Geo-parsing Locations during Natural Disasters for Emergency Intervention Management: A Case Study of Citizen-Led Social Media Posting of the Flood Incidence in Nigeria

Rashidat Sikiru, Olubayo Adekanmbi (PhD), Anthony Soronnadi

Data Scientist Network, Lagos, Nigeria

Abstract

Urban flooding in recent times has resulted in the loss of human lives; the destruction of economic and social infrastructure such as water supply, electricity, roads, and every other means of livelihood. It is estimated that about 23% of the world is directly exposed to flood in a once in a hundred-year flood event thereby increasing the risks to lives and livelihoods. However, the first piece of information that gets broadcasted when a disaster such as flooding happens is, where it does affect. Social media platforms and print media have emerged as significant resources for sharing the study utilize the Spacy model to extract location-specific words from online article gotten from the web. In this research, we proposed a three-step process: data collection, text classification, location extraction, and geocoding and visualization of location on the map. By accurately identifying flooded locations, this method can be adopted by the emergency response teams to be able to prioritize resources and assist people in need more effectively during floods.

1.0 Introduction

Floods are the most widespread, disastrous, and frequently happening natural disasters in the world. Urban flooding in recent times has resulted in the loss of human lives; the destruction of economic and social infrastructure such as water supply, electricity, roads, and every other means of livelihood [1]. It was estimated that 1.81 billion people, which is about 23% of the world's population, are directly exposed to flood depths greater than 0.15 meters in a 1-in-100-year flood event, thereby increasing risk to lives and livelihoods. Of all these estimates, 89% live in low-and middle-income countries.[1]

Flooding in developing countries occurred as a result of changes in climate, excess precipitation, building on waterways, rise in sea level, operation in dams along the borders, rapid population, poor management, and lack of political will [2]. The flooding that happened recently in Nigeria between June and October 2022 caused the death of over 600 people and displaced 1.3 million from their various homes.[2]

However, one of the first pieces of information that gets broadcasted when a disaster such as flooding happens is, where it does affect. Afterward, Further details such as, where exactly is affected? Which suburb, street, building, or area? regarding this announcement is given. Therefore, extracting location information during a disaster (such as flooding) is important for keeping people and authorities informed [3].

Geoparsing refers to a process of natural language processing (NLP) used to identify geographical entities and to encode them in coordinates [4]. This simply refers to, analyzing a language to obtain the name of places and to visualize their locations on maps [4]. Geoparsing has been discussed in numerous studies [7]. Two major approaches have been identified:(a) toponym recognition: Toponym recognition can be considered as a subset of a more general problem studied in natural language processing, called named-entity recognition (NER). Where toponym recognition involves finding entities in the text that correspond to geographic location names, named-entity recognition involves finding locations, as well as entities of other types like names of people and organizations. [10] and (b) toponym resolution is the task of disambiguating toponyms in natural

[1] https://blogs.worldbank.org/climatechange/
[2] https://www.weforum.org/agenda/2022/10/nigeria-flood-rain-climate

language contexts to geographic locations, which plays an essential role in automated geographic indexing and information retrieval [11].

Several research has been carried out on geoparsing of text scraped from Twitter to determine the extent or location of disasters, For instance, [4] concentrated on identifying references to the location in a Twitter text, with further work on toponym resolution, (Yandong et al 2015) [12] analyzed the Sina Weibo (a Chinese microblogging platform) during and after the 2012 Beijing Rainstorm to carry out a topic-based classification of the social media text streams leaving for further research the relationship between the trend of topics under discussion and the stage of emergency events. Also, several research papers have been written to explore the impact of flooding in Nigeria [2], analyses of flood risk [13], and analyses of media articles to enhance the understanding of contemporary environmental challenges [14]. To enhance quick response and solution to flooded areas, this study will therefore;

a. identify references to the locations mentioned during a flood in the social media platforms(Twitter) and/ or well as new websites, that is, Toponym Recognition;
b. assign geographic coordinates as latitude and longitude to the identified locations on the map, that is Toponym resolution

2.0 Methodology

The geoparsing pipeline enables the identification and visualization of the flooded location mentioned in tweets from Twitter and the Newspaper from an online news website. The pipeline includes three main steps (figure 1)

