

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 explore a way 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 browser 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, and performs named entity extraction via Natural Language Processing (NLP) using a remote NLP Web service⁴ on each of the tweets. The extracted named entities are then displayed below each tweet, as can be seen in Figure 1(a). Finally the extracted named entities are sent to Google Analytics to compute trends by pivoting named entities by Analytics data, like users’ geographic locations.

3 Evaluation – Raw Statistics

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 entity occurrences can be easily tracked over time. As an example, Figure 1(b) shows the occurrences of 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}

“tsunami”. Japan was hit by an earthquake followed by a tsunami on March 11, exactly where the peak is on the graph. In general the occurrence graphs also for other examples indeed correspond to what we would expect, albeit the data is not statistically significant. One more example is “iPad” with a peak on March 2, the release date of the iPad 2.

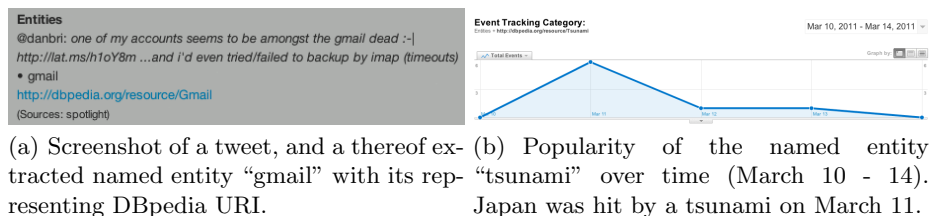


Fig. 1. Twitter Swarm NLP sample output and popularity of a named entity over time.

4 Evaluation – Pivoting Named Entities By Country

Japan was hit by an earthquake followed by a tsunami on March 11. Table 1 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, for example, 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 is no ambiguity, and no language barriers, as we compare named entities on the level of the representing URIs, which are language-independent.

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.

5 Conclusion

In this paper, we measured the “trendiness” of named entities in tweets being consumed using Google Analytics via an unobtrusive browser plug-in. This allows for even richer insights into, for example, the location of users interested in a certain named entity, sliceable back to any point (or period of time) in history where there is data available. This could be very attractive for market research.