Crisis Monitor

Following Unfolding Crises Worldwide

CSE 6242

Fall 2019

Team 134

Savannah Edwards

Robert O’Dowd

Swapna Gupta

Samraj Jeyachandran

Nawida Hussaini

Aron Tharp

**Introduction - Motivation:**

With the recent explosion of social media use, research has focused on extracting insights thereof. The ability to understand and predict the future from this data is lucrative, the benefits being material and moral. Our project visualizes crises as they unfold, where a population is suddenly at risk of mortal peril, by analyzing Twitter data. To facilitate quick response, a real-time dashboard monitors worldwide crises to enable aid organizations, journalists, supply chains, and others to mitigate harm, bodily or financial, and loss of life.

**Overview / Problem Definition:**

The goal is to create a software application to identify political instability or crisis where a population is in danger, briefly explain the crisis, and monitor it. The application is fed Tweets captured in near real-time. End-users can access it through an intuitive and interactive visualization to view ongoing crises worldwide.

NGOs and governments provide periodic reporting. It relies on manual efforts, is at risk of bias, and compromises on speed-to-market. In OECD countries, media outlets provide fast crisis reporting; however, non-OECD countries’ media cannot guarantee the same. Not having an accurate, up-to-date understanding of crises as they unfold is dangerous for those affected. Many projects currently create classification systems focused on historic data but not towards prediction or forecasting. They generally have narrower scope, whereas we seek to identify crisis more broadly.

We seek to derive insights from real-time information, relying on Twitter instead of static reporting. Our success hinges on filling a different niche than historical reporting. We use Tweets from all English-speaking sources to minimize bias and maximize firsthand data.

Many organizations and individuals need to instantaneously know when crises occur and be able to respond immediately. Benefits include timely evacuations, quick police or military response, and minimal supply chain disruption. We measure our software’s accuracy by comparing true positives to false positives for crises identified. We can also measure success after release to the public by tracking repeat users, page views, and user satisfaction surveys. Unfortunately, it is not feasible to measure success through counting lives and dollars saved.

**Survey:**

*Social Media/Social Change*

Social media has revolutionized how we receive news [9] and [20] confirmed that extracting insights holds potential in improving and saving lives. Over 40% of the world’s population shares, creates, likes, and follows streams of information [9]. Verhulst et al. [9] state that companies who own data from the public, such as Twitter, need to become Data Stewards and collaborate for public good [8] – this is our exact goal! They go on to highlight risks and challenges like privacy, security violation and uneven demographic representation [9] while [20] highlights that this data can expose identities.

Social media’s potential is in its speed. Twitter reports breaking news before journalists or media [2]. Hu et al. [2] report that Bin Laden’s death was known before White House announced it [2]. Analysis of these Tweets found that certainty of the event was around 55% before mass media reporting [2] . In a life-threatening situation, time is critical. In [2] they only covered one major crisis whereas we will attempt to broaden the application.

*Detection/Classification/GeoLocation*

Processing big data with MapReduce is discussed in [8] using distributed computing power but it’s not real time which is what we need. Identifying a “reliable Tweet” is debated by Phuvipadawat [1] by tallying followers. Reliability is critical to avoid misinformation. However, reliable breaking news can come from accounts with fewer followers [BNTT]. Many users who Tweeted Bin Laden’s death in [2] had few followers.

Classification methods are discussed in [5], [3], [18] ETC. Burel et al. [5] built a “crisis event identification and classification” API that uses Convolutional Neural Networks to classify Tweets on event relatedness, event type and information category. Burel [5] had similar success with both relatedness and type using CNN or SVM. However, they had trouble classifying the information category, which we aim to overcome. The US Navy studied [3] linking Tweets to current events using Tweeted URLs. This could obtain specific information but is too focused for us. Atefeh et al. [18] used a Naive Bayes Classifier to separate news from meaningless data and a SVM for event detection. This method [18] puts weight on retweets similar to followers [1] - a shortcoming. The SVM method is effective according to [18], [16] and [4]. Aramaki’s [16] use of SVM gives a .97 correlation during the early flu epidemic stage - better than state-of-the-art methods. Paired with SVM, a polynomial kernel created the optimal balance of time and accuracy in [16 and [4]. SVM outperformed as a reliable classifier to filter Arab language Tweets for political unrest. Neubig et al. [17] developed volunteer rapid response efforts to the 2011 Japan Earthquake applying Natural Language Processing to Tweets to mine safety information. They [17] used classification for 9 possible Tweet topics and applied named entity recognition to find subject and location names. This tagging is unique and should be further explored for our purposes.

