



University of
Massachusetts
Amherst

Politics 2.0 with Facebook – Collecting and Analyzing Public Comments on Facebook for Studying Political Discourses

Item Type	article;article
Authors	Shah, Chirag;Yazdani nia, Tayebeh
Download date	2025-05-24 03:29:20
Link to Item	https://hdl.handle.net/20.500.14394/31316

First Submission: 01/28/2011

Revised Submission: 05/06/2011

Accepted:

RUNNING HEAD: POLITICS 2.0 WITH FACEBOOK

**Politics 2.0 with Facebook – Collecting and Analyzing Public
Comments on Facebook for Studying Political Discourses**

Chirag Shah and Tayebeh Yazdani nia

School of Communication & Information

Rutgers, The State University of New Jersey

New Brunswick NJ 08901 USA

chirags@rutgers.edu, tyazdani@eden.rutgers.edu

Abstract

Analyzing publicly available content on various social media sites such as YouTube and Twitter, as well as social network sites such as Facebook, has become an increasingly popular method for studying socio-political issues. Such public-contributed content, primarily available as comments, let people express their opinions and sentiments on a given topic, news-story, or post, while allowing social and political scientists to extend their analysis of a political discourse to social sphere. We recognize the importance of Facebook in such analysis and present several approaches and observations of collecting and analyzing public comments from it. In particular, we demonstrate what it takes to do this manually, what we could learn from it, and how we can automate this process using a Facebook Harvester tool we have developed. In addition, we show how a hybrid approach can be formed giving us quick and easy data collection, and meaningful data analysis with substantially less effort than a manual approach. We believe these methods and tools will be highly valuable for political scientists in studying various political discourses as they take place in the Web 2.0 world.

Keywords

Facebook; Social media data extraction; Comments analysis; Political discourse.

Introduction

The public messages exchanged by social network site members, sometimes called comments or wall postings are a new type of text-based communication. These messages are unusual in that they are public – either world-visible or visible to all of a members' friends – and can be permanently associated with the identity of the poster, more directly and publicly so than listserv

postings (Thelwall, 2009). Due to the nature of these postings, that is, being public and available for harvesting,¹ they make an excellent choice for a social or political scientist to capture and study them as a proxy for people's voices for a given political discourse.

Thelwall (2009) argued that the widespread use of social network sites (SNS) in many countries (Boyd and Ellison, 2007) makes them an important object of study, and also gives an opportunity to investigate informal interpersonal communication on a larger scale than previously possible. In another study, Thelwall (2010) looked at the role of emotion in SNSs and whether emotion is typically reciprocated, and whether Friends express and/or receive similar levels of emotional expression to each other. His findings indicate statistically significant evidence for a weak correlation between the strength of positive emotion exchanged between Friends and received by Friends. This has larger implications on understanding how information propagates from person to person and source to source using SNS, and what they mean to the receiver of information (Shah, 2010).

In this workbench note, we take the problem of collecting and analyzing public comments from Facebook, one of the most popular SNSs, and demonstrate the pros and cons of various approaches. Specifically, we walk the reader through a series of steps that one may have to take for manually studying a political topic on Facebook, demonstrate a fully-automated process, and then a hybrid approach for collecting and analyzing public comments from Facebook. This note

¹ According to Facebook Statement of Rights and Responsibility: 2.4, "When you publish content or information using the 'everyone' setting, it means that you are allowing everyone, including people off of Facebook, to access and use that information, and to associate it with you (i.e., your name and profile picture)."

provides a brief overview of the relevant literature, describes our methods, and their implications for studying political discourses. The system used for automatic and hybrid approaches is one of the crucial and sustaining contributions of this note.

Background

In the growing body of literature on SNSs, several articles have been published focusing on Facebook in particular (e.g., Mayer and Puller, 2008). These studies examine a diverse array of topics, from social capital (Ellison et al., 2007), to information disclosure (Gross and Acquisti, 2005), to temporal patterns in messaging (Golder et al., 2007).

More recently the Facebook data team (2010) looked at the usage of words in different “word categories” in status updates. This led them to discover some patterns in how people use status updates differently, and how their friends interact with different status updates.

