
Tweets of the Nation: Tool for visualizing and analyzing global tweets

Thatchaphon Klomklao

Department of Computer
Engineering
Faculty of Engineering
Chiang Mai University, Thailand
Thatchaphon_k@cmu.ac.th

Panat Ratanarungrong

Department of Computer
Engineering
Faculty of Engineering
Chiang Mai University, Thailand
panat_r@cmu.ac.th

Santi Phithakkitnukoon*

Excellence Center in
Infrastructure Technology and
Transportation Engineering
(ExCITE) and
Department of Computer
Engineering
Faculty of Engineering
Chiang Mai University, Thailand
santi@eng.cmu.ac.th

*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
UbiComp/ISWC '16 Adjunct, September 12–16, 2016, Heidelberg, Germany.
Copyright 2016 © ACM 978-1-4503-3575-1/15/09...\$15.00.
<http://dx.doi.org/10.1145/2800835.2800975>.

Abstract

In this work, we developed a visualization tool that allows the user to observe the popular tweet hashtags of any selected country around the globe, and on the other hand, the user can also search for any particular tweet hashtag to observe its geographic distribution worldwide. In addition, the tool allows the user to export the result for further analysis. The results are based on geotagged tweets of the previous day, via Twitter API. The tool can be useful in many domains, for example, it can be used as a facilitative component of a recommendation system, or used for exploratory analysis of the information propagation – i.e., how a particular topic of social interest spreads over international boundaries. It can also be another channel of news or catching up with talk of the town, and it can also be a tool for social science research and related fields.

Author Keywords

Twitter, hashtag, tweets, data visualization

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

1 Introduction

Twitter is a today's popular microblogging and online social networking space. It allows the user to share online a short message of 140 characters (aka. Tweet) and read other messages of following people on newsfeed. Opinions, feeling, complaints for something they encounter on daily basis are what people like to share or tweet about.

In addition to a regular tweet, twitter users use the hashtag symbol (#) in front of a relevant keyword or phrase (no space) to categorize those tweets, which make it easy for twitter search. So, hashtag is a keyword or common topic with a specific content such as #ubicomp2016, #purba2016, or #NBAPlayoffs. The user can search for a particular hashtag on Twitter, and a list of all tweets with the searched hashtag will be displayed.

Another function of twitter, which seems to be useful and interesting is geotagging. The user can attach geographic information (geo-location) to the tweet. This adds spatial context to the tweet, which already contains content and temporal information. By utilizing all spatial, temporal, and content elements of the tweets, we believe that these information, statistically and graphically, as a tool, can be very useful for researchers, especially those in non-technical disciplines such as social science and humanity. They can use this tool as a primary observation for emerging trends and patterns, which can help advance their researches and studies.

Hot topics are something of interest to large number of people. Knowing what are the hot topics of countries around the world can be useful for behavioral study,

information management, marketing, and so on. So, in this work, we developed a visualization tool that allows the user to observe top popular tweet hashtags of any selected country worldwide, and geographic distribution of a hashtag of interest. The user may observe the trends in different countries and see how the hot topics spread geographically around the globe, which may lead to understanding of information propagation.

Due to different penetration rates of twitter, the number of tweets varies across different countries. Some countries may not have as much data as other countries for observation. Nonetheless, twitter is the most popular micro-blogging service in the world with users across the globe.

2 Related work

There were a number of data visualization projects that have inspired our work, such as *Superbowl Twitter Map*, *Mobility Profiles for Livelihood Characterisation*, and *Tweet of the Town*,

Superbowl Twitter Map [1] is a visualization of geotagged tweets about Superbowl 2014 between Denver Broncos and Seattle Seahawks, showing where people talking about the playing teams and the halftime show worldwide during the game. *Mobility Profiles for Livelihood Characterisation* [2] is an interactive visualization that displays the migration of people between cities in Senegal, which characterizes the mobility in the context of livelihood zones and seasonal activity patterns in Senegal. *Tweet of the Town* [3] is an interactive visualization tool that allows users to search and observe a geographic distribution of a tweet hashtag. The tool was demonstrated using geotagged tweets in Thailand as a case study.

