FACULTY OF ENGINEERING AND ARCHITECTURE

Covid-19 Sentiment Analysis

Big Data Science – Course Project

Dilawar ALI & Kenzo MILLEVILLE

GHENT UNIVERSITY

imec embracing a better life

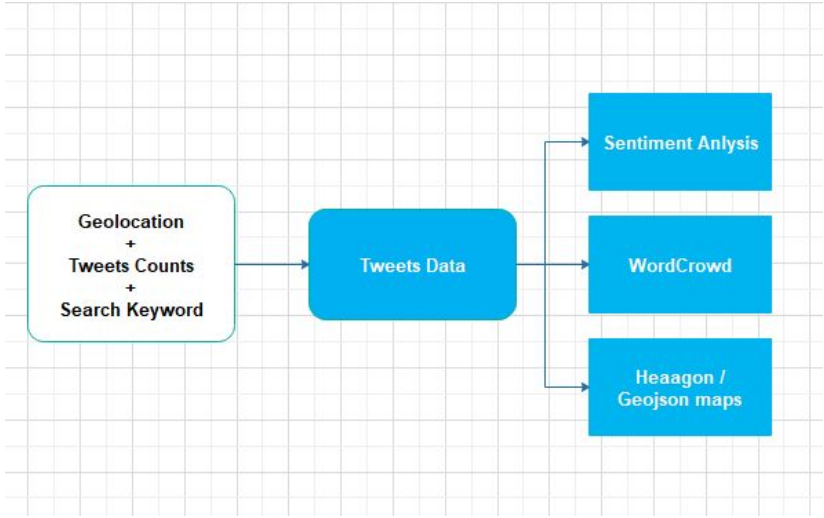
IDLab INTERNET & DATA LAB

Project goals

- The goal of this project is to provide an attractive dashboard for visualization of the coronavirus pandemic using geolocated tweets. These enable a spatiotemporal analysis of the global sentiment and its evolution as the crisis progresses.
- Aim:
  - Gather and analyze recent tweets about the covid pandemic.
  - Dynamic visualization of the data
  - Sentiment analysis (Positive / Negative / Neutral)
  - World cloud representations (Analysis of words most frequently used in particular region while tweeting)
  - Aggregation of data by country
- Link: https://tw06v072.ugent.be/wordcrowd/covid/

Overview

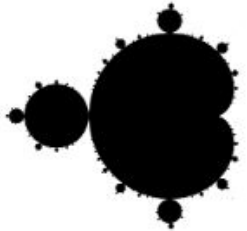
- **Data collection & processing (ETL pipeline)**
  - Twitter API: scraping tweets which have #corona, #coronavirus, covid19, etc
  - Public dataset: “Coronavirus (COVID-19) Geo-tagged Tweets Dataset” [1]
  - Preprocessing & sentiment analysis, dynamic wordcloud generation, Hexagonal maps
- **Infrastructure**
  - Python Django backend server + reactJS/deck.gl frontend
  - Scalable and easy-to-use API
  - The data can easily be updated regularly (e.g. once a day)
- **User Interface**
  - Dynamic dashboard with filter functionality
  - Clustering of tweets on location and country
  - Word cloud for most frequent word presentations



[1] Rabindra Lamsal, "Coronavirus (COVID-19) Geo-tagged Tweets Dataset", IEEE Dataport, 2020. [Online]. Available: http://dx.doi.org/10.21227/fpsb-jz61. Accessed: May. 26, 2020.

Sentiment Analysis

- Preprocessing: Remove links, stop words, etc.
- Libraries we have tried:
  - Flair
  - Textblob
- Textblob performed better and was much faster

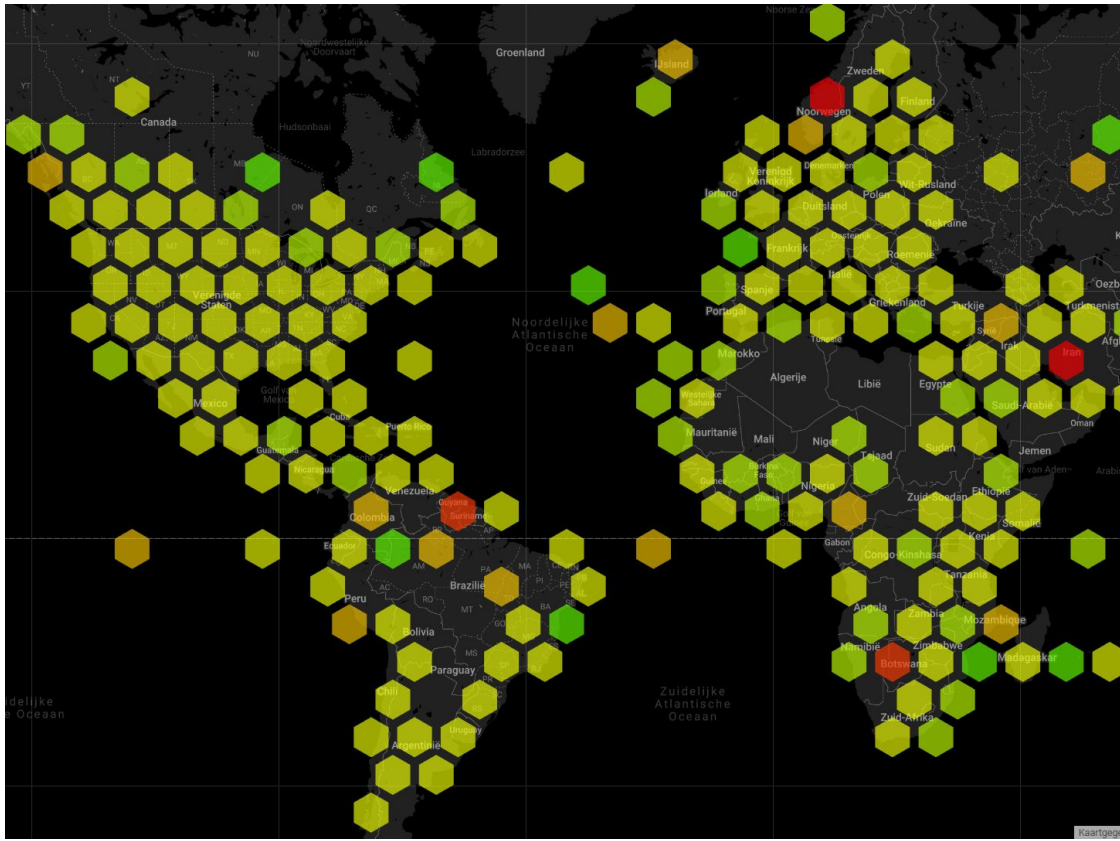


Word clouds

- Remove stop words and perform lemmatization with NLTK library
- Attractive visualization of most frequent words
- Are generated dynamically after optionally filtering the data

Hexagon map

- Scalable and intuitive visualisation of all data
- Can easily be customized
- Clustering is done based on the tweet location



GeoJson map

- Aggregation per country
- Can easily be customized

Conclusions

- We developed a dashboard for sentiment analysis of twitter data for current corona virus epidemic.
- Word clouds used for prompting the most frequent terms.
- Hexagon and Geojson maps for region or per country impact analysis

FACULTY OF ENGINEERING AND ARCHITECTURE

Questions??

dilawar.ali@ugent.be
kenzo.milleville@ugent.be

GHENT UNIVERSITY

imec embracing a better life

IDLab INTERNET & DATA LAB