Wilford Woodruff Paper Foundation – Text Sentiment Analysis

The Customer

The Wilford Woodruff Papers Foundation (WWPF) seeks to digitally preserve and publish Wilford Woodruff's eyewitness account of the Restoration of the gospel of Jesus Christ from 1833 to 1898, making his records universally accessible to inspire all people to study and increase their faith in Jesus Christ, understand and honor sacred temple covenants, and receive the blessings of exaltation with their families.

The Situation

To foster insightful study and user engagement using AI, the client wanted to 1) tag each page of Wilford Woodruff's journal with sentiment and emotion scores to enable users to understand how Wilford Woodruff's emotions and sentiments changed over time and locations, and 2) have a simple recommendation tool that could find similar documents and suggest them to the website's visitors.

The Solution

A data-consulting student team was formed by BYU–I and RBDC. As a final deliverable, the team created a script that sends the pages from the WWPF database to Hume.ai, a research organization that develops high-level emotional analysis. The analysis returns 53 unique emotion scores, which can be displayed individually or categorized as a summary. Each excerpt is also sent to Chat-GPT along with the prompt "Please identify the sentiment and emotions of the following passage." The script then pulls and stores Chat-GPT's responses. After trying many methods and machine-learning models, the team determined that blending various models gave the best range of values.

Scope

The algorithm(s) must perform sentiment analysis for Type 1 and Type 2.

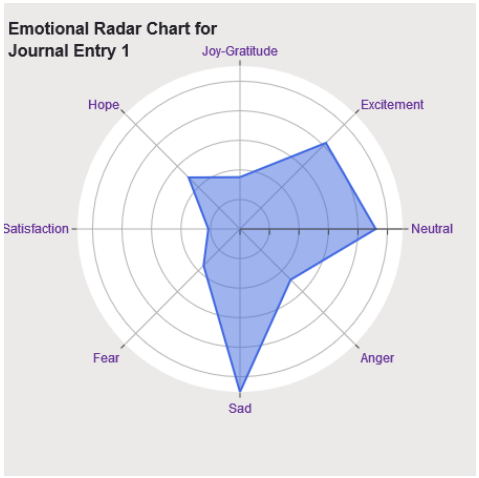
- Type 1 – Determines the polarity of the opinion and can be a simple binary positive/negative sentiment differentiation. A higher dimension such as very positive, positive, neutral, negative, very negative is preferred.
- Type 2 – Identifies signs of specific emotional states presented in the text. Usually, a combination of lexicons and machine-learning algorithms determine what is what and why.

Final Deliverables

- Sentiment analysis algorithm(s). The inputs for sentiment analysis are the WWPF pages and the output is a .csv file.
- A script that can be run on existing or new pages in the WWPF database to determine sentiment.
- User documentation explaining the project, methods, code instructions, and results.
- Visual representations of the sentiment analysis.
- A document comparison tool to make recommendations for website visitors.

Data Visualization

The WWPF wanted to display these emotion scores alongside each document on the website. They asked the team to create a visualization that could be used across each page—one that was complex enough to give details for each individual page while simple enough that a non-technical audience could understand it. The team settled on a radar chart of the grouped emotion categories, giving each page a unique shape that complemented the text.



The WWPF hadn't finalized the exact details they wanted for the visualization, so the team delivered several versions of the code, including ones that combined the various emotion models together and one that kept them separate.

Recommendations

The WWPF hoped that some people seeing these sentiment scores on the website would use the scores as a method to search and explore the journals, so they asked the team to give website visitors recommendations of new entries based on this kind of metadata for each entry. The team took a higher-level approach, adding tags of people, places, and topics to each of these sentiment scores. For each page, this comparison method returns the top three entries that are most similar to the current page, including one page catered towards topical matches and another catered towards people matches. These results will be displayed on the bottom of each page to help visitors explore the documents and connect with Wilford Woodruff more personally.

The final deliverable included a website prototype that demonstrates both the sentiment visualizations and the recommendation suggestions. See the results at https://wilfordsjournal.streamlit.app/.