

IMPLEMENTATION OF WEB SCRAPING FOR E-COMMERCE WEBSITE

Prof. P.S.Gaikwad, Kaushal Parmar, Rohit Yadav, Datta Supekar
Associate Professor ,Student, Student, Student
Computer Engineering Department,

All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, India.

Abstract : In this emerging world of the internet, there is lots of data present and retrieving this data becomes very complicated. As a result web scraping is one of the important method of data gathering. Web scraping is a technique of extracting data from various websites and depending on the tool end-users can access the data in several formats such as spreadsheet, csv, json, xml and database. Web scraping is used in many fields like e-commerce, market research, brand monitoring and etc. Our system proposes a method of fetching product data from e-commerce websites and comparing them. For extracting data different tools are used such as Scrapy, BeautifulSoup, Selenium, etc. Our system uses BeautifulSoup for extracting data. After extraction data is stored into MySQL database. This data is then displayed in a comparable format on our webapp. Visiting websites one by one and comparing product details is time consuming, so to overcome this our system will display all the product details from various websites, which will help the enduser to compare the products.

IndexTerms - Web scraping, E-commerce, Data extraction, Web crawler.

I. INTRODUCTION

The aim of the system is to extract the product details from multiple E-commerce websites and display the product details on a single web interface for comparison. Data extraction is a method of collecting data from various sources. For dynamic websites, it's very complicated to extract data because of its dynamic nature. This system is given input as a product name, which will be given to the various sites. Corresponding input is then searched on each site, relevant product information will be scraped and stored in the database. The stored data is then displayed on a website in a differential format through API. Our system uses MySQL as a database. Connection between python and mysql is done with the help of mysql-connector.

II. MOTIVATION

Our system is being proposed in order to make the user more convenient to compare products from different websites on a single website. The need of Web Scraping in today's world is increasing rapidly, because it allows quick and efficient extraction of data. As the user is not able to visit different websites at a time, it also consumes a lot of time and manual effort, so our system displays all the contents on a single website. We proposed a system which scrapes data from websites and compares the product details on a single website.

III. EASE OF USE



fig 1: Web Scraping

Human copy-and-paste :

A simple web scraping form copies manually and pastes data from a web page into a text file or spreadsheet. Sometimes even advanced web-scraping technology cannot replace personal testing and copying and pasting, and sometimes this can be the only effective solution when removal websites explicitly set up automation protection barriers.

HTTP-requests (HTML code review) :

Using this method, you will be able to find dynamic and static pages by sending HTTP requests to remote servers. This method uses sockets that organize all responses using pre-made data in target packages.

HTML parsing :

Many websites have large collections of pages generated dynamically from a less organized source such as a database. Details of the same category are usually entered on the same pages with a standard script or template.

DOM parsing :

Powerful content is one of the major challenges of web scraping. With embedding a full web browser, such as Internet Explorer or the Mozilla browser controller, programs can retrieve dynamic content created with custom scripts for clients. These browser controls also browse web pages in the DOM tree, depending on which programs can retrieve parts of the pages. Languages like Xpath can be used to scan an emerging DOM tree.

Computer vision web-page analysis :

There are attempts to use machine learning and computer simulations that try to identify and extract information from web pages by translating the pages into a personal appearance.

IV. SYSTEM ARCHITECTURE

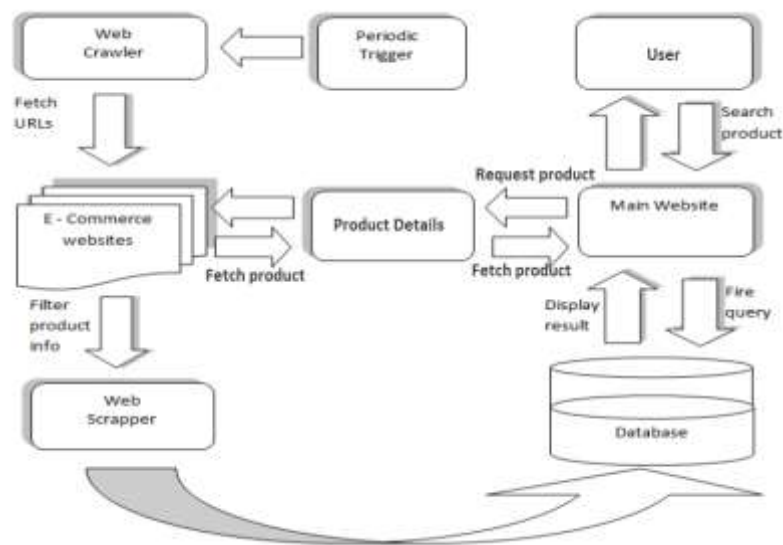


fig 2: SYSTEM ARCHITECTURE

V. IMPLEMENTATION

5.1 Project Modules

5.1.1 BeautifulSoup:

Beautiful Soup is used for extracting data out of HTML and XML les. It works with a parser to provide ways of navigating, searching, and modifying the parse tree.

5.1.2 Selenium:

Selenium is used to scrape a large volume of data such as text and images in a very short time. It uses the WebDriver protocol to control a web browser, like Chrome, Firefox or Opera. For cross-browser and end-to-end testing mostly selenium is used.

