

Social Visualization for Micro-Blogging Analysis

Tanyoung Kim, Hee Young Jeong, Yee Chieh Chew, Matthew Bonner, John Stasko

Georgia Institute of Technology

ABSTRACT

We present a social visualization system of the micro-blogging site 'Twitter'. Named 'Tweesters', our system allows users to explore the social trends of multiple keywords/phases. Tweesters is composed of a multi-view interactive visualization, which features a temporal visualization, a GeoVisualization, co-occurring words and selected tweets. We tested with a dataset of over 600,000 status updates, which suggests a potential for it being used as an analysis tool for marketing, collaborative sensemaking, and understanding social characteristics.

KEYWORDS: Social Visualization, Twitter, Micro-blogging, Social Trend Analysis

INDEX TERMS: H.5.1 [Information Interfaces and Presentation]: Miscellaneous

1 INTRODUCTION

In the past few years, micro-blogging sites have experienced a large gain in popularity [8]. Along this trend, visualizing the various forms of communication happening there has become one of the central topics in the Social Visualization scene. Among those sites, we chose 'Twitter' as the source of data [9] for our visualization system because Twitter has great potential as a meaningful resources in analyzing online social trends as well as individual life style choices as suggested in [1] and [2].

On Twitter, users can leave a message of up to 140 letters, which is called a 'tweet', 'status' or 'update'. They follow other users and check their updates on the home page of Twitter. They also directly reply to their followers or re-tweet their followers' updates, working essentially as a multiple way messaging system. In addition, users can set the properties such as location, one line bios and a profile picture, which are sources of further analysis.

In this poster, we propose a new visualization system named 'Tweesters'. Although the exposed data on Twitter itself are unorganized text-based short messages, we utilize meta information such as time and location visualizing in interactive multiple views. Therefore Tweesters does not only repackage existing content in different graphical forms, but also functions as an analysis tool of social trends.

2 RELATED WORKS

There are many research projects as well as graphic art works [4] that reuse, filter and visualize in different graphical manners than the original data from Twitter: *Twist* provides the trends of users' entered queries with a timeline graph and the tweets

containing the queries [6]. *Twistori* filters and refashions the text of tweets based on emotional and psychological words such as 'love' and 'believe' [5]. *The word on the Tweet* visualizes the distribution of a searched word over the US map [10]. One website called *Graffiter* filters the tweets by only one individual and help track his/her personal lifestyle [7]. In sum, these applications of Twitter visualization/analysis filter data based on one or two properties of the data: pre-defined words (*Twistori*), user-entered words and time (*Twist*), location (*The words of Tweet*) or a specific user and themes (*Graffiter*).

3 DESIGN OF TWEESTERS

3.1 Goal

In contrast to those re-visualization and limited analysis tools, our program aims at featuring multiple properties (time, location and tweets) and visualizing them in one screen (Figure 1). Thus, it provides clues about the popularity of specific terms that users want to investigate .

- 1) timeline integrated with the duration selector
- 2) geo-visualization of US map
- 3) other most frequently appearing terms along the keyword
- 4) real 'tweets' randomly selected.

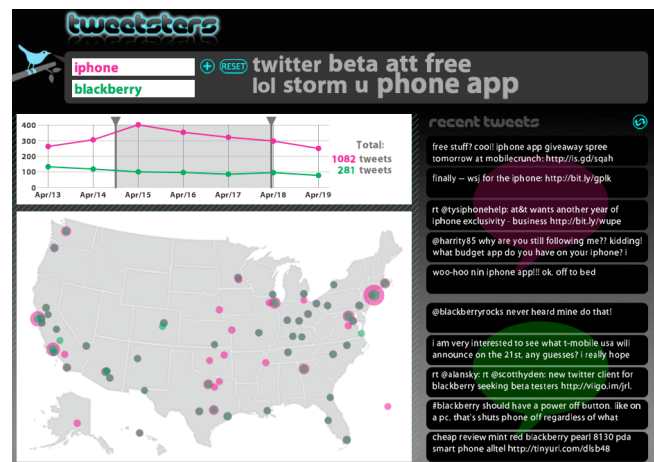


Figure 1. Screenshot of Tweetsters; Comparison Mode

3.2 Visualization Elements and Interaction

In this section, we articulate what each visual element is for, how it interacts with others and how users can explore the system on the prototype.

3.2.1 Input Box

Tweesters is activated by entering a term or phrase that users want to search. A user may enter one term to examine or two terms to compare.

Entering a keyword calls the search engine to find the tweets which contain the keyword. It simultaneously changes all other elements - Trend View, GeoVisualization, Text Clouds, and

Email: {tanykim | hyj | ychew3 | matt.bonner }@gatech.edu, stasko@cc.gatech.edu

Tweets View - to represent the new subset of tweets. When comparing multiple keywords, the results are visually distinguished by color.

3.2.2 Trend View

This view shows the number of tweets over time by day, which helps users understand the overall visibility and popularity of the examined keywords. The Trend View can also be useful for explaining phenomena as well as discovering them.

