Using social media to study social pheonomena

An example using Twitter data

Clayton Whitelaw
REU Student - Dept of Information Systems and Decision Sciences
University of South Florida
Tampa, FL

Manish Agrawal
Professor
Dept of Information Systems and Decision Sciences
University of South Florida
Tampa, Florida

H.R. Rao
Professor
Dept of Management Science and Systems
University of Buffalo
Buffalo, NY

Onook Oh
Visiting Instructor
Dept of Decision Sciences
University of Oregon
Eugene, Oregon

***Abstract*—Twitter is a social networking platform that offers a simple, yet powerful service. The majority of its functionality resides in 140-character-maximum statuses, or 'tweets.' We are developing a Java application that utilizes the service's application programming interfaces (APIs). The purpose is to offer a level of integrated functionality not provided by their website, and to circumvent some of the time constraints of manual analysis. The application provides a convenient way to pull tweets for a particular user or a specified search query into a delimited text document, offering very simple integration with spreadsheet software for analysis. This would be an especially valuable tool for people researching social networking activity, but can also be utilized by anyone to analyze public response to current events or popular culture. An example based on current events is presented.**

Keywords-Twitter; social; networking

I. INTRODUCTION

Telecommunications technology has evolved to a truly impressive level to this date. From telephones, to radios, to television, we have a lot to be proud of in relatively recent history. However, we are the generation of note - We have been lucky enough to witness the birth of some technology that has changed our lives entirely: the internet. We are now living through this societal period of change where people are becoming completely electronically connected, something that will likely continue for a very long time. We have computers, phones, and now cellular phones that perform many of the same functions as a computer. These are all important devices, but the internet is what brought all the public information of the world to our home - or, more recently, into our pockets. However, it's without a doubt that the biggest popular innovation of the web in the 2000s is the advent of social networks.

II. DATA COLLECTION INTERFACE

We developed a Java applet which allows users to search the Twitter database through the applet, which in turn communicates with Twitter's APIs. This applet improves upon the Twitter search, extending the search period from about 6 days to a user-defined period. Additionally, integration of API calls allows the researcher to attempt to work around the issue of not being able to publicly search a topic for longer than a week ago. Finally, the main intent of this program is to print out the details of each user and tweet into a text file, formatted in such a way that it can be easily imported into spreadsheet software for analysis. Figure 1 shows the interface.

The program alone can act as a bridge between Twitter and Microsoft Excel. By using API calls creatively, it attempts to overcome API limits to expand the results of a search by subject. The application allows the researcher to search for a specific topic, and pull statuses or information from any user who appeared in the search results. This leverages the common observation that if someone is posting about a particular topic, then there is a significant chance that the topic could have been mentioned before. Similarly, as people who are friends tend to share common interests, the people any given user talks to (by including @username in their status) may also post about the topic that the client is searching for. The interface of the application has been set up in such a manner that the user has to provide the minimal amount of information to complete the search.

There are two main ways for someone to use this applet: search by username or subject. A username search, as mentioned previously, returns up to 3200 statuses for that user, as well as some information about that user. The subject search can include a general topic, or a mention of a username (@username). Search results are determined purely by Twitter's search engine; this applet just formats the results of the queries, and allows the researcher to perform additional searches to expand the results. The applet can also attempt to translate a limited amount of foreign statuses through the Google Translate API.

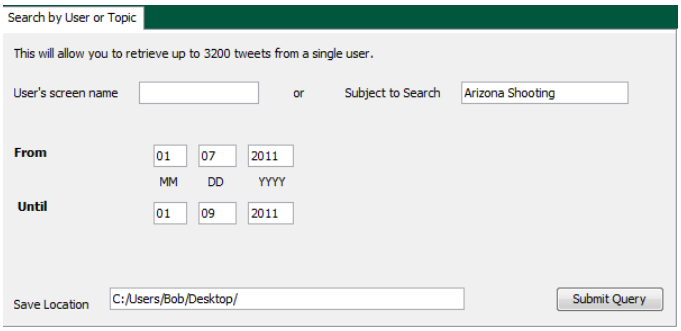

Figure 1 - GUI Sample

This research was funded in part by NSF grants 0926376 and 0926371.
978-1-4577-0161-0/11/$26.00 ©2011 IEEE