Some research in this area has focused on the uses and gratifications of Facebook as Joinson (2008) does. Lampe et al. (2007) meanwhile explored the relationship between profile structure (namely, which fields are completed) and number of friends. Also, Ellison et al. (2007) examined the relationship between use of Facebook, and the formation and maintenance of social capital. They claimed that the site is tightly integrated into the daily media practices of its users; the typical user spends about 20 minutes a day on the site, and two-thirds of users log in at least once a day (Cassidy, 2006). Much of the existing academic research on Facebook has focused on identity presentation and privacy concerns (e.g., Gross & Acquisti, 2005; Stutzman, 2006). Looking at the amount of information Facebook participants provide about themselves, the

relatively open nature of the information, and the lack of privacy controls enacted by the users, Gross and Acquisti (2005) argued that users may be putting themselves at risk both offline (e.g., stalking) and online (e.g., identity theft). Other early Facebook research examined student perceptions of instructor presence and self-disclosure (Hewitt & Forte, 2006; Mazer, Murphy, & Simonds, 2007), temporal patterns of use (Golder, Wilkinson, & Huberman, 2007), and the relationship between profile structure and friendship articulation (Lampe, Ellison, & Steinfield, 2007).

Ringel et al. (2010) explored the phenomenon of using social network status messages to ask questions. They conducted a survey of 624 people, asking them to share the questions they have asked and answered of their online social networks. They presented detailed data on the frequency of this type of question asking, the types of questions asked, and respondents' motivations for asking their social networks rather than using more traditional search tools like Web search engines. They reported on the perceived speed and quality of the answers received, as well as what motivates people to respond to questions seen in their friends' status messages.

When studying the existing literature on Facebook, it becomes clear that there is a lack of research on extracting and analyzing public comments from Facebook. At the same time, collecting and studying such data could be a highly valuable method for researchers looking at various political discourses. This motivated us to take on the task of looking at the comments that appear on various Facebook pages with a focus on studying people's comments and the discussions that follow. In other words, we are interested in the 'meaningful' discussions and debates that take place on Facebook and want to analyze these messages to be able to make sense of them.

Method

We started by asking ourselves – “how would one collect and analyze Facebook comments without any specialized support?” The answer to this is given in the following subsection as a step-wise procedure. We then describe a tool that we have developed to automatically collect a large number of Facebook comments and status updates. Finally, we show how we could combine this automated process for data collection to manual analysis and form a hybrid approach.

Manual data collection and analysis

In order to understand the process of analyzing public comments on Facebook, we visited numerous Facebook pages, and collected and analyzed data (comments) manually. Following are some of our experiences and observations that resulted from this process.

- The layout of most of Facebook pages is set up in a way that the administrator of the page posts a news piece or a bold statement, and that creates a discussions thread that can go on for a while, sometimes days after the original posting. For example, *Reform Immigration FOR America* (<http://www.facebook.com/reformimmigrationforamerica>) is a page that is dedicated to the issue of immigration. Their mission statement as it appears on their info page is: “*the U.S. immigration system no longer works. Fixing it presents a daunting challenge, but action must be taken sooner rather than later. The time is NOW to do the right thing and fight for practical solutions that benefit all of us and are rooted in the restoration of the rule of law, earned citizenship, united families, and fair treatment of workers.*” We chose a wall post with a statement that was a quote from a senator who had spoken on the senate floor. We found 97 comments on this post that we

copied and pasted to a word file in a matter of seconds. Once all the comments were captured, we started going through each one and analyzing them based on relevancy, sentiments, objectivity, and the quality of messages. The objective of this analysis was to find “useful” comments that help us understand people’s reactions to the wall post. Assessing each of the collected comments for these criteria took approximately 30-40 minutes.

- We noticed that in contrast to sports or entertainment pages on Facebook, political pages generate comments that are to the point and for the most part relevant to the given topic at hand. Comments on other pages are sometimes unrelated and completely irrelevant to the page topic. Pages that tackle social/political topics such as healthcare reform or child obesity also generate meaningful discussions that we are able to analyze thoroughly. For example, the popular music group Coldplay has a Facebook page² that is used for reaching out to their fans and informing them of their activities such as tour dates, etc. The layout is very much similar to the other Facebook pages. There are announcements about the band, and each announcement generates hundreds of comments. These comments are overwhelmingly centered around the fans’ passion and devotion to the group. For instance, a link about a Christmas show in Liverpool³ generated 1,310 comments and 18,960 people Liked it. The comments were transferred to a word file and analyzed using the same criteria mentioned before, which took nearly two hours. The majority of the comments were in English and although there were no *real* discussions

² <http://www.facebook.com/coldplay/>

³ <http://www.facebook.com/coldplay/posts/182958395063566>

included, the great many of messages contained strong positive sentiments about the mentioned show.

