RSS Feed Analysis Tool: A Java-Based Approach to Sentiment and Political Leaning Detection

Written by Vuottek Un

Abstract

The RSS Feed Analysis Tool is a Java application that integrates the Stanford CoreNLP library and JSoup HTML parser to assess and report on the sentiment and political orientation of content within RSS feeds. The software provides a graphical user interface (GUI) for users to input RSS feed URLs, analyze textual content, and display the sentiment score alongside the detected political leaning. This paper provides a comprehensive overview of the application's architecture, functionality, and potential use cases, alongside a critical discussion of its design choices, potential biases, and areas for future improvement.

Introduction

The proliferation of online content and the need for its quick assessment for sentiment and bias necessitates tools capable of performing such analyses rapidly and accurately. The RSS Feed Analysis Tool presented in this paper aims to meet this need by providing a real-time analysis of RSS feeds, which are a staple in news dissemination.

System Architecture

The application is built on Java Swing for the frontend, allowing for cross-platform deployment. The backend leverages the Stanford CoreNLP library for natural language processing to compute sentiment scores and uses Jsoup for HTML parsing to extract content from RSS feeds.

Methodology

The tool parses the RSS feed content and processes it using the Stanford CoreNLP pipeline. Sentiment scores are derived for each entry, and the presence of pre-defined politically charged words is used to infer the political leaning of the text. The GUI updates dynamically to display this information.

Features and Functionalities:

- GUI for inputting URLs and displaying results.
- Predefined buttons for popular news RSS feeds.
- Sentiment analysis of RSS entries.
- Detection of political leaning based on word frequency analysis.
- Display of average sentiment and political leaning percentages.
- Customizable analysis depth through an adjustable entry limit.

Evaluation and Results

The tool's effectiveness was evaluated by inputting multiple RSS feeds and comparing its output with expected sentiment and leaning outcomes. The average sentiment score provided a numerical representation of the text's positivity or negativity, while the political leaning was indicated by the percentage of identified politically charged words.

Discussion

While the application performs as intended, there are several considerations:

- The word list-based approach to political leaning detection may introduce bias.
- Sentiment analysis accuracy is heavily model-dependent and may not capture the nuances of political discourse.
- The application's performance has not been tested on extremely large datasets.

Conclusion and Future Work

The RSS Feed Analysis Tool represents a step towards accessible content analysis software. However, future work should focus on improving the sophistication of political leaning detection, optimizing performance, and ensuring ethical and unbiased analyses. Additionally, incorporating machine learning could refine sentiment analysis beyond the current lexicon-based approach.

References

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

J. Hedley. (2021). Jsoup: Java HTML Parser. [Online]. Available: https://jsoup.org/

Oracle. (2021). Java Swing tutorial. [Online]. Available: https://docs.oracle.com/javase/tutorial/uiswing/

Oracle. (2021). Java SE Documentation. [Online]. Available: https://docs.oracle.com/en/java/javase/index.html

The RSS Advisory Board. (2009). RSS 2.0 Specification. [Online]. Available: https://www.rssboard.org/rss-specification