

Randy Joy Magno Ventayen, 2017

Volume 3 Issue 2, pp. 240-249

Date of Publication: 10<sup>th</sup> October 2017

DOI-<https://dx.doi.org/10.20319/mijst.2017.32.240249>

This paper can be cited as: Ventayen, R. (2017). Multilingual Detection and Mapping of Emergency and Disaster-Related Tweets. *MATTER: International Journal of Science and Technology*, 3(2), 240-249.

This work is licensed under the Creative Commons Attribution-Non Commercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

## **MULTILINGUAL DETECTION AND MAPPING OF EMERGENCY AND DISASTER-RELATED TWEETS**

**Randy Joy M. Ventayen**

*Pangasinan State University, Lingayen, Pangasinan, Philippines*

[dayjx@yahoo.com](mailto:dayjx@yahoo.com)

---

### **Abstract**

*The Philippines is considered as a disaster-prone country in Southeast Asia. Today, social media such as twitter serves as a communication outlet and majority of the post are written in English. This is a problem or gap to those who are not well-versed in a foreign language or cannot even read or understand English. This study promotes the use of local language by translating the keyword using the specified language of the identified region. It will enhance and bridge the gap between the major speaking language from the local areas of the country specifically in the Ilocos Region. The tool will search disaster and emergency-related keywords in local language for extraction. Social network's API and tools will be used for community detection and extraction of data. This shall analyze the properties of the community structure detected from Filipino social media users who posted about the disaster in the local language. This study will determine the geolocation and community structure of the disaster and emergency-related post based on the tweet's coordinates, and analyze community structure formed and compare it to actual patterns of disaster-affected areas. Maps will be utilized as a crowdsource to identify the disaster and emergency-related tweets in Ilocos Region. This will also improve the development and use of the tool through the multilingual Twitter data and in real-time detection of disaster-*

*related tweets so that appropriate action may be done promptly. This paper presents the possibility of the affected community which gives bigger changes of possible projecting the exact location for a reliable report to the government for a faster response.*

## **Keywords**

Disaster-Related Tweets, Multilingual, Social Media, Natural Language Processing

---

## **1. Introduction**

### **1.1 Context of the Study**

Social networking sites such as Twitter is one of the most widely used as a source of news and information, and sometimes it is ahead than other media, because of its information feeds from known and unknown which is sent by users.

Typhoon Karen and Lawin hit the Philippines last October 2016 in the northern part of the Philippines specifically the Ilocos Region, it is one of the strongest typhoons with Category 5 status same as the typhoon Yolanda which hit the country last 2013. The researcher observed that most of the tweets during the during the typhoon are related to disaster and emergency in the local language. Twitter is one of the most used social media platforms in the country and could be utilized to determine the location and to minimize the injury during emergencies and disaster.

There are 9 major languages in the Philippines and 2 of the major language and 1 minor dialect are in Region 1 (Csunedu, 2017). The region is one of the biggest regions in the Philippines and not all constituents can understand and speak English and some have little knowledge in speaking Tagalog despite that they are Filipinos. There are many studies related to detection and analysis of tweets while those study is focusing on Tagalog and English only, from these current studies, the proponent realized that there is a need of equal treatment among those who speak those speaking local language and dialect. Since we are talking about life, the life of those speaking in national language is as important as those speaking in the local language.

Moreover, the Philippines is a country that is very attuned to social media, and it is even named as the Networking Capital of the world (Asiancorrespondentcom, 2017). In reality, some government agency in the country has social media accounts for faster information dissemination. To study the nutritional health of consumers.

### **1.2 Research Objectives**

The Objective of this paper is to determine the proper words from different language and

dialect for the disaster and emergency-related tweets. Second, gather real-time multilingual tweets related to disaster and emergency and lastly create geolocation to identify the location of the emergency or disaster tweets to publish a report for a government agency and to the public for rescue. This paper will answer the how to determine the proper words from different language and dialect for the disaster and emergency keywords? how to gather and extract real-time multilingual tweets? and how to identify the location of multilingual tweets to create a web-based geo location heat map with google maps integration?

### **1.3 Scope and Limitation**

The scope of identification of disaster and emergency location is the four provinces of Region 1. While this study focuses on multilingual tweets initially in Region 1, it also gathers tweets from Tagalog and English language in the same region.

### **1.4 Significance of the Study**

Life is valuable whatever language you speak. This study gives bigger chances of rescue for those who are not speaking the major language when disaster strikes even from the isolated areas of the region. This study also encourages us to understand the relationship in the context of the disaster in the region and the tweets from the region about disaster and emergency. This will also help us to identify and provide a closer look in the region which is disaster-prone areas by using the tweets extracted from users. The proponent aims that this study will also help the organization and government agency such as the NDRRMC in its disaster management plans and expand future researchers in multi-language.