In contrast with those mentioned projects, our developed tool is an interactive visualization that allows the user to select a country of interest and displays the most popular tweet hashtags of that country. Our tool also allows users to search on a particular hashtag to observe its worldwide geographic distribution, and allows the user to export the result for further analysis.

3 Tweets-of-the-Nation System

Tweets-of-the-Nation system consists of three parts: data acquisition, data visualization, and data export. The data acquisition part was developed using Python¹. The data visualization and data export were developed using Processing Development Environment (PDE)². Processing is a programming language and environment built on Java, suitable for data visualization and interactive applications.

3.1 Data Acquisition

The Twitter API³ was utilized to acquire the raw data from Twitter. The data is kept in our database as a CSV file. Twitter allows its data be harvested only for the previous seven days (from the date of the query submission). Our system acquires the data only for the previous one day, so that the result is the very most current popular topics.

There are a number of available libraries for connecting and accessing the twitter API. Our system uses tweepy⁴, which is based on Python language, and it

¹ <https://www.python.org>

² <https://processing.org>

³ <https://dev.twitter.com/rest/public>

⁴ <http://www.tweepy.org>

also allows our system to filter the tweets based on the country name, which is very suitable for our system functionality. An example of part of tweepy query code with "Germany" being the value of the country name query is following:

```
places = api.geo_search(query= "Germany",
                        granularity="country")
```

An example of the obtained tweet data (the first few tweets) by using "Germany" as the specified country name is shown in Fig. 1.

Hashtags	Screen name	Coordinates[0]	Coordinates[1]	Created at
germany	Aly Tadros	52.26805422	13.1199934	2209
nowplaying	BB RADIO Playlist	52.381905	13.1199934	2208
heidelberg	princess k-la	49.40889	8.70724	2208
regensburg	Jesus Flores	49.02107166	12.09824885	2206
Endomondo	Ronny Neumann	53.5897953	10.1435998	2206
coldrain	CHOCOBRO #2	51.49637	7.001179	2205
doublefeature	LSDchen_RT	48.48333333	9.21666667	2205
nowplaying	Radio TEDDY Playlist	51.3658954	7.6523215	2204
MVG	Ben	48.15363	11.55311	2203
gettingready	Samira Barnard	47.7333333	10.31666667	2202
Abendspaziergang	Cinderella	49.4833	10.7167	2201

Figure 1: An example of tweet data obtained from twitter API using tweepy library.

Each returned data for a given country name contains tweet hashtags, user ID, user geo-location (latitude, longitude), and timestamp. To get the most popular tweet hashtags of each country, our system then ranks the hashtags according to their counts, which happens in the data visualization process.

3.2 Data Visualization

In our data visualization process, the system loads the obtained tweet data from twitter API, then the data is processed and displayed accordingly. This stage is developed using PDE and so it includes the user

interface and graphic display. A snapshot of the main initial screen (home screen) is shown in Fig. 2.

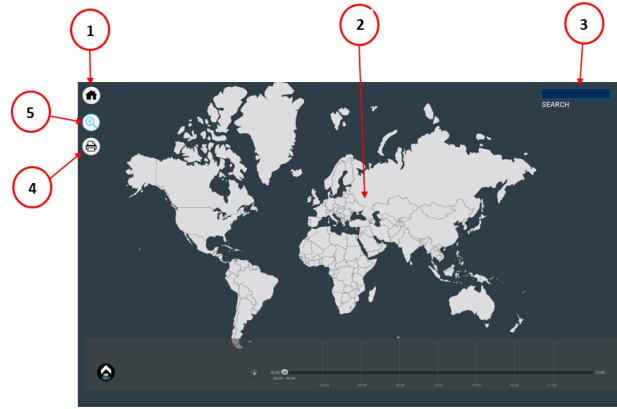


Figure 2: A snapshot of the home screen.

