TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief

Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu
School of Computing, Informatics, and Decision Systems Engineering,
Arizona State University,
{shamanth.kumar, geoffrey.barbier, ali.abbasi, huan.liu}@asu.edu

Abstract

Social media is becoming popular as a key source of information on disasters and crisis situations. Humanitarian Aid and Disaster Relief (HADR) respondents can gain valuable insights and situational awareness by monitoring social media-based feeds. Specifically, the use of microblogs (i.e., Twitter) has been shown to provide new information not otherwise attainable. In this paper, we present a new application designed to help HADR relief organizations to track, analyze, and monitor tweets. The purpose of this tool is to help these first responders gain situational awareness immediately after a disaster or crisis. The tool is capable of monitoring and analyzing location and keyword specific Tweets with near-real-time trending, data reduction, historical review, and integrated data mining tools. In this paper, we discuss the utility of this tool through a case study on tweets related to the Cholera crisis in Haiti.

Introduction

Micro-blogs provide a relatively new means of communication used by millions around the world. Twitter[1], a popular micro-blog service reports its users generate over 95 million micro-blogs, more commonly known as "Tweets," per day. Users can broadcast just about anything using the 140 character space available. Professionals have studied Tweets to gain insights on everything from the Superbowl (Bloch and Carte 2011) to crisis events (Goolsby 2010; O'Reilly and Milstein 2009; Shklovski, Palen, and Sutton 2008).

While current applications and mashups provide some useful tools and visualizations of information, no applications exist that optimize microblog tracking and analysis to support Humanitarian Aid and Disaster Relief (HADR) operations. These operations require real time monitoring, in most cases within moments of a disaster, to get situational awareness. Geographically identified tweets from the disaster location can assist with HADR planning and coordination efforts. TweetTracker is designed to support both functions with integrated tools to analyze data from a combination of temporal, geo-spatial, and topical perspectives.

[1]http://twitter.com/

Related Work

Web sites such as TrendsMap[2], Twitalyzer[3], and Geotwitterous[4] provide interesting mashup applications motivated by a variety of goals to visualize and analyze tweets.

Ushahidi[5] is a crowdsourcing application that has been used for a variety of event map applications including crisis maps for disasters like the 2010 Haiti earthquake (Goolsby 2010). Ushahidi's swift river application provides "filtering and verification of real time data for channels like Twitter, SMS, e-mail, and RSS feeds."

The University of Colorado at Boulder's Project EPIC[6] is working to develop methodologies and technologies to improve "computer mediated communication" during a crisis. This project includes research efforts to standardize tweets to enable more efficient machine processing as well as mapping tweeted reports about disasters.

TweetTracker

The application is designed to be intuitive and focuses on features useful to HADR relief organizations for monitoring and analyzing pertinent tweets from different perspectives. In this section, we briefly describe different system components of the tool and then present a case study using tweets that discuss the Cholera outbreak in Haiti.

System Architecture

The tool consists of 4 major components. An overview of the system is provided in Figure 1

- *Twitter Stream Reader*: Using the Twitter Streaming API, specifically the *statuses/filter*, we constantly retrieve tweets and store them for future analysis. The tweets obtained are filtered based on user specified keywords, hashtags, and geolocation as well to provide complete user control over data. These constraints can also be modified at any time through an administration module.
- *DataStore*: Our datastore consists of a back-end database engine that constantly stores the incoming data from the

[2]http://trendsmap.com/
[3]http://twitalyzer.com/
[4]http://ouseful.open.ac.uk/geotwitterous/
[5]http://www.ushahidi.com/
[6]http://epic.cs.colorado.edu/

