Daniel Vincent
CS 6320.001
Project Report

Text Summarizer

Project Methods
Running the file from the terminal with proper arguments will call main(), which parses the given arguments. The argument for the file to summarize is passed to get_file_contents(), which opens the file and reads it. The file contents are passed to summarize_text(), along with the optional arguments, and the text is summarized. The summarized text is then passed to write_summary_to_file(), which writes the summary to a .txt file that has the same name as the given file with the prefix 'summary_' attached.

Improvements from Proposal
- Reads from and writes to files instead of reading from an internal string variable and dumping the summary to the terminal
- Utilizes an LLM to improve efficiency and accuracy
- Supports multiple file types
- Allows for customization of summary using different summary length and LLMs as arguments

Contributions
Daniel Vincent
80 points - significant exploration beyond baseline
30 points - Innovation or Creativity: Initially had no knowledge of the subject
10 points - discussion of lessons learned and potential improvements

Challenges and Lessons
- Ran into issues where files would not be read correctly. Issue stemmed from corrupt metadata and file reading functions. Fixed by changing function syntax and creating sample files in the proper programs instead of vscode.
- Ran into issues where pdfs would not be read correctly. Switched to pdfplumber for file reading and text extraction instead of PDF2.