U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Historical Conflicts Search Engine

Information Processing and Retrieval

Dinis Sousa (up202006303)
Gabriel Ferreira (up201906072)
João Matos (up202006280)
João Pinheiro (up202008133)

Table of contents

01. Introduction

02. Datasets

03. Pipeline

04. Model

05. Characterization

06. Search Scenarios

07. Conclusions

01

Introduction

Introduction

In this milestone we explore historical conflicts. We gathered data about armed human disputes from all around the world from various periods in history.

Datasets

Datasets

Wikidata

We gathered a list of 9134 “historical events” using a SPARQL query

Wikipedia

For each event we added the corresponding summary from its Wikipedia page

Pipeline

You can enter a subtitle here if you need it

Pipeline

Collect list of events from Wikidata website

Output to a json file

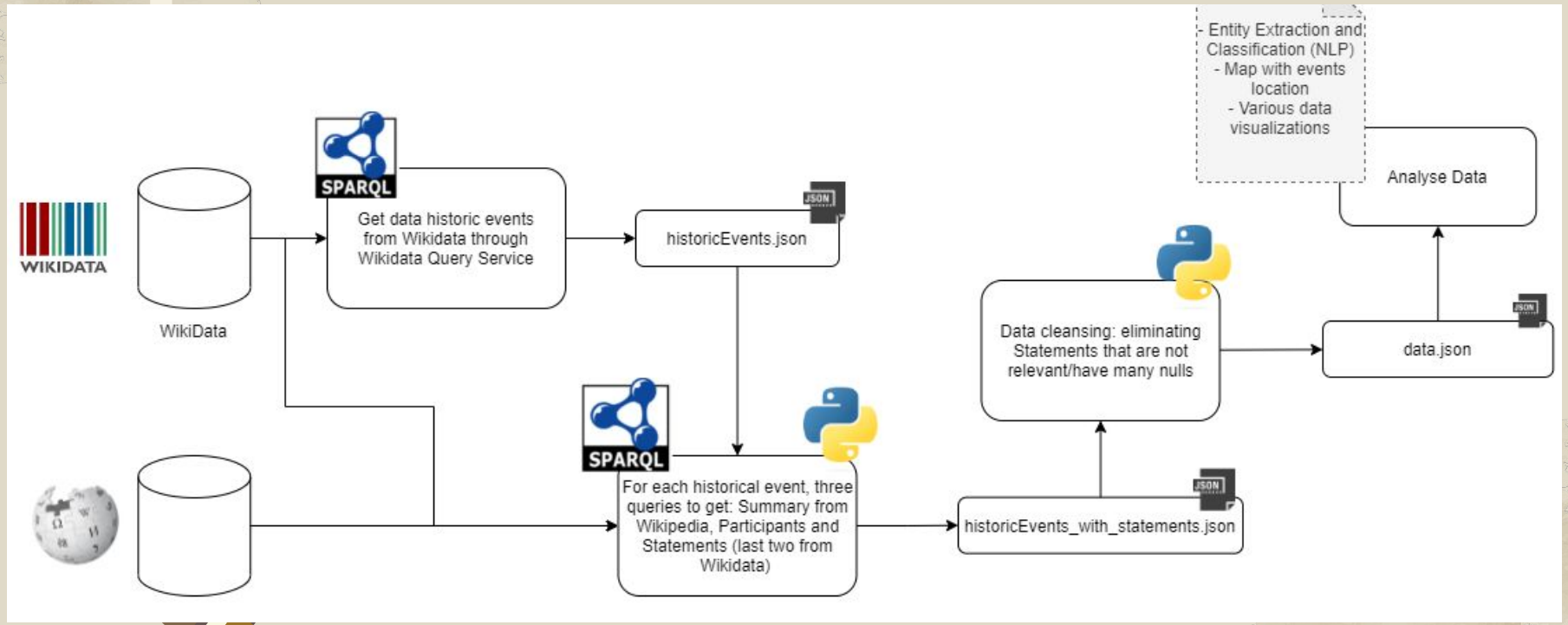
Add summary and more statements from Wikidata

In *fill_events.py*, for each event the Wikipedia page summary is retrieved and we query Wikidata for more information