- A good example for the social/political page is *Join the Coffee Party Movement* page on Facebook⁴, which carries the same layout as the other pages. One of the links on this page that we examined was a statement made by the administrator, which has generated a great discussion thread, and 67 people had commented and 764 had Liked. By analyzing the messages (about 40 minutes), we found that a great majority of comments were relevant to the subject of statement and also the majority could be assessed based on sentiment. The sentiments were a mixture of negative/positive and the discussions were for the most part *meaningful*. There were agreements and strong disagreements with the statement and each created replies by the other commentators.
- When it comes to Facebook pages for corporations such as Starbucks, we found that there were fewer discussions and more open-ended opinionated comments that center on Starbucks as a brand. By looking at about 200 comments from this page (about 1 hour), we can see that the majority of commentators were expressing their love or loyalty for the brand and sharing stories that centered on Starbucks. We found no *meaningful* discussion threads, but only sentiments that were overwhelmingly positive towards the brand.⁵

⁴ <http://www.facebook.com/coffeeparty/posts/144485285605017>

⁵ Note that the sentiment analysis was done by a single individual, which may have biased such subjective judgment. For a more complete analysis, one may want to involve multiple coders in the process.

Given that we wanted to analyze a comment based on its objectivity, sentiment, relevance and other criteria, it would be simply hard to do this only via the automatic process. For example, if we are to determine whether a comment holds sentiments, we must read it to be able to decide whether it is negative/positive. Sometimes a comment is issued in a sarcastic tone and holds the opposite sentiment of what it appears to show. This can only be detected through manual data analysis. The same argument can be made for objectivity and relevance. The most significant disadvantage of this approach, however, is the great amount of time it takes to evaluate large collections of data. Several of our analyses took about one hour for less than 100 comments. Given that many interesting and important political topics generate thousands of comments, and that these comments keep coming constantly, it becomes prohibitively expensive to study many of these political discourses using public comments. The following subsection demonstrates how we could collect a large amount of data from Facebook and start analyzing it with very little effort.

Automatic data collection and analysis

We have developed a Facebook Harvester to quickly and effectively collect a large amount of data from a Facebook page. This data includes the status updates as well as the wall postings. The harvester uses newly introduced Facebook Open Graph APIs⁶. Figures 1 to 3 show the working of this web-based harvester using screenshots.

⁶ <http://developers.facebook.com/docs/opengraph>

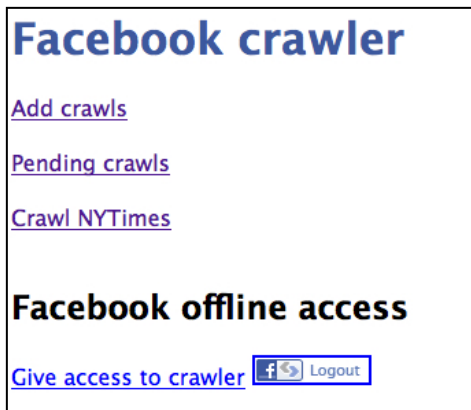


Figure 1: Main menu for the Facebook Harvester.



Figure 2: Starting a new harvesting process requires a Facebook page ID.

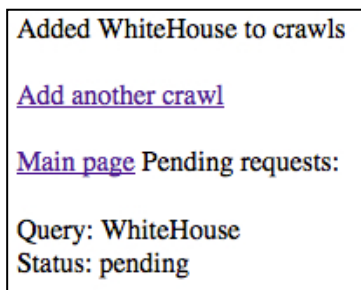


Figure 3: The harvesting process is now running in the background.

The result of running this harvesting process was thousands of messages (status updates and wall comments) within a few minutes. The data collected with these processes is stored in structured

format using MySQL. One could easily export this data in other structured formats, such as comma-separated values (CSV), or XML for further analysis. One could also run SQL queries on the MySQL database directly to filter, sort, and analyze the data.

We ran a similar harvesting process for a climate change group's Facebook page available at <http://www.facebook.com/pages/Climate-Change/>. The results of the data collection are shown in Figure 4 as a partial snapshot obtained using Sequel Pro.⁷ We have also developed a front-end web-based interface to display the collected data from harvesting processes (Figure 5).

