CS261-Data Structure and Algorithms 

Mid Project Proposal (Fall 2021)

**Proposer Details**

|  |  |
| --- | --- |
| Group Number | 62 |
| Registration Number of Group Members | 2020-CS-58  2020-CS-89 |

**Proposal Details**

|  |  |
| --- | --- |
| ***Project*** |  |
| Proposed Project Title | Cine-Scrap Project |
| Executive Summary | This project is entirely based upon the scrapping of movies. It works in a way that it extracts the relevant data provided by the user, scrapes it, and will move it on the list shown on GUI. The purpose of web scrappers is to collect as much data as it can. Same is the case with this project, It will scrap data of almost 1 million movies. It facilitates the user as much as it can i.e. It will ask user for the genre of movie:   * Comedy * Sci- Fi * Horror * Comedy-Romance * Fantasy etc.   Every movie has separate attributes through which 7 of them will be scrapped. The attributes for scrapping are:   * Title * Director * Ratings * Duration * Year * Type (Language) * Genre * Cast   Name of the movie/show  It is GUI based project. The Libraries used for this purpose will be:   * **Pandas** (in case to assemble the data in DataFrame) * **BeautifulSoap** ( that enables us to parse the HTML files ) * **WebDriver** ( that simply allows us to automate the chrome) * **Chrome.Options** (allows to manipulate the different properties of chrome)   The Library for GUI is **PyQt (**This allows flexibility for handling GUI events and makes the coding easier). There are different types of cinemas and theatres. This project will ask the name of the Cinema in which movie is being theatred. Cinema can be DS or Marvel or any required. |
| ***Business Case*** |  |
| Outline the business need for the project | Entertainment companies will be at benefit by movies scrapping |
| End user of the product | *Entertainment companies will use it in many different ways* |
| Motivation for Project | *This is an interesting project. analyze the primary factors that influence a movie’s success, measured by movie rating*. |
| State the level of impact expected should the project proceed and implications of not proceeding | *[State whether the implementation would have an impact at an operational level and/or strategic level and state the impact(s) in 2-3 lines]* |
| ***Technical Details*** |  |
| Name of Entity | movies |
| Attributes of Entity  (Minimum seven attributes/rows can be increased) | |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | *Title* | *Director* | *ratings* | *Cast* | *Year* | *Genre* | *Type* | *duration* | |
| Sample of Scrapping Source | *[Display the sample web page image from which information will be scrapped. Highlight the section in the page from where your each attribute will be extracted]* |
| Github Repository Link | https://github.com/eshatanvr/CS261F21PID62 |
| Sorting Algorithms | Selection, insertion, merge Sort, bubble, quick sort, k-select, |
| |  |  | | --- | --- | | **Algorithm Name** | **Description(Each algorithm in 2-3 lines)** | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | | |
| Searching Algorithms | *[List down the names of searching algorithms with description of each in 2-3 lines]* |
| Searching Filters for each data type | *If string : we can give a string and ask to search names having this string in title*  *Or by director’s name*  *Or genre*  *If integer: we can ask to display list of movies of that year* |
| Multi-Level Sorting | For example: first it will sorted by titles and then by it will get the movies sorted by years |
| Any other features | [Describe details of any other feature that you want to implement, or any bonus task] |
| ***Interfaces for your project*** |  |
| *[Draw layouts in the pencil tool. For each picture of the UI, provide the following table.]*   |  |  |  | | --- | --- | --- | | UI Component Name | Type of UI component | Purpose of UI Component/Other details | |  |  |  | |  |  |  | |  |  |  | |  |  |  | |  |  |  | | |