Program Overview:

Text Extraction and Analysis (Python)

A Python script, extract_words.py, processes a novel downloaded from Project Gutenberg. The script converts all text to lowercase and removes non-alphabetic characters to ensure consistent word analysis. It generates three text files:

The python script extract_words.py processes a novel and converts all the text to lowercase and removes non-alphabetic characters to ensure consistent word analysis. It then generates three text files:

- allwords.txt, containing every word in the novel (including duplicates), with one word per line.
- uniquewords.txt, containing words that appear exactly once in the novel. - wordfrequency.txt, which maps word frequency to the number of words that occur at that frequency, sorted in increasing order.

Unique Word Visualization

The DisplayUniqueWords function loads data from uniquewords.txt and displays a randomly selected subset of unique words on a fixed-size canvas. A custom font, fixed font size, and a palette of three colors are used to create a visually cohesive word cloud. Words are rendered with consistent spacing and constrained within the window to prevent overlap or overflow. Clicking the screen regenerates a new random selection of unique words.

We ran into some trouble with this one because of how the function was set up, each frame it would regenerate the graphics and take all the words away. To fix this we tried to take out the line that clears the GraphicsDevice each time, but this caused weird visual issues like the previous frame staying behind faintly. I tried for hours to fix it, but no luck. The function works as intended; it's just displayed weirdly.

Word Frequency Visualizatio

The DisplayWordFrequency function reads data from the wordfrequency.txt and generates a graphic depicting the relationship between word frequency and number of words with that frequency. This visualization highlights which words appear more or less often.

The MonoGame program starts by displaying the unique words visualization. Pressing the Enter key toggles between the unique word display and the word frequency visualization. To prevent rapid toggling, the program detects only the initial key press rather than continuous key holding.

Software and Libraries Used:

- Python3
- MonoGame
- .NET SDK

- Word Clouds

Workflow Distribution:

Caden - Worked Monogame function DisplayUniqueWords
Jacob H - Wrote Python script to extract text from files
Sahir - Worked Monogame function DisplayWordsFreauency
Jacob E - Wrote README and pdf