⁷ <http://www.sequelpro.com/>

published	likes	comments_count	message	type
2010-05-29T16:03:55+0000	2	3	@ petite buenconsejo yes, i might have a re-write of our bataan camp bak...	status
2010-05-29T15:49:49+0000	0	3	Sayang di ka nakasama. baka meron pang susunod sama ka na. - Tito Andrei	status
2010-05-29T15:43:42+0000	2	0	Congrats sa successful clinic sa Bataan! -Dani â¥	status
2010-05-20T04:51:45+0000	0	0	To all the members of Climate Change, our rehearsals before the gig is at ...	status
2010-05-19T15:35:48+0000	3	0	See you tomorrow! Thursday, May 20. ;) SaGuijo Cafe + Bar. Makati ;)	status
2010-05-17T13:41:46+0000	1	0	To all member's Of Climate Change please be reminded of our rehearsals t...	status
2010-05-03T15:41:01+0000	0	2	Paging Mary, please practice Nana song and Anesthesia for recording this week	status
2010-04-24T11:55:34+0000	0	0	Agree ko Pastor, I will talk to Master para matapus na natin to, para maka...	status
2010-04-24T07:15:20+0000	3	3	Napansin ko ang bilis dumami ng sumusuporta sa atin. Kailangan na talag...	status
2010-04-20T04:53:11+0000	1	4	Mga friends, maglaro tayo ng konti. Paki rate naman ninyo ang mga kanta ...	status
2010-04-17T18:18:52+0000	0	0	Congrats to "Climate Change" ang ganda ng tunog dahil kay master	status
2010-04-16T13:53:38+0000	3	2	Sa open air pala bukas ang gig...kasama ang rivermaya sa mga mag perfor...	status
2010-04-16T05:33:01+0000	6	0	I would like to welcome our new bassist. Gian Pineda, Welcome Gian to Cli...	status
2010-04-06T13:17:52+0000	0	0	Paging members of Climate Change Band, we have practice tomorrow sa G...	status
2011-01-04T10:49:44+0000	0	0	http://konsyltaciai.com	wallpost
2011-01-04T10:48:30+0000	0	0	Prospective economics, energy management systems, new technology ma...	wallpost
2011-01-04T10:46:54+0000	0	0	Prospective economics, energy management systems, new technology ma...	wallpost
2011-01-04T05:28:46+0000	4	2	Coming Soon... >:)	wallpost
2011-01-02T04:26:38+0000	4	1	HAPPY NEW YEAR PO SA LAHAT! :)	wallpost
2011-01-01T14:06:35+0000	1	0	Climate Change will be performing tomorrow sa Gospel Jam together with ...	wallpost
2011-01-04T00:57:35+0000	4	2	Lagayan! Hehe! Happy New Year! =)	wallpost
2010-12-31T10:43:14+0000	1	0	Before the karaoke machines and fireworks go full-blast, a HAPPY NEW YE...	wallpost
2010-12-30T23:28:32+0000	0	1	klan na po labas ng album nio ?	wallpost
2010-12-30T19:26:23+0000	4	0	Happy New Year Everyone, Our call time on Gospel Jam on Sunday is 4:30pm.	wallpost
2010-12-25T01:12:53+0000	1	0	Maligayang Pasko po sa inyong lahat at maraming salamat po sa lahat ng ...	wallpost
2010-12-24T12:36:50+0000	3	0	Let's all take the time to remember that this holiday is not about receiving ...	wallpost
2010-12-24T07:09:53+0000	3	2	MERRY CHRISTMAS sa lahat ng Climate Changers and most especially sa m...	wallpost
2010-12-27T10:01:33+0000	4	3	Climate Change LIVE @ Mel & Joey	wallpost
2010-12-19T16:44:25+0000	0	0	Andrei Dionisio, Jr. on TV! â¥	wallpost
2010-12-19T14:52:07+0000	4	2	Salamat sa mga sumuporta sa guesting ni Hazel sa Mel & Joey. God bless u all!	wallpost
2010-12-19T12:45:43+0000	4	2	Mel & Joey na!!! Please watch. :-)	wallpost

Figure 4: A snapshot (back-end) of the data collected from the Climate Change Facebook page.