Proximity to crises is another important consideration per [7] and [19] and geolocating is a challenge for Twitter. Cecinati et al. [7] examined Tweets about heatwaves by filtering for an Indian state to approximate location. Compton et al. [19] had 80% accuracy in geolocation by defining small Twitter communities on mentions and using the “mention-er’s” public location as a proxy for the original account. Compton [19] developed a network model with users as vertices and mentions as edges, weighted by bidirectional mentions. We’ll go further and include hashtags.

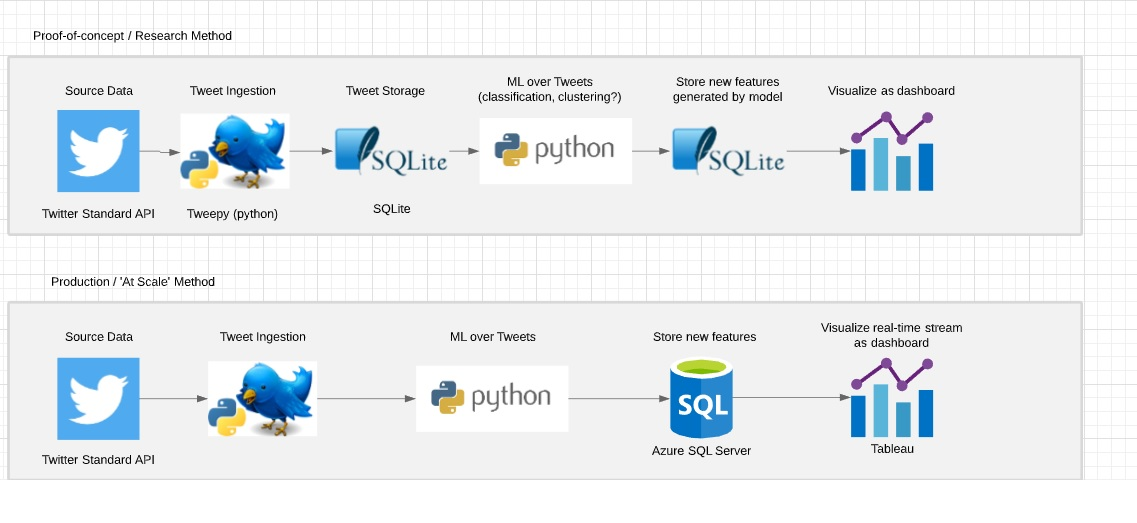
*Sentiment Analysis*

Several sentiment analysis techniques were reviewed using [10], [12], [13] and [15]. In [10] their combination of Unigram plus Senti-features performed best. Agarwal et al. [10] introduced new resources for pre-processing Tweets: emoticon and acronym dictionaries, found to add marginal value. Kaushik et al.’s [12] sentiment analysis method, known as “opinion mining”, examines sentiment to consumer products with supervised machine learning, unsupervised learning and case-based reasoning. The core of [12] is opinion classification, done in direct opinions and comparison. Shulz et al. [13] enhanced the standard 2-3 emotional classes to 7, significantly increasing accuracy to ground situations. Mozetič et al. [15] studied multilingual sentiment classification and found that classification before translation was critical. Unfortunately, their manual methods are not scalable for us.

*Visualization*

Making data easily digestible is discussed in [11], [6], [14] and [21]. Cui et al, [11] introduce blending a trend chart with a word cloud to illustrate content evolution. This would allow us to show crises unfolding (relevance). Anderson et al. [6] use JavaScript to cluster Tweets in a dashboard. Addressing Twitter’s API geo-data limitations, Anderson [6] added geographic context to a user’s Tweet stream using Mapbox-GL. Calderon et al. [14] combined classified data with heatmaps to portray degradation or improvement of a situation in real time. Stojanovski et al. [21] developed user-oriented and keyword-oriented visualizations. The former indicates a change to the norm and the latter shows spikes in a topic/word.

