Namal Institute Mianwali
Computer Science Department
Web Technology
Spring 22

Document:	Web Semester Project	Date:	02-06-2022
Prepared by:	Jawad Ahmed (BSCS-2019-59)	Prepared for:	Sir Ikramullah Khan Niazi
Document Details:	This document contains the details about how the project works and what its functionality is.		

Project Name:

Scraping Amazon Products Reviews System

Introduction:
- This project is a web based system for those who want to do online shopping which will provide simple and easy user interface to operate, which can be used by any online buyers by just copying the amazon product link and then paste it in the scraping amazon product review system so it will give collect all the reviewers review in one page and it will also provide the most frequents words used in the review text, so from this the online buyer can easily decide that shall I buy the product or not. The system will also keep the record of the product reviews that were searched by the buyer.

Problem Statement:
- Nowadays if we search for the product on amazon or some other ecommerce website there are sometimes a lot of product reviews, so reading each and every review is difficult, so it may be difficult for the online buyers to decide

whether I buy the product or not. Therefore, I will go scraping amazon products in order to make their life easy.

Acknowledgement:
- I would like to express my special thanks to my teacher Sir Ikram Khan Niazi who gave me the golden opportunity to do this wonderful project on the Amazon Product Scraping, it increased my practical knowledge in development.
- Another important thing is that I acknowledged that this project belongs to me, there is no such type of project available over the internet so I am the only owner of this project.

Project Details:
- **Functionality:**
  - In this project, There is a home page which is available for guest users and it can be used to scrape the product.
  - A user can be registered.
  - A user can be logged in.
  - On user dashboard the product searched will be stored in database
  - There is another functionality of displaying 15 frequent words used in the product reviews on the user Dashboard.
  - There is an admin dashboard which can be used to see how many users are registered and it can delete a user.

- **Flow Of Working:**
  - I used the MVC directory structure based on the ORM for the project.
  - The Server is separately run having no render template used.
  - There is a separate FrontEnd directory for the html pages.

  - **Run The Server:**
    - After running the server it shows below message on the browser:

    -

  - **Home Page:**
    - Open the home page from the frontend folder. It shows like:

Dashboard

Registration >

Login >

Welcome Guest User

☀ 18ᶜ Mianwali
Pakistan

Kindly Registered And Login
For Saving Your Results

Paste Amazon Product Link Below:

URL:

- Let's see if the **scraping** is working or not.
- First I do not paste any url and I click a scrape button so it give below message which is come from the server:

Paste Amazon Product Link Below:

URL:

Scrape

Missing data! OR Invalid Amazon Product Link!

-
- On the server side I also validate if there is other than amazon.com link then it will show the same message as shown below

URL:

https://realpython.com/the-model-view-controller-mvc-paradigm-summarized-with-legos/

Scrape

Missing data! OR Invalid Amazon Product Link!

○

- Now when I enter any amazon product link first it will show a wait message like:

URL:

https://www.amazon.com/HP-Generation-i5-1135G7-Graphics-15-dy2024nr/dp/B09FXFDGN3/ref=psdc_13896

Scrape

Wait Product Is Scraping

-
- After successfully scraping it give below result:

Product Sraped Successfully

Product Name:

HP 15-inch Laptop,

Scraped Product Reviews:

#	Reviewer Name	Reviews Text
0	Fuse	My job is Senior Business Intelligence Developer and Analyst, and I'm working remotely. To say this laptop i
1	jsand145	Let me preface this by saying, I normally use a work laptop for everything so I know absolutely zero about c
2	Ginger	I've had this laptop for less than 24 hours and there is no Wi-Fi capacity only VPN or Ethernet! In 2022 that

- So the server is scraping and sending the result in the json and then showing its result using ajax.

- Now lets see the **registration** is working or not:

Scraping Product Reviews

Sign Up With The Following Credentials

First Name

Tauqeer

Last Name

Ahmed

Username

Tauqeer2002

Email Address

tauqeer@gmail.com

Password

••••••••••

Register

Already have an account? Login

- After registration it will go for login using window.href in ajax below is the result in the database:

· Options
←T→ | ▼ | id | firstName | lastName | userName | email | password

		id	firstName	lastName	userName	email	password
☐ 🖉 Edit ⯮ Copy ⊖ Delete		3	Danish	Khan	danish2002	danish2002@gmail.com	pbkdf2:sha256:260000$totOMo3X2
☐ 🖉 Edit ⯮ Copy ⊖ Delete		5	jawad	ahmed	jawad2002	jawad@gmail.com	pbkdf2:sha256:260000$q4TQJDPs
☐ 🖉 Edit ⯮ Copy ⊖ Delete		6	Tauqeer	Ahmed	Tauqeer2002	tauqeer@gmail.com	pbkdf2:sha256:260000$Dgl6X6jhh

-
- I used the online free mysql database for this project. As we see above the entry is successfully registered with hashed/encrypted password generated by the server.
- Now **Login** the same user we enter using below interface:

Scraping Product Reviews

Sign in with

Email Address

tauqeer@gmail.com

Password

··········

☐ Remember me

Login

Don't have an account? Register

- After successful login it will go for the user dashboard which shows like below:

SPR   ≡

Dashboard
Home   >
Words   >
My Products   >
User Profile   >

Welcome User

☀ 18c Mianwali
Pakistan

Total Product Searched

10

Paste Amazon Product Link Below:

URL:

- Here the product will be scraped the same as shown in the above for the guest user but here after searching the product will be saved in the database which can be shown later in the my product section.
- Now lets finding the **frequent words** in the product review using the word section in the user dashboard, its interface is look like below:

- After putting the amazon product link it give below result:

