Using social media to study social phenomena
An example using Twitter data

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

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
Twitter activity 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.

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. For example, Figure 2 shows a sample 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 counties 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 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...
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 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.