Text-Based Location Extraction

PROBLEM STATEMENT

To Develop a Text-based Geolocation Extraction system to extract the geographical location or origin of given text content. The goal is to associate textual data, such as social media posts or news articles, with specific geographic locations, ultimately facilitating applications in content recommendation and regional linguistic analysis.

ASSUMPTIONS

1. **Text-Location Relationship:** The model assumes a discernible relationship between the textual content and the associated geographical location. It relies on patterns and linguistic cues within the text to make accurate predictions.

2. **Availability of Labeled Data:** The model assumes access to a labeled dataset containing text samples with corresponding accurate geographical locations. The quality and diversity of this dataset significantly impact the model's performance.

3. **Relevance of Textual Features:** The model assumes that relevant features for geolocation extraction are present in the text. It may not perform well on extremely short or ambiguous texts lacking sufficient context.

4. **Stationarity of Language Patterns:** The model assumes a certain level of stability in language patterns across different time periods. Drastic linguistic changes over time may affect its accuracy.

APPROACH

In our approach, we leveraged textual data enriched with annotations to train a **Named Entity Recognition (NER) model** using **SpaCy**, a powerful natural language processing library in Python. Our primary objective was to extract location information from the text. Through meticulous annotation of the textual corpus, we provided the model with labeled examples of locations, allowing it to learn patterns and characteristics indicative of geographic entities.

Then, we utilized **Folium** to mark these **locations on a map**, providing a visual representation of geographic entities extracted from the text.

MODELS /ALGORITHMS USED

We developed a Named-Entity Recognition (NER) model by harnessing the power of spaCy's NER model and pipeline. Our training process involved leveraging 319 meticulously curated text-annotation pairs to ensure the model's proficiency and accuracy.

For Visualization, we used the Python Folium Library which offers Open-Source Map facilities and location Marking, using OpenStream Map API.
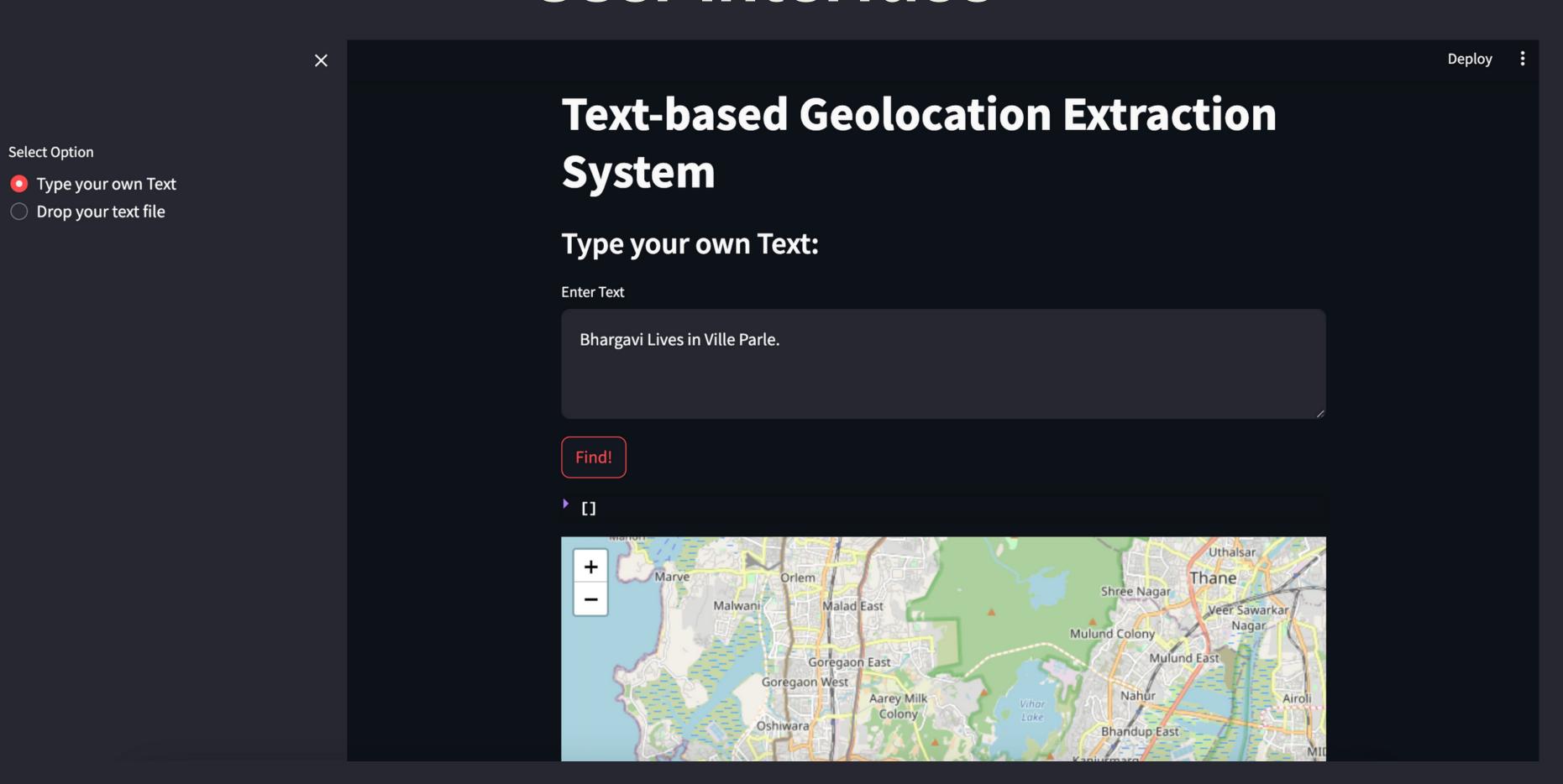
CONCLUSION

In conclusion, our project effectively demonstrated the synergy between **advanced natural language processing** techniques and **interactive mapping technologies.**

This combined approach not only enhanced our understanding of the spatial distribution of information but also showcased the potential for leveraging NLP and mapping technologies in various applications, such as **geospatial analysis**, **information retrieval**, and **data visualization**. Moving forward, our project serves as a testament to the power of interdisciplinary collaboration in unlocking insights from complex datasets and advancing research in both natural language processing and geographic information systems.

User Interface

User Interface

Text-based Geolocation Extraction System

Select Option
- ◯ Type your own Text
- ⬤ Drop your text file

Drop your text file:

Upload a .txt file

☁⬆ **Drag and drop file here**
Limit 200MB per file • TXT, DOC

Browse files

📄 **0.txt** 4.6KB ✕

Uploaded Text Content:

Find!

Content

TITLE: Krewe du Vieux 2018: Take a virtual tour of the route with King Richard Campanella LINK:
http://www.nola.com/mardi_gras_nola/2018/01/take_a_virtual_tour_of_the_kre.html
This area formed the heart of the plantation of Bernard Xavier Philippe de Marigny de Mandeville,
who lived in a mansion where the electrical substation now stands. Expecting that the Louisiana

▶ []

THANK YOU

Alan George
Bhargavi Vaidya
Srijan Upadhyay