

INFO 5731- Computational Methods

Exercise-2

Amulya Bodempudi

11645805

ParseHub:

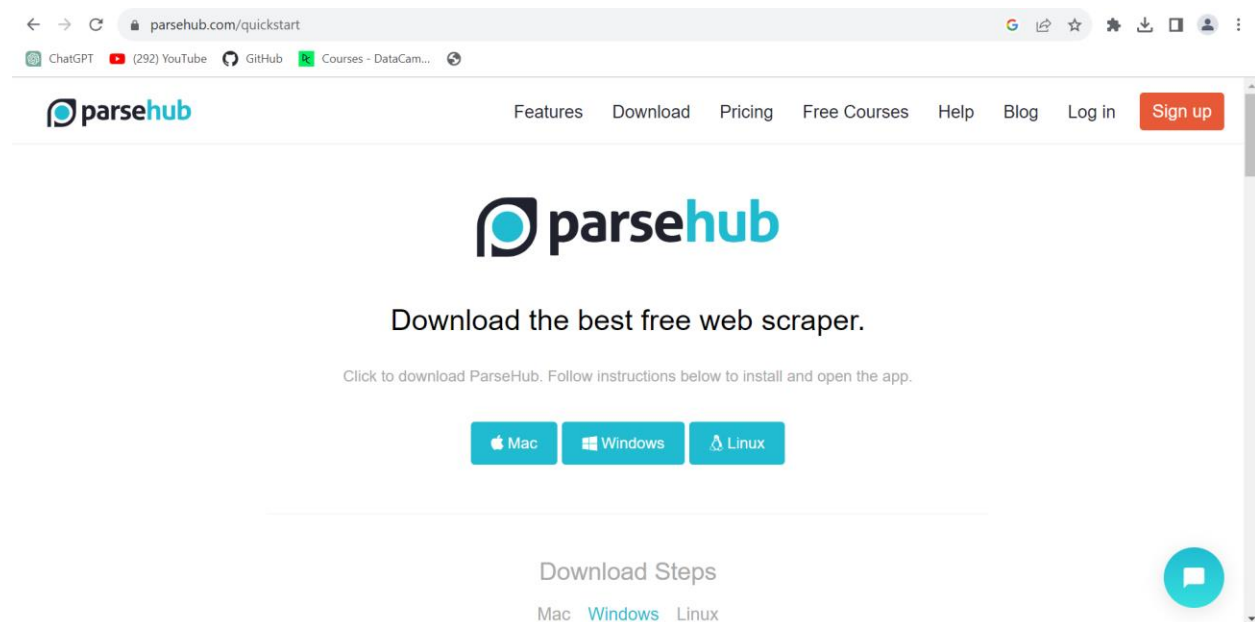
A user-friendly and visual online scraping application called ParseHub makes it possible for both non-programmers and developers to extract data from websites. In this document, I'll give an overview of ParseHub, lay out the procedures to be followed for web scraping, and present the results of the data that ParseHub has acquired.

The effective online scraping application ParseHub offers a point-and-click user interface for collecting data from websites. It is intended for users who do not want to spend a lot of time scripting but want to collect structured data from the web. In addition to a commercial version with more comprehensive features, ParseHub also offers a free version with constrained functionality.

Steps to Follow for Web Scraping with ParseHub:

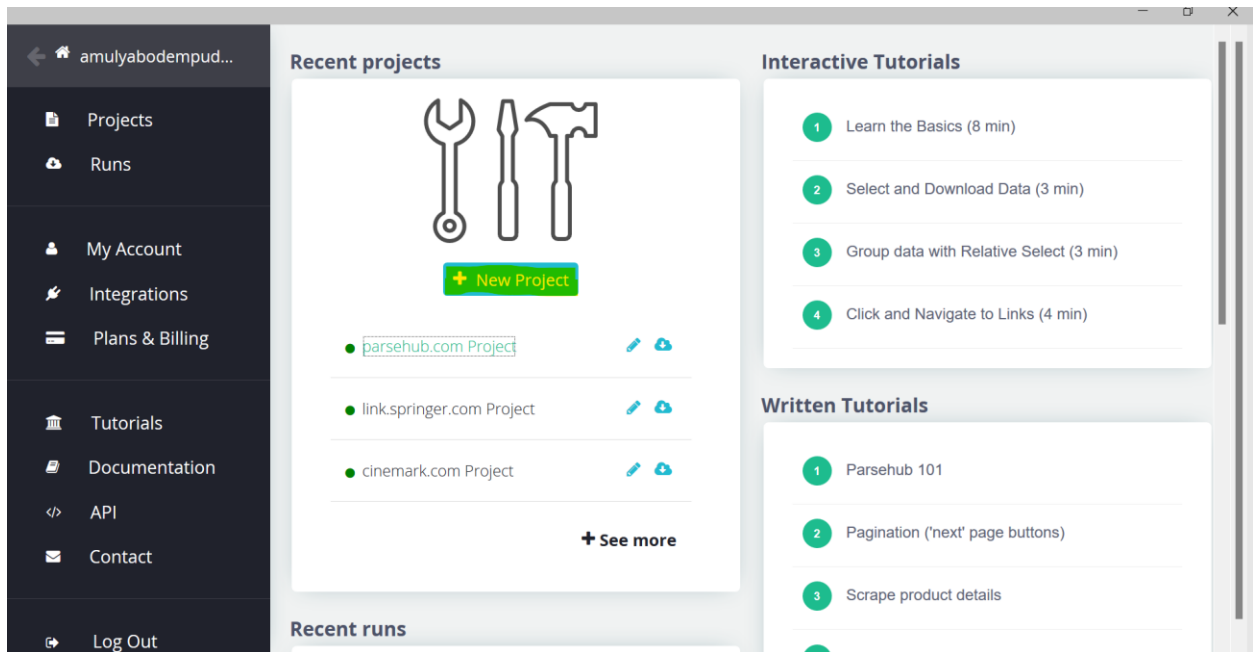
Here are the general steps to follow when using ParseHub for web scraping:

