Covid-19 Sentiment Analysis

Big Data Science – Course Project

Kenzo MILLEVILLE & Dilawar ALI

GHENT
UNIVERSITY

imec
embracing a better life

IDLab
INTERNET & DATA LAB

Project goals

▷ Visualize the spatiotemporal global sentiment of the Covid-19 pandemic in a web application

▷ Gather tweets mentioning the pandemic, preprocess them and perform sentiment analysis

▷ Geocode tweets which are not geolocated, aggregate them on location and country

▷ Visualize results in an intuitive and scalable web application, by using WebGL-powered maps and word clouds

GHENT
UNIVERSITY

Application architecture

▷ Data collection from open dataset and Twitter API

▷ Preprocessed via NLTK, sentiment analysis via TextBlob

▷ Data is stored with Django in an SQL database, with spatial operations via the GEOS interface



NLTK
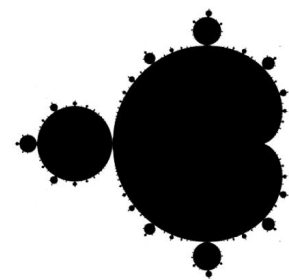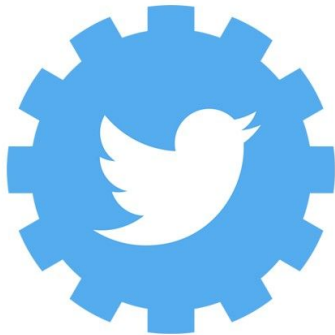
TextBlob

For NLP Preprocessing

GHENT UNIVERSITY

Application architecture

▷ Public API exposes data manipulations to frontend web application or third-party apps

▷ Frontend is made with ReactJs and deck.gl, to provide a scalable and dynamic interface

▷ Data can be filtered and analysed in real time
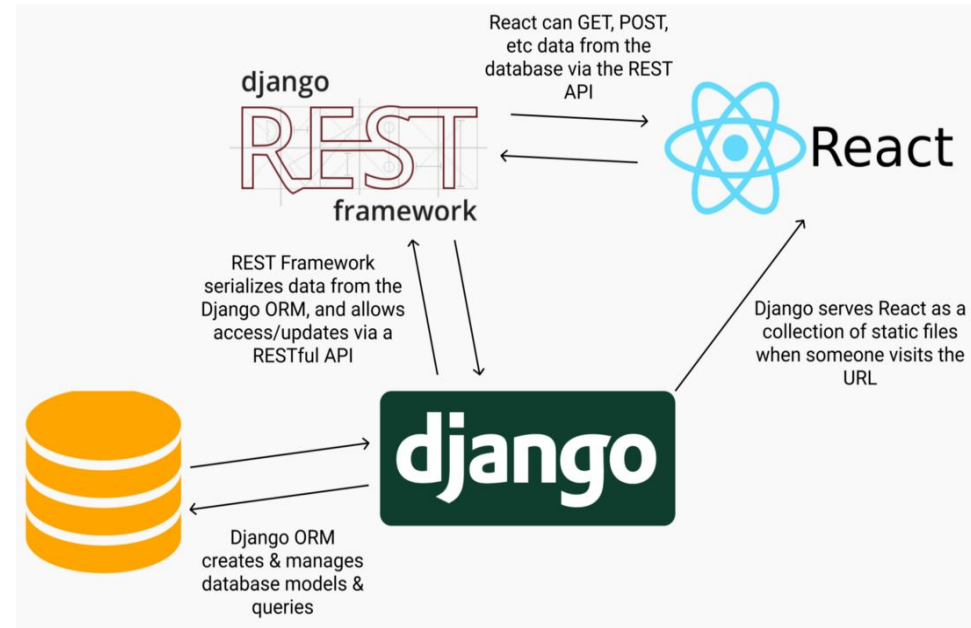
Map visualizations

▷ On application load, a request is made to get all of the tweets (without the text) and they are clustered in a hex map.

▷ When hovering over the map, information about tweets in that area is shown

Amount of Tweets: 546
Polarity: 0.12
Sentiment: Relatively Positive
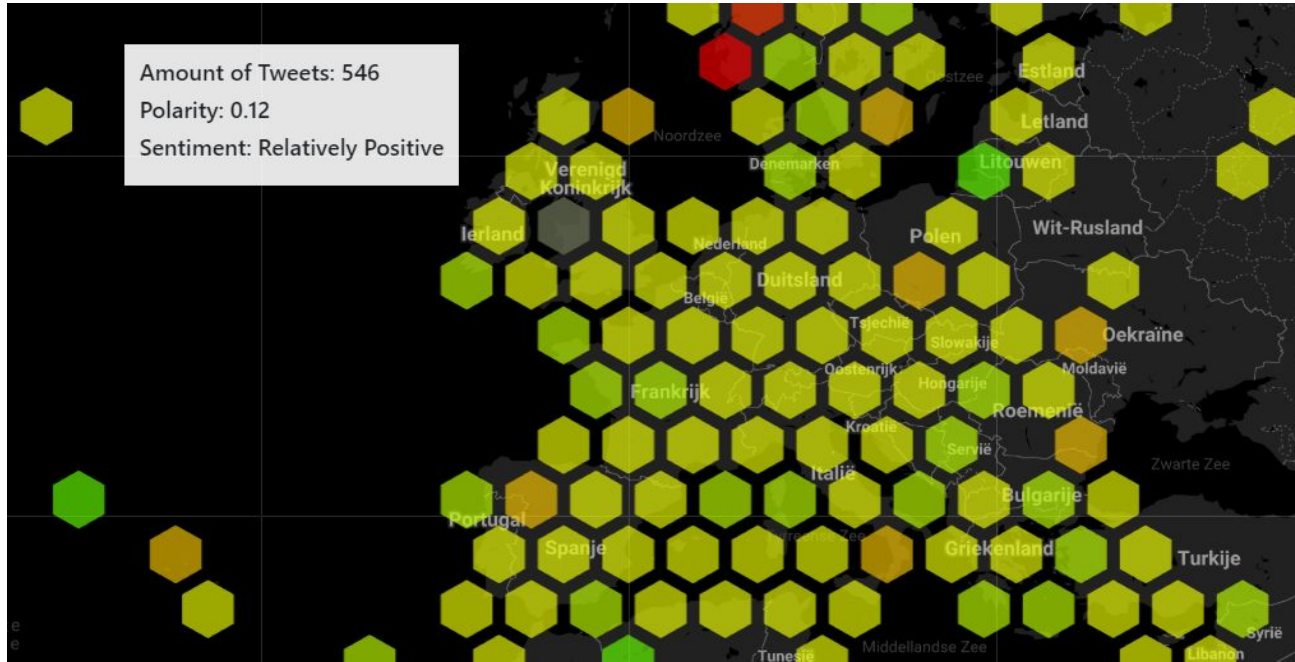
Map visualizations

