

Quotes Analysis Tool

This Python script is designed to analyze a TSV file containing quotes from various authors. It provides several features for analyzing and extracting information from the dataset.

Getting Started

These instructions will help you run the script and utilize its features.

Prerequisites

- Python 3.x
- Pandas (for data handling)
- NLTK (Natural Language Toolkit, for text processing)
- TSV file containing quotes (e.g., quotes.tsv)

You can install the required Python packages using the following commands:

```
pip install pandas  
pip install nltk
```

Before running the script, you'll need to download additional NLTK resources. Open a Python interactive session or a Python script and run the following commands:

```
import nltk  
# Download the 'punkt' tokenizer models (used for text tokenization).  
nltk.download('punkt')  
# Download the 'stopwords' corpus (common English stopwords for text processing).  
nltk.download('stopwords')
```

Running the Script

1. Place your TSV file containing quotes (e.g., quotes.tsv) in the same directory as the Python script.
2. Open a terminal or command prompt.
3. Navigate to the directory containing the Python script and your TSV file.
4. Run the script by entering the following command:

```
python quotes_analysis.py
```

Replace `quotes_analysis.py` with the name of your Python script if it's different.

Functionality

1. Count Quotes by Author:

- Enter the author's name to find out how many quotes they have in the dataset.

2. Longest Quote:

- Find the author with the longest quote and display the quote itself.

3. Authors with a Specific Word:

- Search for a specific word and discover which authors have quotes containing that word.

4. Word Frequency Analysis:

- Analyze the most and least frequent words used in all the quotes.

Customization

You can customize the script by editing the `quotes_analysis.py` file. Each feature has its own function and can be modified to suit your specific requirements.

Acknowledgments

- The script uses the Pandas library for data handling and NLTK for text processing.