The **Word Frequency Analyzer** is designed to process textual data and extract insights about word usage. The main goals are:

1. Preprocess text (cleaning, tokenization, stopword removal, lemmatization).
2. Compute word frequencies.
3. Visualize word frequency using bar charts and word clouds.
4. Save the results for further analysis or reporting.

Tools and Libraries

- **Python Libraries:** nltk, pandas, matplotlib, wordcloud, re, collections
- **NLTK Resources:** Stopwords, WordNet Lemmatizer (no punkt dependency required with Colab-safe regex tokenization)
- **Environment:** Google Colab

Pipeline Overview

1. **Text Preprocessing**
    - Convert text to lowercase.
    - Remove punctuation, numbers, and special characters.
    - Tokenize using regex (re.findall) for Colab compatibility.
    - Remove common stopwords (e.g., “the”, “is”, “and”).
    - Apply lemmatization to reduce words to their base forms (e.g., “processing” → “process”).
2. **Word Frequency Analysis**
    - Count the occurrence of each unique word using collections.Counter.
    - Store results in a **CSV file** for external use.
3. **Visualization**
    - **Bar Chart:** Shows top N frequent words for easy analysis.
    - **Word Cloud:** Provides a visual representation of word prominence.

4. **Reusable Function**
    - analyze_text(text, top_n=20, save_prefix="output") allows processing any text input with automatic frequency analysis and visualization.

Sample Output

Input Text:

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI)

that deals with the interaction between computers and humans using natural language.

Text preprocessing is an essential step in NLP tasks.

Word Frequency Table (Top 5):

Word	**Count**
Natural	2
Language	2
Nlp	2
Processing	1
subfield	1

Visualizations:

- **Bar Chart:** Highlights “natural”, “language”, and “nlp” as the most frequent words.
- **Word Cloud:** Visually emphasizes frequently occurring words for quick insights.

Key Features

- Colab-compatible preprocessing (avoids punkt errors).
- Stopwords removal and lemmatization for more meaningful analysis.
- Exports results to CSV for reporting and further processing.
- Modular design allows integration into larger NLP pipelines.

Future Improvements

1. Extend to **multiple documents or CSV datasets**.
2. Add **bigram or trigram analysis** for phrase frequency.
3. Implement **interactive visualizations** with Plotly or Bokeh.
4. Include **custom stopword lists** for domain-specific texts.

Conclusion

The Word Frequency Analyzer provides a robust and reusable tool for analyzing text data, generating frequency statistics, and visualizing key terms. The Colab-safe design ensures reliability and ease of use for any text dataset.