3.2.1 Home screen

A snapshot of the home screen shown in Fig. 2, here are the descriptions of the user interface:

- (1) *Home icon*: this icon brings the user back to the home screen (Fig. 2).
- (2) *World map space*: the user can choose any country by clicking on the country area.
- (3) *Search bar*: another way for choosing the country is by typing in the country name from the keyboard and press Enter.
- (4) *Export icon*: the user can export the result as a PDF file (explained in Section 3.3)
- (5) *Search icon*: by clicking this icon, the user can enter a searched hashtag using the keyboard, and the result of the geographic distribution of tweet will be shown on the map (explained in Section 3.2.3)

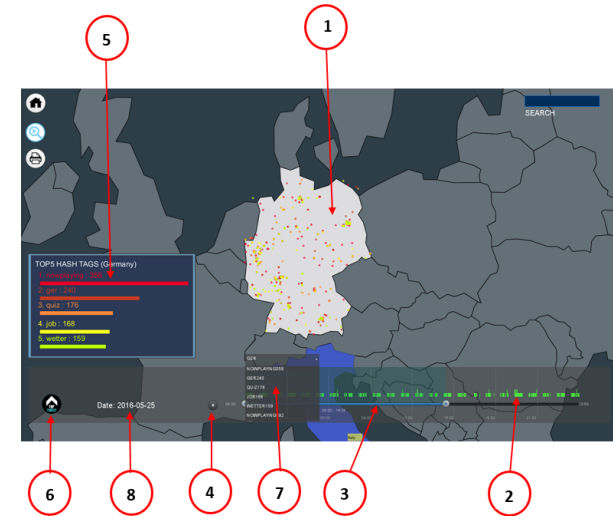


Figure 3: A snapshot of our system when a country is selected and the top tweet information is displayed.

3.2.2 Graphics

By selecting the country, the top tweets of that country is shown both statistically (top 5) as well as geographic locations of those top 5 tweet hashtags in that country. Here are descriptions of the graphics displayed in our system (Fig. 3):

- (1) *Geo-tagged tweets*: geo-tagged tweets are displayed in the geographic location of the twitter (i.e., the user).
- (2) *Frequency graph*: horizontal bar chart displaying the counts of tweet over the period of time (24 hours).
- (3) *Adjustable observation time window*: the user can adjust time window for the displayed tweet result by moving the scroll bar limit of time period. For example, if the time window is between 9pm and

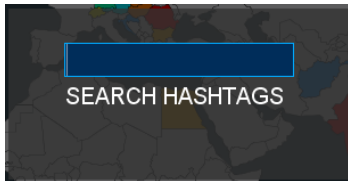


Figure 5: The search box

- 10pm, the result will be based on the tweets generated during this time period only.
- (4) *Play button*: the user can start or pause the animation by clicking this button. When playing, the result is shown and changing over time (24 hours).
 - (5) *Top tweet dashboard*: a list of most popular tweet hashtags based on their counts is shown along with their respectively counts during the selected period of observation time window.
 - (6) *Top tweet display button*: the user can choose to show or hide the top tweet dashboard by clicking this button.
 - (7) *Hashtag display selection*: the user can choose to display the result of any particular hashtag of interest from the list (scrolling menu).
 - (8) *Date*: the system displays the dates of coverage i.e., previous one day.

3.2.3 Global geographic distribution of tweet

In addition to the top tweet feature of the system, the visualization tool also offers another feature that allows the user to observe the geographic distribution of a searched tweet hashtag. A snapshot of this feature is shown in Fig. 4. Here are descriptions of the graphics displayed in this feature:

- (1) *Search button*: a search box (Fig. 5) will appear when the user clicks on this search button. The user can enter any hashtag keyword to be searched in this search box.
- (2) *Hashtag bar*: it displays the searched hashtag.
- (3) *Color bar*: it shows the popularity level (ranking order) for the searched hashtag.

