Design Laboratory (CS69202)

Spring Semester 2023

Assignment 3: Web Crawling, Extracting Information

Assignment date: January 10, 2024 and January 12, 2024

Important Instructions:

1. You need to design a menu-driven program to resolve user queries. We leave the design of the menu up to you, but keep in mind that your menu should be easy to use and address all queries. The user should also be able to go back to the previous menu.
2. Write Python code using PLY to extract the above fields. Your program should show all the possible query fields a user can ask for (from the above list items).
3. You must think correctly about what kind of errors can come in the process and try to handle them. Use the PLY package in python. PLY ref: https://www.dabeaz.com/ply/
4. You must NOT use any other parsing tools apart from PLY (ex: Beautiful Soup is a strict no or any other framework) . Should anyone not adhere to this instruction, they will be awarded ZERO marks.
5. Your code should address the objectives using PLY. Anyone found addressing the objective with no such use of PLY will be awarded ZERO marks.
6. Not adhering to these instructions can incur a penalty (worst case being 0 marks).
7. You can write a readme file to provide any particular instructions related to program execution steps, input format, or anything that you might think is useful for the evaluator while evaluating the assignment.
8. Create a log file to keep track of all input queries
9. Plagiarism in any form is not allowed. Students found copying/sharing code will be awarded 0 marks. You may discuss ideas, share your logic etc but you must not share/copy code at all costs.
10. All errors should be handled properly.
11. In case you make any design assumptions/choice, write a report along with the codes clearly stating the reason for your choice.
12. Submit the Assignments in <Roll>_CL2_A3.py and task as <Roll>_CL2_TS.py
13. Also submit a README file.
14. Save this in a folder named in the format: <Roll No.>_CL2_A3. Compress this folder to zip format, creating a compressed file <Roll No.>_CL2_A3.zip. Upload this compressed file to moodle. Example: If your roll no. is 22CS60R05, the folder should be 22CS60R05_CL2_A3, and the compressed file should be 22CS60R05_CL2_A3.zip.
15. Not adhering to these instructions can incur a penalty.

Day-2 (12/01/24)

Home Task:- Make a gpt api, tutorial link - https://docs.google.com/presentation/d/1rje4Jy72SbKFJVgHX80B92ICpF6bvNC5kZcI8bVapaQ/edit?usp=sharing [Send beforehand after 10/01/24 class]

Tutorial - 1:- using gpt-3.5 api for text generation. [2:00-2:40], template link - https://colab.research.google.com/drive/1WBKgknguWnbExCUd5L77X0VgsQ1OB27B?usp=sharing

Task - 1:- Use gpt-3.5 to generate a summary of a document. [2:40 - 3:10]

Problem:- You are given a document (max. 4000 words) in a text file {file path taken from command line}. Use GPT-3.5 to generate a summary of the document and store the summary in a text file. You also need to find the percentage reduction in summary. **Percentage reduction in summary** is the ratio between the number of words in summary and number of words in the actual document.