**Project Brief: Daily Report Generation for Kanopy Platform**

|  |  |
| --- | --- |
| Project Name | Daily Report Generation for Kanopy Platform |
| Date | 07/09/2024 |
| Project Client | Brian Young |
| Project Overview | This project involves developing a script designed to operate in the background. The script's primary function is to automatically download daily reports from the Kanopy platform for specified date ranges. Upon completion of the download process, the script consolidates all downloaded documents to compile comprehensive information.  By executing this script, users can streamline the aggregation of daily reports from Kanopy, facilitating efficient data compilation and analysis. |
| Goals  & Objectives | The primary goal of this project is to automate the retrieval and compilation of daily reports from the Kanopy platform. The objectives are:   1. Reduce Human Interaction: Minimize or eliminate manual intervention when interacting with Kanopy, specifically by automating the process of entering dates and downloading reports. 2. Automate Data Retrieval: Enable the script to autonomously retrieve daily reports from Kanopy for specified date ranges. 3. Enhance Efficiency: Streamline the process of compiling daily reports by automating both the download and consolidation tasks. 4. Ensure Accuracy: Improve the accuracy of report compilation by leveraging automated processes, reducing the potential for human error. 5. Facilitate Analysis: Provide a consolidated dataset of daily reports suitable for further analysis and decision-making purposes. |
| Constraints  & Assumptions | **Constraints:**   1. **User-Initiated Execution:** The script operates on a user-triggered basis rather than being scheduled, requiring manual initiation each time it is needed. 2. **Path Dependency:** The script assumes specific paths for storing downloaded reports on the user's system. It may encounter issues if the designated folders are absent or have conflicting names. 3. **Data Integrity:** It assumes consistent data format and availability from the Kanopy platform during each execution session, which may affect the reliability of the script if there are unexpected changes in data structure or format, specifically that the file format for the downloaded file is .xls. 4. **Manual Date Range Entry:** Users are required to manually alter the script to specify the desired date range for downloading reports from the Kanopy Platform. This manual intervention is necessary each time a different date range is needed. 5. **Credential Management:** Any changes in authentication credentials (e.g., username, password) for accessing the Kanopy Platform require corresponding updates in the script. Failure to update credentials may result in authentication failures and hinder report retrieval.   **Assumptions:**   1. **Folder Availability:** It assumes the absence of conflicting folder names in the specified destination path where new folders are created by the script to store downloaded reports. 2. **User Authentication:** It assumes users have valid credentials and permissions to access the Kanopy platform, including the ability to log in and retrieve the necessary reports. 3. **Error Handling:** The script assumes the ability to handle errors gracefully, such as network interruptions or unexpected server responses, to maintain reliable performance during execution. 4. **Execution Environment:** It assumes the script will execute within an environment compatible with its dependencies and configuration settings, ensuring consistent operation as intended by the user. |
| Project Scope | The scope of this project includes developing a script capable of automating the retrieval, downloading, and compilation of daily reports from the Kanopy platform across multiple years of date ranges. Key aspects of the project scope encompass:   1. Script Development: Designing and implementing a robust script that runs on a designated system to interact seamlessly with the Kanopy platform. 2. Functionality:  * Automated Retrieval: Enabling the script to autonomously retrieve daily reports from Kanopy for specified date ranges spanning several years. * Download Management: Ensuring efficient management of the download process, organizing files according to predefined paths on the user's system. * Compilation: Aggregating downloaded reports into a cohesive dataset or document to facilitate comprehensive analysis.  1. **Error Handling:** Incorporating robust error-handling mechanisms to manage exceptions and unexpected scenarios during script execution, ensuring reliable performance over extended periods. 2. **Dependencies and Environment:** Specifying necessary dependencies (e.g., libraries, frameworks) for script operation and ensuring compatibility with various operating system environments. (dependencies mentioned in the requirements.txt file) 3. **Documentation:** Providing comprehensive documentation that includes installation instructions, usage guidelines, and troubleshooting tips tailored to handling multiple years of date ranges effectively. 4. **Limitations:** Clearly defining constraints such as dependency on stable internet connectivity, assumptions about data format consistency over extended periods, and user permissions for accessing Kanopy. 