International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Visualization of Optimal Product Pricing using E-Commerce Data

N Greeshma, C Raghavendra, K Rajendra Prasad

ABSTRACT: ***With the number of e-commerce websites being increasing rapidly, online shopping has become the trend nowadays. Though, online shopping is very easy; however, when it comes to product selection it is a tedious task and time consuming to identify which online site gives us the best price and offers. Comparing products and filtering them from each online site is a very time consuming task for a buyer. This paper uses the techniques of Web Scraping using python libraries like Beautiful Soup, requests, matplotlib for identifying the best prices and for deciding the best product deal to the customer from different online websites. Web scraping is an automated technique of extracting data from websites. In this paper, real time data is extracted from various e-commerce sites and compared automatically. Finally, the results are graphically displayed based on which the customer makes the appropriate decision.***

Keywords: Web Scraping, e-commerce data extraction, python libraries.

I. INTRODUCTION

Generally, a web browser is used to search for information on the internet. Browsers offer a simple and easy way to view different websites and access them. Websites contain huge amounts of data which is in unstructured form. There is a lot of junk with useful data mixed on a website. So, to look only for the useful and appropriate information on the website relevant data extraction has to be done. This can be achieved by using the techniques of web scraping. The method of extracting information from websites is known as "Web scraping" [6].

Web scraping is also termed as "Web Harvesting" or "Web Data Extraction" or "Web Data mining". It is an automated technique used to extract large amounts of data faster and easier. These large amounts of data are collected and stored in a structured format (such as .CSV files, Databases). Across the world few commercial web page administrators describes web scraping is considered as legal and some don't. The legality of using web scraping completely depends on web page administrators only. If they agreed, then they allow people to access the data of a particular website [1].

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
N Greeshma*, Dept. of CSE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: greeshmanalla@gmail.com
C Raghavendra, Asst. Professor, Dept. of CSE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: crg.svch@gmail.com
Dr. K Rajendra Prasad, Professor and Head, Dept. of CSE, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India. Email: krprgm@gmail.com

Fig 1: Data from unstructured to structured format through Web Scraping

In the world, Internet businesses are easy to start and low risky to maintain. People prefer to establish an online store because of low tax, no crowd, more variety and early updates and so many. But the numbers of e-commerce services are increasing, this in turn results in the customers tending to spend a lot of time in deciding price, rating, features of the product and duration for delivery. Nevertheless, 54% of Internet used by people looking for data about merchandise or administrations, 48% data searches for educational purpose, 40% contents is searched for health and clinical data, 28% job seeking actions, and 24% data are searched for government and law administrative organizations [2]. This paper discusses one of the ways of extracting the data from the e-commerce websites and revealing to customer screen which helps them to sort out huge amount of irrelevant data. Web scraping can be implemented through many programming languages like Python, Node.js, PHP, Ruby, C++, etc. This paper uses the implementation of python language for Web Scraping, as python is more adaptive to further data processing; it is easy in implementation and also has many open source frameworks and libraries such as Beautiful Soup, Requests, Pandas, Matplotlib, etc.

II. METHODS

Python is an open source general-purpose language with great interactive environment. It is Object Oriented, Procedural and Functional which supports large number of modules and libraries [3].

Requests [10] library is used for making HTTP requests in python for accessing web pages. We can get the raw HTML of webpages which can then be parsed for retrieving the data.

Beautiful Soup [9] is a popular python library that parses a web page and provides a convenient interface to navigate the content. It pulls out data out of HTML and XML files. By simple commands, Beautiful Soup could parse content from within the HTML container [7]. It is considered as the advanced library for web scraping and can be installed in python by issuing 'pip install beautifulsoup4' in command prompt.

Matplotlib is a free, open-source and a friendly visualization library in Python for 2D plots of data arrays. It is a multi-platform data visualization library built on NumPy arrays.

One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in an easily digestible visuals. It has several toolkits to extend Matplots functionality [4].