Install ParseHub Desktop Application:

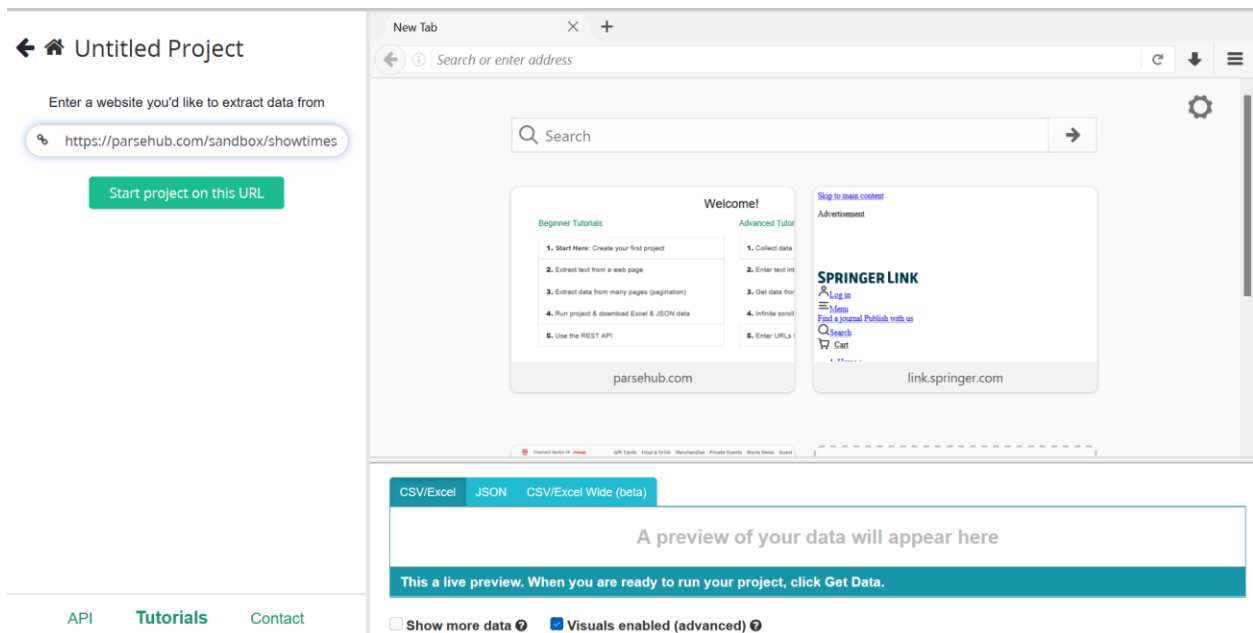


Select which device you are using and start by downloading and installing the ParseHub desktop application on your computer.

Create a New Project:



Open ParseHub, click on "New Project," and provide a name for your project.



Enter the link of any website, you want to scrape.

ParseHub: Showtimes

https://parsehub.com/sandbox/showtimes

parsehub Select Mode

Movie Showtimes (1 - 10 of 80 movies)

The Shawshank Redemption

IMAX 3D: 06:00 PM

Regular: 06:00 PM 06:15 PM 09:30 PM 09:45 PM

Schindler's List

IMAX 3D: 06:15 PM

CSV/Excel JSON CSV/Excel Wide (beta)

A preview of your data will appear here

This is a live preview. When you are ready to run your project, click Get Data.

