Plan Overview

A Data Management Plan created using DMPTool

Title: Analysis of the information cascade in social media on the example of Twitter posts and spread of hashtag #stopukrainizacjipolski

Creator: Mateusz Czerwiński

Affiliation: University of Warsaw (uw.edu.pl)

Funder: Digital Curation Centre (dcc.ac.uk)

Template: Digital Curation Centre

Project abstract:

Project mainly focus on visualization of #stopukrainizacjipolski contagious in social network within Twitter users. Next goal is to empirically measure growth for tweets with examined hashtag, compare patterns for contagious in common follower/following relations.

Start date: 06-30-2022

End date: 01-31-2023

Last modified: 02-01-2023

Analysis of the information cascade in social media on the example of Twitter posts and spread of hashtag #stopukrainizacjipolski

Data Collection

What data will you collect or create?

Data was collected as json respond from twitter API, queried with:

- #stopukrainizacjipolski and next
  - for each user whom used hashtag, list of fellowers was also obtained.

How will the data be collected or created?

Data will be collected via python3 requests library, using Academic Reaserch bearer token which allow full search and have 10_000_000 tweets / month limit.

Documentation and Metadata

What documentation and metadata will accompany the data?

For purpose of the study, not all metadata obtaied from API calls will be stored. Respected meta information were used:

1. Author:
   - username,
   - name,
   - ID;
2. Tweet:
   - timestamp for created_at date,
   - content,
   - entites extracted by twitter by @ symbol
   - hashtags extracted by twitter by # symbol
   - public metrics:
     - likes,
     - replies,
     - retweets,
     - views;
   - author ID.

Ethics and Legal Compliance

How will you manage any ethical issues?

Because of data type and access policy there should not by any ethical issues. Collected data is not sensitive and available to all other users of Twitter

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Data owner is Twitter, for this assay needs, downloaded data was transfered to NoSQL, self-hosted database.

Storage and Backup

How will the data be stored and backed up during the research?

Data is stored in self-hosted NoSQL databse. Because of rare tweets update (weekly), data is as well backuped weekly, on VPS.

How will you manage access and security?

Data is used by one person, so there is no problem with access controls and/or security menagment

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

All of collected data is long-term value, and could be shared.

What is the long-term preservation plan for the dataset?

SPARQL and NoSQL databases.

Data Sharing

How will you share the data?

SPARQL backend with RDFS social-networks vocabulary, frontend by telegram/slack/discord chatbot

Are any restrictions on data sharing required?

Rate limits, authorization

Responsibilities and Resources

Who will be responsible for data management?

Person responsible for collecting and cleaning

What resources will you require to deliver your plan?

- python3
- mongo backend
- Apache Jena backend
- SDK for chatbot frondend
- nginx/caddy middleware