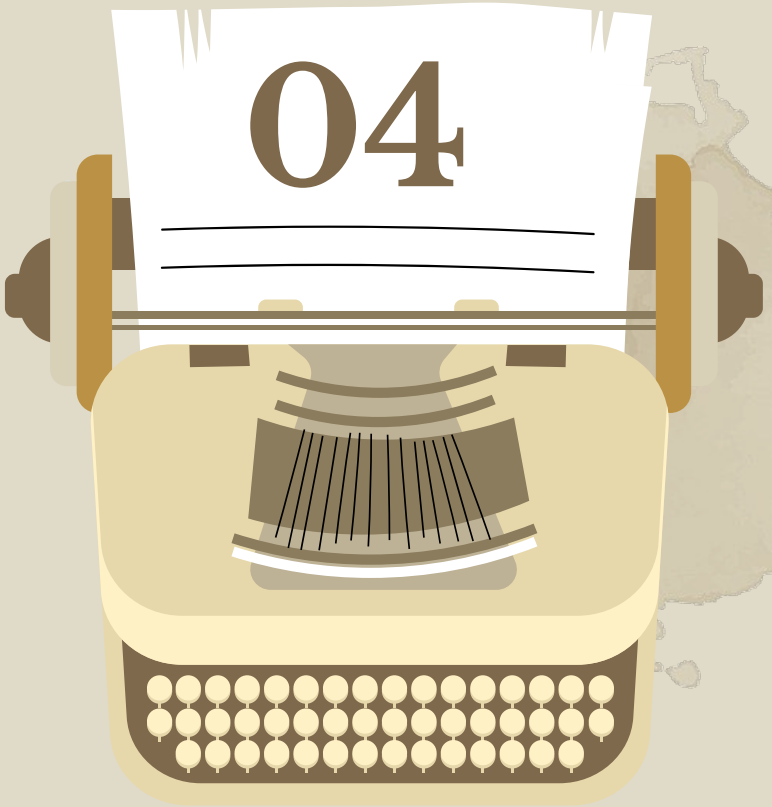
Processing

In *processing.py*, the full data, in a json file, is applied some filters and flattened. Output to both json and csv

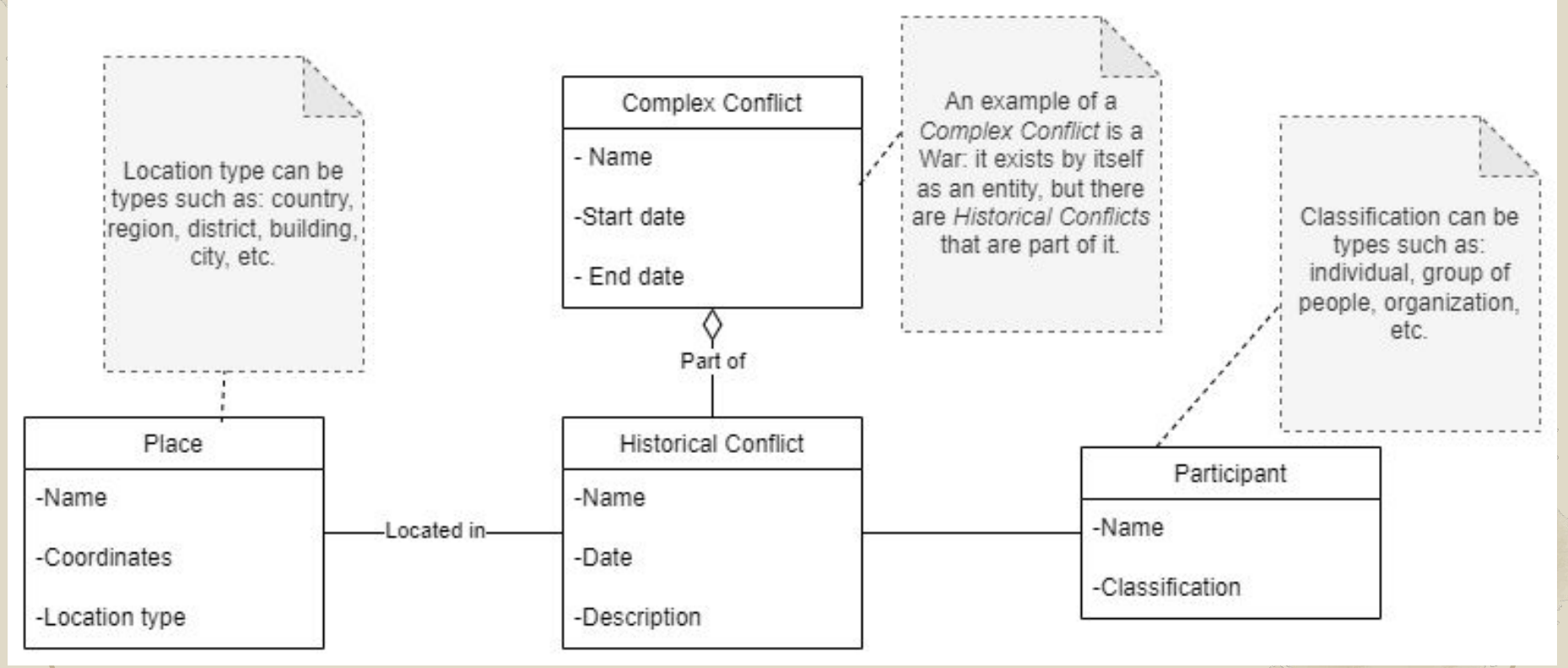
Pipeline


WIKIDATA
WikiData
SPARQL
Get data historic events from Wikidata through Wikidata Query Service
historicEvents.json
SPARQL
For each historical event, three queries to get: Summary from Wikipedia, Participants and Statements (last two from Wikidata)
historicEvents_with_statements.json
Data cleansing: eliminating Statements that are not relevant/have many nulls
data.json
Analyse Data
- Entity Extraction and Classification (NLP)
- Map with events location
- Various data visualizations