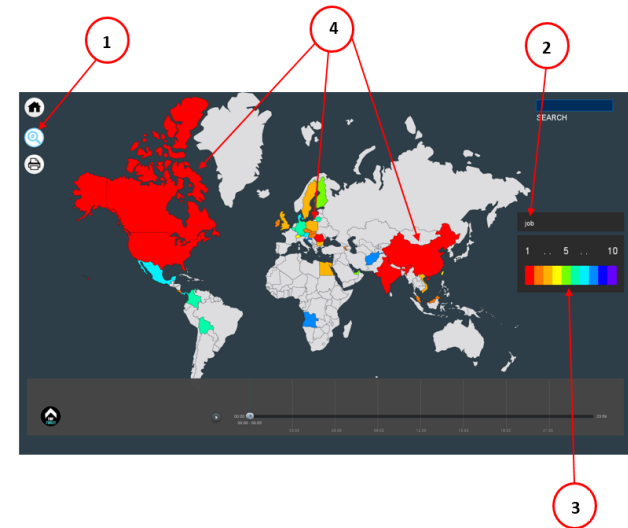


Figure 4: A snapshot of the result of geographic distribution of a searched tweet hashtag.

- (4) *Geographic distribution*: countries are highlighted with different colors according to the ranking order (1st to 10th rank) of the searched hashtag. The ranking is based on the entire data of the whole previous day. .

3.3 Data Export

Some users may be interested in using the result displayed in our system for their further studies. Therefore, our tool allows the user to export the resulting information as a PDF image file, as shown in Fig. 6 (as an example). A report contains information such as top tweet hashtag counts, selected country, time window, date, and geographic distribution of tweets in the selected country.



Figure 7: Save report box

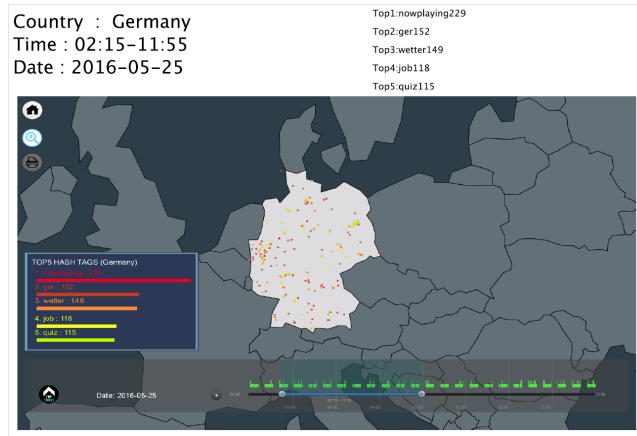


Figure 6: An example of an exported PDF image file.

By clicking the Export icon, the tool will prompt the user to enter a file name for the exported image. A prompting box for exported image file name is shown in Fig. 7. The system will save this image file to a folder named "report" in the same directory where the system is running.

3.4 Demo

For demonstration of how our developed visualization tool works, a video clip showing all features of our tool is available at <https://www.youtube.com/watch?v=2DC4MjHpulc>

4. User Experience

It is important to put our tool to the test with the real users. We wanted to get the feedbacks from the actual users to improve our system. So, we conducted a user experience study, based on the theory of four elements of user experience [4]. We randomly recruited 50 subjects, each was asked to use our tool and

afterwards answer a 1-to-7 Disagree-Agree response scale questionnaire with four statements: It is useful; It is easy to use; It is easy to start using; It is fun and engaging. All 50 subjects were in a mix of different genders, ages, and occupations. Our questionnaire is shown in Fig. 8.

4.1 Overall result

The overall result is shown in Fig. 9. The highest rating is "It is easy to use" with the average scale of 5.40. The second highest rated elements are "It is easy to start using" and "It is useful" with the same average scale of 5.15. The lowest rating is "It is fun and engaging" with the average scale of 4.75.

Figure 8: User experience questionnaire

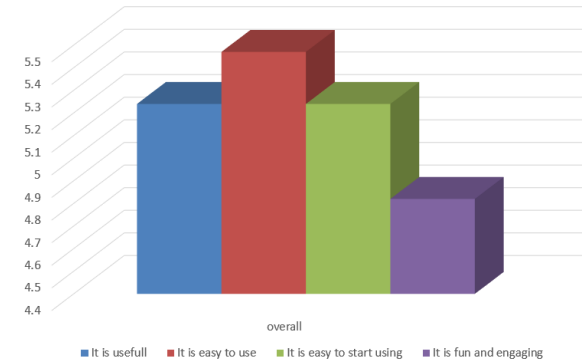


Figure 9: Average agreement scales from 50 subjects

From the rating, it suggests that the tool is easy to use. Some subjects felt that the tool is not useful enough. For example, a comment from one of the subjects was "It is useful but for some specific users, but not for all". The overall result and the subjects' comments suggest

that our current system needs an improvement on its usefulness and fun-and-engaging elements.