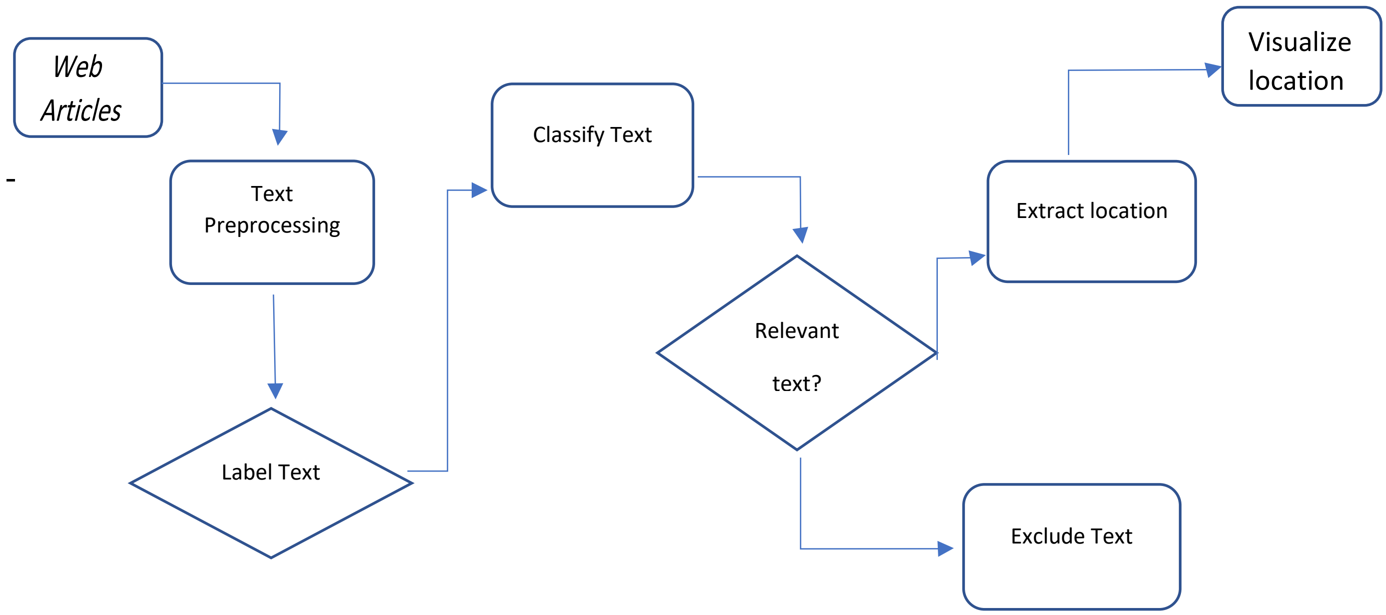

Figure 1: Research work methodology flowchart

2.1 Data Collection

The dataset contains about 978 articles collected from an online newspaper while searching with the keyword "flood". The Python library used was Beautiful Soup and the Request library. This dataset contains flood issues around the globe.

2.2 Text Preprocessing

The datasets crawled were cleaned to remove unwanted characters, emoticons, etc. using regular expressions. Then each article was tokenized into sentences for easy accessibility and the total number added up to about 21,683.

2.3 Text Labelling

A rules-based method of classification was used after manually checking each sentence to look for keywords that best described the situation of floods in each sentence. Keywords such as submerged, flooded, destroyed, killed, lost, losses, submerged, etc. These were then used to label the text as affected or not affected. These datasets were filtered to retain only text that discussed the effect of flooding about 7,000 datasets

2.4 Location Extraction

Using the method proposed by [5,6], location identification was carried out by utilizing features from Part-of-Speech tagging and dependency Parsing in Spacy. This was used to extract locations identified as geo-political entities, Local entities, and some locations identified as organizations. For some sentences, some had more than one location, to make this more easily geocoded, we then applied the explode () function to split this location into different rows. Afterward, the rows without locations were dropped and the locations that are not within Nigeria were dropped.

2.5 Geocoding/Geoparsing

The method of geocoding using the Geopy library. Geopy is a Python client for several popular geocoding web services. Python developers can easily locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources. Each geolocation service you might use, such as Google Maps, Bing Maps, or Nominatim, has its own class in Geopy abstracting the service's API. This paper used Google Maps using the class GoogleV3 to pick the coordinates of locations within Nigeria.[3]

[3] https://geopy.readthedocs.io/en/stable/

3.0 Results and Discussion

3.1 Location extraction

Using POS (Part of Speech) tagging and dependency parsing with the Spacy model to identify the location, our rule-based model was able to identify locations for about 300 sentences split from about 70 complete articles into sentences spanning between 2019 and May 2023. The Spacy model was able to extract about 75% of the location correctly., while other locations were omitted and not recognized. Most of the articles contain more than one location, in other to split them into separate rows, we used the explode functions. See the table below.