## **2. Related Literature**

There are numerous researchers have used social media as a source of data to understand various disasters, with applications such as situational awareness (Vieweg et al., 2013) and understanding the public sentiment (Syaifudin & Puspitasari, 2017).

In the study (Stowe, et. al, 2016) with the title Identifying and Categorizing Disaster-Related Tweets, the tweets during Hurricane Sandy which impacted New York in 2012 was used. The researcher proposes an annotation scheme for identifying tweets, it uses a system for classifying disaster-related twitter tweets. Categories were used to identify disaster-related tweets such as Sentiment, Action, Preparation, Reporting, Information, and movement. Based on its preliminary result, it shows the relevant information that can be extracted automatically via batch processing after the events, and the researcher is exploring possibilities to extend the approach to

real-time processing.

According to the twitter India blog from @twiterindia , they used Twitter and worked with NGOs and another private sector with the participation of the citizen towards a strategy of disaster relief operations. They realize the usefulness of the social media during disaster relief during the Kashmir floods of 2014 and the work was replicated in 2015 when Chennai was hit with a flood. The outcome a team up and collaboration with NGOs, citizens, government agencies for disaster relief operations.

Another study was conducted in Automatic Classification of Disaster-Related Tweets (Parilla-Ferrer, et. al. 2016) which research about the classification of disaster-related tweets in metro Manila last 2012. The tweets were labeled as information and uninformative to check the reliability of the information posted. A machine learning algorithm was used which is the Naïve Bayes and Support Vector Machine (SVM). Based on the result of the study, SVM has a better result than the other, and it revealed that there are more uninformative tweets than the informative tweets, while the informative tweets were more likely to retweeted thus provide awareness to the public.

A system was tested (Regalado et. al., 2016) entitled FILIET: An Information Extraction System For Filipino Disaster-Related Tweets. The study acknowledges the problem in the extraction of Filipino language, therefore, it creates a system that could extract relevant information from Filipino disaster-related tweets which they call it FILIET: Filipino Information Extraction Tool for Twiter. While the goal of this study is not just about the local Filipino language, the proponent also concerns the location.

Unlike the previous studies, this study will focus on multilingual tweets and identify locations in the region which directly provide where the disaster is happening.

### **3. Methodology**

#### **3.1 Overview of the Process**

Typhoon Karen and Lawin hit the Philippines last October 2016. The typhoon brought weeks of torrential rain which caused flooding, landslide, damages that cause another emergency in several areas. During the disaster and its aftermath, subscribers of Twitter used this social medium to tweet information about the disaster in local languages such as Pangasinan, Iloko, and Bolinas.

The first thing to do is to classify the disaster and emergency-related keywords and

acquire the current translation of keyword for disaster and emergency using a local dictionary and to interview. Tweets about disaster and emergencies were gathered first via Twitter Search API and extract live tweets related to disaster and emergency. Tweets will be filtered from 2 languages and 1 dialect according to the gathered translation data. Maps will be used as a crowdsource to identify the disaster and emergency-related tweets in Ilocos Region. Determine the geolocation and community structure of the disaster and emergency-related post in the local language with. Create an information website for information dissemination and rescue.

### **3.2 Data Acquisition**

Translation for the different disaster and emergency-related tweets will be done manually by research, interview, and local dictionary. The native and elderly are one of the targets to interview since most of them are well-versed in the local language.

### **3.3 Data Processing**

This paper will use Twitter streaming API, and the Google Maps API to create a Node.js web app that displays the provenance of tweets on a map in real time and will deploy in Bluemix. The web application will identify the geographical origin of tweets that mention disaster-related tweets. Node.js development environment will be utilized with the used of Eclipse with the Nodeclipse plugin to develop the app.

To start this, the acquired translated keyword related to disaster and emergency will be used to search specified keyword on Twitter. A twitter streaming API tool will be used to search for a specified keyword, and hashtag phrase in the public Twitter universe whenever a new tweet is posted. When a tweet is posted that matches the provided search criteria, the tweet is passed through the node and identify its coordinates.

After obtaining the coordinates, the location will be mapped using integrated maps with the tool Google Maps API. Maps shall be used as a crowdsourcing to identify the disaster and emergency-related tweets in Ilocos Region which gives bigger changes of possible projecting the exact location (Catanghal et. al, 2017).

The website will also serve as an application to validate the data based on the number of tweets, The number of tweets in the location, will determine the validity of the tweet. Information from the government agencies which areas are disaster and emergency prone areas will help us also to identify the validity of the tweets. The website will be open for government agencies, LGU and to the public.