4.2 User genders

If subjects are separated by gender; 28 males and 22 females, the result is shown in Fig. 10.

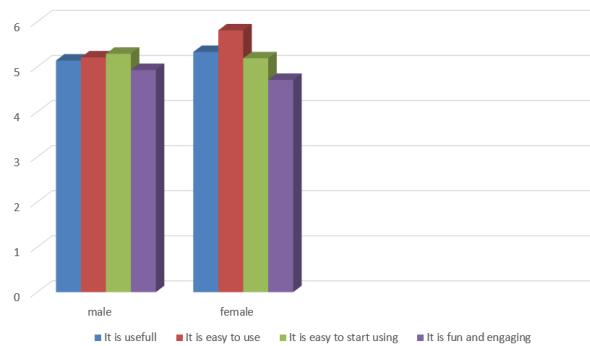


Figure 10: Average agreement scales separated by gender

For male subjects, the highest rating is “It is easy to start using” with the average scale of 5.29. The runner-up is “It is easy to use” with the average scale of 5.21. The third rank is “It is useful” with the average scale of 5.14 and the lowest rating is “It is fun and engaging” with the average scale of 4.93. Male subjects have given all four elements with similar rating. The fun-and-engaging element is also the element to be improved in our visualization tool.

On the other hand, the highest rating of female subjects is “It is easy to use” with the average scale of 5.81. The second rank is “It is useful” following by “It is easy to start using” with the average rating of 5.33 and

5.19, respectively. The lowest average scale is 4.71 for “It is fun and engaging”. The female subjects mostly find that our tool was easy to use, while the fun-and-engaging element was least impressed and needed to improve. One female subject’s comment concerning fun-and-engaging element is “Looks interesting but it needs more features for more user interactability”. Moreover, there also needs to be an improvement on artistic and aesthetic aspects of our tool, as one of the female subjects’ comments is “It needs more beautiful and attractive graphics”.

4.3 User ages

Our subjects were in five different age groups; 13 subjects were 11-20 years old, 28 subjects were 21-30 years old, four subjects were 31-40 years old, one subject was 41-50 years old, and four subjects were more than 50 years old. The result is shown in Fig. 11.

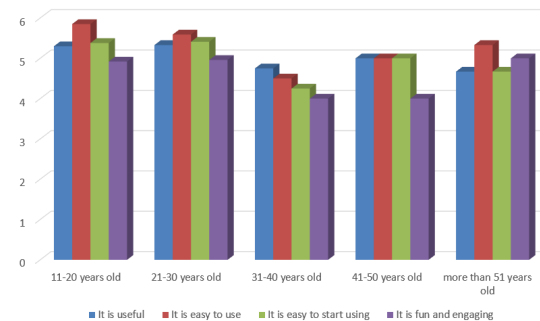


Figure 11: Average agreement scales separated by age range

The highest rated element in most age group (except for the 31-40 age group) is “It is easy to use”. This apparently is a standing point of our tool. That’s

probably because our tool does not require any complicated user interaction or input, but simple and common user interface with which general ICT users are familiar.

The subjects in the age groups of 11-20 years old and 21-30 years old had highly rated "It is useful".. This seems to suggest that our tool is more attractive (in terms of usefulness) to young users, who are more active users of ITC and especially social media in their everyday life than other age groups. As they are highly connected to digital technology and social media, this age group users therefore perhaps see more potential applications of our tool and so more appreciate its usefulness.

4.4 User occupations

There were a mix of occupations of our subjects. Most of our subjects were students (35 subjects), and there were two business owners, two government officers, two company employees, one academic, one self-employed, one university staff, one retiree, and five others. The result grouped by occupation is shown in Fig. 12.