Model



Model

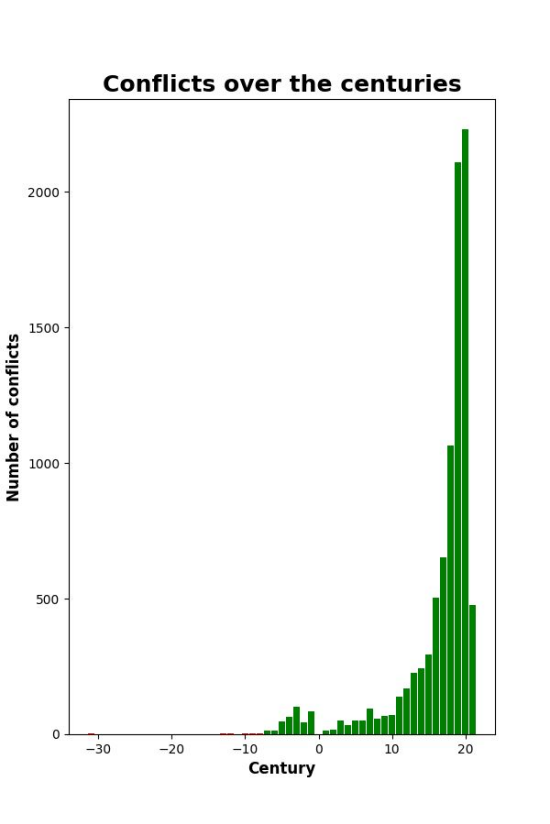

Location type can be types such as: country, region, district, building, city, etc.
Complex Conflict
- Name
-Start date
- End date
An example of a *Complex Conflict* is a War: it exists by itself as an entity, but there are *Historical Conflicts* that are part of it.
Classification can be types such as: individual, group of people, organization, etc.
Part of
Place
-Name
-Coordinates
-Location type
Located in
Historical Conflict
-Name
-Date
-Description
Participant
-Name
-Classification