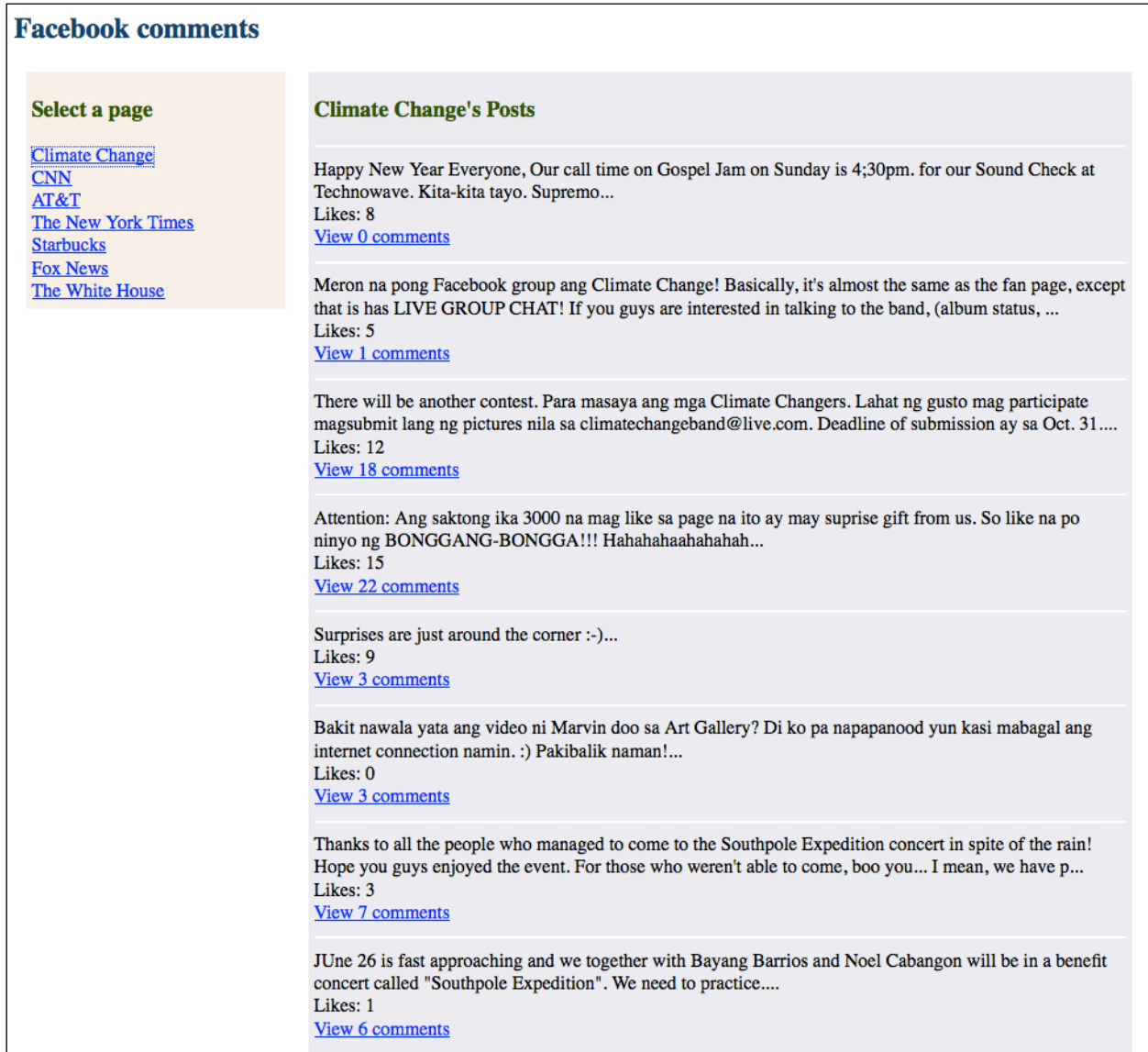


Figure 5: A snapshot (front-end) showing various harvesting jobs run and comments collected for 'Climate Change'.

Hybrid approach

At times and depending on the objective (whether it is sentiment analysis or relevance), it is both easier and more accurate to analyze the comments by simply reading them and going through them one by one. However, the manual approach may not be the most practical when it comes to

larger size data. When we are faced with thousands of comments, it can be difficult and we are looking to analyze them without the intention of investigating the details of each message, the automatic approach is our best solution.

Here we provide a hybrid approach, in which the data collection is done automatically and the analysis is facilitated with the help of sorting and filtering features of the system. Following on the same example of the previous subsection, we have collected a large number of comments from a climate change Facebook page using automated processes. This itself saved us enormous amount of time, but now we are left with thousands of these comments and it could take days to go through them.