### 3.4 Development Methodology

The researcher chooses Agile Software Development for its simplicity, it is a methodology for the creative process. Agile software development focuses on keeping the code simple, testing often and delivering functional software as soon they are ready. Its goal is to develop a simple yet reliable system.

## 4. Discussion and Results

### 4.1 Discussion

With the use of interview from the elderly and with the translation app tool, the researcher found the translation from Iloko, Pangasinan, and Bolinao which is the language and dialect spoken in the region.

**Table 4.1:** Shows the disaster-related keyword in 2 languages and 1 dialect

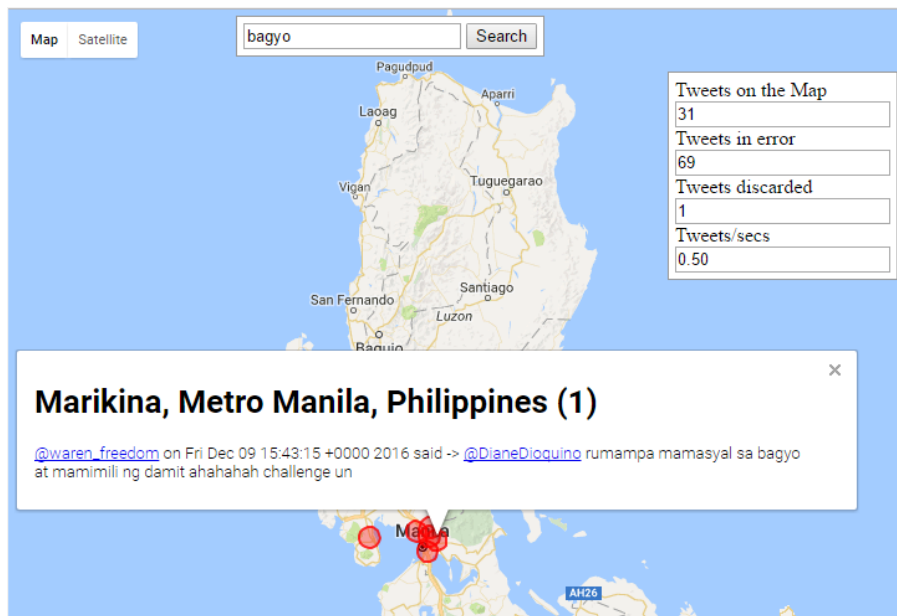
Disaster-Related Keywords	ILOKO	PANGASINAN	BOLINAO
Typhoons	Bagyo	Bagyo	Bagyo
Flood	Layos	Delap	Layos
Earthquakes	Gingined	Yegyeg	Rayon
Volcanoes	Bulkano	Bulkan	Bulkan
Heavy Rain	Tudo	Uran	Rapeg

**Table 4.2:** Shows the emergency-related keyword in 2 languages and 1 dialect

Emergency related keywords	ILOKO	PANGASINAN	BOLINAO
Fire	Apuy	Pool	Apuy
Bridge	Rangtay	Taytay	Taytay
Rescue	Tulong	Tulong	Tulong

The researcher conducted a simple testing tool of keyword searching using TOTEM (Tweets on The Map) app which is deployed in Bluemix, it is a simple UI in which you type a query (as you can do with Twitter) and start seeing matching tweets pop onto the map at their place of origin, announced by a bouncing tweet icon. As a first common filter prior to classification, the app defined a keyword-based query using the Twitter streaming API. The researcher tested the keyword “bagyo” and “delap” a Pangasinan word for typhoon and flood. As discussed in the Methodologies, the app that is deployed in Bluemix, a cloud flat form that can

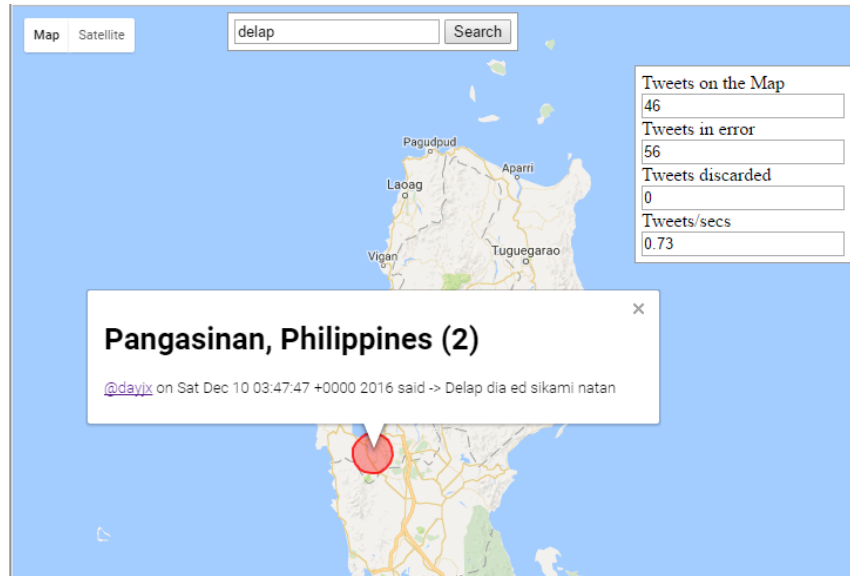
build, run, deploy and manage applications on the cloud. The app will show the location of the tweet and mark with a circle that could increase the size as the number of tweets increases. Clicking the circle will show you the name of the place, the number of related tweets. Upon testing the keyword “bagyo”, the app displays real time tweets related to the word “bagyo”, upon randomly checking the tweets related to bagyo, the researcher found out the content of the tweet is not related to emergency or disaster, instead, the tweet is related to a place.