id	screen name	date	time	status
37786162668118000	con_te	2/16/2011	08:11:37Z	hey...wassup Egypt?
37786161216757700	opt2success	2/16/2011	08:11:37Z	BREAKING NEWS>> Egypt-style protesters rally in Bahrain MAN
37786157546868700	omarzarka	2/16/2011	08:11:36Z	RT @bencnn: As Arab regimes fall, we'll discover lots of dirty l
37786148143112100	Egypt20110125	2/16/2011	08:11:34Z	RT @monaeltahawy: RT @IceQueer I'm an atheist homosexual
37786140413145000	MariettaStire54	2/16/2011	08:11:32Z	@benthegosselink can not believe what is happening in #egyp
37786134016819200	amftrust	2/16/2011	08:11:31Z	RT @NoochMoney: Fascinating article on Egypt and the role of
37786127016394700	razzzletoff	2/16/2011	08:11:29Z	Anti-government protests spread to Libya (AP): AP - Hundre
37786126987034600	razzzletoff	2/16/2011	08:11:29Z	amzn.to/dZLP1 Anti-government protests spread to Libya (A
37786126911545300	AlyaaGad	2/16/2011	08:11:29Z	The Egyptian government (in exile & denial) has terribly violat
37786123409440700	angienassar	2/16/2011	08:11:28Z	RT @BloggerSeif: #Lebanon, enough with the corruption, enou
37786117457580000	sayedastudio	2/16/2011	08:11:27Z	i was also harassad on the very day of the friday celebrations f
37786114265718700	tvheadlines	2/16/2011	08:11:26Z	Horror! 60 Minutes Correspondent Lara Logan Sexually Assault
37786103247421400	ElSult	2/16/2011	08:11:23Z	RT @Eyousry: @Sandmonkey pls spread the word so everyone
37786102450491300	josedclara	2/16/2011	08:11:23Z	Egypt is now ruled by a gaggle of generals: Are Egyptians celeb
37786062977761200	NewpolicyonIran	2/16/2011	08:11:14Z	RT As Egypt uprising inspires Middle East, Iran sees biggest pro
37786059756544000	travel_hubs	2/16/2011	08:11:13Z	Latest #hubpages | What do you need to know before you visit
37786055402856400	danicoopers	2/16/2011	08:11:12Z	RT @Sandmonkey: Wow, Egypt responded quickly :) http://ww

Figure 2 – Sample statuses

Twitter activity it is a great example of how new innovations in telecommunications technologies can be utilized to offer a simple, yet powerful service. Twitter, more recently, has even implemented geolocation features, which is sure to be another area with which this applet can integrate in the future. As telecommunications technology improves, websites like Twitter are sure to continue to utilize all the new features it can, and this applet can evolve with it. Many of these innovations can be used to study social phenomena as the example in this paper shows.

III. USE CASE EXAMPLE

This application can be utilized by anyone who wants to find out what is happening right now. Current events in news media are going to result in tweets posted about these events. If someone is curious as to what others are sharing about what he or she just saw in the news, then it's just as simple as performing a search through the applet. This applet would provide a great interface to pull as much data as possible about the topic.

For example, we show the use of the application to gather information about the recent situation in the Middle East. Egypt is the center of attention, so doing a simple search will pull up to 1500 of the latest and/or popular statuses on the topic. Figure 2 shows an example of the tweets collected from the application.

