

A Tweet Consumers' Look At Twitter Trends

Thomas Steiner¹, Arnaud Brousseau^{1,2*}, and Raphaël Troncy²

¹ Google Germany GmbH, ABC-Str. 19, 20354 Hamburg, Germany
`{tomac,arnaudb}@google.com`

² EURECOM, Sophia Antipolis, France
`raphael.troncy@eurecom.fr`

1 Introduction

Twitter Trends allows for a global or local view on “what’s happening in my world right now” from a tweet producers’ point of view. In this paper, we show the possibility to complete Twitter Trends by having a closer look at the other side: the tweet consumers’ point of view. While Twitter Trends works by analyzing the frequency of terms and their velocity of appearance in tweets being written, our approach is based on the popularity of extracted named entities in tweets being read.

2 Twitter Swarm NLP Extension

We have developed a Google Chrome extension called Twitter Swarm NLP³ that injects JavaScript code into the Twitter.com homepage. The extension first checks if the user is logged in to Twitter.com, and if so, retrieves the tweets of the current user’s timeline, search result page, or profile page on a one-by-one basis, and performs Named Entity Extraction (NEE) via Natural Language Processing (NLP) using a remote NLP Web service⁴ on each of the tweets. The extracted entities are then displayed below each tweet, as can be seen in Figure 1(a), and finally are sent to Google Analytics to compute trends by pivoting the reported named entities by Google Analytics data, like users’ geographic locations.

3 Evaluation

We examined the period from February 24 to March 11, 2011. The extension reached *1,009 pageviews* as reported by Analytics, and had *35 all-time users* and *28 seven-day active users*. All in all, the extension has detected *1,533 unique different named entities* in total.

Using Google Analytics, named entities can be easily tracked over time. As an example, Figure 1(a) shows the occurrences over time for the named entity

* The author is a graduate student at EURECOM, Sophia Antipolis (France), and currently works as an intern at Google Germany GmbH.

³ <https://chrome.google.com/webstore/detail/dpbphenfakflfmdlanimlemacankjol>

⁴ http://tomayac.no.de/entity-extraction/combined/{text_to_be_analyzed}

“iPad”. Albeit the numbers are not statistically significant, the peak of interest is on March 2, the day where the iPad 2 was revealed, and the second highest peak is on March 1, the day before the announcement, where media was atwitter with expectation of the device. Another example is given in Figure 1(b), which shows the occurrences of the named entity “tsunami”. Japan was hit by an earthquake followed by a tsunami on March 11, exactly where the peak is on the graph. Hence, the occurrence graphs indeed correspond to what we would expect.

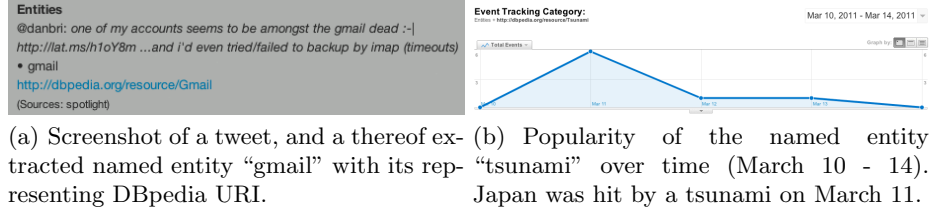


Fig. 1. Twitter Swarm NLP sample output and popularity of a named entity over time.

Japan was hit by an earthquake followed by a tsunami on March 11. As one would expect, this was reflected in the most read-about named entities for the period March 11 to March 14. Table ?? shows the occurrences distribution pivoted by country. Our data set is too small to be statistically significant, however, the potential for this data to reveal new insights is promising. Given enough data, we could, e.g., provide an answer to the question whether among Twitter users the tsunami caused more interest in the American, or the European continent. As we use URIs as named entity identifiers, there are no ambiguity, and no language barriers.

Entity	Total	Germany	Finland	United States	Chile	India	Netherlands	Italy
dbp:Tsunami	8	3	2	1	1	0	1	0

Table 1. The top named entity “tsunami” (Mar. 10 - 14) pivoted by country.

4 Conclusion

In this paper, rather than measuring the “trendiness” of terms/hashtags in tweets being produced, we measure the “trendiness” of named entities in tweets being consumed using Google Analytics. This allows for even richer insights into, e.g., the location of users interested in a certain trend, sliceable back to any point (or period of time) in history where there is data available.