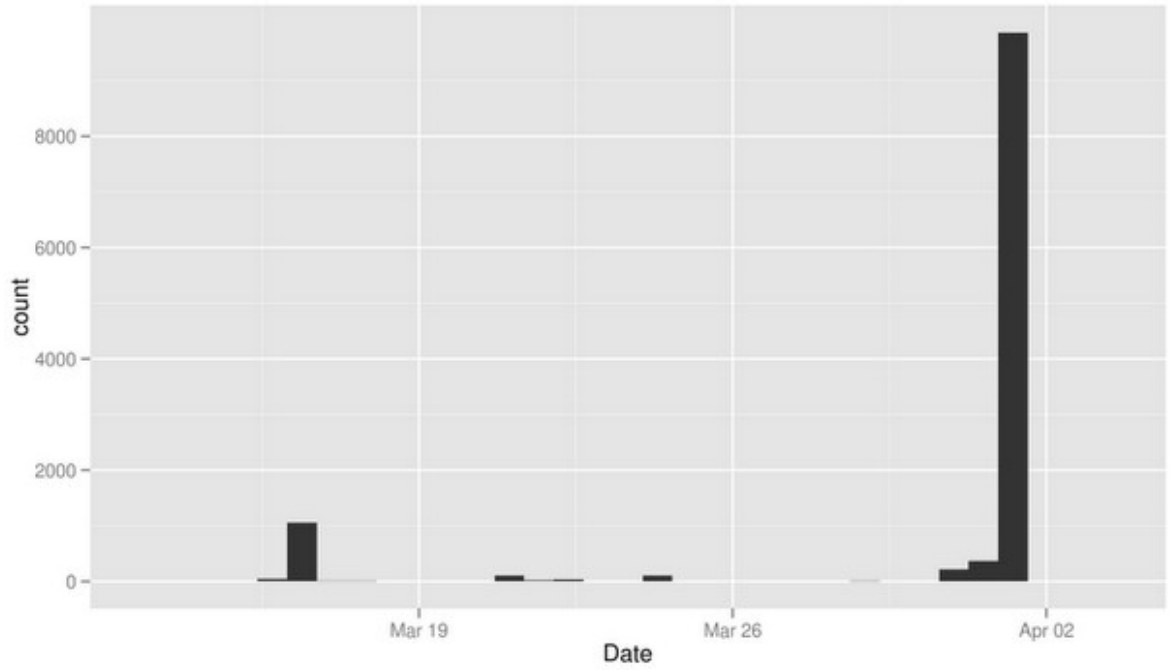


Illustration : Voting Requests by Date

1. These are the public votes displayed on the contestant's page through the thumbs up icon

2. These are http requests to the voting link against each video when the user clicked on the thumbs up icon.

3. These are http requests which are collectively related to the video page (page view). A normal human user would browse through a page first, which downloads some other urls, such as the HTML for the page, javascript, images, and so on. A normal vote request would be included collectively. ~~A direct http request to the voting link would such as that via a script or through none-browser means would not download any other parts of the page (this sentence needs re-writing).~~A direct http request to the voting link on the other hand does not do this, and only makes a specific request to vote without downloading the other parts that make up the page.

4. A normal human vote count should be the same or less than the number of page views. Only three videos highlighted show abnormal behavior and it is recommended these be adjusted to the page view counts.

Are you saying contestants cheated?

While the use of programming scripts to accrue votes is no new tactic and we should, in fact, have a more robust mechanism to monitor such activity during a contest, we cannot prove the culpability of the human agents. The contestants might be innocent actors with overzealous friends or colleagues who ran the voting scripts. As of now, since there is no way to ascertain their part in this irregularity, it's best we give them the benefit of the doubt. What comes through

loud and clear is that once you do away with the scripted votes, four contestants still manage to have enough votes to maintain their positions in the final five. In the fifth position, we now have a contestant from the top ten finalists, who has secured the requisite votes (after vote adjustment) to propel him into the final five.

Recommendation

Digital Dance (Cijo Abraham), From the Wild Into the Digital World (John Musila) and Digital Coverage in a Digital World (T.J. Burks) had additional vote url counts than page views. It is recommended that the total votes for these videos be adjusted to the page view counts, and not the actual vote counts as displayed on their individual webpages (thumbs up icon) during the voting period.

The rankings of the adjusted voting would now read as:

1. Top Votes: Connecting Souls, Bridging Dreams – Marie Jude Bendiola (1113)
2. Second: From The Wild Into The Digital World - John Musila (810)
3. Third: With No Distinction - T.J. K. M. (369)
4. Fourth: I Am A Ghetto Digital Native – James Mlambo (321)
5. Fifth: Digital Dance – Cijo Abraham (198)

Transparency at CIS

The Digital Natives with a Cause? research inquiry has always been shaped around concerns of transparency, equity and community accountability. In our research methods as well as in outputs of the different activities, we have always maintained a complete transparency of decision making processes as well as in depending upon the incredible people we work with to help us learn, grow and reflect openly on the concerns that we have been engaged with. We strive to follow this method and in publishing these statistics, we want to ensure that there is complete transparency about the votes that were accrued and how the final winners were selected. We also take this opportunity as a learning experience to re-think the question of the non-human actor in our networks and further about the nature of participation and reputation online. We hope that the publishing of these results will help answer any inquiries on how the process unfolded.

View Logs and Source Code

- All logs from the web server for this period. (24.7MB) Identical IPs are from caching server.
- R script to evaluate data for table

What next?

Since we spotted the error in time, we haven't disbursed the prize money of EUR 500 to each of the Top 5 contestants. They will now receive the prize along with a chance to participate in the Digital Native workshop-cum-Webinar, slated to be held in July 2012~~mid-June~~. The top 10 videos will be showcased in this event.