Mylavarapu Saketh
11596861

INFO 5731-Computational methods

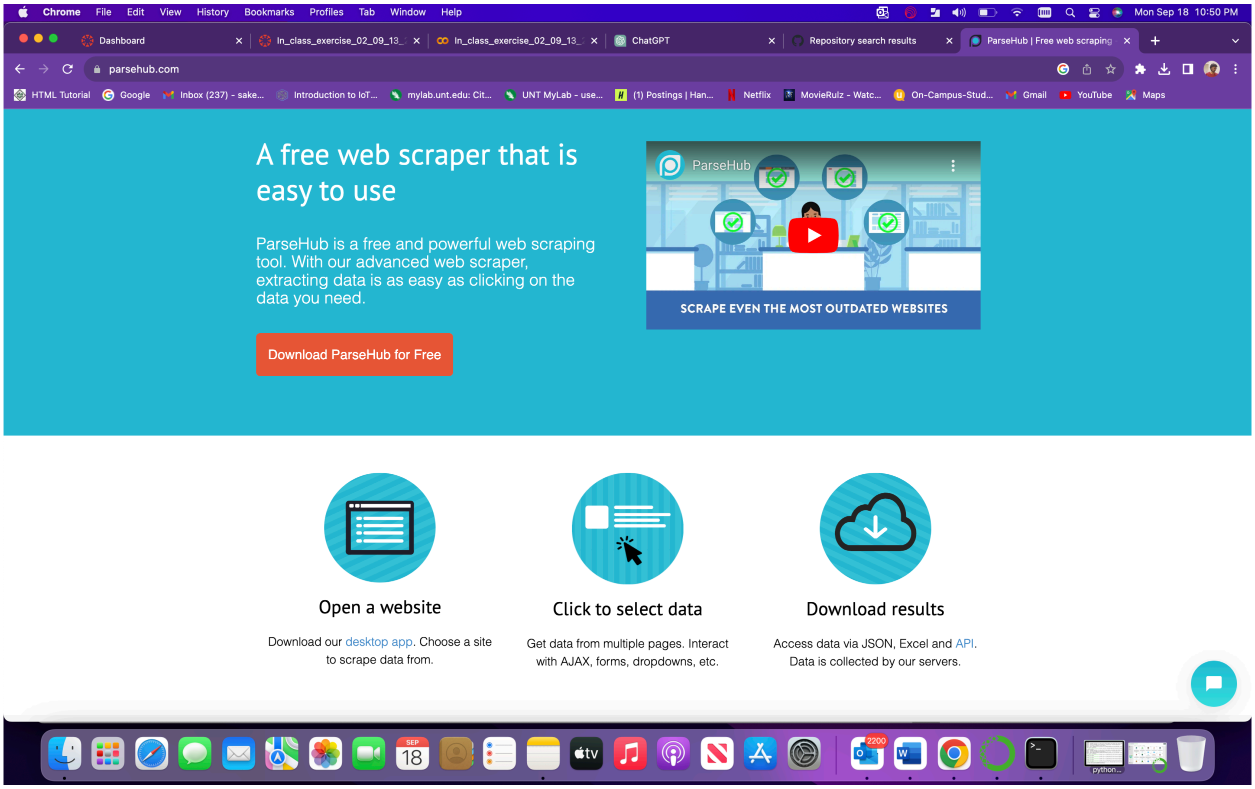
Exercise 02

Question 04:

ParseHub is a powerful and user-friendly web scraping tool that allows you to extract data from websites without the need for coding. It offers a point-and-click interface that makes it accessible to users with varying levels of technical expertise. In this guide, I'll provide you with an introduction to ParseHub, step-by-step instructions on how to use it for web scraping, and details on the final output of the collected data.

Introduction to ParseHub:

ParseHub is a web scraping tool that allows you to turn web pages into structured data. It is particularly useful for extracting data from websites with complex structures or dynamic content. Some key features of ParseHub include:



A free web scraper that is easy to use

ParseHub is a free and powerful web scraping tool. With our advanced web scraper, extracting data is as easy as clicking on the data you need.

Open a website

Download our desktop app. Choose a site to scrape data from.

Click to select data

Get data from multiple pages. Interact with AJAX, forms, dropdowns, etc.

Download results

Access data via JSON, Excel and API. Data is collected by our servers.

- Point-and-Click Interface: ParseHub's visual interface makes it easy to select the data you want to scrape from a website. You simply click on the elements you want to extract.

- Data Preview: You can preview the data as you scrape it, ensuring that you're getting the information you need.

- Scheduled Scraping: ParseHub allows you to schedule scrapes, so you can automate data extraction at regular intervals.

- Export Options: You can save the scraped data in various formats, including CSV, Excel, JSON, and more.

- Steps to Follow for Web Scraping with ParseHub:

- Install ParseHub: Start by downloading and installing ParseHub on your computer. It's available for both Windows and macOS.

Create a New Project:

- Open ParseHub and click on "New Project."
- Enter the URL of the website you want to scrape and click "Start Project."
- Select Data to Scrape:
- Use the point-and-click interface to select the data you want to extract. You can click on elements like text, images, tables, and links.
- ParseHub will automatically generate a sample of the data to confirm that you've selected the correct elements.
- Refine and Train Your Data Selection:

- If necessary, you can further refine your selection by adding more data points or using advanced features like conditional logic.
- Run the Scrape:

- Click on the green "Run" button to start the scraping process. ParseHub will visit the website and extract the data based on your selections.
- Preview and Modify Data:

- After the scrape is complete, you can preview the extracted data and make any necessary modifications.
- Export Data:

- Click on the "Download" button to export the data. You can choose to save it in formats like CSV, Excel, JSON, or others.
- Schedule Scrape (Optional):

- If you want to automate data extraction, you can set up a schedule to run the scrape at specific intervals.
- Final Output of Collected Data:

- The final output of the collected data depends on the format you choose during the export process. Commonly used formats include:

- CSV (Comma-Separated Values): Data is saved in a plain text format with values separated by commas, making it easy to open and analyze in spreadsheet software like Microsoft Excel.

- Excel: Data can be saved directly as an Excel spreadsheet, maintaining the structure and formatting of the original data.

- JSON (JavaScript Object Notation): Data is exported in a machine-readable format, which is useful for further data processing and integration with other applications.

- Other formats: Depending on your needs, you can also export data in XML, HTML, or other formats supported by ParseHub.

In conclusion, ParseHub is a versatile tool for web scraping that offers a user-friendly interface, and it allows you to easily extract and save data from websites in various formats to meet your specific requirements. Whether you're gathering data for research, analysis, or business purposes, ParseHub can streamline the process of web scraping without requiring coding skills.