TeleScan: Analyzing Telegram Messages for Toxicity and Information Spread

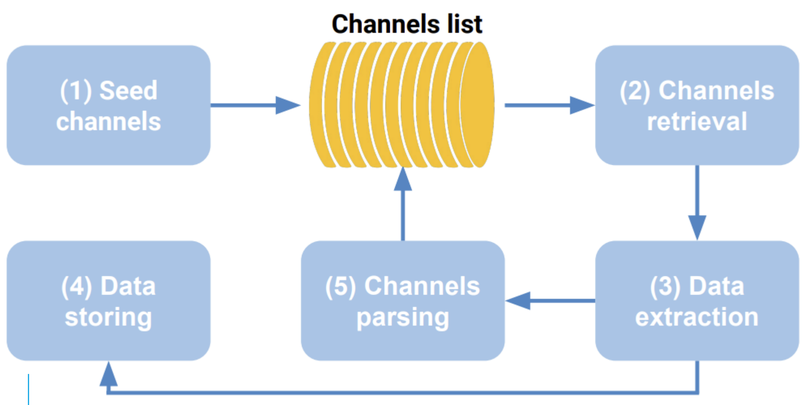
Susmita Gangopadhyay
Gesis -- Leibniz Institute of Social Science

Abstract

Telegram is one of the most popular instant messaging apps in today's digital age. Telegram channels are public broadcast channels on the Telegram messaging platform where creators can share messages, media, and updates with a large audience. We introduce a continuous long-term collection of messages from Telegram channels and offer analysis capabilities to discover insights into language distribution, entities, toxicity across messages, and potential applications for understanding Telegram's ecosystem and tracking the spread of information.

The Collection Framework

- We begin with a set of seed channels related to Citation, Reach and Subscribers and then extend our dataset by including the source channel of each forwarded message.
- Snowball approach



Statistics

- 100,167,504 messages from 14,530 channels downloaded
- 559,292 channels discovered

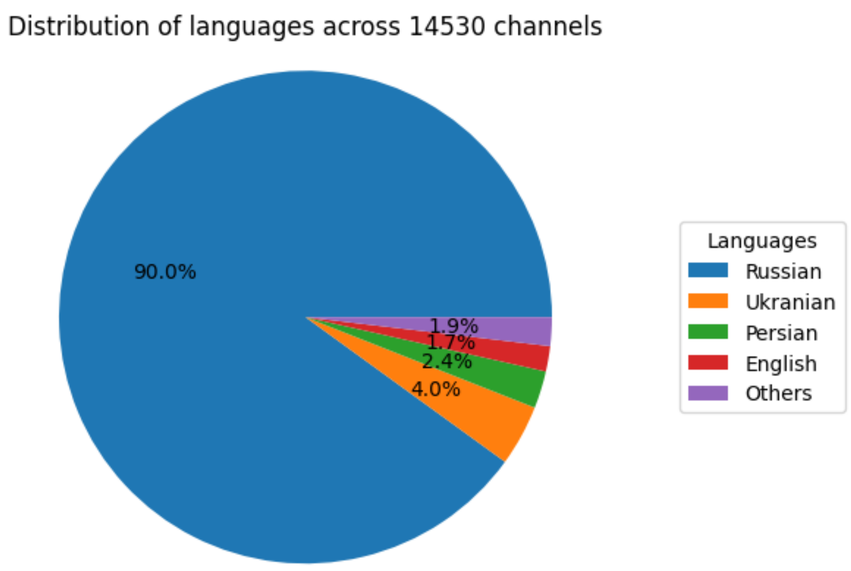
Research Questions

- RQ1: What languages are most commonly used for messaging on Telegram?
- RQ2: Do specific channels act as hubs of information spread? Is this information toxic?
- RQ3: Determine the role of seed lists in Telegram

Results

RQ1

- We analyze the channels and the languages that these channels belong to
- We find that more than 90% of these channels have the working language of Russian



RQ2

- We create a channel-to-channel graph and analyze the centrality scores of the nodes.
- Analysis of degree and eigen centrality shows Russian channels dominate, are central and influential parts of the network, and contribute to the top 350 channels.
- A particular Russian channel came out to be the most influential channel in both degree and eigen centrality that did not come from any of the seed lists.

R3

- We perform an analysis of which seed list results in a bigger graph span

Work In progress

- Assigning toxicity score to each message
- Assigning sentiment scores to each message
- determines if the most forwarded messages have a certain characteristic (e.g., toxic/non-toxic, positive/negative) investigating the spread on the graph.

Use Cases

- Spread of information and misinformation
- Spread of rumors, claims, toxicity, propaganda,
- Comparison with other social media platforms
- Domain specific Named Entities

.

*** Feel free to reach out for ideas, suggestions, and brainstorming**

Kontakt: Susmita Gangopadhyay (Susmita.Gangopadhyay@gesis.org)