

# Research Workshops - NLP

## Project presentation

Jakub Koziół, Tomasz Nocoń, Jan Skwarek

Faculty of Mathematics and Information Science  
Warsaw University of Technology

May 30, 2022

# Sections

- 1 Introduction
- 2 Collecting data
- 3 Exploratory data analysis
- 4 Topic modelling
- 5 Streamlit app
- 6 Further research

# Introduction



**Figure:** Ukrainian service members look for unexploded shells after fighting with Russian troops in Kyiv, Ukraine, on February 26. Sergei Supinsky/AFP via Getty Images

# Introduction



**Figure:** A child looks out of a bus window with drawings on it as civilians are evacuated from Irpin, Ukraine. Vadim Ghirda/AP

# Sections

- 1 Introduction
- 2 Collecting data
- 3 Exploratory data analysis
- 4 Topic modelling
- 5 Streamlit app
- 6 Further research

# Collecting data

The screenshot shows a PyCharm IDE window titled 'WS-NLP-project - main.py'. The left sidebar displays a project structure with files like '02\_march.csv', '21\_feb.csv', '22\_feb.csv', and '03\_march.csv'. The main editor shows a Python script that uses the 'smtwitter' library to scrape tweets. The script initializes a list 'tweets\_list1', sets a start time, and iterates through tweets from March 2, 2022, to March 3, 2022. It prints the index and date of each tweet and appends it to the list. The bottom panel shows the 'Run' output, displaying a list of 1000 tweets with their dates and times.

```

7 tweets_list1 = []
8
9 start = time.time()
10 # Using TwitterSearchScrapper to scrape data and append tweets to list
11 for i, tweet in enumerate(smtwitter.TwitterSearchScrapper
12
13     'Ukraine since:2022-03-02 until:2022-03-03 lang=en'
14     .get_items()):
15
16     # if i > 400: # number of tweets you want to scrape
17     #     break
18
19     if i % 1000 == 0:
20         print(i, tweet.date)
21
22     tweets_list1.append(
23         [tweet.date,
24          tweet.id,
25          tweet.content,
26          tweet.user_username,
27          ]
28     for i, tweet in enumerate(smtwitter.TwitterSearchScrapper
  
```

Run: main

```

1000 2022-03-02 23:57:15+00:00
2000 2022-03-02 23:54:26+00:00
3000 2022-03-02 23:51:37+00:00
4000 2022-03-02 23:48:56+00:00
5000 2022-03-02 23:46:06+00:00
6000 2022-03-02 23:43:19+00:00
7000 2022-03-02 23:40:40+00:00
8000 2022-03-02 23:38:06+00:00
9000 2022-03-02 23:35:29+00:00
10000 2022-03-02 23:32:44+00:00
  
```

**Figure:** Snsrape is a scraper for social networking services (SNS). It scrapes things like user profiles, hashtags, or searches and returns the discovered items, e.g. the relevant posts.

# Collecting data

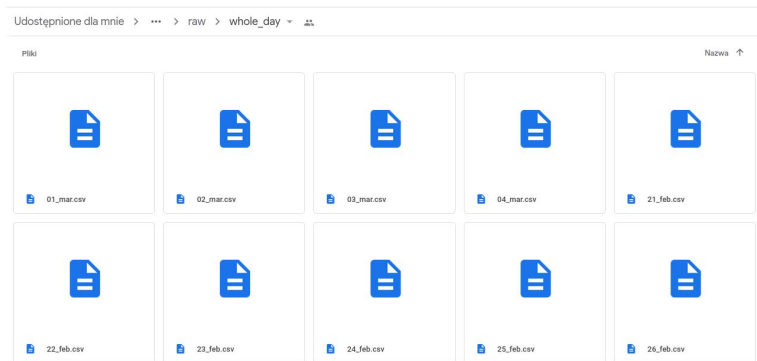


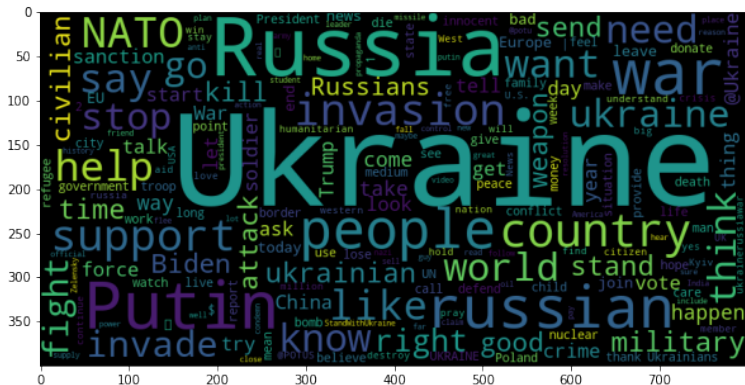
Figure: We scraped over 10 million tweets in less than a week using snsrape.