To aid us in this process, we can use querying, sorting, and filtering using the Sequel Pro or a similar tool for database access and manipulation. This processing may differ depending on the objective of the. For instance, we may only want to look at messages that contain a certain word to see how many people are using a negative/positive term to address an issue and at the same time want to know the number of likes to those comments. Figure 6 shows a snapshot of the data with messages containing 'climate' word in them, and sorted by the number of comments to that message. This simple restructuring of the data was obtained by an SQL query taking only a fraction of a second to run. The obtained data is now a small subset (a few dozen messages), and more suitable and manageable for analyzing only the messages that explicitly talk about the climate, with the processing prioritized using the number of comments posted on a given message. This allows for a more thorough examination of discussions that take place on a given topic.

published	likes	comments_count	message	type
2010-10-27T07:38:57+0000	12	18	There will be another contest. Para masaya ang mga Climate Changers. La...	status
2010-06-03T03:28:31+0000	1	5	To all the members of Climate Change. we are all invited at Dani's party o...	status
2010-12-27T10:01:33+0000	4	3	Climate Change LIVE @ Mel & Joey	wallpost
2010-12-22T00:35:21+0000	4	3	Salamat po sa lahat ng patuloy na sumusuporta sa Climate Change pati na...	wallpost
2010-12-24T07:09:53+0000	3	2	MERRY CHRISTMAS sa lahat ng Climate Changers and most especially sa m...	wallpost
2010-12-18T17:39:30+0000	4	2	Malapit na ang album launching ng Climate Change...Isama po ninyo sa pa...	wallpost
2010-12-15T17:17:18+0000	4	2	Ilang tulog na lang available na ang first album ng Climate Change band. ...	wallpost
2010-11-01T15:46:13+0000	5	1	Meron na pong Facebook group ang Climate Change! Basically, it's almost ...	status
2010-05-20T04:51:45+0000	0	0	To all the members of Climate Change, our rehearsals before the gig is at ...	status
2010-05-17T13:41:46+0000	1	0	To all member's Of Climate Change please be reminded of our rehearsals t...	status
2010-04-17T18:18:52+0000	0	0	Congrats to "Climate Change" ang ganda ng tunog dahil kay master	status
2010-04-16T05:33:01+0000	6	0	I would like to welcome our new bassist. Gian Pineda, Welcome Gian to Cli...	status
2010-04-06T13:17:52+0000	0	0	Paging members of Climate Change Band, we have practice tomorrow sa G...	status
2011-01-04T10:48:30+0000	0	0	Prospective economics, energy management systems, new technology ma...	wallpost
2011-01-04T10:46:54+0000	0	0	Prospective economics, energy management systems, new technology ma...	wallpost
2011-01-01T14:06:35+0000	1	0	Climate Change will be performing tomorrow sa Gospel Jam together with ...	wallpost
2010-12-31T10:43:14+0000	1	0	Before the karaoke machines and fireworks go full-blast, a HAPPY NEW YE...	wallpost
2010-12-24T12:36:50+0000	3	0	Let's all take the time to remember that this holiday is not about receiving ...	wallpost
2010-12-19T16:44:25+0000	0	0	Andrei Dionisio, Jr. on TV! â¥	wallpost
2010-12-15T05:46:16+0000	0	0	Promote: Guys nood kau ng Mel And Joey sa linggo .. December 19 .. mag...	wallpost
2010-12-12T05:19:08+0000	0	0	Iwagayway Mo - Climate Change	wallpost

Figure 6: Filtered data filtered for messages with 'climate' in them, and sorted using the comments counts in the ascending order.

Another way of prioritizing message processing could be by using the length of a given message.

We observed that more *meaningful* messages tend to be lengthier than those without useful or interesting critique. Figure 7 shows a snapshot of the results ordered by the message length.

