

Task 2

Introduction

Task 2 is to write a program that scraps three web pages to get the top three authors that have the most papers published in the years 2021 – 23 and then save the results in an Excel spreadsheet in the following format:

| | Person 1 | Person 2 | Person 3 |
|-------|----------|----------|----------|
| 2021 | 10 | 9 | 8 |
| 2022 | 9 | 8 | 7 |
| 2023 | 8 | 7 | 6 |
| Total | 27 | 24 | 21 |

Table of Functions

| Function Prototype | Description of its Purpose |
|--------------------|---|
| check_webpage | Takes a URL in the form of a string to get the text of the webpage. It returns the text of the webpage. If the webpage is not valid or does not exist, it terminates the program and displays an error message. |
| get_auth_pubs | Takes a list of author names and a list of authors for a given year. Returns the number of publications for the given year per author in the provided list. |
| get_clean_list | It takes the input webpage and returns a cleaned list of author names from the provided website. |
| new_wb | Takes a dictionary of authors and publications, writes them to a new Excel workbook, and saves the workbook to file. Returns the active sheet (openpyxl object) |
| get_top_auths | It takes a list of author names and returns the top three authors for the given years. |
| | |

Snapshots of Program Execution

None, the program has no visible anything when it runs. The program only outputs the Excel file.

Group Members

Josh Manchester - All roles