5. **Testing and Validation:** Conducting thorough testing to validate script functionality across diverse date ranges and scenarios, ensuring accuracy and reliability in report retrieval and compilation. 6. **Deployment:** Facilitating straightforward deployment of the script on the user's system, ensuring it operates efficiently and effectively within the intended scope and environment. 7. **Maintenance and Support:** Outlining plans for ongoing maintenance, updates, and user support to address issues, enhance functionality, and adapt to changes in Kanopy's platform or user requirements over extended periods.   By encompassing the capability to handle multiple years of date ranges within the project scope, stakeholders can be assured of the script's versatility and suitability for comprehensive and sustained reporting needs from the Kanopy platform. |
| Technology Used | Technology Used:   * Programming Language: Python   Libraries Used:   1. os:  * Provides functions for interacting with the operating system, used for managing file paths. * After downloading the reports, change the name of the files according to the date it is associated to, and move the renamed filed into the daily report folders.  1. xlrd:  * Allows Python to read data and formatting from Excel files, facilitating the extraction of information from specific types of documents.  1. pandas:  * Offers data structures and operations for manipulating numerical tables and time series, utilized for data manipulation and analysis. * To insert the new date field in the downloaded report and thus then combine all the excel files into one combined report.  1. selenium:  * Automates web browsers to facilitate interaction with web elements and retrieve information from web pages.  1. webdriver:  * Specific to Selenium, used here with Chrome to initiate and control interactions with the Chrome browser.  1. common exceptions and support utilities:  * Used to handle various scenarios and ensure robust interaction with web elements.  1. time:  * Provides various time-related functions, used for adding delays or timeouts during script execution. * Added sleep times before each operation for the web page to load, and avoid any errors related to finding any elements on the webpage, as well as sleep times before every time a file is downloaded to avoid any overlaps or increasing overhead on the web browser.  1. pathlib:  * Offers classes and methods to interact with filesystem paths, used for managing and navigating file paths in a platform-independent manner.   Webdriver Used:   * Chrome WebDriver: Specifically chosen to automate interactions with the Chrome browser, enabling scripted navigation and data retrieval from web pages.   The folder also contains a requirements.txt file as well as a text file with the commands sed to execute the script. |
| Target Audience | The target audience for this project includes individuals within the collection department of the Library of the University of Mississippi who require data related to videos streamed on the Kanopy Platform. |
| Success Criteria | * Automated Data Retrieval: The script successfully retrieves daily reports from the Kanopy Platform without manual intervention, demonstrating automation of data retrieval processes. * Accuracy of Data: Ensures that the downloaded data accurately reflects the videos streamed on the Kanopy Platform for the specified date ranges, verified through comparison with platform records or through spot checks. * Efficiency Gains: Using manual efforts, this task would have taken approximately 36 hours (1 month = 1 hour) to enter data and download reports over a three-year period. In contrast, the script completed the task of retrieving and compiling three years of daily reports from the Kanopy platform in just 1.5 hours. Therefore, using the script saved approximately 34.5 hours of manual effort over the three-year period compared to performing the task manually, representing a significant reduction in time of approximately 95.8%. This substantial reduction underscores the efficiency gained by automating the retrieval and compilation of Kanopy platform reports with the script. * Reliability: Demonstrates consistent performance in retrieving and compiling data over extended periods, with minimal errors or interruptions. * User Feedback: Positive feedback from the collection department users indicating ease of use, effectiveness in retrieving required data, and reliability of the script. * Adaptability: Ability of the script to handle changes in date ranges and credentials effectively, minimizing the need for manual adjustments. * Documentation Completeness: Comprehensive documentation that is clear, understandable, and includes installation instructions, usage guidelines, and troubleshooting tips, ensuring ease of deployment and maintenance. * Compliance: Ensures the script complies with organizational policies and data security measures when handling sensitive information from the Kanopy Platform. |
| Budget | No monetary budget needed. |
| Timeline | No specified timeline defined prior to the project.  Timeline of the project: 07/07/2024 – 07/09/2024 |