**Figure 1:** Shows the experiment output of the keyword "bagyo"

The main reason why the keyword “bagyo” doesn’t shows an expected result because it is also a tagalog word, which is the national language of the Philippines, and it is expected to have multiple results in the map of the Philippines. During the current testing, there are 31 tweets on the map but there are no tweets from region 1.





**Figure 2:** Shows the experiment output of the keyword "delap"

On the experiment of the keyword “delap”, it shows a result from Region 1, specifically in Pangasinan, where there is a tweet from a user that there is a flood in the place.

The main objective of our experiment is to identify the location of multilingual tweets for possible rescue. While it is obvious that the result is good, validation of the data should be done using another method such as validating the information from the government agencies if the area is disaster-prone, or there is an ongoing disaster.

**Table 4.3:** Shows the source of verification for disaster and emergency cases

Disaster and Emergency Related Keywords	Source of Verification
Typhoons	PAGASA
Flood	Flood-prone areas
Earthquakes	Philvolcs
Volcanoes	Philvolcs
Heavy Rain	PAGASA
Fire	Local Government Unit
Bridge	Local Government Unit
Rescue	Local Government Unit

## 4.2 Results

The result of the experiment is impressive, using the Twitter API and Google maps API we will retrieve the user’s location which displayed on the map. These will help us to determine the possibility of the affected community which gives bigger changes of possible projecting the



exact location for faster response.

## 5. Conclusion and Recommendation

The extraction of location from multilingual disaster and emergency-related tweets is important is interesting study because the life of a person which speaks a very rare dialect is important as the same as the person speaking a major language. In the future, further study should be conducted not limiting in Region 1 but also in other language and dialect. Further study in the investigation on how to extract location information that is hidden in hashtags and from all other languages not just in the region could be an important study to determine. While this study is still ongoing, the tool for mapping tweets will improve the development will continue.

## References

- Asiancorrespondentcom. (2017). Asian Correspondent. Retrieved 25 September, 2017, from <http://asiancorrespondent.com/54475/philippines-named-these-social-networking-capital-of-the-world-indonesia-malaysiaamongst-top-10/>
- Catanghal Jr, R. A., Palaoag, T. D., & Malicdem, A. R. (2017). Crowdsourcing Approach For Disaster Response Assessment. *MATTER: International Journal of Science and Technology*, 3(1). <https://dx.doi.org/10.20319/Mijst.2017.31.5969>
- Csunedu. (2017). Csunedu. Retrieved 25 September, 2017, from <http://www.csun.edu/~lan56728/majorlanguages.htm>
- Parilla-Ferrer, B. E., Fernandez Jr, P. L., & Ballena IV, J. T. (2014, December). Automatic Classification of Disaster-Related Tweets. In Proc. International conference on Innovative Engineering Technologies (ICIET) (p. 62).
- Regalado, R. V. J., Kyle Mc Hale, B., Garcia, J. P. F., Kalaw, K. M. D. F., & Lu, V. E. (2015). FILIET: An Information Extraction System For Filipino Disaster-Related Tweets.
- Stowe, K., Paul, M. J., Palmer, M., Palen, L., & Anderson, K. (2016). Identifying and Categorizing Disaster-Related Tweets. In Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media (pp. 1-6). <https://doi.org/10.18653/v1/W16-6201>
- Syaifudin, Y. W., & Puspitasari, D. (2017). Twitter Data Mining for Sentiment Analysis on Peoples Feedback Against Government Public Policy. *MATTER: International Journal of*

Science and Technology, 3(1). <https://dx.doi.org/10.20319/Mijst.2017.31.110122>

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010, April). Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1079-1088). ACM. <https://doi.org/10.1145/1753326.1753486>