Detecting Destruction: Disaster Recovery Reporting

Samantha Scaglione
B.S. Computer Science '19
University of Notre Dame
Notre Dame, IN
sscaglio@nd.edu

Mara Staines
B.S. Computer Science '18
University of Notre Dame
Notre Dame, IN
mstaines@nd.edu

Maggie Thomann
B.S. Computer Science '18
University of Notre Dame
Notre Dame, IN
Margaret.M.Thomann.1@nd.edu

***Abstract*—We look to further develop the use of Twitter in disaster management by looking more closely at tweet content. By selecting tweets about a disaster then analyzing those that describe damage, we will provide important insight to citizens and public agencies.**

Keywords—Twitter, tweet, social media, disaster, damage, clustering, reliability

I. Overview

When a natural disaster hits, damage can be enormous. Roads may become impassable, traffic jams form during evacuations, and utilities are interrupted. For both citizens and public agencies, there isn't a fast, reliable way to track and act on information about damage. The team believes that citizens affected by a disaster share valuable information on Twitter about the damage a disaster has caused. In this project, the team will track the fallout and damage from natural disasters by analyzing data from Twitter. This information may include injuries, deaths, property damage, or infrastructure breakdowns. If this fallout can be tracked in near real time, both public agencies and citizens will be able to make better decisions in the wake of a disaster.

If public agencies have access to knowledge of what infrastructure has been compromised, they can work strategically on fixing the most important issues. For example, 4 roads may be blocked by debris, and the city may only know about 3. When the city sends out the first road crew, they will send it to the most important road out of those 3. If the other road is much more heavily traveled or leads to a hospital, it should have been the one prioritized. Because the city was not aware of damage to an important road, it wasn't prioritized. However, if citizens who live near that road or who tried to travel on it are active Twitter users, they are likely to share the information that the road is impassable. If Twitter users describe damage, our project should ensure that public agencies are aware of that damage.

Citizen knowledge of the current state of damage is also important. If a main road leading into town is closed, most citizens would still attempt to use that road before learning of its condition through a source like the local news. This could cause large backups and even prevent road crews from reaching the site of the damage. Citizens can also make smarter choices if they have real-time information on flooding, power outages, or which gas stations remain open. This information is often the subject of tweets during and immediately after a disaster, making Twitter an ideal data collection source.

II. Related Work

There has been research into Twitter data during and after natural disasters. Some of this research focuses on detecting natural disasters as quickly as possible[4]. This paper provides a possible strategy for identifying tweets related to a disaster maximizing speed and accuracy. Another study used the volume of Tweets during hurricane Sandy to predict monetary damage[3]. Extent of damage is part of our project goal, but the team would also like to classify tweets further to identify what kind of damage has occurred. In 2016, classification of tweets into categories such as information, emotion, support, or disaster relief was done with regard to Typhoon Haiyan[1]. This yielded conclusions about how Twitter use evolves after a disaster, such as which categories of tweets are most common as time progresses after the disaster. However, this study manually coded its dataset into tweet categories, while the team will be looking to automate this tweet clustering. Additionally, this analysis was done after all of the data was collected, while the team will implement a real time system. Finally, the current body of work most similar to our project is titled Floodtags[2]. This group works to provide real time detection and mapping of flooding using Twitter. Similarly to Floodtags, the team will be identifying damage in sub-areas of the affected region. While Floodtags uses this data to track one parameter (a probabilistic flood map), the team aims to identify multiple types of damage (flood, road closure, power outage, etc.). Overall, a solid groundwork has been laid in the field of Twitter disaster analysis. The team believes they will be expanding on this research significantly by using Twitter data to identify specific instances of damage.

III. Methods

In order to report on the effects of natural disasters, the team will first gather tweets in real-time from a user's specific location and classify them as whether they are related to a disaster or not. Any tweets that aren't relevant to the disaster will be filtered out. The credibility of the remaining tweets will be measured to determine which of these remaining tweets to prioritize in our application. The team will classify these tweets by ensuring that they relate to damage created directly from the disaster and may use the location of the tweet as a metric to

determine how reliable the claim is. This classification will also break down these tweets into what type of damage may have occurred. Finally, the team plans to display this information in a web application through which users or disaster recovery forces in the affected area can gather information to better proceed after the disaster.

any challenges they face to aid in production of deliverables. The schedule was developed to allow for a full week of delay after the Mid-Term Presentation so the team can respond appropriately to class feedback as well as any unforeseen challenges.

