

# SCRAPE THE PRODUCT DETAILS FROM E-COMMERCE WEBSITES

**Dr. M. Rajeswari<sup>1</sup>,**

Assistant Professor, Department of B. Com (Business Analytics),  
PSGR Krishnammal College for Women, Coimbatore, India.

[rajeshwarim@psgrkcw.ac.in](mailto:rajeshwarim@psgrkcw.ac.in)

**D. Pavyaasree<sup>2</sup>,**

UG Scholar, Department of B. Com (Business Analytics),  
PSGR Krishnammal College for Women, Coimbatore, India.

[pavyaasree067@gmail.com](mailto:pavyaasree067@gmail.com)

## ABSTRACT

The aim of this paper is to extract data from e-commerce websites for comparison using python. In this rising world of the internet, there are many eCommerce websites available it becomes hard to extract product details from these websites. Web scraping is one of the essential procedures for information amassing. Web scraping is a method of extracting information from numerous websites and depending on the tool quit-customers can get admission to the information in several codecs such as spreadsheet, CSV, JSON, XML, and database. Web scraping is used in lots of fields like e-commerce, market studies, logo monitoring and etc. This system proposes a method of fetching product records from eCommerce websites and evaluating them.

**Keywords:** Web Scraping, e-Commerce Websites, Beautifulsoup, Mining, HTML, Python.

## I. INTRODUCTION

The purpose of the paper is to extract the product info from a couple of E-commerce websites and display the product information 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 due to its dynamic nature. The popular idea in the back of internet scratching is to get better records that exist on a domain and convert it right into a configuration this is usable for analysis. Python is one of the usually applied programming dialects for data science ventures. As an end result of lots of facts overwhelming the Internet with enough facts, retrieving those facts turns very complex. Web scraping is an automatic method to obtain big amounts of records from websites. Most of these records are unstructured statistics in an HTML layout which are then transformed into established statistics in a spreadsheet or a database so that they can be used in various applications. Through this paper, we are using BeautifulSoup libraries to extract the data from the eCommerce websites. Utilizing python with BeautifulSoup makes web scrapping easier This will help information researchers gather and store information from site pages effortlessly without investing an excess of energy in getting ready datasets.

## OBJECTIVE

- To extract the required product details from the ecommerce sites using web scraping.
- The data collected are structured in a skilled data and then stored in the CSV format.

## II. RELATED WORKS

B. Liu and Y. Zhai, 2006, Web data extraction system, automatically and repeatedly extract data from internet pages and can supply the extracted facts to a database or a few other software. Web data extraction system is a wrapper generator that in particular designed with essentially three processes [2].

Broucke and Baesens, 2018, Web scraping became the only manner for an application to gain statistics from the internet till the latest emergence of APIs, which can be particular interfaces designed to make verbal exchange simpler for packages and servers. API are incredible tools that can provide the information in a prepared manner. API can be accessed for diverse records kinds as Wikipedia articles, tweets on Twitter, and plenty of greater [3].

Even though some websites give API, not all API is loose to use and additionally, and they're very constrained in terms of the way a lot of statistics they offer and what statistics they provide. Furthermore, the internet site developer's precedence is to focus on and keep the frontend interface over the backend API. To summarize, we cannot depend upon APIs for obtaining the net facts we want, so we want to use internet scraping techniques with the purpose of ensuring that the records we are becoming suit what the business or enterprise wishes.

E. Vargiu and M. Urru, 2012, Numerous scrapers had been written in various programming languages and frameworks are being used for retrieving internet information, which includes lovely soup, scrapy, Java, and Ruby, Beautiful soup is used to extracting banner ads from one-of-a-kind websites [5].

G. Boeing and P. Waddell, 2017, The information amassed from reserving business and big residences no longer reflect the present marketplace pastime and consequently, hides the latest information of rental markets in the US. The scraper has been designed and advanced to bridge this gap [6].

Hillen, 2019, Web scraping is already being used for scientific and business programs because it permits for the improvement of big, custom-designed information sets at minimum costs. It is supposed to update old strategies of records series as more corporations are looking to trap up on new developments that are observed by using clients [8].