▷ An alternative visualization aggregates the data per country, for some countries we have chosen and gathered data for (gathered tweets in a radius around major cities of each country)

▷ By using geojson data of the countries, we can visualize them as well
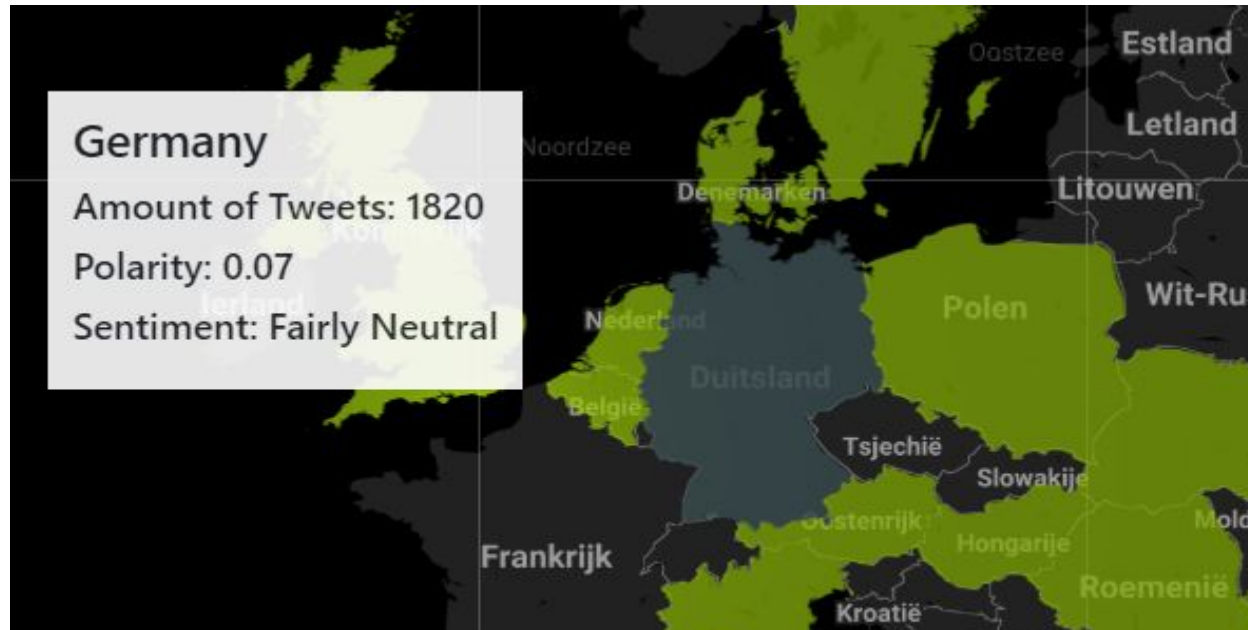
Interactivity

▷ The currently visible data can be filtered (client side) and the user can swap between both visualisations

▷ By filtering on the client side, we avoid additional requests

▷ When zooming in/out, the hexagons are recalculated dynamically

▷ Hovering over a feature shows the amount of tweets and average sentiment for that area, based on the current filter options

Total tweets: 68097

Filter data

Language:

en

Start date: 04/28/2020

End date: 05/26/2020

Filter data   Toggle Countries

GHENT
UNIVERSITY

Interactivity

▷ By clicking on a feature on the map, a word cloud is shown for all tweets inside that area

▷ This is done via a POST request to the server, with all the IDs of tweets inside that feature

▷ For the countries visualization, we show 3 word clouds (All, Positive, and Negative tweets)

Scalability

▷ The Django database with a NGINX server is highly scalable (e.g. Instagram)

▷ Deck.gl map overlays can support up to millions of points by using GPU acceleration

▷ The main problem is downloading the data from the server, but this can be done incrementally or the tweets can be clustered beforehand (e.g. per 50 or 100 km)

▷

GHENT
UNIVERSITY

Customizability

▷ By using high-level libraries, each part of the application can be customized and optimized freely

▷ The application logic consists of loosely coupled functions and components, which can be adjusted individually

▷ By using ReactJS and Django, the strengths of Python (data manipulation) and Javascript (interactivity) are combined

GHENT
UNIVERSITY

Conclusion

▷ With open data and the Twitter API we were able to perform sentiment analysis on a global scale

▷ The data is visualized in a dynamic spatiotemporal map, with filter functionalities

▷ The application can easily be customized and extended, the exposed API can also be used to power third-party applications

▷ Link to live demo: https://tw06v072.ugent.be/wordcrowd/covid/

GHENT
UNIVERSITY

Questions??

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