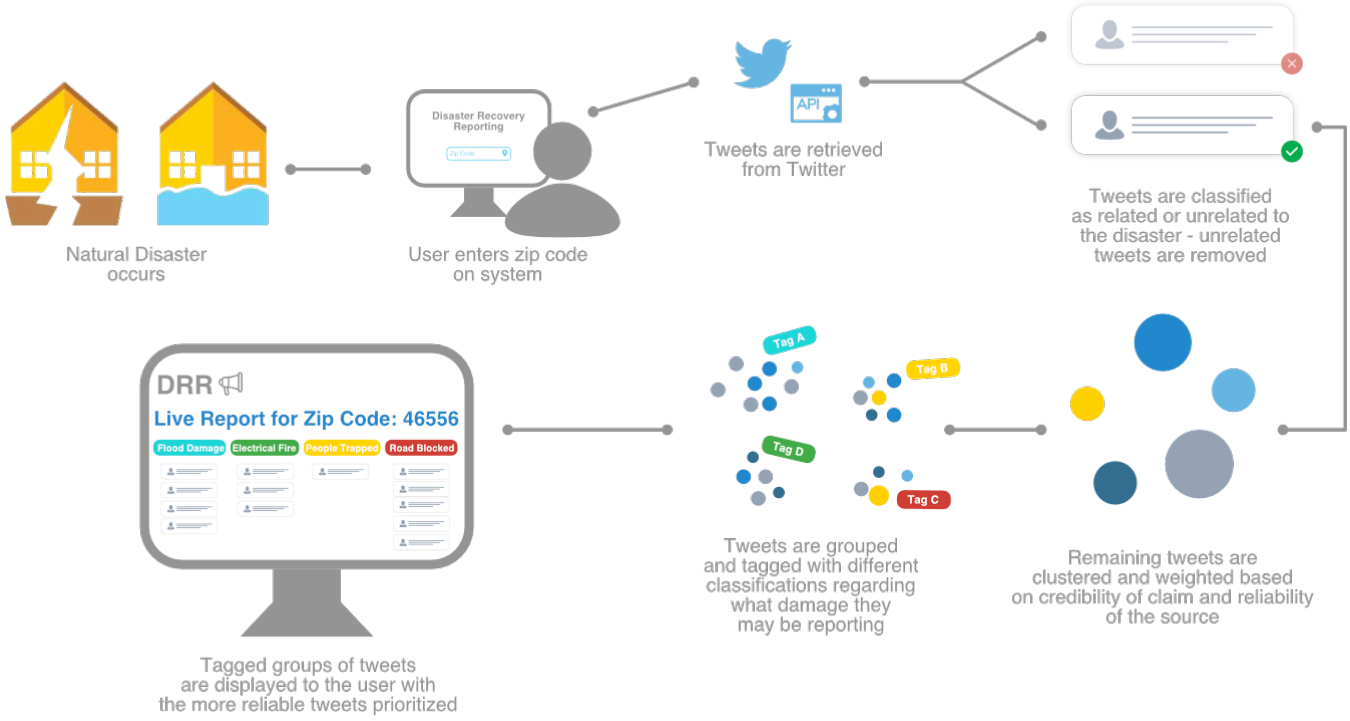


Fig. 1. Methods Graphic

A. *Future Milestones*

- Collect tweets from past event or stream tweets now to cover an event
  - Choose appropriate database program
  - Select search terms
- Successfully remove tweets not about disaster destruction
- Cluster remaining tweets on content
- Measure reliability of claims about damage
- Report reliable claims in web application

B. *Plan of Action*

The team's Plan of Action is detailed by the table, Figure 2, and corresponding Gantt Chart, Figure 3, below. The Plan of Action Gantt Chart is color-coordinated to separate the three main components of our project. The blue tasks are concerned with twitter-scraping and analyzing data, the darker green tasks are concerned with web-application development, and the lighter green tasks are concerned with developing deliverables. Although not explicitly shown on the Plan of Action Gantt Chart, the team will keep a detailed log of their progress and