**Proposed Method**



*Innovation*

Our project goes above and beyond existing, state of the art products in predictive speed, global reach and diminished bias. We are applying our models towards rapid event detection. Existing models prioritize accuracy over speed and can take days or weeks to confirm a crisis outbreak. They also rely extensively on manual validation to ensure accuracy. To maximize speed, our application analyzes a real-time stream of Tweets.

We are also improving on existing models by minimizing sources of bias. To achieve this, we are intentionally basing our analysis exclusively on a free, open product with near global purview. There is no institutional bias in our data since we give no preference to media sources which may be heavily influenced by the government; all users are treated equally.

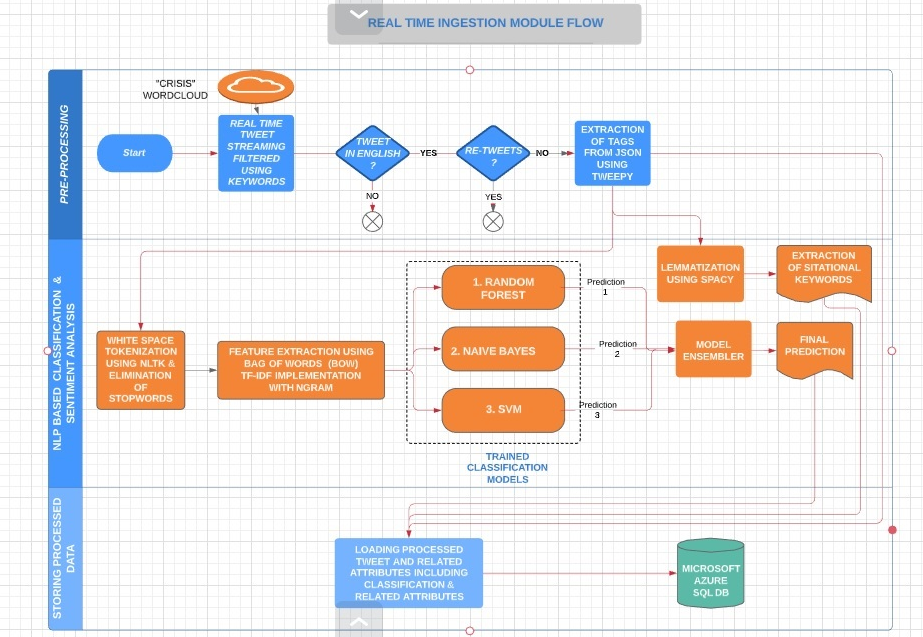
Another advantage is our geographical span. Many existing tools limit themselves to a specific location and a user may be required to toggle between numerous tools to have a global view. Our reach greatly exceeds this, attaining a global scope except in the 3% of countries which limit or ban Twitter, most notably China.

*Description of Approaches*

Data collection occurs solely through the Twitter API using the Tweepy library in Python. We filter Tweets for the English language and apply a filter using a lexicon of 380 crisis words [23]. The following JSON elements of the Tweets are saved to a relational database: timestamp, text, coordinates, place name, place type, and country. Initially, we are initially excluding retweets to collect only unique information, but further analysis is worthwhile to understand its effect on performance.