Data from social media is very useful because it can reveal interesting patterns. For example, Figure 3 shows the distribution of countries identified in the search for Egypt as the data source. We can see that at this particular time there was considerable interest in Libya among the people posting status regarding Egypt. The user can then decide whether he or

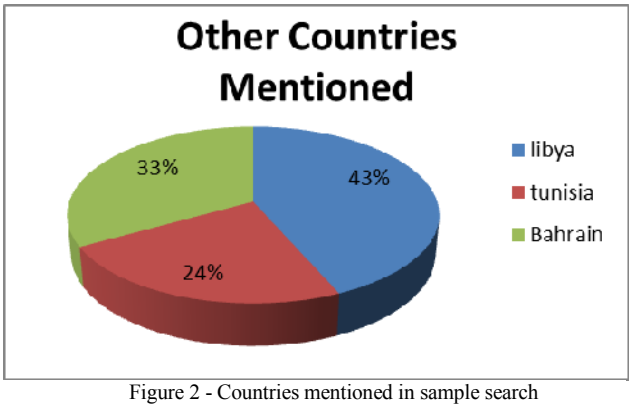

Figure 2 - Countries mentioned in sample search

ID	Username	Real Name	Total Twe	Location	Language	Friends	Followers
245937087	DuneSurf	Joseph	148	Earthly place w/sand & lizards	en	17	75
177983366	New_Pho	Alex	3445	USA	en	1662	1990
53053958	Kulwinde	Kulwinder Singh	823	India	en	200	413
247035482	yyrdtbbvv	Cardea Mccaffery	148	null	en	10	0
245320475	Power_of	Damien Hannah	66	Ontario, Canada	en	24	146
243528393	abcwayde	ABC DEF	47	null	en	1	2
252892800	northtexa	NorthTexasDREAM	4	Dallas, TX	en	4	2
20723642	worldfina	World Finance	111036	The Netherlands	en	10475	11528
243665696	shabmasr	shab masry	7	null	en	0	9
25869543	Amnamar	Amna Amer	8408	Cairo	en	1093	464
49472202	Sedaye13	Savalan	5169	Earth	en	375	111
59017015	JasenSok	Jasen Sokol	101	Berea, OH	en	97	209
91726993	wutever0	Fadi H. Rezq	2594	Planet Earth	en	103	168
240658920	JasonForc	Jason House	3296	null	en	121	0
17864199	11autos	11autos	3339	san francisco	en	66	137
17864345	22autos	22autos	3414	Amsterdam	en	104	137
19571016	magatopi	James Hubbs	53512	United States	en	2178	417
152628884	m_karem	Mahmoud Abu ElM	196	Egypt	en	46	106
109533060	Bieber_be	Bieber_Lover	1030	DownToEarth	en	1098	1202
68764611	Bebhbhy	Trissye	4	null	en	5	5
25527440	edmank	Edward W. Mank	1915	Augusta, Maine, USA	en	78	41
61989293	elsa7er	ana wa7ed	145	qatar	en	14	101
181010408	lifeofalaz	enuhski	2346	in your iDevice	en	96	86
17268481	annfinste	annfinster	21523	Phoenix, Arizona	en	1249	1999
223341335	Dareensa	Dareensawan	461	United States of America	en	60	163
108514029	SHEPFM	SHEP FM	12132	Madison, WI	en	304	500
49402901	revolucio	Green Revolution	1750	Tehran	en	113	100
134496245	SonarFX	Sonar FX	3862	null	en	891	1260
230617467	nourmous	Nour Mousa	149	null	en	17	39
16555398	MichaelTE	Michael Bishop	1282	Boise, Idaho / Global	en	502	280

Figure 4 – Sample user information

she would like to search further into Libya, or any other subject the user wishes.

Performing a search, specifically, will result in two files returned – information about the tweets, as shown above, and information about a limited number of users. Figure 4 displays a sample of the latter file regarding some of the users in the Egypt search.

Researchers in social media are very interested in patterns of followers. Users with a large number of followers are likely to be influential opinion-makers in the community. As an example, the graph in figure 5 shows the followers as a function of the number of posts users have made. As we can see, very few posters have greater than 2000 followers. But there is one user with about 11,500 followers. This particular user is World Finance. Since that user appeared in the search results, and has a large number of tweets, there is a considerable chance that the user might have some more information on the subject. If desired, the researcher can decide to pull the latest tweets from that user for more information. Since Twitter only maintains 3,200 tweets from a user, an online tool such as that proposed in this paper is helpful in collecting all the tweets from such prolific users.

Figure 6 shows some tweets made by the worldfinance user. Researchers may categorize these tweets to determine the factors that drive other users to follow this user.