API Tutorials Contact

Show more data Visuals enabled (advanced)

Select Data to Scrape:

Use the point-and-click interface to select the data elements you want to extract from the web page, so that ParseHub will create a selection. You can also create multiple selections for different data points also.

ParseHub: Showtimes

https://parsehub.com/sandbox/showtimes

parsehub Select Mode

Movie Showtimes (1 - 10 of 80 movies)

The Shawshank Redemption

IMAX 3D: 06:00 PM

Regular: 06:00 PM 06:15 PM 09:30 PM 09:45 PM

Schindler's List

IMAX 3D: 06:15 PM

Regular: 06:15 PM 06:30 PM 09:45 PM 10:00 PM

CSV/Excel JSON CSV/Excel Wide (beta)

selection1_name selection1_image

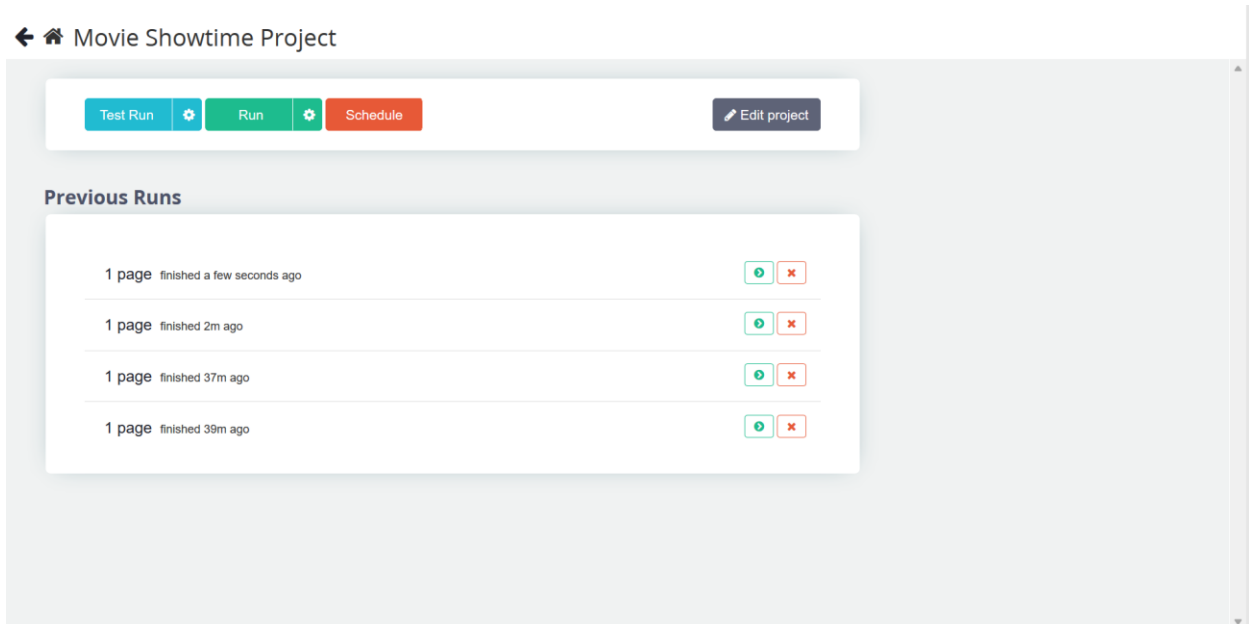
This is a live preview. When you are ready to run your project, click Get Data.

API Tutorials Contact

Show more data Visuals enabled (advanced)

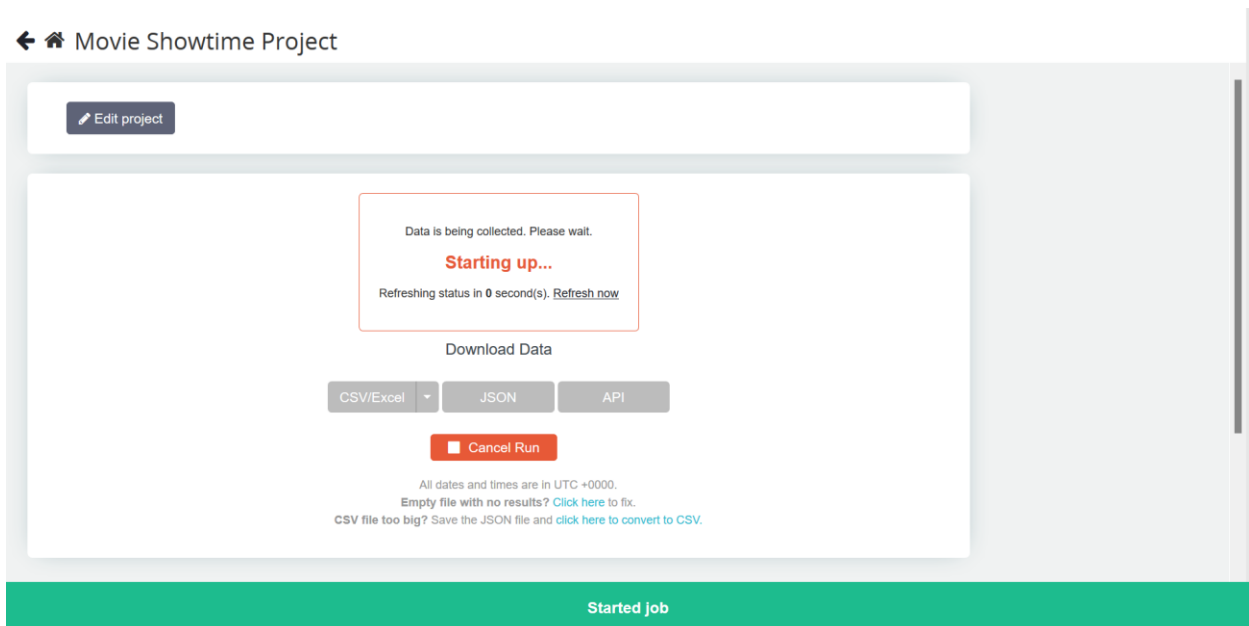
Run the Scraping:

