

Final Project Proposal: Textual Analysis of U.S. Presidential Speeches

The purpose of this project is to examine the development of themes and sentiments in U.S. presidential speeches throughout history. By analyzing these discourses, we intend to highlight patterns and practices that reflect shifts in political priorities, public sentiment, and rhetorical strategies.

The Main Question

How have themes and sentiments in presidential speeches changed over time, and what are the implications of these changes for the social and political landscape of different historical periods?

Purpose

Presidential speeches provide a unique perspective on the political and social climate of their time. Analysis of these discourses can reveal changes in political content, priorities, and the impact of major events. This work will enhance our understanding of how presidential rhetoric has changed in the context of changing national and global contexts.

Data Sources

The primary data source will be publicly available transcripts of U.S. presidential speeches. The main sources are:

- The American Presidency Project

Methodology

Data Collection

- **Source:** Collect transcripts from several presidents at various times in history, focusing on key speeches such as inaugural addresses and State of the Union addresses.
- **Procedure:** Organize transcripts in a structured format (CSV) with metadata, including lecture date, president, and description.

Data Preprocessing

- **Cleanup:** Remove comments (e.g., [applause], [laughter]), change text to lowercase, and remove punctuation.
- **Tokenization:** Split the text into individual words or tokens.
- **Normalization:** Remove stop words and lemmatize words to their base forms.

Sentiment Analysis

- **Methodology:** Apply VADER (Valence Aware Dictionary and sentiment Reasoner, here's a [link](#) to it) to classify the emotions in each speech as positive, negative, or neutral.
- **Objective:** Analyze how the overall sentiment of the speeches has changed over time.

Visualization

- **Sentiment Over Time:** Draw a graph of sentiment scores over time.
- **Comparison Between Presidents:** Create bar charts that compare each president's sentiment scores (positive, negative, neutral, compound).

Challenges and Solutions

- **Data Quality:** Ensure the accuracy and completeness of speech transcripts by cross-referencing multiple sources.
- **Historical Language Differences:** Address changes in language usage over time to ensure accurate analysis by standardizing text and using advanced NLP techniques.
- **Processing Efficiency:** Use optimized algorithms and data structures effectively.

Evaluation Metrics

- **Accuracy of Sentiment Analysis:** Use cross-validation methods to assess the accuracy of sentiment classification.
- **Clear Graphics:** Ensure the graphics are clear, informative, and effectively communicate the findings.

Expected Results

- A comprehensive study of the trends in themes and sentiments in presidential speeches.
- Insights into how presidential rhetoric was shaped by historical events and social and political changes.
- Visualizations clearly showing observations and trends over time.

Conclusion

This project will provide a comprehensive analysis of presidential speeches, offering valuable insights into the political landscape in the United States. By analyzing both sentiments and themes, we aim to highlight how presidents have addressed the nation's challenges and aspirations over time.