Mitchell, 2018, Web scraping API which is a web scraping provider is on a larger scale, based on customized requests net scraping. It offers establishments with structured statistics with the aid of accessing scraped statistics in their purchaser the use of API [9].

Riya Shah, Karishma Pathan, Anand Masurkar, Shweta Rewatkar, Prof. (Ms.) P.N.Vengurlekar, Comparison of E-commerce products the use of web mining permits users to research expenses and get the preferred product at minimum price. Users can also pick out a couple of products that belong to the same class for comparing their functions. To acquire the best offers from e-commerce websites web crawlers and net scraping strategies are used to fetch certain facts [10].

## III. METHODOLOGY

Web scraping work comprises 3 stages. The first phase is the fetching section, the desired website that contains the applicable information has to be accessed – Flipkart and Amazon. The second segment of the work is to scrap the product information like product name, product rate, and ratings from eCommerce websites. The third section transforms these data right into a structured format.

### 1. Fetching Phase:

In this phase, the preferred website that contains relevant data has to be accessed. This is done through the HTTP protocol used to send request and receive responses. Content libraries are used to

get access to web page content. These libraries can be used in this phase by sending an HTTP GET request to the desired URL for getting the HTML document sent back in response.

**2. Web Scrapping:**

The utilization of Web Scrapping is to extract HTML data from URLs. In this work python library requests and beautifulsoup4 are used for performing web scrapping. To parse HTML pages a python library BeautifulSoup4 is used. The product information from the various website is scrapped and saved in csv format with the help of the library.

**3. Transforming Phase:**

In this phase, the unstructured data are transformed to a structured format. These structured data can be stored in the CSV format.

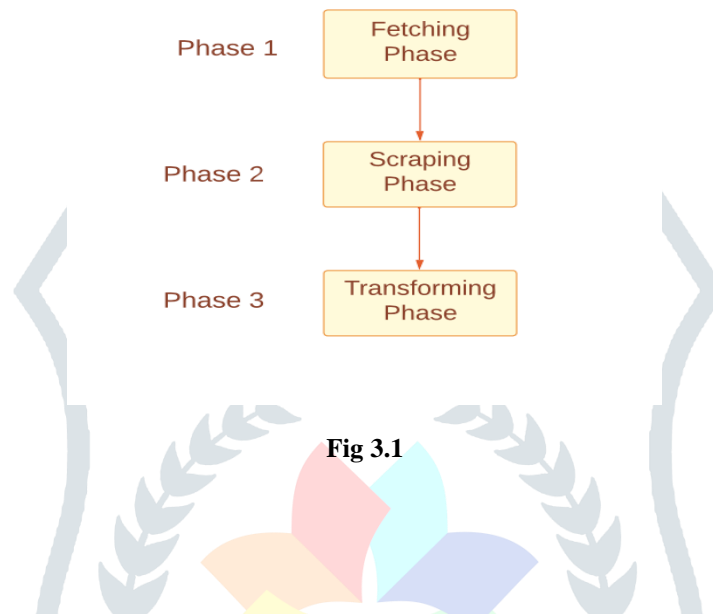


Fig 3.1

**WORK FLOW**

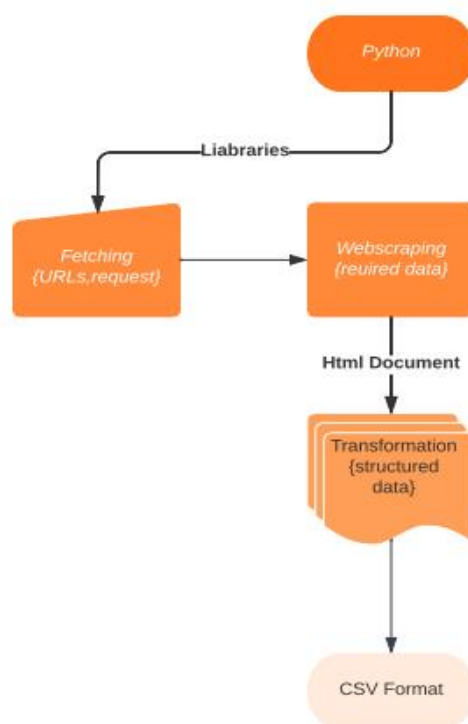


Fig 3.2

To extract records the use of internet scraping, these are the basic steps to be followed:

Step 1: In today’s online business development, there are numerous Ecommerce websites available on internet. In this step find the URL of the required e-commerce websites that will be scrapped.