05

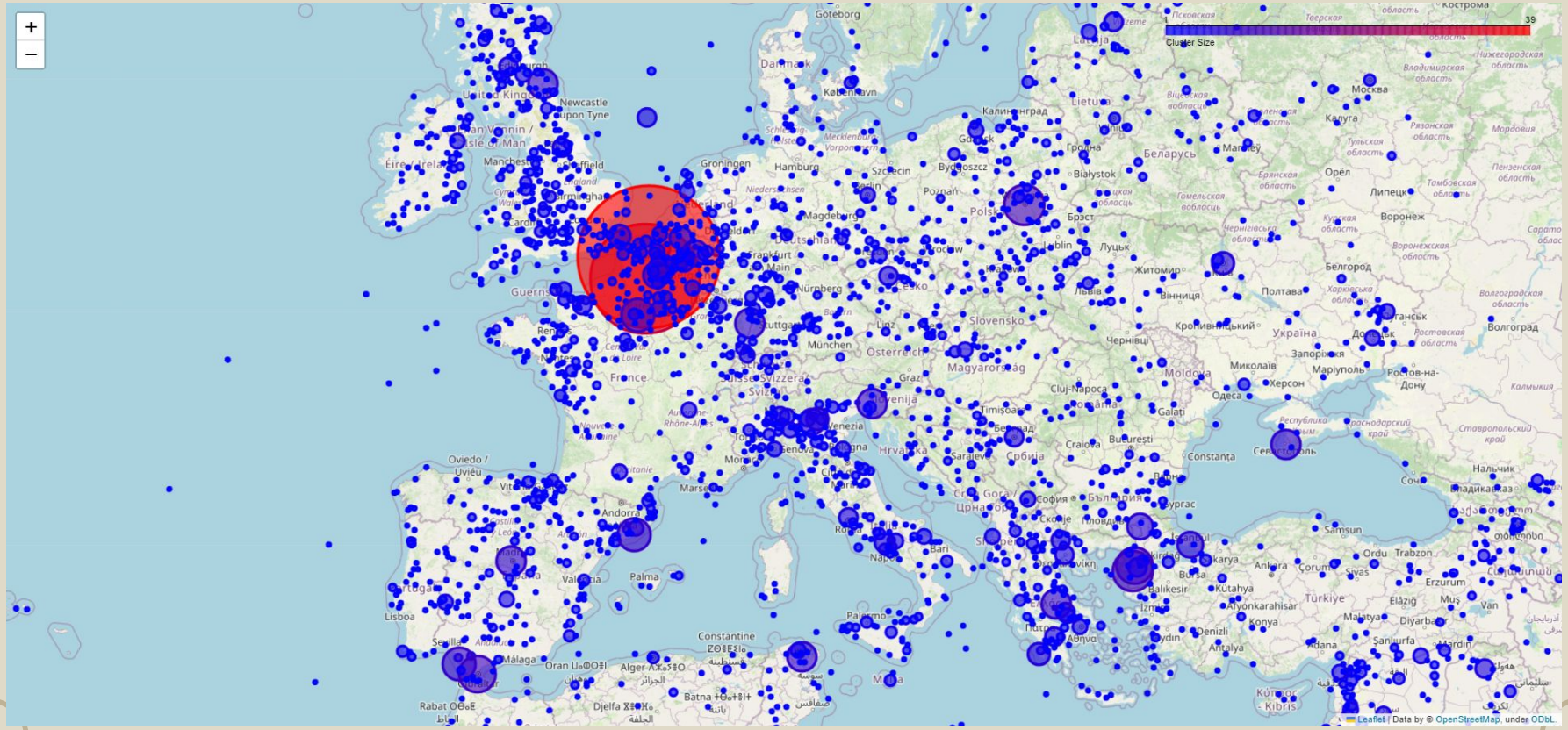
Characterization

Characterization

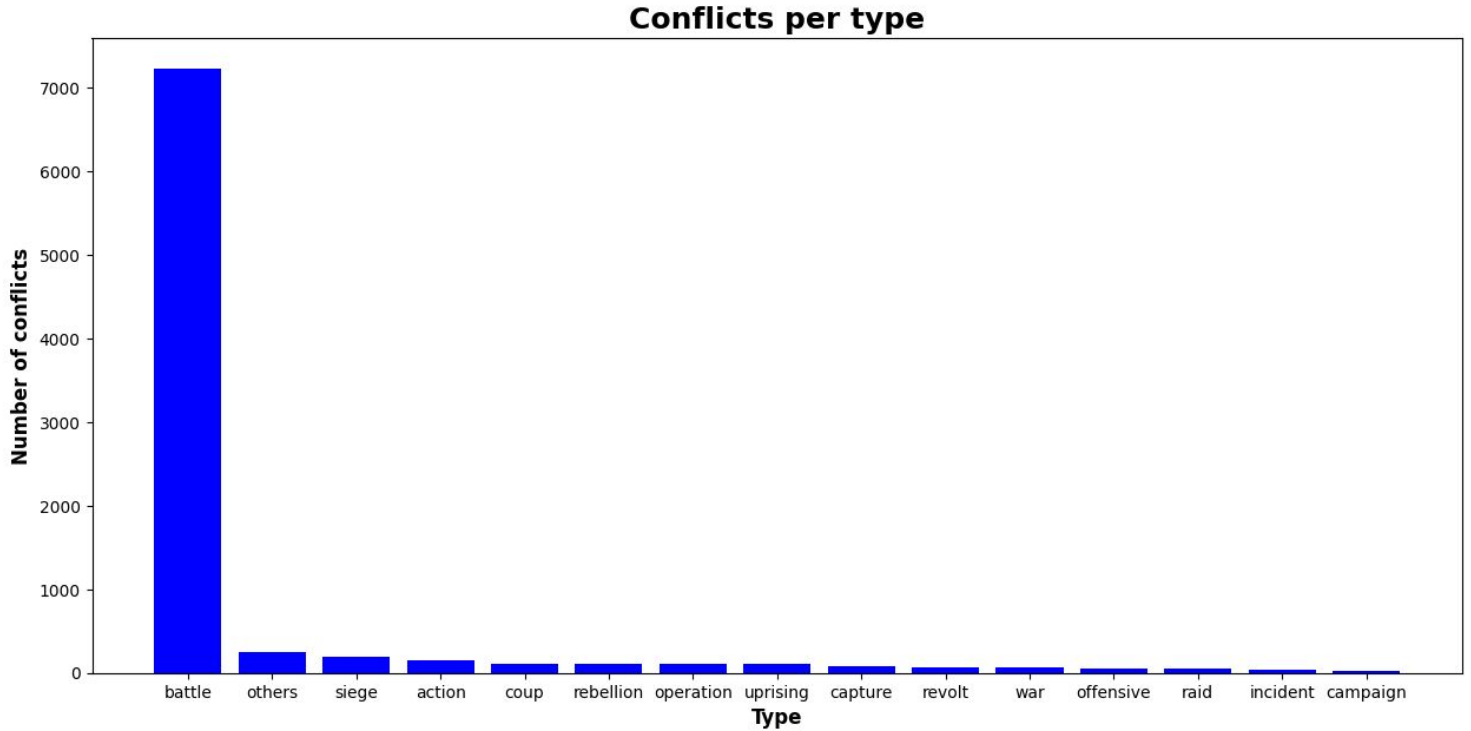
Oldest event: **Battle of Banquan in 30.000 BC**.