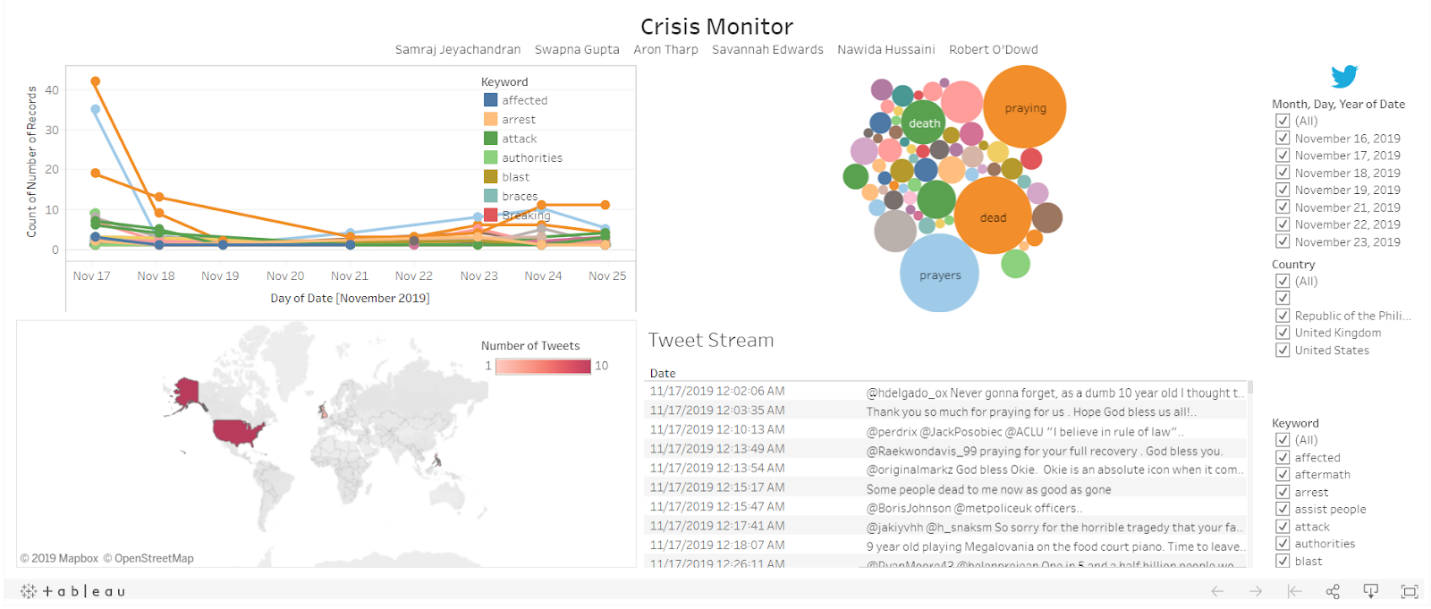
Data is stored using SQLite which can hold approximately 1 million Tweets before encountering performance issues. Our current parameters, assuming no major crisis, would account for 6 weeks of Tweets before performance is a concern. SQLite’s portability and ease of setup made it a clear choice. In the future, we can scale our project to a cloud-based relational database or message queue such as Azure SQL or Event Hub.

The final model, detailed extensively in the experiment section and reflected in the table below, is first trained on a preprocessed dataset [22] and then applied to new Tweets. The modeling focuses on Singular Value Decomposition via a bag of words implementation. Using TFIDF logic, we reduce our model to the top 5,000 words to avoid overfitting and improve performance. Our prediction value is whether a Tweet is crisis related. We then aggregate our models into an ensemble indicator for final prediction to maximize accuracy.



The analyzed results are visualized in a Tableau dashboard, utilizing word maps to show the topic of the crisis, trend charts to show evolution of crises, heat maps to show where the biggest crises are taking place, and tooltips to inform the user of more specific details. These features bring our data to life, allowing users to easily digest and interact with the information and keep up-to-date and safe. Further, Tableau allows a user to set custom filters if they wish to only view information relevant to them.

Our Crisis Monitoring dashboard is hosted by Tableau Online where anyone can utilize our data to keep themselves safe. With several charts on the dashboard, there are a variety of ways to utilize it depending on the user’s need. As an example, if a user is in a particular country (or traveling there in the near future), he or she can go to the Country filter on the right hand side and filter down to the country of interest. Upon filtering, he or she can observe at the bubble and trend charts what the most popular keywords are in this country currently, and over time. After examining these charts, ‘attack’ may be a keyword of interest. In this case, the user would go to the keyword filter at the bottom right and filter down to attack. Now, all Tweets of interest can be easily and clearly viewed in the Tweet Stream portion of the dashboard.

