Geosemantic Entity Recognition and Mapping of Electoral Violence Hotspots from News Reports

Flora O.
Research and Innovation Department
Data Science Nigeria
Lagos, Nigeria
flora@datasciencenigeria.ai

Olubayo A.
Data Science Nigeria
Lagos, Nigeria
olubayo@datasciencenigeria.ai

Anthony S.
Data Science Nigeria
Lagos, Nigeria
anthony@datasciencenigeria.ai

Abstract

In Nigeria, electoral activities are marred by violence, contrary to democratic ideals. To address this, a Geosemantics approach, employing NLP; Name Entity Recognition models, is utilized. This method allows the extraction of toponym entities from unstructured data, such as news reports using a novel Multilingual Based NER model WikiNeural, identifying geographical hotspots location of electoral violence. By streamlining information extraction, this offers a comprehensive dataset for analysis of violence depth across location offering a way to mitigate these attacks in subsequent elections. This approach not only aids in understanding and mitigating electoral violence but can also be applied to diverse tasks requiring geosemantic entity extraction from textual data.

1 Introduction

Geosemantics is a field that involves the integration of Geospatial entities and Semantics information to enhance location-based data understanding. This field utilizes Natural Language Processing (NLP) and Geospatial techniques for comprehensive analysis. Geosemantics is applicable in addressing societal issues like Electoral violence a prominent problem faced in politics. Electoral violence is usually prevalent in Nigeria as the electoral tension builds up so therefore tracking hotspot location where these violence are known to occur is imperative in taking steps towards mitigating this issue. By linking toponyms with semantic annotations, it optimizes creation of unstructured data into a structured robust dataset. By extracting entities with Name Entity Recognition model and geoparsing unstructured toponym data it enables effective visualization on maps.

2 Related Works

Similar approach has taken place one of which is Geo-Semantic-Parsing a geo-tagging technique used to collect geographic data from free text and extracts the corresponding geographic coordinates Nizzoli et al. (2020). In 2018 a geoparsing algorithm was developed using an OpenStreetMap database, and a geotagging algorithm Middleton et al. (2018). A geoparsing algorithm was developed called TAGGS that extracts and can locate geographical locations referenced in a tweet's text Bruijn et al. (2018). Study was carried out to know how well off-the-shelf software identifies locations,

and what is needed to improve automated location identification, called Geo-parsing Gelernter & Mushegian (2011). A study presented a way to precisely identify geospatial relationships linked to spatial entities, ultimately improving outbreak monitoring and disease surveillance from textual data Syed et al. (2023). Addressing the challenge of maintaining synchronization between a Points of Interest (POI) database and real-world changes jointly extracting POI mentions and identifying their associated accessibility labels from unstructured text Sun et al. (2021).

3 Methodology

This study uses data from Arise News media website based on news reporting on Electoral violence surrounding the 2023 election period in Nigeria. This research methodology uses Natural Language Processing (NLP) and Geospatial analysis techniques as outlined below:

3.1 Data Collection

The study obtained data from Arise News website through web crawling method utilizing the requests library and BeautifulSoup for accessing and parsing the HTML (Hypertext Markup Language) content. The keyword employed for the news article retrieval was "Nigeria Election 2023 Violence." The data collected captured the pre-election, during the election and post-election phases of the Nigeria's electoral violence.

3.2 Entity Recognition and Information Retrieval

A Pre-trained Multilingual Named Entity Recognition model WikiNeural that utilizes in a novel way BERT's multilingual power in a silver data creation process for NER Tedeschi et al. (2021). This was used for the extraction of the location entity. To create a robust dataset other entities that gives more details about the violence that occurred in those locations was retrieved entities like the violence type and political parties affiliated to the electoral violence using SpaCy Dependency Parsing a model that analyzes relationships between words in sentence. Using the SpaCy `en_core_web_sm` model to extract information like the violence type that took place during the election, utilized a rule based method to analyze patterns focusing on verbs like "attacks", "intimidation" and its relationship with nominal subjects (perpetrators) and direct objects (victims) to identify and extract the information. Combining the WikiNeural and SpaCy method serves as a validation in filtering out locations that are not related to electoral violence that the NER model may have retrieved.

3.3 Toponym Geoparsing

The extracted location contained locations at State, Local Government Area (LGA) and street level. A Nigerian admin boundary dataset[1] was used to filter out the street level locations leaving the locations at Street and LGA level.

Toponyms Resolution: The extracted locations were geoparsed to get their longitude and latitude using Geonames API which was able to accurately get the coordinates down to LGA and State level.

3.4 Exploratory Data Analysis

The resulting dataset contained Date, URL, Articles, Locations, Violence type and Political Party. The analysis for the hotspot location of electoral violence was carried out by spatially joining the locations with Nigerian admin boundary data which contains the State and LGA geometry points. Further analysis was carried out to show the trends of this attacks particularly to show its build up around election time while also giving insight on the depth of violence that occurred and the political party affiliated to these reports.