published	likes	comments_count	message	type	msg_length
2011-01-04T10:46:54+0000	0	0	Prospective economics, energy management systems, new technology ma...	wallpost	591
2011-01-04T10:48:30+0000	0	0	Prospective economics, energy management systems, new technology ma...	wallpost	568
2011-01-04T10:49:44+0000	0	0	http://konsyultacii.com	wallpost	444
2010-11-01T15:46:13+0000	5	1	Meron na pong Facebook group ang Climate Change! Basically, it's almost ...	status	405
2010-12-24T12:36:50+0000	3	0	Let's all take the time to remember that this holiday is not about receiving ...	wallpost	391
2010-12-18T17:39:30+0000	4	2	Malapit na ang album launching ng Climate Change...Isama po ninyo sa pa...	wallpost	334
2010-12-24T07:09:53+0000	3	2	MERRY CHRISTMAS sa lahat ng Climate Changers and most especially sa m...	wallpost	330
2010-12-31T10:43:14+0000	1	0	Before the karaoke machines and fireworks go full-blast, a HAPPY NEW YE...	wallpost	319
2010-12-19T16:44:25+0000	0	0	Andrei Dionisio, Jr. on TV! âY	wallpost	315
2010-10-27T07:38:57+0000	12	18	There will be another contest. Para masaya ang mga Climate Changers. La...	status	309
2010-12-25T01:12:53+0000	1	0	Maligayang Pasko po sa inyong lahat at maraming salamat po sa lahat ng ...	wallpost	272
2010-06-26T16:16:28+0000	3	7	Thanks to all the people who managed to come to the Southpole Expeditio...	status	269
2010-04-20T04:53:11+0000	1	4	Mga friends, maglaro tayo ng konti. Paki rate naman ninyo ang mga kanta ...	status	225
2010-04-24T07:15:20+0000	3	3	Napansin ko ang bilis dumami ng sumusuporta sa atin. Kailangan na talag...	status	186
2010-05-20T04:51:45+0000	0	0	To all the members of Climate Change, our rehearsals before the gig is at ...	status	184
2010-12-15T05:46:16+0000	0	0	Promote: Guys nood kau ng Mel And Joey sa linggo .. December 19 .. mag...	wallpost	174
2011-01-01T14:06:35+0000	1	0	Climate Change will be performing tomorrow sa Gospel Jam together with ...	wallpost	168
2010-04-16T13:53:38+0000	3	2	Sa open air pala bukas ang gig...kasama ang rivermaya sa mga mag perfor...	status	165
2010-06-12T10:04:34+0000	1	6	JUne 26 is fast approaching and we together with Bayang Barrios and Noel ...	status	162
2010-06-03T03:28:31+0000	1	5	To all the members of Climate Change. we are all invited at Dani's party o...	status	151
2010-04-06T13:17:52+0000	0	0	Paging members of Climate Change Band, we have practice tomorrow sa G...	status	149
2010-10-27T04:33:17+0000	15	22	Attention: Ang saktong ika 3000 na mag like sa page na ito ay may surpris...	status	146
2010-06-28T15:11:04+0000	0	3	Bakit nawala yata ang video ni Marvin doo sa Art Gallery? Di ko pa napapa...	status	145
2010-12-30T19:26:22+0000	8	0	Happy New Year Everyone, Our call time on Gospel Jam on Sunday is 4:30pm.	status	133

Figure 7: Data ordered by the message length.

Implications and Future Work

We demonstrated how one could go about collecting public comments data from Facebook and analyzing them manually. We pointed out general observations and specific lessons learned using several examples. It was clear that such data can be a valuable asset for studying a political discourse, but very expensive without additional technology support. We then presented Facebook Harvester, a web-based tool we have developed to collect public comments and their attributes from a Facebook page. These comments include status updates and wall posts. Furthermore, we showed how such automated data collection could be combined with simple filtering to provide us significantly less expensive analysis methods. This can be extremely helpful in studying various socio-political issues. A couple of scenarios are presented below.

- *General scenario.* The White House posts constant updates, announcements, and contents on their Facebook page,⁸ which includes pictures and videos. On a given post, there are typically few hundreds to few thousands comments posted by the visitors. One could not only collect these data once, but also keep collecting them at regular interval (e.g., daily) using our harvester. Such automated data collection and a few simple filtering could allow one to monitor White House's official stand on certain issues and people's opinions on them.
- *Specific scenario.* Starbucks recently rolled out a new brand logo, which created quite a bit of stir in the loyal fans and customers. Not surprisingly, they started posting comments on Starbucks' Facebook page expressing their opinions and sentiments. Using our system, one could easily collect Starbucks status updates and wall comments from Facebook to study these opinions and sentiments, as well as Starbucks' own reactions to these comments.

The Facebook Harvester is available for public access and use for free under a Creative-Commons license at <http://www.infoextractor.org/fbh/>. We are currently working on extending this tool to allow collecting data from sites other than Facebook, such as CNN.com and CNNMoney.com, which incorporate Facebook wall on their posts or news-stories for people to comment using their Facebook credentials. We also plan on including other popular sites where people post comments on socio-political issues, such as nytimes.com.