Date	Sentence	Locations Identified
2019-11-15	Matthew Ochei, Asaba Delta State rice farmers have decried the loss of over 300 hectares of rice farms to a ravaging flood at Omeligboma, Oko in Oshimili South Local Government Area of the state	Oshimili South Local Government Area
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Uzanu
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Udaba
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Ofukpo
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Osomogbe

Table 1: Table showing locations extracted from the text

3.2 Geoparsing Locations Extracted

For the locations extracted, we used the Geopy library for geocoding, the GoogleV3 class was able to return the coordinate of about 80% correctly while misplacing the coordinate of some locations for another. For example, in the table below, Ofukpo in the Edo state was misplaced with Otukpo in the Benue state. See table below

Date	**Sentence**	**Locations Identified**	**Address**	**Longitude**	**Latitude**
s2019-11-15	Farmers have decried the loss of over 300 hectares of rice farms to a ravaging flood at Omeligboma, Oko in Oshimili South Local Government Area of the state	Oshimili South Local Government Area	Oshimili South, Delta, Nigeria	6.0698466	6.62106109999999
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Uzanu	312107, Uzanu, Edo, Nigeria	7.1943021	6.6352313
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Udaba	312111, Udaba, Edo, Nigeria	6.85258689999999	6.593355
2019-11-07	The communities affected are: Anegbette, Udaba, Osomegbe, Ugochi, Ofukpo, Agbaburu, Uzanu and Ifeko	Ofukpo	Otukpo, Benue, Nigeria	7.1982133	8.1393179

Table 2: Table showing coordinates of extracted locations

3.3 Visualization of Geocoded Locations

After geocoding each location, we use the folium library to visualize our location on the map for easy identification to non-technical persons who do not have access to our code and cannot understand the technical aspect of the research. Below are the images gotten between the period of 2019 and 2023 (figure 1-2). From the visualization, we can see that we were able to identify the locations as well as even mentioned in the article.

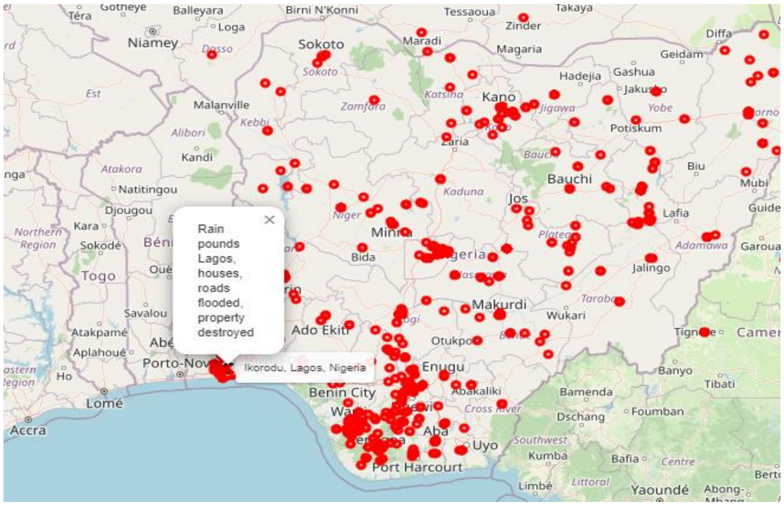

Figure 2: Image showing locations affected by flood in 2019

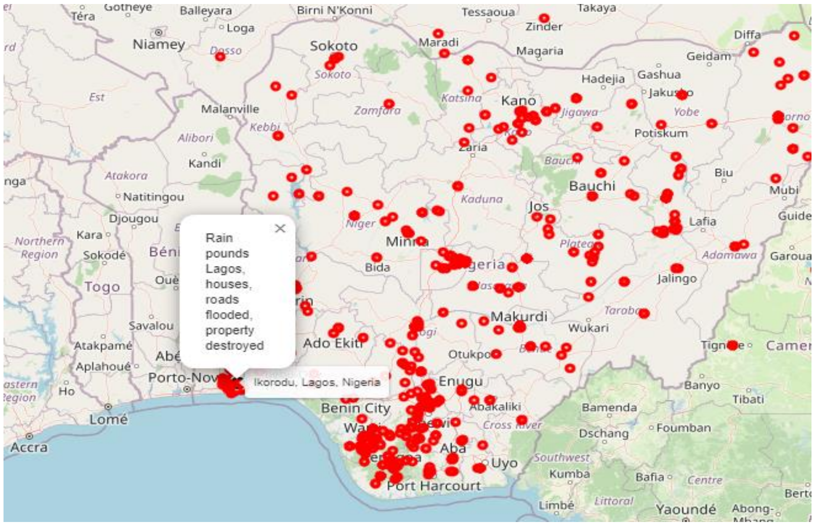

Figure 2: Image showing locations affected by flood in 2022

The visualization shows the majority of the locations mentioned in the text. This can further be developed into a web application that shows a real-time visualization of such locations for emergency response.