For the student users, the highest average scale is "It is easy to use" and the lowest is "It is fun and engaging". This is along the same lines with the result separated by age group. As the student users are relatively younger than subjects of other occupations, so they are more familiar with the ICT and social media and hence find our tool easy to use. However, since the tool may not have many exciting features, the student users may find it not very fun and engaging, which is among other elements that needs be improved in our future development. This may also has to do with the fact that

since student users are actively using ICT and social media and therefore they have seen more sophisticated, interesting, and advanced professional applications on the market, and so don't find quite find our tool so fun and engaging, as evidenced from some student users' comments; "Font is not very attractive and too difficult to read", "It needs exciting and stylish graphics", and "Good tool but there are still plenty of rooms for improvement, especially the artistic aspect".

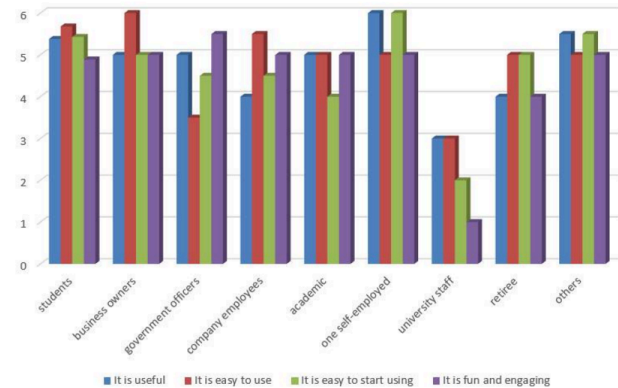


Figure 12: Average agreement scales separated by occupation

For other occupations, there was a small number of samples to draw conclusive observations. Nonetheless, there was a useful comment from a female university staff. She thought that our system was useful, but for specific users for example, researchers who are interested in social media data analysis. She did not see the benefit from using the tool for her or some other people. We find that this is a useful comment, so in the future, we will take this comment among others into consideration when improving our tool..

5 Conclusion

In this work, we developed an interactive visualization tool that allows the users to observe the popular tweet hashtags of any selected country in the world, and the global geographic distribution of a searched top tweet hashtag. The user can also export the displayed result as an image file for further analysis.

To evaluate our tool, we conducted a user experience study, using questionnaire with a group of randomly selected 50 sample users categorized by gender, age, and occupation. Generally, the most users think that our system is easy to use. Some of them suggested that the system should be more useful and engaging. Based on the results of our user experience study, the tool needs to improve on fun-and-engaging element and artistic aspect. The sample users also commented that our visualization tool is useful for some specific users for example, the researchers, but no benefit of use for some other occupations.

In the future, we will improve on the graphics displayed to be more stylish and eye-catching, so it will appear more attractive, and include more exciting features to allow the user to be more engaged and fun to use. Nonetheless, we believe our current developed tool is useful to some extent for the users who are interested in social network analysis as it serves as assistive exploratory analysis of the social media data that can lead to better understanding of social behavior and potentially forming new interesting research questions.

References

1. CartoDB. 2014. Superbowl Twitter Map. Retrieved Sept 10, 2015 from: http://srogers.cartodb.com/viz/1b9b0670-8d15-11e3-8ddf-0edd25b1ac90/embed_map.
2. P. J. Zufiria, D. Pastor-Escuredo, L. Ubada-Medina, M. A. Hernandez-medina, I. Barriales-Valbuena, A. J. Morales, J. Quinn, P. Hidalgo-Sanchis, M. L. Oroz, W. Nkwambi. 2015. Mobility Profiles for Livelihood Characterisation. Retrieved Sept 10, 2015 from: <http://pulselabkampala.ug/d4d-senegal/>
3. K. Piriyawarangkul, K. Rodphotong, S. Phithakkitnukoon. 2015. Tweet of the town: case study of Thailand. In *Adjunct Proceedings of the 2015 ACM Conference on Pervasive and Ubiquitous Computing*, Osaka, Japan. Pages 1117-1121.
4. F. Guo. 2012. More Than Usability: The Four Elements of User Experience. Retrieved April 25, 2016 from <http://www.uxmatters.com/mt/archives/2012/04/more-than-usability-the-four-elements-of-user-experience-part-i.php>