Project Report: TextRulez Chrome Extension

Introduction

In the Fall 2023 CS 410 class at UIUC, the TextRulez team—comprising Timothy Crosling, Deepthi Abraham, Avi Nayak, and Srishti Sharma—developed a groundbreaking tool as part of their project for the Text Information Systems module. The project aimed to empower users by enabling them to extract key topics from web pages effortlessly. Leveraging advanced techniques in text analysis and an inverted index, the team created a Chrome extension with sidebar and popup functionalities. This report delves into the technical intricacies of the project, highlighting the core elements, methodologies, and functionalities implemented.

Technical Overview

Inverted Index and Text Analysis

Central to the project's functionality was the implementation of an inverted index, enabling efficient information retrieval from web pages. This index, coupled with text analysis techniques, facilitated the extraction of key topics. Tokenization, a process of breaking down text into smaller units, was employed along with regular expressions in JavaScript for pre-processing.

Integration of Python and Chrome Extension

A notable technical achievement was the successful integration of a Python script with the Chrome extension. The Python script, encapsulated within the `textAnalyticsServer.py` file, hosts a critical function named "analyze_content." This function harnesses SpaCy-based named-entity recognition to identify important entities within the text. By tallying entity occurrences, the system selects the top 5 entities by importance. Furthermore, it generates descriptions using GPT-based techniques and transmits HTML strings back to the Chrome extension client.

Deployment Instructions for Users

The team meticulously designed a user-friendly deployment process. New users are guided through a sequence of steps to ensure seamless installation and utilization of the extension:
- **Installing Required Libraries**: Users are directed to ensure the presence of necessary Python libraries by executing the command `pip install flask spacy numpy openai`.
- **Starting Python Server**: The Python server is initiated by running `python textAnalyticsServer.py`, ensuring proper functionality.
- **Configuring Extension**: Users navigate to `chrome://extensions`, toggle on 'Developer Mode', and load the unpacked extension from the local directory where the project is stored.

Demo and Code Overview

Video Demo Overview

The video demo showcased the practical application of the TextRulez Chrome extension. It provided a walkthrough for new users, demonstrating the step-by-step process of accessing and utilizing the extension on text-heavy web pages like CNN or Reuters articles. The demo emphasized the plugin's sidebar and popup capabilities, enhancing user experience when browsing content-rich web pages.

Code Overview

The project comprises several files, each serving distinct purposes:
- **textAnalyticsServer.py**: This file houses the core functionality, featuring the `analyze_content` function responsible for SpaCy-based entity recognition, entity counting, and GPT-based description generation.
- **HTML/CSS/JS/PNG Files**: These files are templates used to structure and generate the popup interface within the Chrome Extension

- **manifest.json**: Providing an overview of the extension's components and settings, this file includes permissions and action-based fields/attributes essential for extension functionality.
- **background.js**: Containing necessary listeners and chrome runtime actions, this file facilitates basic administration of the extension installation and communicates information collated from text analytics back to the Chrome extension.

_____

Conclusion

The TextRulez team's project showcases the seamless integration of text analysis techniques, Python functionalities, and Chrome extension development. By implementing an inverted index, leveraging sophisticated text analysis tools, and integrating diverse components, the team succeeded in creating a robust tool that extracts key topics from web pages. The detailed technical walkthrough and user-friendly deployment process ensure that users can maximize the benefits of this innovative Chrome extension, providing significant value in accessing and comprehending information on text-heavy web pages.