README for app.py

Profiling app.py with cProfile and SnakeViz

Overview

This is Movie Plot Information Retrieval Project which is based on the searching of the movie plot in a given dataset.

The goal of our project is to analyze and extract meaningful information from IMDB dataset.csv dataset.

Our information retrieval algorithm returns similar movie titles based on an input query.

This README provides guidance on profiling the `app.py` file using cProfile and visualizing the results with SnakeViz. Profiling is a useful technique to analyze the performance of a Python script and identify areas for optimization.

Install Python and Necessary Libraries:

Use the following command to install the required libraries from the requirements.txt file:

```
pip install -r requirements.txt
```

dataset Link : "https://drive.google.com/file/d/1CKwVcjFXs2Jvbipspa6AWQZcymRRdX6F/view?usp=sharing";

Important:

If "tfidf_matrix.npz," "vocabulary.txt," and "processed_text.csv" files are absent, the project will process the dataset again and server takes some time to start.

If those files are already present then server won't take much time to start

so for those files Link : https://drive.google.com/drive/folders/1ZOTIWhdFp4Wa0UnD64SSlIn4iOB87UE7?usp=sharing

Note : project will run without these file also :"tfidf_matrix.npz," "vocabulary.txt," and "processed_text.csv"

Execute the main.py file to run the application command **"**python main.py**"**

Navigate to **http://127.0.0.1:5000/** to use the Information Retrieval system.

Search Functionality:

Enter a query in the provided input field.

Submit the query to retrieve relevant movie plots.

Example Query:

"A bartender is working at a saloon, serving drinks to customers. After he fills a stereotypically Irish man's bucket with beer, Carrie Nation and her followers burst inside..."

Profiling with cProfile

To profile the `app.py` file, you can use the `cProfile` module. Add the following code at the beginning of your script:

pythonCopy code

```
import cProfile # Your existing code in app.py if __name__ == "__main__": cProfile.run("main()",
filename="prof.out")
```

Replace `"main()"` with the actual function call or the part of your code that you want to profile.

Analyzing Profiling Results

To analyze the profiling results, you can use SnakeViz. Follow these steps:

1. Run the `app.py` script to generate the profiling data file (`prof.out`).
2. Open a new terminal, navigate to the folder containing `prof.out`, and run SnakeViz:

   bashCopy code

   ```
   snakeviz prof.out
   ```

3. SnakeViz will start a web server and open a new tab in your web browser with a graphical representation of the profiling data.
4. Explore the various tabs and visualizations provided by SnakeViz, including graphs, function calls, and cumulative time.

Additional Tips

- SnakeViz provides an interactive way to explore the profiling results. Use the mouse to navigate and zoom in on specific parts of the graph.
- The command `python snakeviz.py` mentioned in the comments is not necessary for profiling `app.py`. Use the instructions above to visualize the profiling data.