The Social Synapse: Unveiling Public Opinion and Elements of Discourse Connections

Bahadır Gezer, Bahri Alabey

Computer Engineering, Boğaziçi University - Advisor: Suzan Üsküdarlı

Motivation

To develop an approach that reveals conversations and public opinion on the internet. Sentiment analysis provides insights but often misses how entities and elements of discourse are connected. This project aims to fill this gap by analyzing both sentiment and the relationships between entities in online conversations.

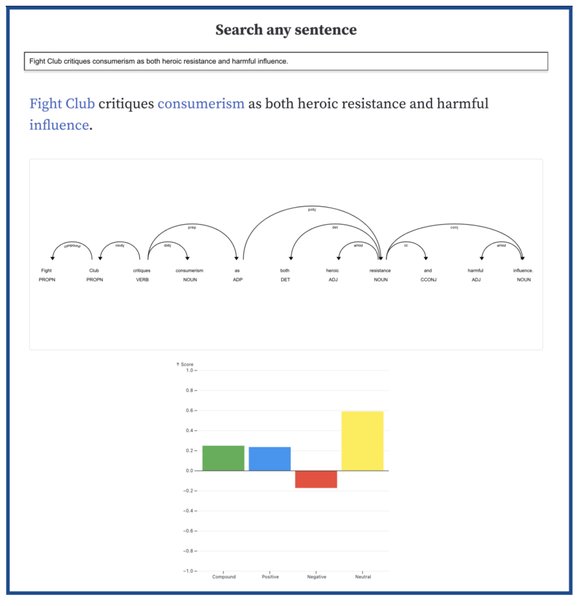

Figure 1:Social Synapse Main Page Interface

Overview

We propose a text-centric approach to uncover relationships between elements of discourse within online conversations. Our system analyzes a large collection of text, focusing on the inherent structure and relationships within the conversations. Named entities are identified and relationships between them are analyzed. This allows us to construct a network of relationships based solely on the information that we've processed.