By examining the chronological distribution of specified terms, researchers can identify how a subject matter of interest

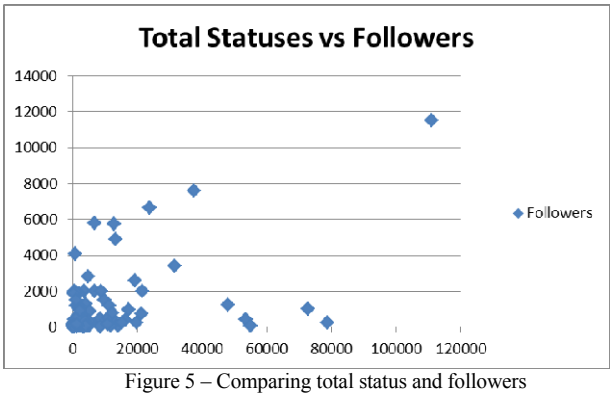

Figure 5 – Comparing total status and followers

worldfinance: World markets sink as protests escalate in Egypt - http://newzfor.me/?cuqh
worldfinance: Narrowing Your Focus While Broadening Your Horizons - How To Strik A Balance - h
worldfinance: :: Daily Digest for January 28th http://bit.ly/glADuT
worldfinance: Official grillings make property price controls a personal issue - http://newzfor.me
worldfinance: :: Furry Creatures Explain Bank Bailouts: 'The Screwing Of The American People' (V
worldfinance: :: Biden To Jobless: 'Hang In There' http://bit.ly/iddAJ8
worldfinance: :: New York City may have to lay off 15,000 teachers: mayor http://bit.ly/fFI8aE
worldfinance: :: U.S. warns Egypt on aid, urges security restraint http://bit.ly/fMrX06
worldfinance: :: Airlines change flight schedules amid Egypt unrest http://bit.ly/fgParf
worldfinance: Corrupt local officials one step ahead in illegal-land-sale game - http://newzfor.me
worldfinance: What is Your 1CWhy 1D? - http://newzfor.me/?cemh
worldfinance: :: Humans left Africa 65,000 years earlier: study http://bit.ly/gWA0E1
worldfinance: :: NASA remembers Challenger, vows to advance in space http://bit.ly/fpRHI3

Figure 6 – Sample user timeline statuses

has evolved over time. For example, figure 7 below illustrates the prominence of Egypt among all the latest posts from the user in Figure 6. The graph is tracking the number of times Egypt has been mentioned for each post before the current one. Since the trend is very linear, the researcher can conclude that this user has been tweeting regularly about this topic at a steady rate. Data collection can continue as long as the researcher desires. The application provides the tools to get the necessary information.

This program can be utilized by anyone who is looking for social opinion on any topic. It can be a valuable tool for any researcher of public events, academic or otherwise.

IV. Further Research

In the near future, we plan to make the tool available as a server application to allow interested people to monitor tweet

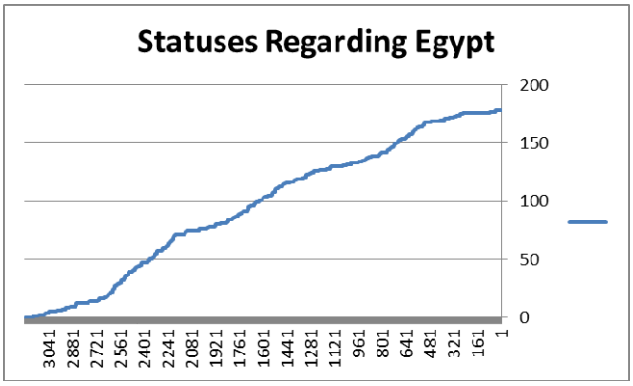

Figure 7 – Statuses regarding Egypt in World Finance's recent Twitter activity history

topics by supplying a search topic and leaving it to the application to continuously grab related tweets and write the information to the file system from where it can be retrieved periodically by the researcher. It is also planned to expand the applications to have more options, allowing each searcher to refine their results and avoid wasting precious API hits on irrelevant data. The challenge at the moment is handling the API rate limits. We are currently looking into adding authentication (through OAuth) to allow users to receive more results.