

Title: Ambassadorial Meeting Observation Summary with Group 6

Date: 1/03/2024

Groups Involved:

- Group 7
- Group 6

Purpose of Meeting: The purpose was to expand my understanding of the various projects and methodologies employed in NLP across different groups.

Attendees:

From Your Group: Nachiketh J

From Group 6: Rasika Vilas Burde, Xinyang Shao ,Rutuja Manohar Pate, Samarth Agrawal, Pradosh Nataraj , Qiang Han

Summary of Proceedings:

I attended the meeting of Group 6 to gain insights into their methodologies and approaches concerning their NLP project. The key discussions are summarized below:

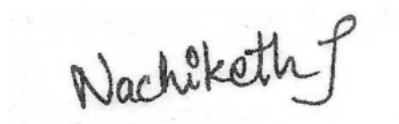
1. Overview of Group 6's Methodology:

- Data Collection and Processing:
Group 6 is collecting data from COVID-19-related discussions on subreddits like r/DebateVaccines and r/Coronavirus. They use the Python Requests library for data scraping, ensuring adherence to Reddit API guidelines by managing request rates and including user-agent headers.
- Data Cleaning Techniques:
They employ pandas for data cleaning, where they focus on removing duplicates, handling missing data, and normalizing text to ensure uniformity for analysis.
- Analytical Techniques:
The group uses NLTK for tokenization, transforming raw text into structured formats suitable for NLP tasks. They apply scikit-learn for TF-IDF analysis to identify key terms and phrases, and gensim for Latent Dirichlet Allocation (LDA) to discover prevalent topics within the discussion data.
- Sentiment Analysis:
Sentiment analysis is performed using VADER from the NLTK library and TextBlob, to assess the emotional content within the community discussions, providing insights into the prevailing sentiments across different threads.

- Comparative and Contextual Analysis:
They utilize both TF-IDF and LDA to analyse linguistic styles and thematic concerns across different subreddits, allowing them to compare and contrast how discussions vary across communities.
- Visualization and Further Analysis:
Techniques like t-SNE for dimensionality reduction and K-means clustering are used to visualize data clusters and explore patterns. Additionally, they use LIWC to delve deeper into the psychological and emotional underpinnings of text data, enriching their analysis by revealing more about the social dynamics within the online discussions.
- Observations:
While Group 6's project is intricate and their methodologies are well-structured, much of their implementation phase had not started by the time of the meeting. The theoretical framework and planned analysis techniques are robust, aiming to uncover nuanced insights into social dynamics and identity formation through language use in online discussions.
- Conclusions:
The meeting was highly informative, providing me with a broader perspective on the use of NLP in analysing social media data. Although their specific methodologies are not directly applicable to our project, understanding their approach helps in broadening my knowledge base and might inform future projects or interdisciplinary collaborations.

Signatures from Your Group:

Nachiketh J



Signature from group 6-

Member 1: Rasika Vilas Burde




Member 2: Samarth Agrawal

Member 3: Rutuja Manohar Pate



Member 4: Xinyang Shao



Member 5: Pradosh Nataraj



Member 6: Qiang Han

