AutoDash

Team Members

- Bo Hu (bo12@illinois.edu)
- Bo Tian (botian3@illinois.edu)
- Weikun Wu (weikunw@illinois.edu)
- Yangliang Lu (yl164@illinois.edu) (Team lead)

What is our free topic?

Our free topic is the development of a dynamic dashboard for automated text information retrieval and sentiment analysis based on user-defined keywords.

Detailed Description and Task

The task involves creating an interactive web-based dashboard where users can input specific keywords like "Apple" or "Climate Change." The system will then automatically query a search engine to scrape the top relevant websites, preprocess the text data, perform sentiment analysis, generate text summaries, and provide insightful visualizations.

Evaluation

To evaluate the completeness of the project and if it meets the initial expectation. We can break it down into several components.

GREEN - DONE
RED - DID NOT COMPLETE
YELLOW - MODIFICATION

1. **Text Retrieval**: Use a search engine API or text database like Elasticsearch to find the top web pages based on the user-input keyword.
    - Create a prompt that initiates user input
    - Connect search engine API to scrape top results

2. **Web Scraping**: Extract textual content from these web pages.
3. **Data Preprocessing**: Perform text cleaning and normalization.
    - Preprocessing, Filtering
4. **NLP Analysis**: Conduct sentiment analysis and text summarization.
    - Topic modeling for keyword and topic analysis
        - The document may contain multiple topics
        - Each topic might have a unique keyword relevant to the overall topic.
    - Binary sentiment classification (positive or negative)
        - Identify what is the overall sentiment
        - Identify what words are more negative or positive
    - Summarization
        - A quick summary of the documents based on this search
        - We can also apply an overview per topic
5. **Visualization**: Implement a dashboard using Python's Dash framework to display the results.
        - Interactive keyword graph
        - A graph shows the overall sentiment(positive or negative)
        - A chart shows the sentiment distribution of the top words
        - A general summarization based on this search

Summary

Throughout the development process, the project underwent several visual modifications for enhanced functionality and user experience. The following adjustments were made to the original plan:

- The initial concept of an interactive keyword graph was refined to present a dedicated keyword graph for each identified topic.
- Rather than providing a general summary, it was determined that offering a distinct summarization for each document would yield more precise insights.
- It was concluded that an overall sentiment indicator would sufficiently reflect prevailing trends, making the analysis of individual word sentiments superfluous.
- To deliver more granular information, a summary for each linked page was implemented, eliminating the need for a broad general summarization.