Crowdsourced Twitter Data for Emergency Management

<u>Abstract</u>

Despite warnings to take shelter or evacuate from a potential disaster site, many people remain in place and share messages and photographs of a given disaster using social media platforms such as Twitter. During a number of recent weather-related emergencies, thousands of Twitter users sent messages from locations under notice for mandatory evacuation. While the authors do not condone this behavior, they chose to develop a software application that will allow emergency managers to use these messages and images for the greater good. Crowdsourced Twitter Data for Emergency Management (CTD-EM) is a web-based application that provides emergency managers with exact-location images and messages chronicling the effects of a natural or synthetic disaster. The data comprise all geo-coded Twitter posts sent from a specified catchment area. Emergency mangers can access these data in real-time, often hours before first responders can safely enter the disaster zone. The accompanying lat/lon metadata allow for the exact location of an image or message to be geographically pinpointed and verified. These time-critical data can help emergency managers better plan their relief and rescue efforts, and by doing so, save lives.

<u>Project Overview</u>

CTD-EM was developed to provide emergency managers with a bi-directional communication system for identifying and responding to important Twitter-based messages and images sent by the general public during weather-related emergencies. A simulated, real-time, beta test was conducted using geo-coded images sent through Twitter during Hurricane Sandy. Thousands of images depicting storm damage and flooding in the coastal areas of Connecticut, New York, and New Jersey during Hurricane Sandy were captured through a direct connection to Twitter's full data stream (see Figures 1 & 2).

Figure 1

Image of Damage Captured from Twitter

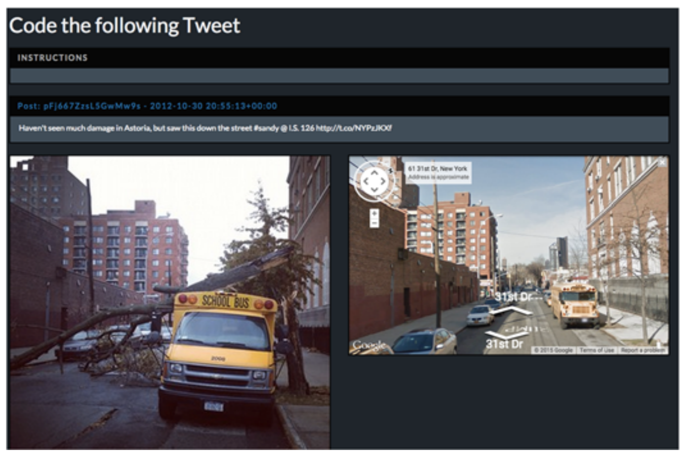
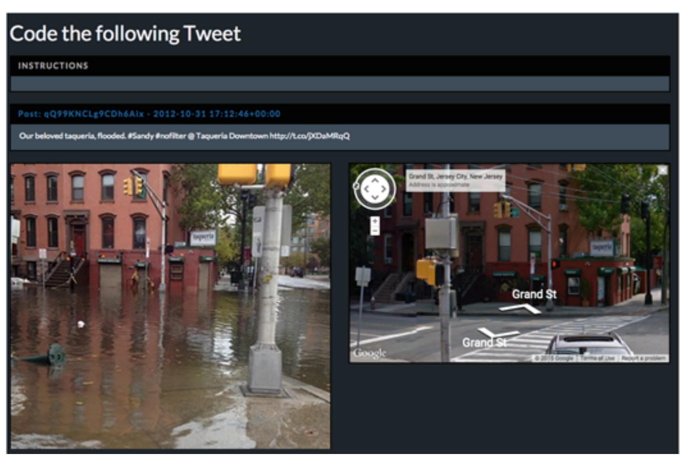