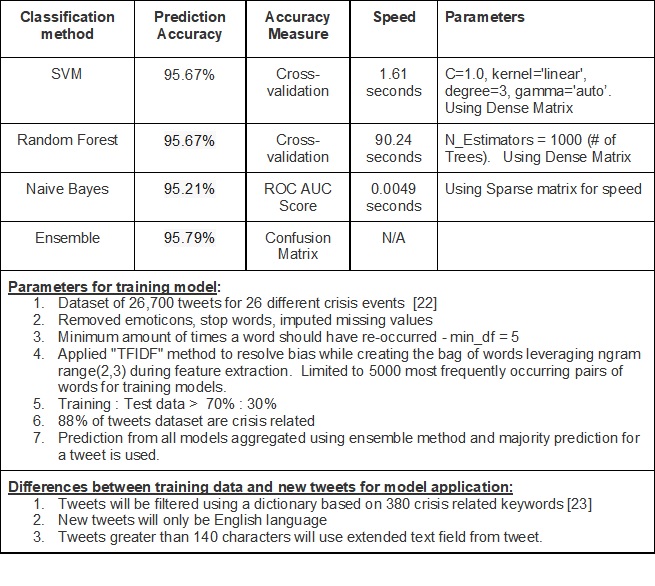


Due to the data limitation of Twitter’s standard API, the dashboard’s historical data is not extensive. Over time, the dashboard will become more useful to users. Certain countries currently lack data, but this will be filled in naturally.

The Tableau Mobile app allows for a portable view of this visualization. It converts the dashboard into a mobile-friendly view and maintains the same interactivity. Another benefit is that it downloads a local copy when connected to the internet, so the dashboard can still be used when the user is offline. This is ideal for users with spotty connections, not uncommon for users in unstable crisis regions.

Additionally, Tableau can email alerts to users when crises in their region are detected, enabling instant distribution of the information, and in turn, keeping our end users safe.

**Experiments/ Evaluation**

****

*Experiment Design*

We sought to answer several questions in our analysis. Foremost is whether we can correctly classify a Tweet as crisis related or not. Next, can we identify the evolution of a specific crisis. Third is whether we can map out developing crises. Fourth is whether a time-series analysis of the data can support early detection with regularity. Fifth is whether Twitter is a useful data source to detect a myriad of crises. Last is if Twitter qualifies as a less biased source than traditional media.

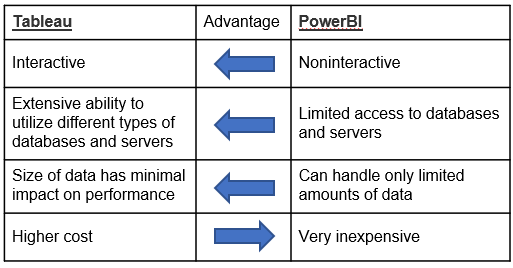
*Details of the Experiments*

The Sklearn Python package was used for two primary reasons: Sklearn has a wide variety of machine learning algorithms to feed the preprocessed data into and specific Sklearn functions allow for the bag-of-words model to be easy and efficient. While other packages offer a broad range of tools to process unstructured text, Sklearn coupled with NLTK for preprocessing was the ideal approach.

All models classified Tweets with reasonable accuracy, therefore we chose an ensemble method to average the prediction models and maximize classification accuracy for each Tweet. While the random forest implementation is quite slow to train the model, speed isn’t critical since retraining models weekly should be sufficient. The key is the model accuracy and application to real-time Tweets. Additionally, the Naive Bayes method supports a confidence score. We tested different values for this and found 97.5 to be the best for reducing false positives in our data. This was necessary due to the high number of false positives stemming from Twitter users’ liberal usage of crisis terminology in non-crisis situations.

*Visualization Tool*

We compared Tableau and PowerBI to determine how to best communicate the project findings. Important considerations for our visualization are country-level aggregation and applying a gradient to the frequency of crisis Tweets. Additionally, our visualization must be able to update in real-time.  Further considerations which have not yet been explored is whether we can portray the severity of the crisis and whether we can categorize the crises. The results are found in the table below.