Web Scraping generally works with the following steps:

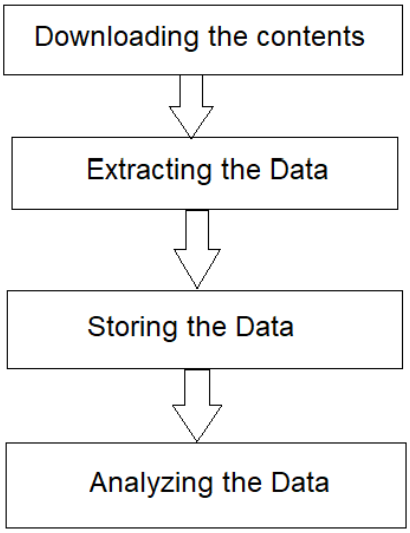

Fig 1: Web Scraping work flow

Step 1: Downloading contents from web pages

Find the URL of the web page to be scrapped. It can be any active website link on which data extraction can be performed. Required contents can be downloaded by requesting the contents from the webpage.

Step 2: Extracting the Data

The downloaded contents are HTML and mostly unstructured form. Hence, the library Beautiful Soup is used in parsing the data and extracts structured data from the downloaded contents.

Step 3: Storing the Data

The extracted data is then stored and saved in any structured format like CSV, Database or JSON.

Step 4: Analysing the Data

The web scraper will analyze the data that is stored. The stored data can be transformed by writing the code using different python libraries.

III. IMPLEMENTATION

A case study:

In this paper, two e-commerce websites are chosen for scraping the prices of a product (smart phones, tablets, laptops, etc.). As the data is usually nested in HTML tags, analyzation of e-commerce website is done using the Inspect element (In chrome browser, the shortcut key for inspecting a web page is Ctrl+Shift+I). So, inspect the web page to see under which tag the data we require to scrape is present. For the data to be scraped from a website we implement few lines of code using python. Libraries like Requests, Beautiful Soup are imported and also some other data analytic libraries like numpy, Pandas, etc. are also used in the code for data manipulations. Python Requests module library is used for making HTTP requests that are sent to a website to get the raw data (HTML) from the web page. Then the Beautiful Soup library is used for parsing the HTML data. Finally, Data extracted from the code can be stored in a required format such as CSV files, databases, etc. The data from these can be easily retrieved at any point of time as it is structured and stored in local database or CSV file.

In any e-commerce site data changes continuously take place i.e., prices of products may vary at any moment or some products may be out of stock too. This frequent data changes should also be considered dynamically in the python code.

Large data is extracted dynamically and stored in CSV. The data stored in CSV format is then plotted by using the matplotlib library in python for visualization of the extracted data. In this implementation, only a sample of data is taken and considered for plotting the graph. Hence, a comparison bar plot is made from these two e-commerce sites by taking prices on ordinate and product names on Abscissa as coordinates in fig 2.

The data can be visualized through various ways like bar graphs, scatter plots, line plots, pie charts, etc. In this current study, bar graph based price comparison is made [11].

The optimum results are of the website1 and website2 are tabulated as below. By the following table customer can decide the best available price on different websites of a particular product

#	Product_Name	Website1_Price	Website2_Price
1	Realme C2 (Diamond Black| 16 GB)	Rs.5,999	Rs.4,599
2	Realme 3i (Diamond Blue| 32 GB)	Rs.7,999	Rs.7,599
3	Realme 5 (Crystal Blue| 32 GB)	Rs.9,999	Rs.8,999
4	Realme 3i (Diamond Red| 64 GB)	Rs.9,999	Rs.9,999
5	Realme 3 (Diamond Red| 32 GB)	Rs.9,999	Rs.9,999
6	Realme 2 Pro (Blue Ocean| 64 GB)	Rs.10,599	Rs.9,999
7	Realme 5 (Crystal Purple| 64 GB)	Rs.10,999	Rs.10,999
8	Realme 5 (Crystal Blue| 128 GB)	Rs.11,999	Rs.12,999
9	Realme X (Space Blue| 128 GB)	Rs.16,999	Rs.16,699

