Project proposal

Team Members: Zhiyang Zhou, Yushun Cheng, Hui Yin Lam, Trang Dang

Source of data

- The following link with the built-in search engine will be used to extract the links for news articles relating to the election 2021
  https://www.cbc.ca/search?q=2021%20election&section=news&sortOrder=date&media=all

Document Structure

- The title, description, and date of news included in the above link will be extracted, they were written in standard English.
- The genre is political reporting. In terms of register, it employs formal language suitable for conveying factual information and analysis on political topics.
- For each document, the title will be a sentence consisting of around 10 words, and the description will be a sentence consisting of 10 to 40 words

Corpus Structure

- The corpus will be focusing on political news related to the 2021 Canadian Federal Election, the coverage will be from the official kickoff of the election campaigns (Sept 2, 2021) to the day before the election (Sept 19, 2021)

Annotations

- The topics (economy, environment, social, health, etc.)
- The sentiments (positive, neutral, negative)
- The types of sources (party statements, analysis, poll results)
- The political parties involved
- The geographical information
- The meta-information about the reporter

Corpus Storage Format

- JSON format will be used as the length of different types of annotations might be different

Extensive Usage

- The corpus can contribute to the analysis of political trends, such as shifts in voter preferences, regional differences in voting behaviors, and the impact of campaigns on election outcomes.
- It can also be used by media researchers to analyze media coverage of the election which includes framing of issues, bias in reporting, and the role of social media in shaping public opinion.