# Cochrane Library System

**Shital V. Moradiya**

**01 November, 2022**

The Cochrane Library Scraper was developed to extract URLs and metadata from Cochrane Library reviews. The review information is formatted and saved to a file..

## Design Analysis

1.The Cochrane Library is made to be run from the command line and to be compiled using Maven. It scrapes the Cochrane Library website and collects the URLs and metadata from the reviews under each topic using a Runner class to call the static function scrape() of the Cochrane Library Scraper class.

2. This system Scraper sends a http GET request to the Cochrane Library topics page via this URL (https://www.cochranelibrary.com/cdsr/reviews/topics) using the Apache Http library. The JSoup library is used to parse the response String and scrape the topic names and URLs for each topic's search result page once it receives a positive response.

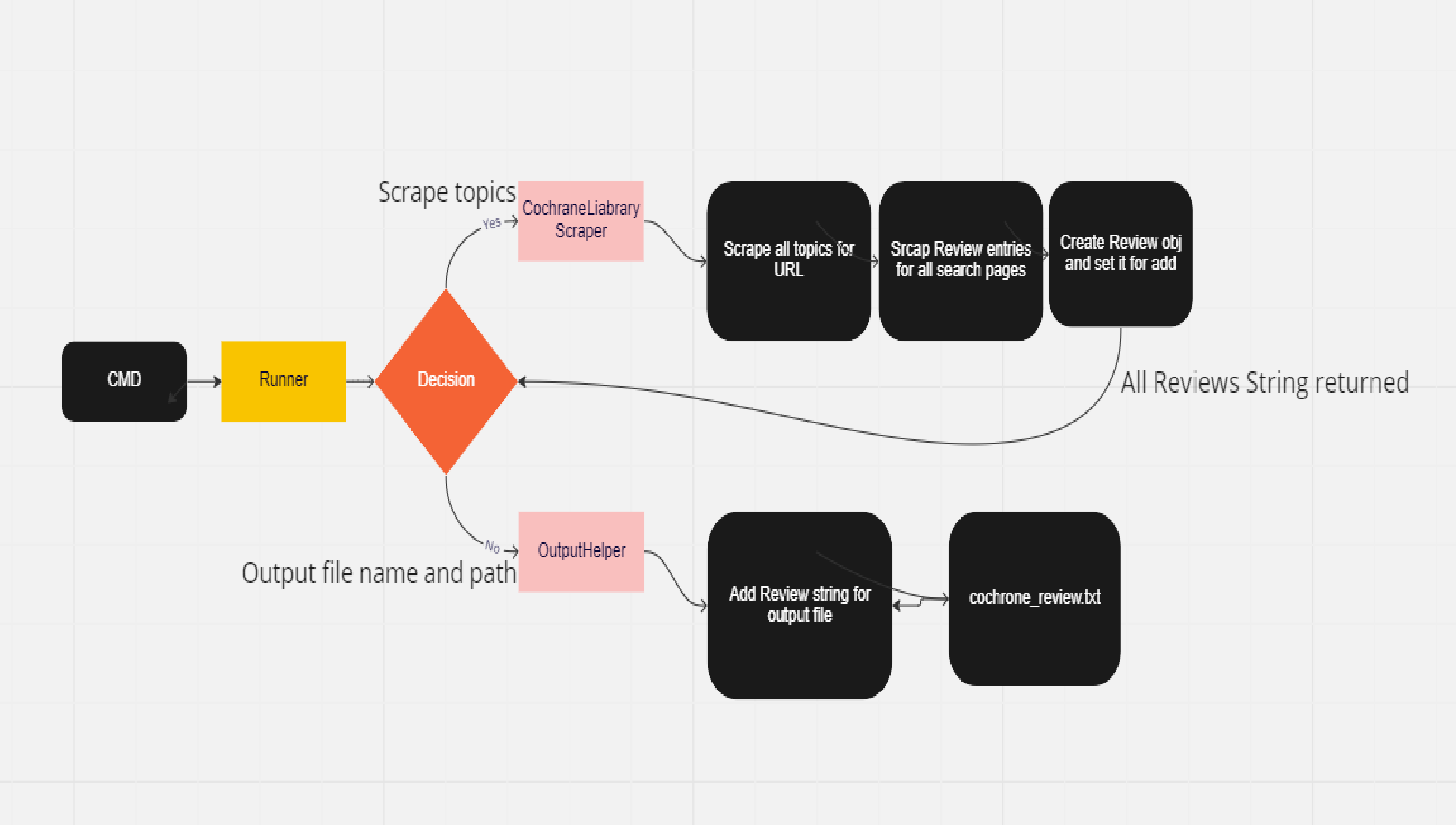
3. I picked the JSoup library because of how user-friendly it is, how clearly it selects HTML elements, and how easily it integrates with the Apache Http library.

4 . Each time a http GET request is made to the topic search results page, it loops through each topic (or, if multithreading is enabled, creates a new thread for each topic and executes the following logic on each thread). When it receives a favorable response, it will scrape the URL and other information from each review on the page (Each page by default has up to 25 reviews). Title, author names, publication date, and the name of the current topic are all included in the metadata. After each page's metadata and URLs have been extracted, review objects are created and added to a synchronized set. The scrape() method shuts down all connections and resources once all reviews from each topic have been collected.

5.The formatted review data String is returned to the Runner class, where it is passed as an argument to the OutputHelper class's static toFile() function, along with the other two command line arguments indicating the output file path and name. The formatted review data String is added to or replaced with the output file at the specified path by the toFile() function.

6.The program uses a custom Logger class to log information, debug information, and error information to the command line. This is used in place of System prints in order to be replaced by a more robust Logger class in whatever program it is used with.

# Flow Chart


Scrape topics
CochraneLiabrary Scraper
Yes
Scrape all topics for URL
Srcap Review entries for all search pages
Create Review obj and set it for add
CMD
Runner
Decision
All Reviews String returned
No
OutputHelper
Output file name and path
Add Review string for output file
cochrone_review.txt