Click on the green "Get Data" button to start the scraping process. So that the ParseHub will visit the web pages that to extract the data you specified, and compile it into a structured format.



Review and Export Data:

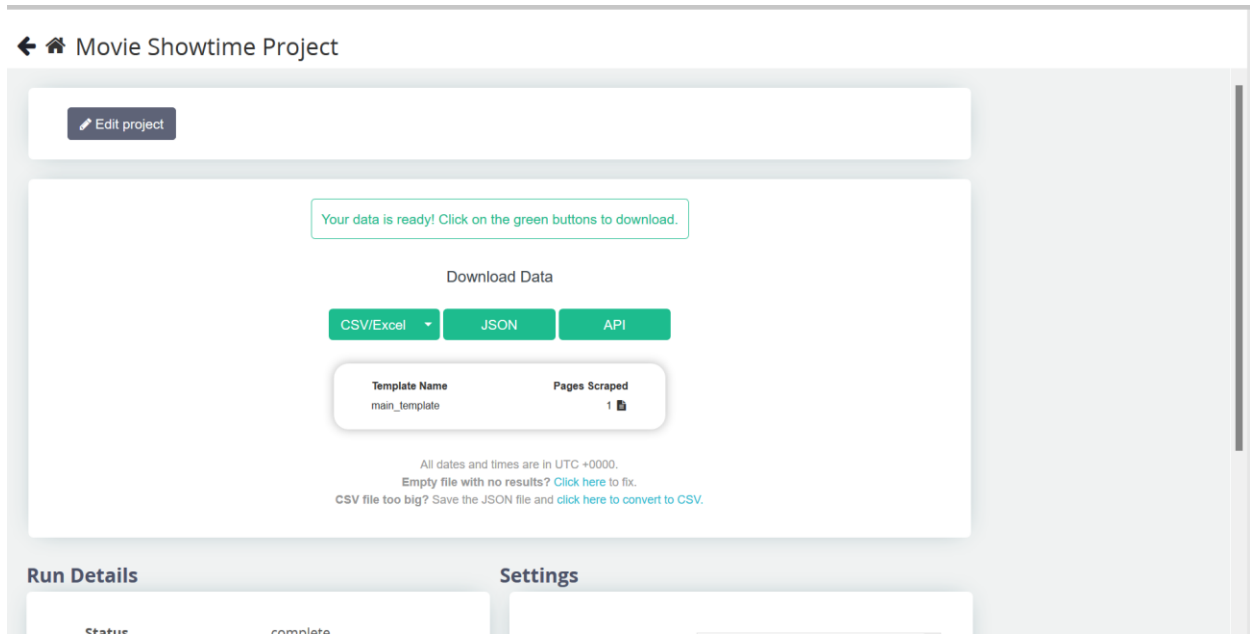
Once the scraping is completed, ParseHub will display the data in a table format.



You can Review the extracted data to ensure accuracy. Click on "Save & Run" to save the data as a CSV file.

Final Output:

A CSV file containing the structured data you took from the website will be the output's final form. The CSV file's columns correspond to the how many attributes you choose to scrape, and each row represents a single data point. This CSV file can be opened in Microsoft Excel or Google Sheets for more analysis, or you can utilize it for a variety of data-related applications.



By removing the need for scripting and offering a user-friendly interface to effectively extract data from websites, ParseHub makes web scraping simpler. Anyone wishing to get data from the web without having the technical know-how needed for conventional programming-based scraping can benefit from using it.