Step 2: The data on the website is typically nested in tags. The selected eCommerce website pages are inspected to find underneath that the desired product falls.

Step 3: In this step, the usage of Request and BeautifulSoup libraries are used to extract the required data from the websites.

Step 4: Run the code to scrap and extract only the required information from the website.

Step 5: After extracting the information from the desired website, the extracted data can be saved in the CSV format.

IV. EVALUATION

In this work, the Flipkart and Amazon websites are taken as the two different e-commerce websites for web scraping. The end result of internet scraping is the extracted information of digital gadgets. Through web scraping only the required fields are scraped; they are product name, price and rating. Table 1, Table 2 and Table 3 show the scraped data from two different websites and they are stored as CSV files.

Table with columns: Product, Rating, Rating, Col Price, Product URL. It contains a list of product listings from Flipkart, including details like Dell Inspi, HP Core i, and ASUS Vnc.

Table with columns: Product, Rating, Rating, Col Price, Product URL. It contains a list of product listings from Amazon, including items like 2022 ASUS, Acer Aspire, and HP Stream.

Fig 4.1

V. CONCLUSIONS AND FUTURE WORK

Comparison of E-commerce products using web mining is a web-based system that will help users in decision-making while buying products online. Web scraping structures assure that the accrued records are exact, dependable, and have a better classification as the information present is in a gigantic which is hard to observe and keep up with. In this work, the eCommerce websites are scraped to extract the laptop product details and the extracted data are stored in CSV format. In the future, attempt to increase the

accuracy of product details. The extracted data has been visualized. With the help of Customized scrapping tools, the data collection and aggregation methods are easy. The developments of the tool are possible with open-source languages.

## REFERNCE:

- [1] Asma, “The Future of Web Scraping Services”. Retrieved from Information Transformation Services: <https://it-s.com/the-future-of-web-scrapping-services/> 29 September, 2021.
- [2] B.Liu and Y. Zhai, “Structured Data Extraction from the Web Based on Partial Tree Alignment”, IEEE Trans. Knowledge and Data Eng., vol. 18, no. 12, pp. 1614-1628, Dec. 2006.
- [3] Broucke, S. V., Baesens, B. (2018). “Practical Web Scraping for Data Science: Best Practices and Examples with Python”. (1st, Ed.) Apress.
- [4] David Mathew Thomas, Sandeep Mathur “Data Analysis by Web Scraping using Python”-Third International Conference on Electronics Communication and Aerospace Technology, ICECA, 2019.
- [5] E.Vargiu and M. Urru, (2012) “Exploiting web scraping in a collaborative filtering-based approach to web advertising,” Artificial Intelligence Research.
- [6] G.Boeing and P. Waddell, “New insights into rental housing markets across the United States: web scraping and analyzing craigslist rental listings,” Journal of Planning Education and Research, 2017.
- [7] Habib Ullah, Zahid Ullah, Shahid Maqsood, and Abdul Hafeez, “Web Scraper Revealing Trends of Target Products and New Insights in Online Shopping Websites”.
- [8] Hillen, J. (2019, November), “Web scraping for food price research”. British Food Journal, 121, 3350–3361.
- [9] Mitchell, R. (2018). “Web scraping with Python: Collecting more data from the modern web”. O'Reilly Media, Inc.
- [10] Riya Shah, Karishma Pathan, Anand Masurkar, Shweta Rewatkar, Prof. (Ms.) P.N.Vengurlekar, (2016, May) “Comparison of E-commerce Products using web mining”, International Journal of Scientific and Research Publications.