# Sections

- 1 Introduction
- 2 Collecting data
- 3 Exploratory data analysis**
- 4 Topic modelling
- 5 Streamlit app
- 6 Further research



## Exploratory data analysis



**Figure:** Wordcloud with the most popular words from Twitter on the 2nd of March (from tweets with keyword 'Ukraine').

# Exploratory data analysis

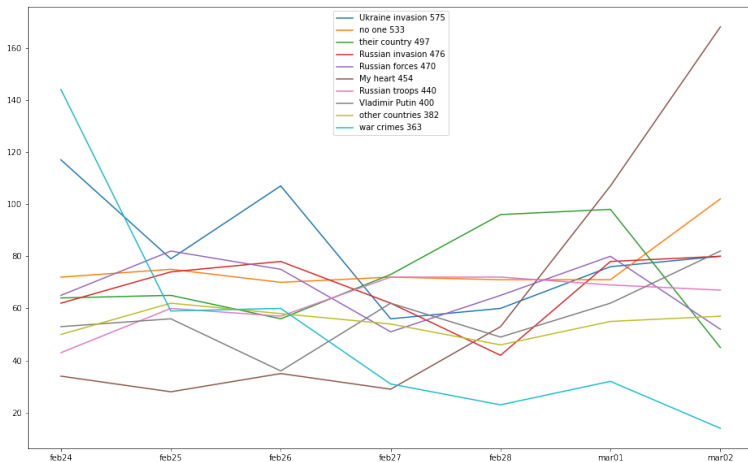
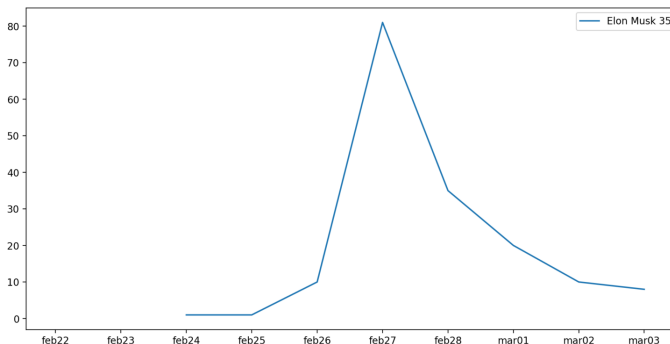


Figure: Plot over time with 2-grams based on frequency.

# Exploratory data analysis



*'Elon Musk Comes To The Rescue After Ukraine Government Pleads For His Help - Ukraine highlighted the tweet, writing: "Dear @elonmusk, Ukraine needs your support. Your stance and your actions matter. An appeal by our Minister of Digital Transformation here."'*

# Exploratory data analysis



*"From today, we are removing thousands of bottles of Russian vodka and beer from our shelves in response to Russia's escalating aggression towards Ukraine"*

*"OFF THE SHELVES: Some liquor stores in Brooklyn are pulling Russian vodka from their shelves in solidarity with Ukraine."*

# Sections

- 1 Introduction
- 2 Collecting data
- 3 Exploratory data analysis
- 4 Topic modelling**
- 5 Streamlit app
- 6 Further research

# Topic modelling

**Topic Word Scores**



Figure: BERTopic - topic word scores

# Topic modelling

*"VitalikButerin very upset by Putin 's decision to abandon the possibility of a peaceful solution to the dispute with Ukraine and go to war instead this be a crime against the ukrainian and russian people I want to wish everyone security although I know that there will be no security "*

*"Ukraine accept donation in BTC eth and USDT"*

# Topic modelling

*"vincentwym97 defenceu not all chinese citizen support Russia  
many citizen in China also support Ukraine"*

*"nntaleb even if Putin take Ukraine this will be a big loss overall  
for he it unite NATO reinforce ukrainian identity isolate Russia  
economically from the west Russia will be force to turn more and  
more towards China ultimately end up as a chinese vassal"*



# Topic modelling

Topics over Time

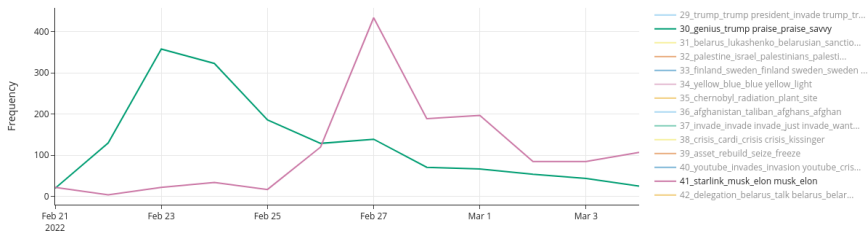


Figure: BERTopic - topics over time

# Sections

- 1 Introduction
- 2 Collecting data
- 3 Exploratory data analysis
- 4 Topic modelling
- 5 Streamlit app**
- 6 Further research

# Streamlit app

Figure: video Streamlit

# Sections

- 1 Introduction
- 2 Collecting data
- 3 Exploratory data analysis
- 4 Topic modelling
- 5 Streamlit app
- 6 Further research**

# Further research

- Application maintenance and development.
- Further topic modelling.
- Advanced visualizations of correlations between Twitter data and geopolitical events.
- Research on Russian tweets.
- Semantics analysis.

# Thanks for your attention!



Figure: QR code for our GitHub repository