5.1.3 Mysql-connector:

It basically enables the python programmers to connect to the Mysql database.

5.2 Algorithm

1. Take product name as a input from user
2. Scrapes the product details
3. Stores details in database
4. Displays the information on the users window

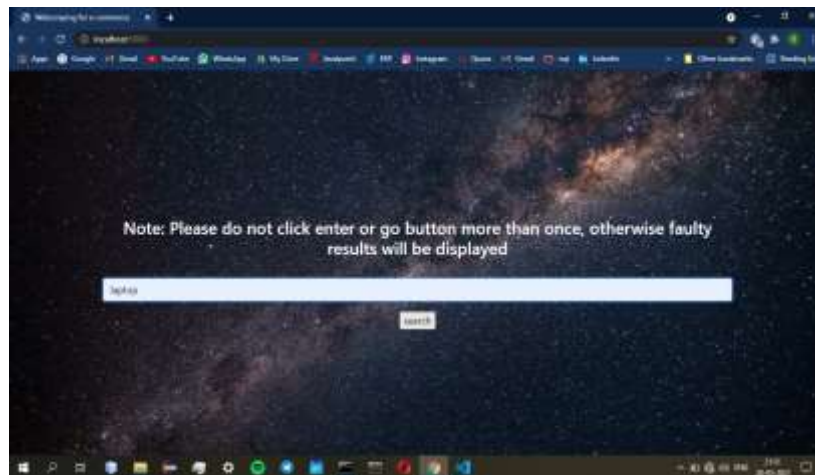
Algorithm Description:

While taking the product name as an input from the user, the input is then provided to the different E-commerce websites from which the data is scraped/extracted. As the data cannot be fetched directly from any website, so for that we require different web scraping tools/libraries. Our system uses Selenium and BeautifulSoup for extraction. Selenium is used for automating web based applications. Selenium has a webdriver which is used to control a web browser. Once the product name is entered by the user, using webdriver it automates to the particular website. As every website consists of HTML tags through which we can fetch the required data, for this purpose BeautifulSoup is used. When the web page is loaded on every website using webdriver, BeautifulSoup fetches the HTML tags of that page using class name and the required data of the respective tag is extracted.

The scraped data then is stored in Mysql database. Python and MySQL are connected using the mysql-connector module. The product details stored in the database are then displayed to the users screen.

VI. RESULTS

Comparison of product prices from different ecommerce websites and the result is displayed on a single website. The system also allows the user to analyze and compare product specifications of the products. To achieve this result the web miners download the required product details and the web scraper is used to extract the details of these products available from different websites. The system will allow users to redirect to the actual website of the user-selected product.



Name	Price(₹)	Image(Img)
Lenovo Ideapad Slim 1 150H Core i7(14th Gen) (15.6")	39999	
Lenovo Ideapad Slim 1 150H Core i5(14th Gen) (15.6")	44999	
Lenovo Ideapad Slim 1 150H Core i7(14th Gen) (15.6")	29999	
HP 15s Dual Core i5(11th Gen) (15.6") Laptop (MSRP)	28999	

Name	Price(₹)	Image(Img)
HP Pavilion Gaming Laptop (15.6") Core i7(11th Gen) (15.6")	124999	
HP Pavilion Gaming Laptop (15.6") Core i5(11th Gen) (15.6")	79999	
HP Pavilion Gaming Laptop (15.6") Core i7(11th Gen) (15.6")	114999	
HP Pavilion Gaming Laptop (15.6") Core i5(11th Gen) (15.6")	69999	

VII. CONCLUSION

Our system scrapes the data from different websites and compares the product detail on a website.

VIII. FUTURE WORK

1. Instead of displaying limited products, multiple products can be displayed per website.
2. Attempt to increase the accuracy of product details.
3. In the future it should be redirected to the specific products website when clicked on it.

IX. ACKNOWLEDGMENT

It gives us immense pleasure in presenting the research paper on "IMPLEMENTATION OF WEB SCRAPING FOR E-COMMERCE WEBSITES". The success and the outcome of this paper required a lot of guidance. We are very grateful to our guide Prof. P.S.Gaikwad who has provided expertise and encouragement. We thank mam who provided vision and knowledge that was very helpful throughout the research. All that we have done is only due to the great guidance.

REFERENCES

[1]Rabiyatou DIOUF, Edouard Ngor SARR,Ousmane SALL,Babiga BIRREGAH,Mamadou BOUSSO,Sény Ndiaye MBAYE,"Web Scraping: State-of-the-Art and Areas of Application " -2019 IEEE International Conference on Big Data (Big Data).

[2]David Mathew Thomas, Sandeep Mathur"Data Analysis by Web Scraping using Python"-Third International Conference on Electronics Communication and Aerospace Technology [ICECA 2019].

[3]Riya Shah, Karishma Pathan, Anand Masurkar, Shweta Rewatkar, Prof. (Ms.) P.N.Vengurlekar "Comparison of E-commerce Products using web mining"-International Journal of Scientific and Research Publications, Volume 6, Issue 5, May 2016.

[4] https://en.wikipedia.org/wiki/Web_scraping

[5] Fig.1: <https://images.app.goo.gl/WeSEFBWg6zeXn1tq9>