By adjusting two sliders at the starting and ending dates, a user is able to adjust the duration of tweets. This filtering also affects the GeoVisualization and Tweets View.

3.2.3 GeoVisualization

This view allows a user to see where a given search term is appearing most frequently. Using the location information of each tweet, we create a map view of the continental US. In this view, the top 70 most populous US cities were used as location markers in order to simplify our prototype's dataset. The frequency of tweets in a location is denoted by the size of a circle over the city. The size corresponds to the number of overall tweets on the city relative to the other locations.

When a user moves their mouse over the circle, a display bubble will also appear, showing the number of relevant tweets in that location. Similarly on Trend View, filtering based on city is also possible; a user's city selection prompts the program to narrow down the results within only that city.

3.2.4 Text Clouds

Text Clouds help the user discover what is discussed along with their search term. This can help explain phenomena, offer interesting contextual information or suggest a new search. Text Clouds consists of 5-10 words. Similar to 'tag clouds' [3], these words will vary in size according to the frequency that they co-occur with the keywords.

Clicking on a word in the Text Cloud starts a new search, replacing the keyword in Input Box with it and triggering immediate changes in other elements to reflect the new dataset.

3.2.5 Tweets View

Tweets View provides additional context for keywords. In this view, up to five random tweets that contain each keyword appear. This can prove invaluable to understanding trends. When comparing multiple keywords, the five random tweets of each keyword are visually divided with the same two color sets used in the other elements. Random tweets are regenerated by clicking the refresh button.

4 IMPLEMENTATION AND USAGE EXAMPLES

We developed our prototype using Processing, a Java-based open-source programming environment for animations and interactions. Our initial dataset was gathered by connecting to Twitter's data-mining feed for one week (April 13 – 18, 2009), resulting in a ~6,000,000 raw tweet dataset. Among this dataset, we eliminated about 50% of tweets which do not have location information, and others in ambiguous locations or written in a language other than English. After filtering the tweets with incorrect location information, the final size of the dataset was ~600,000 tweets with which we tested our prototype. Here we briefly enumerate several usage examples we found with the dataset: temporal and regional analysis, and information from Text Clouds.

The keyword 'tax' was most popular on April 15th, a traditional due-date for US Taxes: among a total of 956 tweets containing the word 'tax' during the week, 697 occurred on April 15th.

Tweet terms can often be sharply segregated by city. For example, the term 'Boston Celtics' was prevalent in Boston and also in Chicago, the Celtic's playoff opponents at that time. Keywords about weather are appropriate to show regional differences as well as temporal trends. 'Tornado' was most popular in the city of Nashville (6 out of 9), especially on April 13th and 14th.

A keyword 'iPhone' leads to 'twitter' on Text Clouds while 'Blackberry' does not. A user may hypothesize that people are interested in using iPhones rather than Blackberries to access Twitter.

5 DISCUSSION & FUTURE WORKS

Evolving from the simple examples above, Tweetsters has the potential to be a social trend analysis tool. For instance, it can be used for tracking the popularity of specific products or brands, sense-making during crisis, and understanding social characteristics of cultural and political issues.

Tweetsters, while an effective prototype, would benefit from additional metadata. For example, semantic analysis to distinguish between positive and negative tweets would be especially valuable to marketers and politicians. Also integrating with external data sources may provide more clues to explain a search term's popularity in one city or a spike in the Trend View's timeline.

We have so far utilized time, location and actual tweets. For future enhancement, we may visualize the connectivity of users and including 'reply' status among users. We expect it would offer a new social dimension to explore.

6 CONCLUSION

We presented a social visualization system 'Tweetsters'. It represents unstructured data of micro messages into organized information. Through its dual filtering of time and location and the additional text, it facilitates obtaining insights into social trends happening on micro-blogging sites, which are otherwise scattered and inaccessible.. In conclusion, Tweetsters may be a supporting tool for 1) marketers who desire to track the popularity of their products or services according to time and location, 2) healthcare professionals who need to monitor ongoing crisis such as flu spread, and 3) people who are curious about the trends of controversial issues.

REFERENCES

- [1] D. Giustini and M. Wright Twitter: an introduction to microblogging for health librarians. *Journal of the Canadian Health Libraries Association* Vol. 30, pp. 11–17, 2009.
- [2] A. Java, et al. Why we twitter: Understanding microblogging usage and communities. In *Proc. of the 9th WebKDD and 1st SNA- KDD 2007 workshop on Web mining and social network analysis*, pp. 56–65, 2007.
- [3] F. Viegas and M. Wattenberg, Tag Clouds and the Case for Vernacular Visualization. *Interactions* 15, 4, pp. 49–52, 2008.
- [4] N. Yau, 17 Ways to Visualize the Twitter Universe. Accessed at <http://flowingdata.com/2008/03/12/17-ways-to-visualize-the-twitter-universe/>
- [5] <http://twistori.com>
- [6] <http://twist.flaptor.com>
- [7] <http://www.grafitter.com/>
- [8] <http://www.guardian.co.uk/technology/2008/dec/18/internet-website>
- [9] <http://www.twitter.com>
- [10] <http://www.uuworld.com/wordontweet>