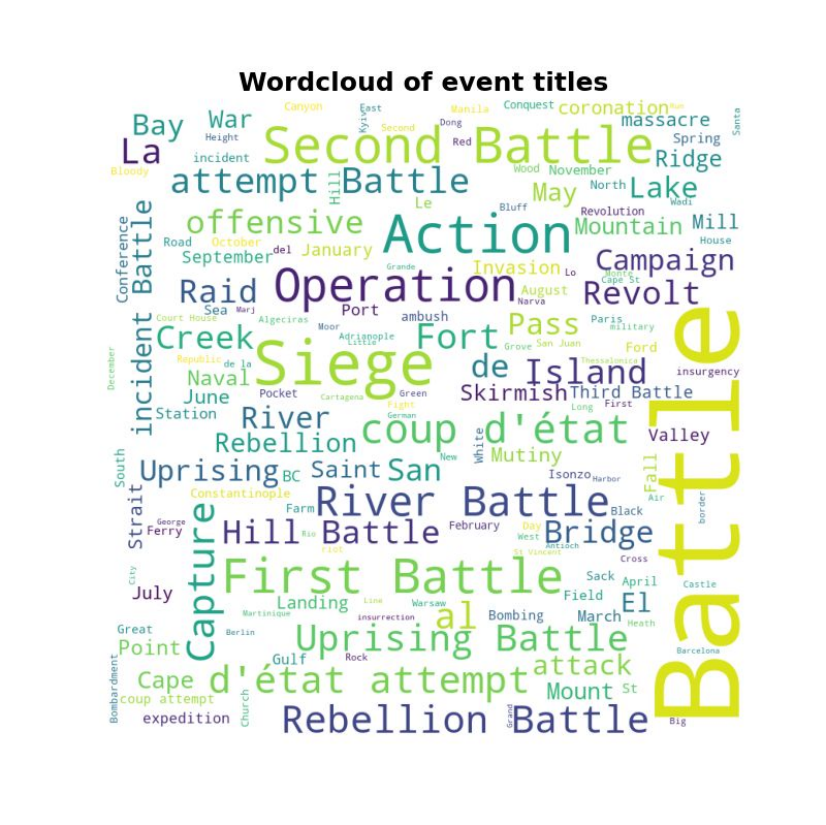
Characterization



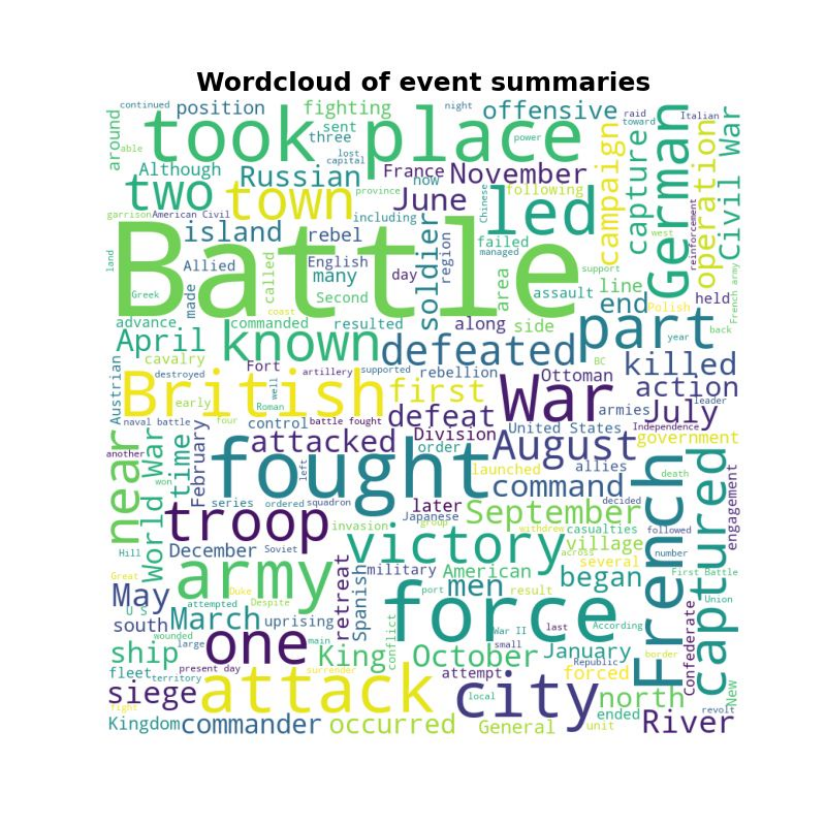
Characterization



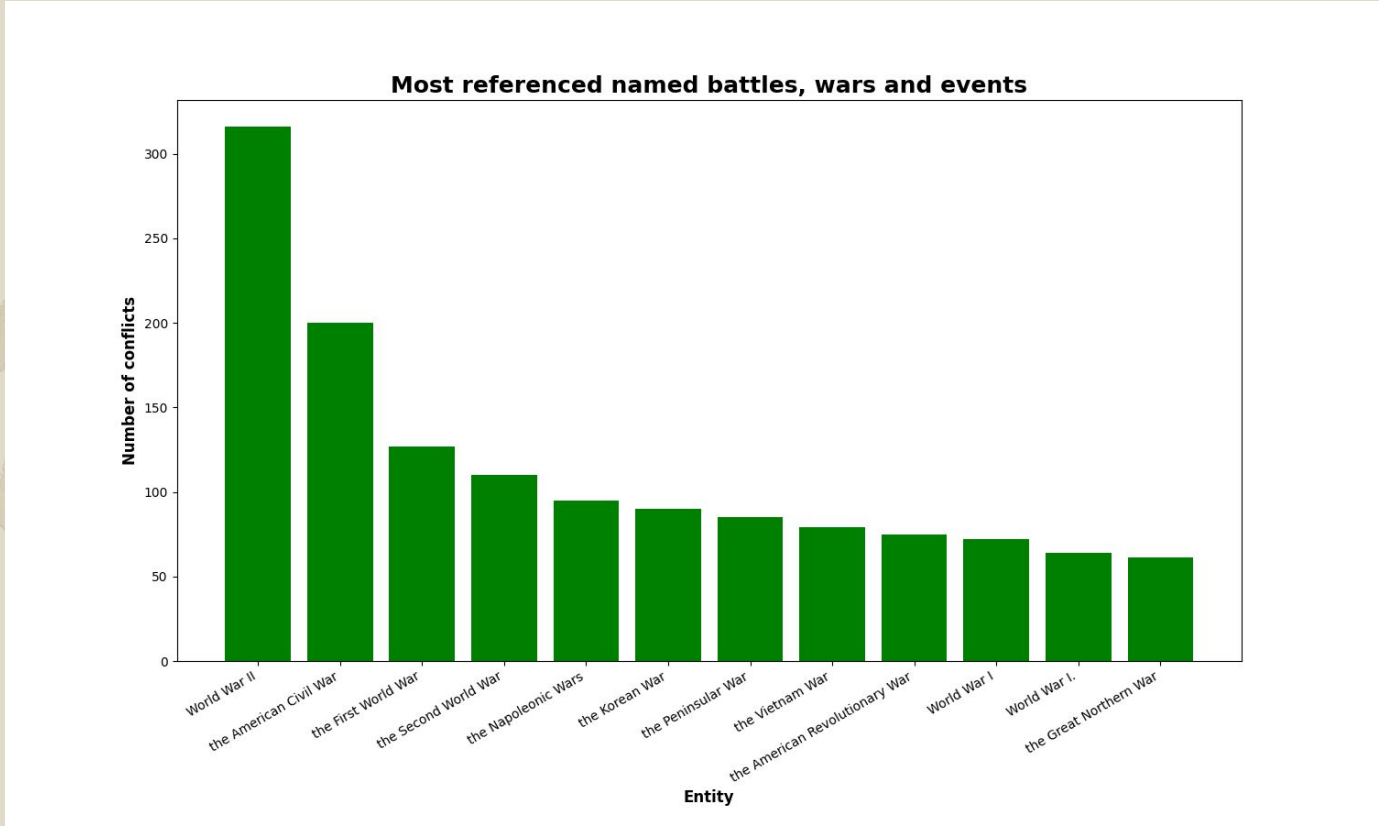
Characterization



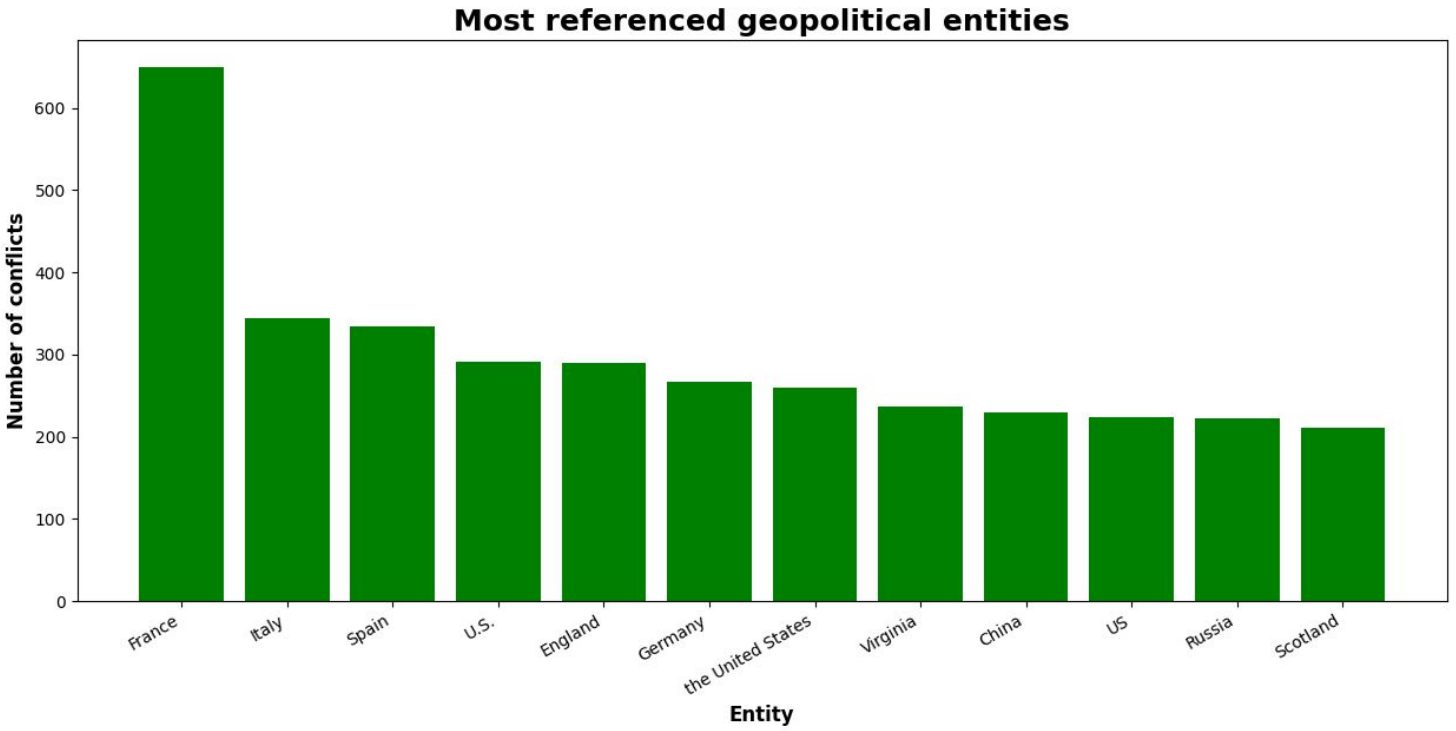
Characterization



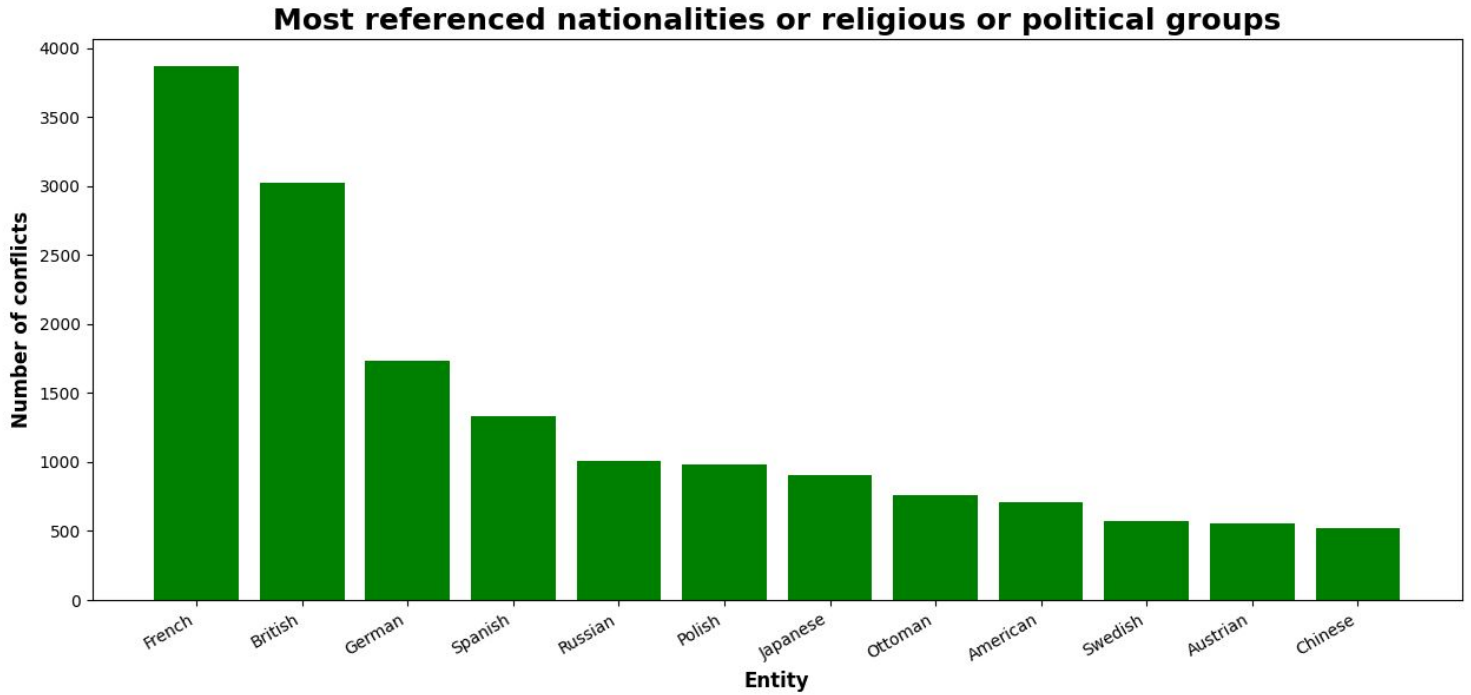
Characterization