- Now it's time to check the searched product details:
- Below is the interface for it:
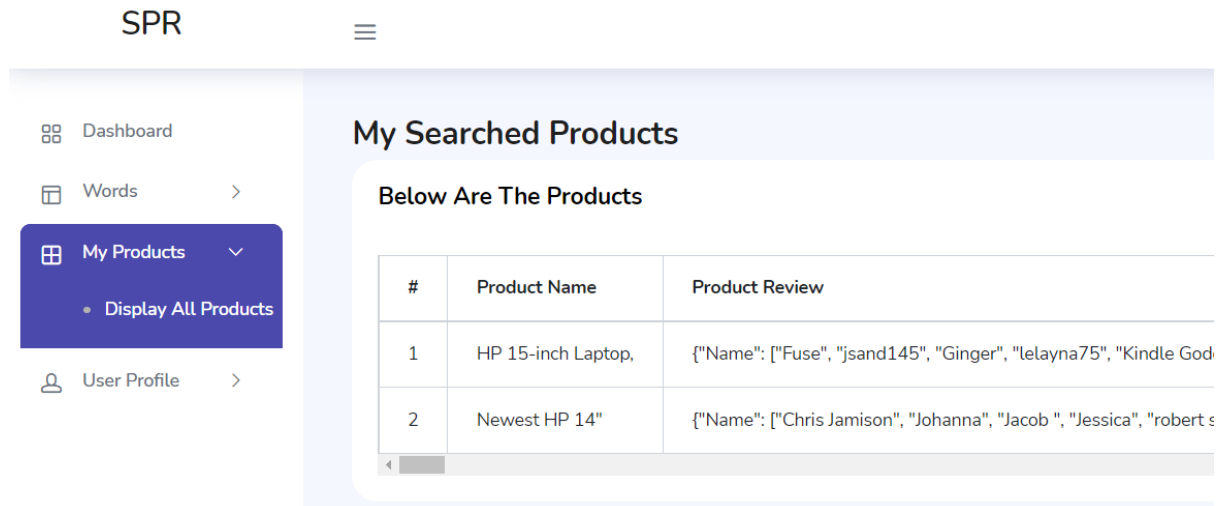
-
- When this page is open and the server is on then it will automatically get the product details saved in the database.

Now getting towards the **admin dashboard** in there is a admin account when I entered it in the login page so it will redirect to the admin dashboard which looks like below:
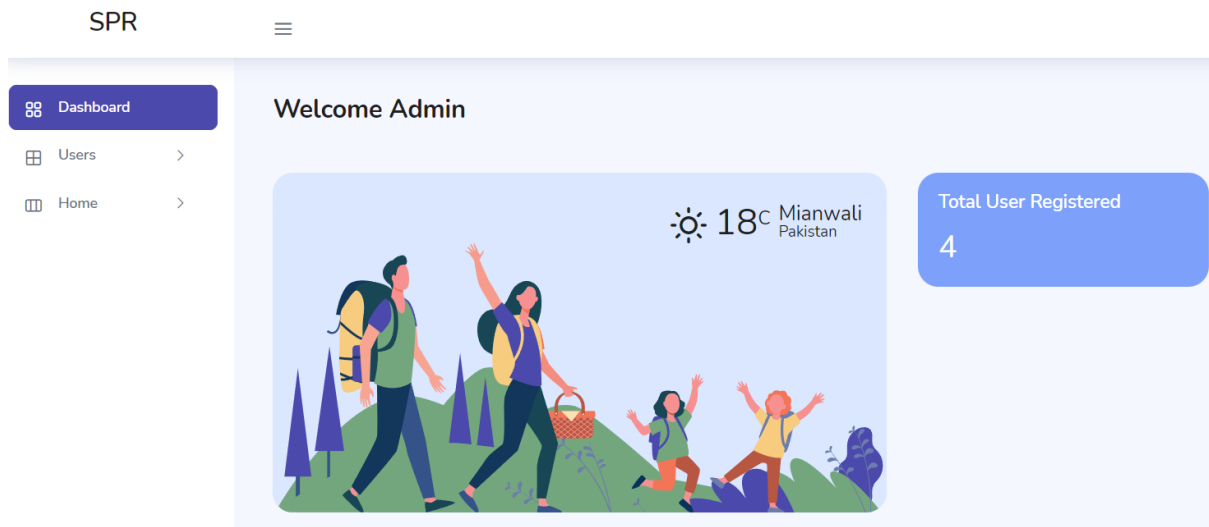
Now from the admin dashboard when I get towards the **Users** section so it looks like below which automatically get the registered user details when server is active:

SPR ≡

Dashboard

Users ⌄

• Display All Users

Following Are The User Registered

Below Are The Users Name And Email

#	First Name	Last Name	Email	Delete Btn
3	Danish	Khan	danish2002@gmail.com	Click Here To Delete
5	jawad	ahmed	jawad@gmail.com	Click Here To Delete
6	Tauqeer	Ahmed	tauqeer@gmail.com	Click Here To Delete
7	Jawad	Ahmed	jawad2019@namal.edu.pk	Click Here To Delete

- Now when I click the delete link it will delete the record from the database. But there is a little issue with page reloading although the data is removed from the database.

Conclusion:

- In this project I learned how to make the server and the client and how to send requests from the client towards the server using the ajax.
- Also I learned how to scrape the product reviews from the amazon website.
- There are also still issues with the project but it is almost completed.
- I also try with sessions and cookies but I was stuck because I used the server and the client separately.
- Overall the purpose of this project is achieved.

Confidential:

- **This project code and the report belong to me.**

Signature: Jawad Ahmed