⁸ <http://www.facebook.com/WhiteHouse>

References

- Boyd, D., Ellison, N. (2007), "Social network sites: definition, history, and scholarship", *Journal of Computer-Mediated Communication*, available at:
<http://jcmc.indiana.edu/vol2013/issue2001/Boyd.ellison.html>, Vol. 13 No.1.
- Cassidy, J. (2006, May 15). Me media. *The New Yorker*, 50–59.
- Ellison, N.B., Steinfield, C., Lampe, C. (2007). The benefits of Facebook “friends:” social capital and college students’ use of online social network sites. *Journal of Computer-Mediated Communication* 12 (article 1).
- Ellison, N. B., Steinfield, C. and Lampe, C. Charles Steinfield, A familiar face(book): profile elements as signals in an online social network. *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '07*
- Ellison et al. (2007), The Benefits of Facebook “Friends:” Social Capital and College Students’ Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, 12: 1143–1168. doi: 10.1111/j.1083-6101.2007.00367.x
- Facebook Data Team (2010). What’s on your mind? Retrieved from:
<http://www.facebook.com/notes/facebook-data-team/whats-on-your-mind/477517358858>
- Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks, *Proceedings of WPES'05. ACM Alexandria, VA (2005)*, pp. 71–80.
- Golder, S.A. ,Wilkinson D. and Huberman, B.A. (2007). Rhythms of social interaction:

- messaging within a massive online network. In: C. Steinfield, B. Pentland, M. Ackerman and N. Contractor, Editors, *Proceedings of Third International Conference on Communities and Technology*, Springer, London, pp. 41–66.
- Hewitt, A., & Forte, A. (2006, November). *Crossing boundaries: Identity management and student/faculty relationships on the Facebook*. Paper presented at CSCW, Banff, Alberta, Canada.
- Joinson, A. N. Looking at, looking up or keeping up with people?: motives and use of facebook, *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* , CHI '08.
- Mayer, A. and Puller, S.L. (2008). The old boy (and girl) network: social network formation on university campuses, *Journal of Public Economics* 92, pp. 329–347.
- Mazer, J. P., Murphy, R. E., & Simonds, C. J. (2007). I'll see you on "Facebook:" The effects of computer-mediated teacher self-disclosure on student motivation, affective learning, and classroom climate. *Communication Education*, 56 (1), 1-17.
- Ringel et al. (2010). What do people ask their social networks, and why?: a survey study of status message q&a behavior. *Proceedings of the 28th international conference on Human factors in computing systems*. CHI '10.
- Shah, C. (2010). Information derivatives – a new way to examine information propagation. *Proceedings of Workshop on Human Computer Interaction and Retrieval (HCIR) 2010*. August 22, 2010. New Brunswick, NJ.

Stutzman, F. (2006). *An evaluation of identity-sharing behavior in social network communities*.

Paper presented at the iDMAa and IMS Code Conference, Oxford, Ohio.

Theilwall, M. (2009). MySpace Comments. *Online Information Review*, 33 (1), 58-76.

Theilwall, M. (2010). Emotion homophily in social network site messages. *First Monday* [Online],

15 (4). Available from

<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2897/2483>.

Author Note

Chirag Shah

School of Communication and Information, Rutgers University

Chirag Shah is an Assistant Professor in the School of Communication and Information at Rutgers University. He received his PhD in Information and Library Science from the University of North Carolina (UNC) at Chapel Hill. He received his MS in Computer Science from UMass Amherst, where he worked with Bruce Croft and James Allan on high accuracy retrieval, and topic detection and tracking. At UNC, he worked with Gary Marchionini and Diane Kelly on various issues concerning exploratory information seeking and interactive information retrieval. He has also worked at many world-renowned research laboratories, such as FXPAL in California and National Institute of Informatics in Tokyo, Japan. His dissertation is focused on collaborative information seeking. He is also interested in social search and question-answering, digital preservation, and contextual information extraction. He has developed several tools for exploratory information seeking and extraction, including "Coagmento" for collaborative information seeking and the award-winning "ContextMiner" for capturing contextual information from multiple online sources.

Correspondence concerning this article should be addressed to Chirag Shah, Rutgers University, 4 Huntington St. New Brunswick, NJ 08901 or to chirags@rutgers.edu.

Tayebeh Yazdani nia

School of Communication and Information, Rutgers University

Tayebeh Yazdani nia is a student of Master in Library & Information Science (MLIS) program in the Dept. of Library & Information Science within the School of Communication & Information (SC&I) at Rutgers University. She is interested in studying social media usage for understanding various socio-political issues.