4.0 Conclusion

This research proposed a process of visualizing locations using the Spacy small models for location extraction and using the Geopy library to identify the coordinate of the extracted locations. The coordinates were then visualized on the map using the Folium library. The identification of these locations can be further used by emergency responders, government, and non-governmental organizations to allocate resources carefully and the rescue team can as well save lives as soon as possible.

However, this result can be further developed into a web application that can be used for real-time purposes. In the course of this research, our access to Twitter to extract data was a bit challenging and we resulted to using articles from online newspapers. Also just like [6] mentioned the use of a trie form gazetteer, we couldn’t use that as a result of the unavailability of a gazetteer peculiar to Nigeria. However, future research can be used to gather names of locations in a gazetteer which may be used to improve the model to robustly identify locations within the region. Also, research can also look into geoparsing locations from the local languages like Yoruba, Igbo, Hausa, and Pidgin.

References

1 Agbonkhese et al. (2014) Flood Menace in Nigeria: Impacts, Remedial and Management Strategies. Civil and environmental research ISSN 2224-5790(Paper) ISSN 2225-0514 (Online) Vol. 6, No 4, 2014

2 Nura Umar and Alison Gray. Flooding in Nigeria: a review of its occurrence and impacts and approaches to modeling flood data. International Journal of Environmental Studies. https://doi.org/10.1080/00207233.2022.2081471

3 Lingad, J., Karimi, S., & Yin, J. (2013). Location extraction from disaster-related microblogs. Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion. doi:10.1145/2487788.2488108

4 Edwin et al. (2020). Adaptive Geoparsing Method for Toponym Recognition and Resolution in Unstructured Text. Advances in Remote Sensing and Geographic Information Science and Their Uses in Geointelligence 12(18), 3041; https://doi.org/10.3390/rs12183041

5 Gelernter, J., & Balaji, S. (2013). An algorithm for local geoparsing of microtext. GeoInformatica, 17(4), 635–667. doi:10.1007/s10707-012-0173-8

6 Dhavase, N., & Bagade, A. M. (2014). Location identification for crime & disaster events by geoparsing Twitter. International Conference for Convergence for Technology-2014. doi:10.1109/i2ct.2014.7092336

7 de Bruijn, J.A., de Moel, H., Jongman, B. et al. TAGGS: Grouping Tweets to Improve Global Geoparsing for Disaster Response. J geovis spat anal 2, 2 (2018). https://doi.org/10.1007/s41651-017-0010-6

8 Fohringer, J., Dransch, D., Kreibich, H., and Schröter, K.: Social media as an information source for rapid flood inundation mapping, Nat. Hazards Earth Syst. Sci., 15, 2725–2738, https://doi.org/10.5194/nhess-15-2725-2015, 2015.

9 Ghahremanlou, L., Sherchan, W., & Thom, J. A. (2014). Geotagging Twitter Messages in Crisis Management. The Computer Journal, 58(9), 1937–1954. doi:10.1093/comjnl/bxu034

10 Lieberman, M. D., & Samet, H. (2012). Adaptive context features for toponym resolution in streaming news. Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR ’12. doi:10.1145/2348283.2348381

11 Lieberman, M. D., & Samet, H. (2011). Multifaceted Toponym recognition for streaming news. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information - SIGIR ’11. doi:10.1145/2009916.201002

12 Yandong Wang et al. (2015). Using Social Media for Emergency Response and Urban Sustainability: A Case Study of the 2012 Beijing Rainstorm. Sustainability 2016,8, 25; doi:10.3390/su8010025

13 Akinola Adesuji Komolafe et al. / American Journal of Environmental Sciences 2015, 11 (3): 157.166. DOI: 10.3844/ajessp.2015.157.166

14 Olalekan Adekola and Jessica Lamond. (2018). 6 - Systems thinking toward climate resilience. Investing in Disaster Risk Reduction for Resilience Design, Methods, and Knowledge in the Face of Climate Change 2022, Pages 141-162