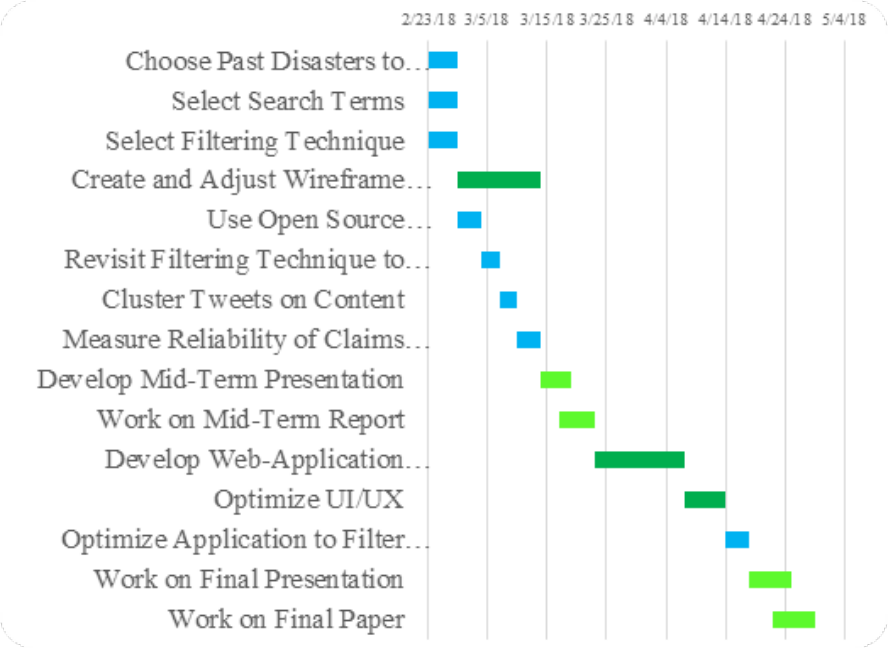
TABLE I. PLAN OF ACTION

Plan of Action			
Task	***Start***	***End***	***Duration***
Choose Past Disasters to Gather Data From	02/23/2018	02/28/2018	5
Select Search Terms	02/23/2018	02/28/2018	5
Select Filtering Technique	02/23/2018	02/28/2018	5
Create and Adjust Wireframe for Web Application	02/28/2018	03/14/2018	14
Use Open Source GetOldTweets-python to collect past data	02/28/2018	03/04/2018	4
Revisit Filtering Technique to Fix/Optimize	03/04/2018	03/07/2018	7
Cluster Tweets on Content	03/07/2018	03/10/2018	3
Measure Reliability of Claims about Damage	03/10/2018	03/14/2018	4
Develop Mid-Term Presentation	03/14/2018	03/19/2018	5
Work on Mid-Term Report	03/17/2018	03/23/2018	6
Develop Web-Application Basic Functionality	03/23/2018	04/07/2018	15
Optimize UI/UX	04/07/2018	04/14/2018	7

Plan of Action			
Task	***Start***	***End***	***Duration***
Optimize Application to Filter Relevant Real-Time Data	04/14/2018	04/18/2018	4
Develop on Final Presentation	04/18/2018	04/25/2018	7
Work on Final Paper	04/22/2018	04/29/2018	7

Fig. 2. Plan of Action Table

Fig. 3. Plan of Action Gantt Chart



REFERENCES

[1] David, Clarissa C., Jonathan Corpus Ong, and Erika Fille T. Legara. "Tweeting Supertyphoon Haiyan: Evolving functions of Twitter during and after a disaster event." PloS one 11.3 (2016): e0150190.

[2] Eilander, Dirk, et al. "Harvesting social media for generation of near real-time flood maps." *Procedia Engineering* 154 (2016): 176-183.

[3] Kryvasheyeu, Yury, et al. "Rapid assessment of disaster damage using social media activity." *Science advances* 2.3 (2016): e1500779.

[4] Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. "Earthquake shakes Twitter users: real-time event detection by social sensors." *Proceedings of the 19th international conference on World wide web*. ACM, 2010.