### 

### **Given Assignment**

Here is the Assignment Details:

-make a selenium script that will do the below automation

- visit rokomari website

- click on লেখক Subcategory

- collect all লেখক লিস্ট

- collect all book details (title, price, description, review) including comments and faq of each লেখক।

- save the data into excel

- please use multithread if you need this.

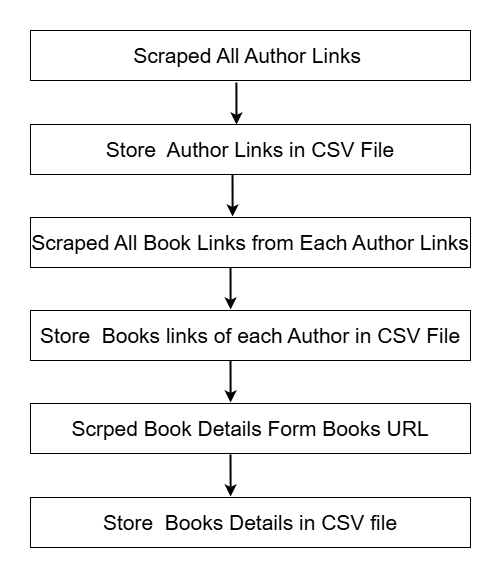
Submission Guideline:

- share your github

- make presentation video of your code with proper explanation in bangla.

### **Brain BrainStorming**

Visit: <https://www.rokomari.com/>



### **System Design**

1. **Headless Browser Setup** The system uses a headless Chrome browser to scrape data without opening a visible browser window. This improves performance and is suitable for automated environments like servers or CI pipelines.
2. **Multithreading for Fast Scrapping**
   * **Author Book Link Scraping:** Multiple author pages are processed concurrently using multithreading to collect book URLs faster.
   * **Book Details Scraping:** Multiple book URLs are processed in parallel to extract book details more efficiently.
3. **Book URL Status Tracking**
   * Book URLs are stored in a CSV file (book\_urls.csv) with an additional column named **Status**.
   * Initially, all book URLs are saved with the status **Pending**.
4. **Status-Based Processing**
   * The system loads only the book URLs with status **Pending** every time the script runs.
   * After successfully scraping a book's details, its status in the CSV is updated to **Completed**.
5. **Dynamic Scrolling for Book Details**
   * While scraping book detail pages, the script uses scrollHeight to perform dynamic page scrolling.  
      This ensures all dynamically loaded content (e.g., Q&A, reviews) is visible and retrievable by the scraper.
6. **Data Storage & Merge**
   * Book details (title, price, summary, comments, and Q&A) are saved to books\_data.csv.
   * New data scraped in subsequent runs are **appended** to the existing file, preserving previous results.
7. **Error Handling & Retry Logic**
   * Each book scraping attempt includes retry logic in case of failure.
   * Graceful error handling ensures the script continues running even if some books fail to load.

### 

### **Code**

<https://colab.research.google.com/drive/1QTUIORhxmStYIuxpAU6edfBH_bvoOxIY?usp=sharing>

### **Output**