Assignment 1: Language Processing with NLTK, spaCy, and Sentiment Analysis

Objective

The objective of this assignment is to familiarize students with basic text processing tasks using Python, the Natural Language Toolkit (NLTK), and spaCy.

Prerequisites

- Python programming

- Basic understanding of text processing

- NLTK and spaCy libraries installed

Total Points

This assignment is worth 20 points.

Tasks and Questions

Task 1: Text Exploration (2 points)

1. Import the text corpus "Sense and Sensibility" by Jane Austen from the `nltk.corpus.gutenberg` module. (0.5 point)

2. Calculate and print the number of words and sentences in the text. (0.5 point)

3. Find the 20 most frequent words in the text and their frequencies. (1 point)

Task 2: Concordance and Similar Words (2 points)

4. Use the `concordance` method to find occurrences of the word "love" in the "Sense and Sensibility" text. (1 point)

5. Use the `similar` method to find words that appear in the same context as "love." (1 point)

Task 3: Dispersion Plot (2 points)

6. Create a dispersion plot for the words "love", "money", "marriage", and "society" in the text of "Sense and Sensibility." (2 points)

Task 4: Sentiment Analysis using TextBlob (4 points)

7. Extract the first 100 sentences from "Sense and Sensibility." (1 point)

8. Perform sentiment analysis on each sentence using TextBlob. (2 points)

9. Create a bar chart to visualize the number of sentences that are categorized as Positive, Neutral, or Negative based on the sentiment analysis. (1 point)

Task 5: Named Entity Recognition using NLTK (5 points)

10. Extract a paragraph of at least 100 words from "Sense and Sensibility." (1 point)

11. Identify and list the named entities in the paragraph using the `nltk` library. (2 points)

12. Categorize the named entities into types such as Person, Organization, Location, etc. (2 points)

Task 6: Named Entity Recognition using spaCy (5 points)

13. Extract another paragraph of at least 100 words from "Sense and Sensibility." (1 point)

14. Identify and list the named entities in the paragraph using the `spaCy` library. (2 points)

15. Categorize the named entities into types such as Person, Organization, Location, etc. (2 points)

Submission Guidelines

- Submit your code as a Python script (.py) or Jupyter Notebook (.ipynb).

- Include comments to explain your code.

- Make sure to test your code before submission.