Figure 2

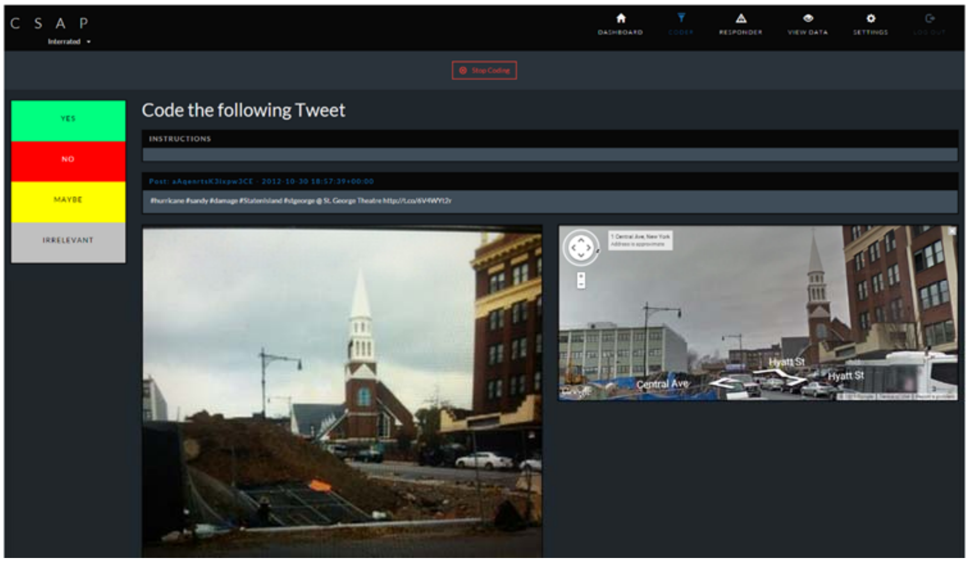
Image of Flooding Captured from Twitter



Since the science of machine-based image recognition has not progressed to the level of accurately identifying images of storm damage or the depth of floodwater, the application relies on human judgment to codify the pertinent images. It is also critical for the human codifier to determine whether the exact location of the image, as indicated by its accompanying lat/lon coordinates, can be accurately identified. To this end, CTD-EM is designed to provide a team of human codifiers with a stream of real-time images obtained from Twitter, juxtaposed with an image from Google Street View that shares the same lat/lon coordinates as the Twitter image (see Figure 3).

Figure 3

Codifier Module



Once the vast amount of data has been automatically culled using key search terms such as "damage" and "flood," the human codifiers are provided with four coding options: YES, NO, MAYBE, and IRRELEVANT. A Punnett-style square is used to assist with the coding process (see Figure 4). For each image, the codifiers are expected to consider two questions:

- Is there visible storm damage or floodwater?
- Can the location depicted in the image be verified (using the Google Street View pane)?

As the visual data is filtered in real-time, all images that have been marked as pertinent for review are queued for presentation through the Emergency Manager Module (see Figure 5). Both the Twitter image and its corresponding Google Street View image are clearly presented. Additionally, the street address corresponding to the image's lat/lon coordinates is provided. This information is obtained through the "Latitude/Longitude Lookup" feature of Google's Geo-coding API. To further facilitate the decision-making process of an emergency manager using this system, the text of the tweet from which the image was obtained is also presented.

