Martin Luther King Jr. Speech Word Frequency Analysis

This project is a **basic Python web scraping and text analysis program** that extracts and processes the text of Dr. Martin Luther King Jr.'s speech from a webpage. It cleans and structures the text, analyzes word frequency, and exports the results as a CSV file.

Features

- **Web Scraping**: Fetches the text of Dr. King's speech from a public URL using requests and BeautifulSoup.
- **Text Processing**: Cleans and preprocesses the text to remove punctuation and standardize casing.
- **Word Frequency Analysis**: Breaks down the text into individual words and counts occurrences.
- **Data Export**: Saves the word count analysis as a CSV file.

Project Workflow

1. **Extract Speech Text**:
    - Uses requests to load the webpage and BeautifulSoup to parse HTML.
    - Collects all paragraph tags (<p>) where the speech text is located.
2. **Data Cleaning**:
    - Removes newline characters and punctuation.
    - Converts all words to lowercase for consistency.
3. **Word Frequency Calculation**:
    - Splits the cleaned text into individual words.
    - Uses pandas to count occurrences of each word and structure the data into a DataFrame.
4. **Save Results**:
    - Exports the DataFrame to a CSV file with columns for each unique word and its count.

Requirements

- **Python 3.x**
- **Libraries**:
    - requests
    - BeautifulSoup from bs4
    - pandas

Usage

1. **Run the Script**:
    - Make sure the required libraries are installed using pip install requests beautifulsoup4 pandas.
    - Execute the script, which will output a CSV file containing the word frequency count.
2. **View Results**:
    - Open the MLKJ_Speech_Counts.csv file to view each unique word and its count in the speech text.

Example Output

Word	Count
freedom	15
justice	10
dream	20
equality	8

The output CSV file MLKJ_Speech_Counts.csv shows each word along with its frequency in the speech.