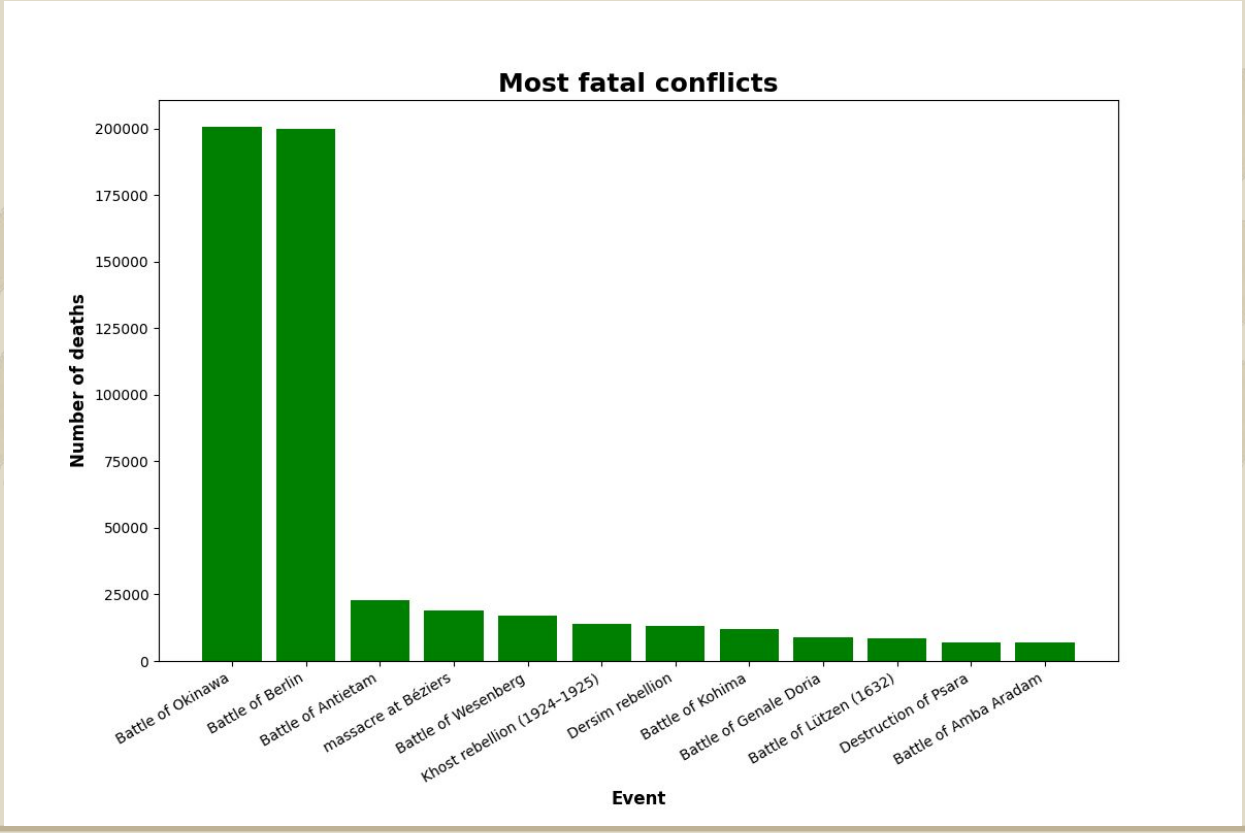
Characterization



Characterization



Characterization



Search Scenarios

Search Scenarios

1. Who participated in the biggest battles of the 19th century?
2. What were the deadliest battles in Europe related to World War I?
3. In which battles did the Portuguese participate between 1300 and 1800?
4. What were the consequences of the Spanish Civil War?

Conclusions

Conclusion in numbers

9134

Events gathered from Wikidata

8998

Events with summary suitable for use

557025

Number of registered deaths in the dataset

5664

Contain an image

3240

Contain a list of participants

Conclusion

We may not have collected every past conflict or not even all battles that we can think of, but we believe to have gathered and processed a decent sample into a coherent dataset.The text description of each conflict will allow us to further extract relevant and interesting information in our future search engine.

Thanks

Do you have any questions?