Figure 4

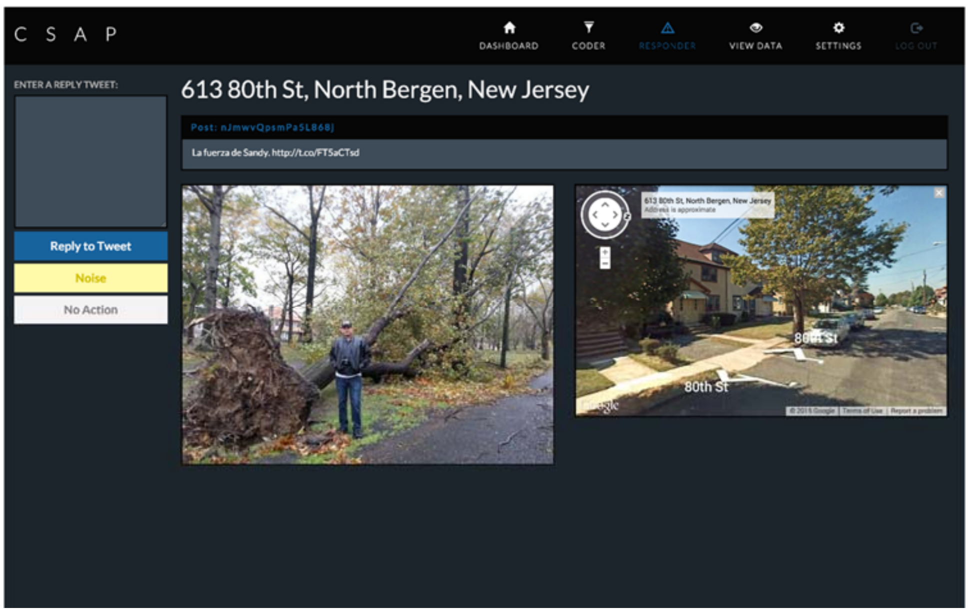
Codifying Schema

		Damage Visible	
		Yes	No
Location Identified	Yes	YES	NO
	No	MAYBE	IRRELEVANT

The street address, Twitter image, Google Street View image, and the text of the tweet associated with the image allow the emergency manager to make an informative decision of whether or not the social media-based communication warrants a response and/or other action, such as dispatching first responders to the location. Directly from this module, the emergency manager can send a response to the Twitter user who posted the image. This response takes the form of a “reply” to the original tweet. The Twitter user and all of his or her followers can see the emergency manager’s reply, thereby facilitating bi-directional communication between the emergency manager and the Twitter user who provided the valuable information. For example, the emergency manager can request that the Twitter user provide more information about the location by tweeting more descriptive text or sending more images. Essentially, the system has spawned a critical communication link between an emergency manager and a bystander in the field. The bystander serves as a human sensor, continuously assessing the environmental conditions for a very specific location.

Figure 5

Emergency Manager Module

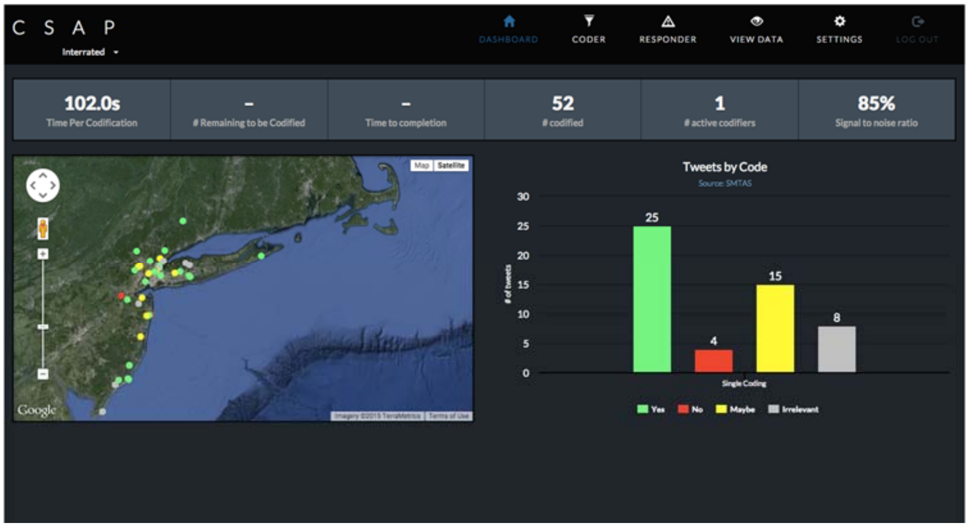


Additional monitoring of the system data can be viewed though the Data Collection Module (see Figure 6). This module provides an overview of the data stream, a map indicating where storm damage or flooding has been observed, and multiple measures of coding efficiency. The module indicates the total number of images that have been processed with a graphical representation of how many images fall into each of the four coding categories. Coding efficiency is also measured by the: a) average time for coding each image, b) total number of images in queue, c) estimated time for completely coding a batch of data, 4) signal to noise ratio, and 5) total number of individuals actively codifying the data.

As the images are being coded, a real-time map of the data collection catchment area displays points where images indicate the presence of damage or flooding (green map points in Figure 6). Equally important, the map displays points where no damage or flooding has been observed (red map points in Figure 6). As the number of coded images increases, the map continuously populates with more indicators of damage/flooding versus no-damage/no-flooding. With enough data, the color-coded map points begin to cluster, providing a broad overview of large swaths of damaged areas or flooded land.

Figure 6

Data Collection Module



<u>Technical Specifications</u>

CTD-EM is based on a real-time, web-application model utilizing state-of-art frameworks. It relies on a structure-less, non-relational database. The utilization of a schema-less storage mechanism enables extensibility of the software to other social media datasets. The real-time functionality of the software provides the time-critical reactivity needed for the data to be codified and presented to the emergency managers. The front end of the application is compatible with multiple web-browsers and accessible from any Internet connected machine.

<u>Funding Source</u>

National Oceanic and Atmospheric Administration's Coastal Storm Awareness Program, Connecticut Sea Grant (R/CSAP-8-CT). Research conducted from January 2014 through June 2015. (http://seagrant.noaa.gov/fundingfellowships/coastalstormsawarenessprogram.aspx).

<u>Contact</u>

John F. Edwards, Ph.D.
Associate Research Professor &
Coordinator of Data Laboratories
Social Science Research Center
Mississippi State University
One Research Blvd., Suite 103
Starkville, MS 39759
je@ssrc.msstate.edu
662.325.9726