**Further Opportunities**

To anyone seeking to further this work to better address their needs, the following considerations were explored project and are recommended to someone seeking a fuller implementation. Our crisis lexicon is nearly exhaustive, however we observed significant variance in the precision of individual terms. Analyzing this variance and reducing the lexicon could reduce the false positives in our model.

To glean more understanding from individual Tweets, incorporating Python’s SpaCy library would add significant context for targeting crises. While Twitter has a built-in geotagging feature, it is inherently limited in usefulness since it is not applied by default and a user can tag any location they choose. The library extracts context from the text of the Tweet, identifying facilities, locations, organizations and more, improving users’ understanding of exactly what is happening. To increase accuracy, n-grams could also be considered, where the order of words matters, instead of a pure SVD analysis. Python’s Pickle library provides many useful tools for this.

Having time series data via the premium Twitter API would allow for additional analysis by examining the volume of crisis Tweets. Change detection algorithms, like CUSUM, could be used to mark the beginning of a crisis and prediction algorithms, like ARIMA, could be used for predicting their occurrence.

We observed potential overfitting of training data. We would address this by manually classifying tens of thousands of Tweets. A more complete training data would include a significant number of Tweets irrelevant to a crisis. We anticipate that this would improve the model fit and result in a better classification of real time tweets.

The group acknowledges that a global model may not be appropriate for all end-users, and the tradeoff in accuracy when including additional regions may not be useful for everyone. A regional implementation may prove more valuable in certain cases which may increase precision due to regional differences in the use of the English language.

Due to constraints in the Twitter API and short timeframe of this project, we do not believe that we collected enough data to answer some of the experiment’s questions. Several more months of data would support further time-series analysis and allow us to better understand the evolution of crises. Additionally, a longer list of identified crises could be compared with media reporting to determine differences in bias, speed and accuracy.

To an end-user looking to implement our approach, costs to consider include a premium Twitter API, Tableau licensing, Azure infrastructure and, if the visualization requires, web hosting.

**Conclusion and Risk**

The ability to identify crises quickly, across the globe, and without bias of media, could provide humanitarian and financial payoff. Unfortunately, there is potential for abuse by authoritarian governments looking to exacerbate crises. There is also risk if inaccuracies exist in our model, providing a false sense of security or causing costly action when none was needed. Some of our assumptions, such as focusing on English Tweets only, may amplify our false negatives, especially in non-English speaking countries.

**Distribution of team member effort**

All team members have contributed a similar amount of effort.

## Works Cited

1. Phuvipadawat, S., & Murata, T. (2010). Breaking News Detection and Tracking in Twitter. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. doi: 10.1109/wi-iat.2010.205

2. Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., & Ma, K.-L. (2012). Breaking News on Twitter. *CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2751–2754. doi: 10.1145/2207676.2208672

3. Overbey, L., Batson, S., Lyle, J., Williams, C., Regal, R., & Williams, L. (2016). Correlating Twitter Sentiment and Event Data to Monitor Social and Political Unrest. *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation.* Retrieved from http://sbp-brims.org/2016/proceedings/CP\_120.pdf

4. Mishler, A., Wonus, K., Chambers, W., & Bloodgood, M. (2017). Filtering Tweets for Social Unrest. *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*. doi: 10.1109/icsc.2017.75

5. Burel, G., & Alani, H. (2018). Crisis Event Extraction Service (CREES) - Automatic Detection and Classification of Crisis-related Content on Social Media. *15th International Conference on Information Systems for Crisis Response and Management*. Retrieved from http://oro.open.ac.uk/55139/1/gburel\_iscram18.pdf

6. Anderson, J., Saez, G. C., Anderson, K., Palen, L., & Morss, R. (2019). Incorporating Context and Location into Social Media Analysis: A Scalable, Cloud-Based Approach for More Powerful Data Science. *Proceedings of the 52nd Hawaii International Conference on System Sciences*. doi: 10.24251/hicss.2019.275

7. Cecinati, F., Matthews, T., Natarajan, S., Mccullen, N., & Coley, D. (2019). Mining Social Media to Identify Heat Waves. *International Journal of Environmental Research and Public Health*, *16*(5), 762. doi: 10.3390/ijerph16050762

8. Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. *OSDI ’04: 6th Symposium on Operating Systems Design and Implementation.* Retrieved from https://www.usenix.org/legacy/event/osdi04/tech/full\_papers/dean/dean.pdf

9. Verhulst, S. G., & Young, A. (2017). The Potential of Social Media Intelligence to Improve People’s Lives: Social Media Data for Good. *Data Collaboratives*. Retrieved from http://datacollaboratives.org/static/files/social-media-data.pdf

10. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. *Proc. ACL 2011 Workshop on Languages in Social Media,* 30–38. Retrieved from http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf

11. Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M. X., & Qu, H. (2010). Context preserving dynamic word cloud visualization. *2010 IEEE Pacific Visualization Symposium (PacificVis)*. doi: 10.1109/pacificvis.2010.5429600

12. Kaushik, A., Kaushik, A., & Naithani, S. (2015). A Study on Sentiment Analysis: Methods and Tools. *International Journal of Science and Research (IJSR)*, *4*(12), 287–292. doi: 10.21275/v4i12.nov151832

13. Schulz, A., Paulheim, H., Thanh, T. D., & Schweizer, I. (2013). A Fine-Grained Sentiment Analysis Approach for Detecting Crisis Related Microposts. *Proceedings of the 10th International ISCRAM Conference.* Retrieved from https://pdfs.semanticscholar.org/4c4a/8ee0ba1a76beb10c2d3739772478dbb5aea7.pdf

14. Calderon, N. A., Arias-Hernandez, R., & Fisher, B. (2014). Studying Animation for Real-Time Visual Analytics: A Design Study of Social Media Analytics in Emergency Management. *2014 47th Hawaii International Conference on System Sciences*. doi: 10.1109/hicss.2014.176

15. Mozetič, I., Grčar, M., & Smailović, J. (2016). Multilingual Twitter Sentiment Classification: The Role of Human Annotators. *Plos One*, *11*(5). doi: 10.1371/journal.pone.0155036

16. Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter Catches the Flu: Detecting Influenza Epidemics using Twitter. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1568–1576. Retrieved from https://www.aclweb.org/anthology/D11-1145.pdf

17. Neubig, G., Matsubayashi, Y., Hagiwara, M., & Murakami, K. (2011). Safety Information Mining — What can NLP do in a disaster. *Proceedings of the 5th International Joint Conference on Natural Language Processing*, 965–973. Retrieved from https://www.aclweb.org/anthology/I11-1108.pdf

18. Atefeh, F., & Khreich, W. (2013). A Survey of Techniques for Event Detection in Twitter. *Computational Intelligence*, *31*(1), 132–164. doi: 10.1111/coin.12017

19. Compton, R., Jurgens, D., & Allen, D. (2014). Geotagging one hundred million Twitter accounts with total variation minimization. *2014 IEEE International Conference on Big Data (Big Data)*. doi: 10.1109/bigdata.2014.7004256

20. Bellagio Big Data Workshop Participants. (2014). Big data and positive social change in the developing world: A white paper for practitioners and researchers. *Oxford: Oxford Internet Institute*. Retrieved from https://assets.rockefellerfoundation.org/app/uploads/20140507222100/Big-Data-and-Positive-Social-Change-in-the-Developing-World.pdf

21. Stojanovski, D., Dimitrovski, I., & Madjarov, G. (2014). TweetViz: Twitter Data Visualization. *Conference on Data Mining and Data Warehouses*. Retrieved from https://www.researchgate.net/publication/288427532\_TweetViz\_Twitter\_Data\_Visualization

22. A. Olteanu, S. Vieweg, C. Castillo. 2015. What to Expect When the Unexpected Happens: Social Media Communications Across Crises.  *Proceedings of the ACM 2015 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '15)*. Retrieved from http://crisislex.org/papers/cscw2015\_transversal\_study.pdf

23. A. Olteanu, C. Castillo, F. Diaz, S. Vieweg (2014). CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. *Proceedings of the AAAI Conference on Weblogs and Social Media (ICWSM'14).* Retrieved from https://crisislex.org/icwsm2014\_crisislex.pdf