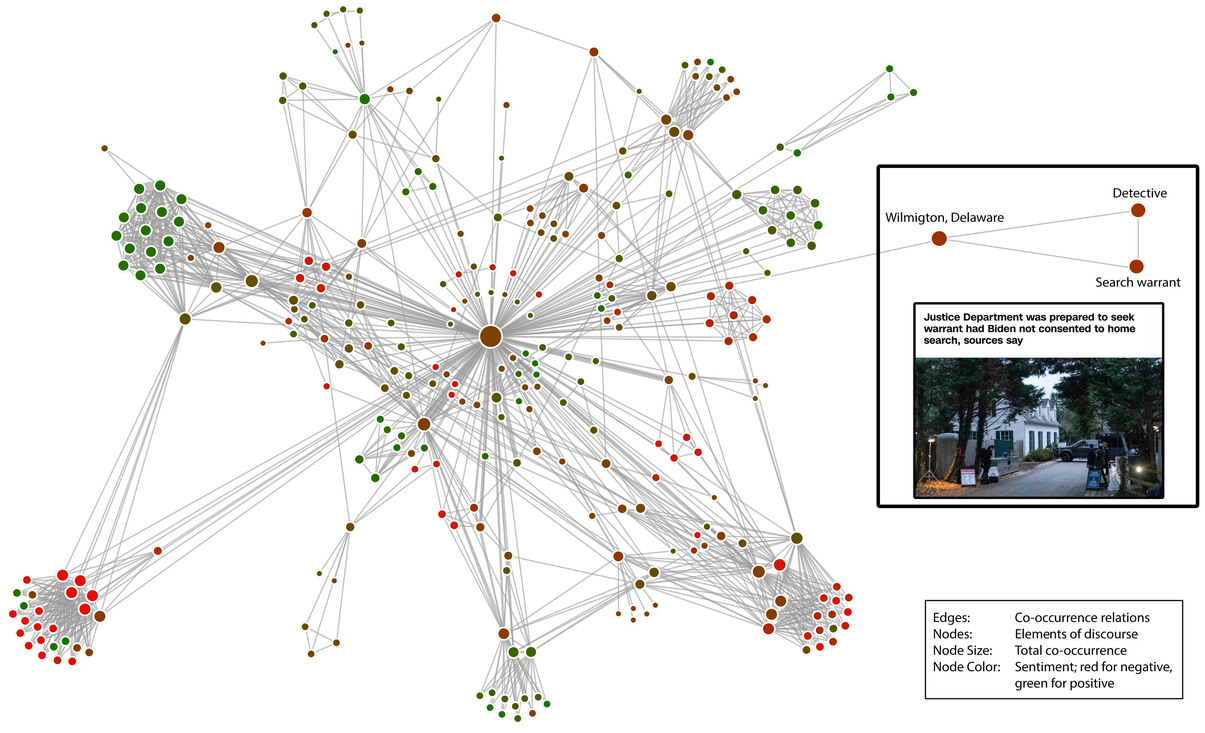

Figure 2:Topic Graph for Joe Biden

Text Processing Pipeline

1. Fetch data from online conversations.
2. Clean data (punctuation, links, etc.).
3. Link elements of discourse (people, locations, etc.) with named entity linking (NEL).
4. Create context-windows using sentence dependency graphs from Spacy [1] to analyze co-occurrences.
5. Perform sentiment analysis to each context-window using NLTK's [2] VADER lexicon.
6. Store processed information as a fixed-structure JSON object.

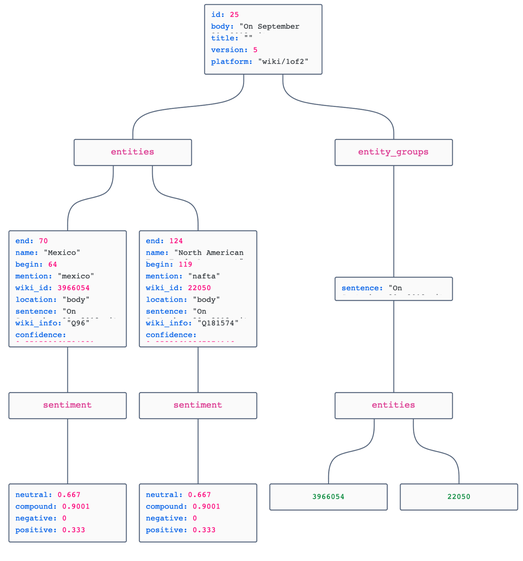

Figure 3:Example Processed JSON Structure

Results and Future Work

We developed a web-based tool that effectively visualizes the relationships and sentiment for different elements of discourse. We examined this tool with datasets from various domains and as it stands the data includes around 160k unique entities, across 500k online conversations. Future work includes incorporating time-series analysis, graph-based search, and exploring applications in SEO optimization.

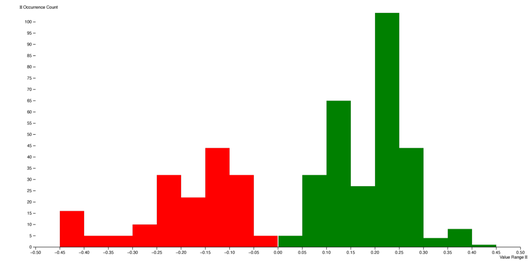

Figure 4:Sentiment Histogram for Donald Trump

References

[1] Van Landegham Boyd Honnibal, Montani. spacy: Industrial-strength natural language processing in python. 2020.

[2] Edward Loper and Steven Bird. Nltk: The natural language toolkit, 2002.

Acknowledgements

The production of this work has been enabled by means of the TagMe service offered via the D4Science Services Gateway operated by D4Science.org. We would like to thank our advisor, Suzan Üsküdarlı, for her diligence and attentiveness throughout our project.

Contact:

- Web: hocamsimdi.com.tr
- Email: cmpe492@hotmail.com