4 Results

The outcome of the Geospatial analysis revealed the focal points of Electoral violence during the election period. The map emphasizes Lagos state as having the highest incidence of electoral violence

---

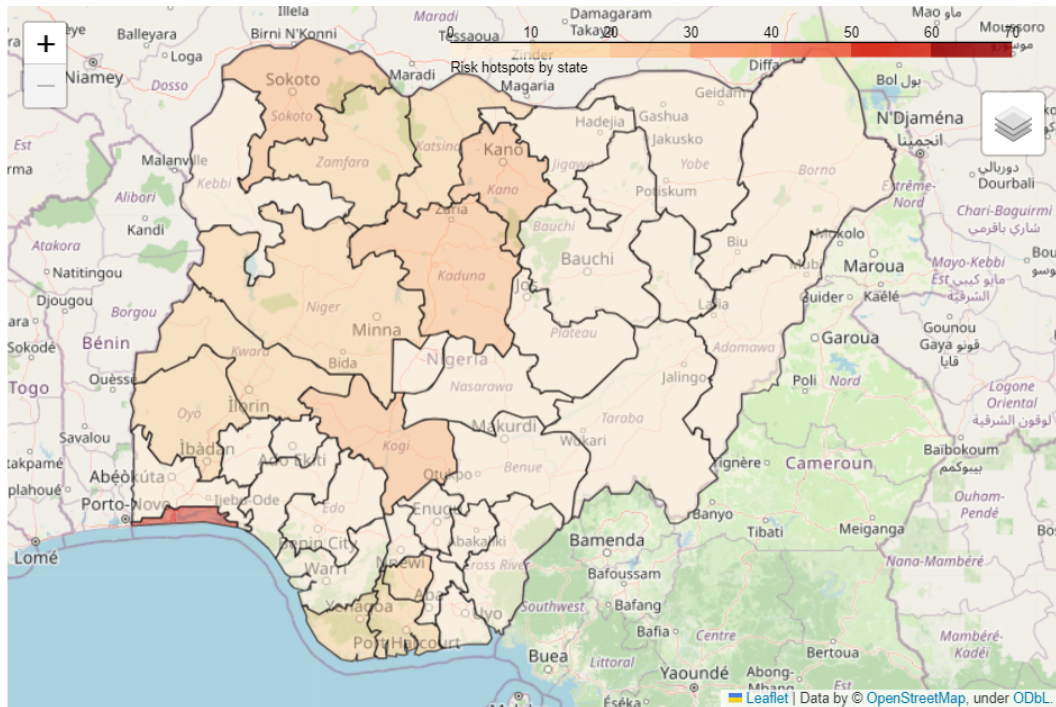[1] `Nigeria-AdministrativeBoundaries|DataCatalog(worldbank.org)`

Figure 1: Hotspot Location of Electoral Violence in Nigeria

in Nigeria. This may be influenced by various factors, including Lagos being one of the most densely populated states in Nigeria, characterized by a diverse population with conflicting preferences among different ethnic groups. Also, Lagos is often seen as a key region for political power and influence in Nigeria.

The outcome of the other analysis carried out also indicates the temporal pattern of electoral violence, with the peak occurring in February 2023, aligning with the period of the last Presidential Election in Nigeria. The period spanning from February to April 2023, encompassing the Presidential, Governorship, and other elections, witnessed elevated levels of electoral violence.

Additional analyses reveals the varieties of violence across these locations and the political affiliations.

5   Conclusion

The aim of this project was achieved in using NLP techniques as a means of location retrieval from unstructured data generating a robust dataset for electoral violence in Nigeria. This approach is useful and can be replicated to mitigate future electoral by manning these hotspot locations with adequate security as election tension builds up. Future enhancements to this project involve constructing a model capable of identifying locations entities not only from textual data but also from audio, video, and other unstructured data forms across different languages.

References

De Bruijn, H. de Moel, B. Jongman, J. Wagemaker, and CJH Aerts. Taggs: Grouping tweets to improve global geoparsing for disaster response. *Journal of Geovisualization and Spatial Analysis*, pp. 1–14, 2018. doi: 10.1007/s41651-017-0010-6.

Gelernter and N. Mushegian. Geo-parsing messages from microtext. *Transactions in GIS*, 6, 2011. doi: 10.1111/j.1467-9671.2011.01294.x.

Middleton, G. Kordopatis-Zilos, S. Papadopoulos, and Y. Kompatsiaris. Location extraction from social media: Geoparsing, location disambiguation, and geotagging. *ACM Transactions on Information Systems*, pp. 1–27, 2018. doi: 10.1145/3202662.

Nizzoli, M. Avvenuti, M. Tesconi, and S. Cresci. Geo-semantic-parsing: Ai-powered geoparsing by traversing semantic knowledge graphs. *Decision Support Systems*, 136, 2020. doi: 10.1016/j.dss.2020.113346.

Sun, J. Huang, C. Yuan, M. Fun, H. Wang, M. Liu, and B. Qin. Gedit: geographic-enhanced and dependency-guided tagging for joint poi and accessibility extraction at baidu maps. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pp. 4135–4144. ACM, 2021. doi: 10.1145/3459637.3481924.

Syed, E. Arsevska, M. Roche, and M. Teisseire. Geospatre: extraction and geocoding of spatial relation entities in textual documents. *Cartography and Geographic Information Science*, 2023. doi: 10.1080/15230406.2023.2264753.

Simone Tedeschi, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pp. 2521–2533, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. URL https://aclanthology.org/2021.findings-emnlp.215.