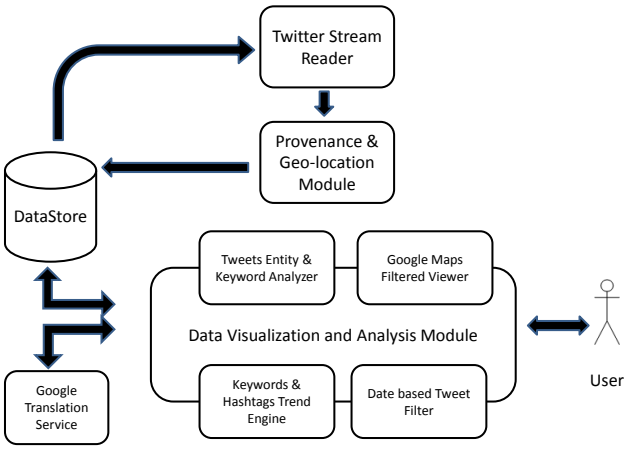

Figure 1: System architecture of TweetTracker

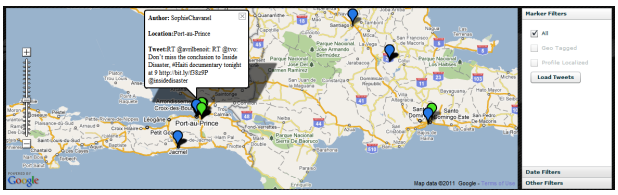

Figure 2: Geolocated and normal tweets

stream reader, while supplying the visualization module with requested data.

- *Visualization and Analysis Module*: This module is web based and is the single point of user interaction with the system. It is comprised of the following:
  - The map includes filters to focus on tweets of interest.
  - The entity and keyword analyzer ranks entities contained within tweets including hashtags, hyperlinks, and screen names. The analyzer generates frequent keywords from tweets visualized as tag clouds.
  - Date-based tweet filtering supports the analysis of older tweets. The play mode helps visualize changes in tweets through hashtags and entities.
  - The keywords trending engine is used to generate trends of keywords specified by the user.

TweetTracker in Action

The application constantly monitors Twitter streaming feed [7] using specific keyword and hashtag filters related to the Haiti cholera outbreak, for example #cholera and haiti cholera.

The primary mechanism for monitoring tweets is through a map. Geotagged tweets are displayed on the map coordinates corresponding to the tweet. Tweets can be visualized on the map with specific color coding for geo-located tweets (*Green*) and normal tweets (*Blue*) as shown in Figure 2. Tweets from different languages can also translated

[7] http://dev.twitter.com/doc/post/statuses/filter

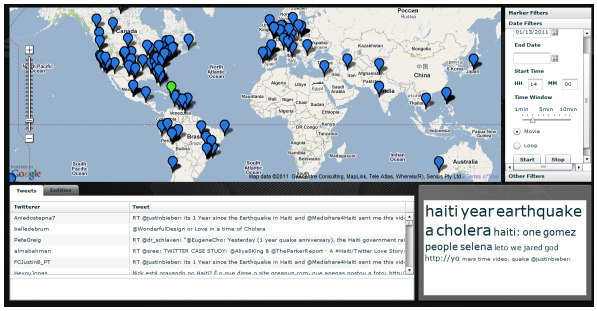

Figure 3: Different analysis options, an example with tag-clouds

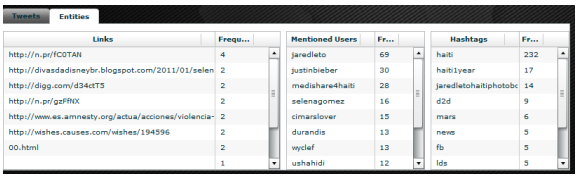

Figure 4: Entity references in tweets

through Google Translate to aid user understanding. For analysis, users choose from different views to allow streaming, streaming from a specific point in time, playback, and playback with loop as shown in Figure 3. Emerging trends are identified using tag cloud created from tweets. Entities which may be referenced in tweets: relevant agencies/users, additional hashtags, and hyperlinks that convey user sentiment can be previewed, as seen in Figure 4, to help users understand who or what is being discussed.

Future Work

Future efforts include optimizing displays, adding analysis features, and integrating additional microblog sources. Feedback from relevant HADR agencies will serve as a basis for making improvements to the tool.

Acknowledgments

This work was supported, in part, by the Office of Naval Research.

References

Bloch, M., and Carte, S. 2011. Twitter chatter during the super bowl.

Goolsby, R. 2010. Social media as crisis platform: The future of community maps/crisis maps. *ACM Trans. Intell. Syst. Technol.* 1(1):1–11.

O'Reilly, T., and Milstein, S. 2009. *The Twitter Book*. O'Reilly.

Shklovski, I.; Palen, L.; and Sutton, J. 2008. Finding community through information and communication technology in disaster response. In *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, 127–136. ACM.