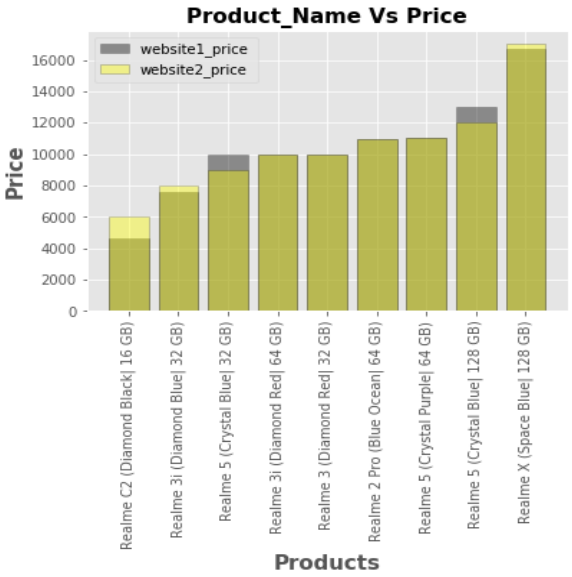

Fig 2: Price comparison of a product from different e-commerce websites

In the above obtained graph, Black color indicate price data from ecommerce website1 and Yellow color indicate price data from e-commerce website2. If the price in e-commerce site1 is greater than e-commerce site2 the top of the bar is black in the graph as black represents price of website1. The Yellow color on the top of the bars represent that website2 prices are greater than website1. In another case, if the price for a product is the same in both the e-commerce sites then the colors black and yellow are mixed and shows a different color i.e., green(having same bar plot for both the e-commerce sites). In this way, price comparison visualization is done by automated method through web scraping between two e-commerce sites which lets the user decide the optimum cost [12].

IV. CONCLUSION

This work presents a novel collaborative Price filtering by using Web scraping techniques to shopping sites. To show how the system works in practice, a suitable case study is explained, i.e., how price comparison is done from e-commerce sites. Suggest price comparison from different online sites. Not only from websites, data can also be scraped from social media, job boards, travel websites, etc. Data availability on online is consistently increasing day-by-day. This huge amount of data is stored in computers, databases, etc., for the convenience and easy retrieval process. As massive amounts of data available on the web, Web Scraping and data analysis plays a vital role in the coming future for data extraction. But the people who do scraping should take into account that they are not breaking any kind of law which could make them liable for any offence. So far with best knowledge the data extracted in this paper is done in a completely legal way and from open source websites.

REFERENCES

1. Legality and Ethics of Web Scraping; Twenty-fourth Americas Conference on Information Systems, New Orleans, 2018.
2. Shakra Mehak, Rabia Zafar, Sharaz Aslam, Sohail Masood Bhatti. "Exploiting Filtering approach with Web Scrapping for Smart Online Shopping: Penny Wise: A wise Tool for Online Shopping", 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2019.
3. Python – The Fastest Growing Programming Language (IRJET) Volume: 04 Issue: 12, Dec-2017 e-ISSN: 2395-0056.
4. https://www.geeksforgeeks.org/python-introduction-matplotlib/#targetText= Matplotlib% 20is%20 an%20amazing% 20visualization, Hunter%20in%20the% 20year%202002.
5. Increased Information Retrieval Capabilities on e-Commerce Websites using Scraping Techniques 978-1-5386-2182-0/17/ IEEE , 2017 International Conference on Sustainable Information Engineering and Technology (SIET).
6. https://www.edureka.co/blog/web-scraping-with-python/
7. C. Raghavendra, S. Sivasubramanian, A. Kumaravel, "Improved image compression using effective lossless compression technique", Cluster Computing, March 2018.
8. https://en.wikipedia.org/wiki/Web_scraping
9. https://www.crummy.com/software/ BeautifulSoup/bs4/doc/
10. https://realpython.com/python-requests/
11. A. Kumaravel, S. Sivasubramanian, C. Raghavendra "Features Subset Selection using Improved Teaching Learning based Optimization (ITLBO) Algorithms for Iris Recognition", Indian Journal of Science and Technology, Vol 10, Issue 34, 1-12, September 2017.
12. K Rajendra Prasad, C Raghavendra, A Kumaravel, "Study on Secure Health Care Data using Big Data Analytics", Jour of Adv. Research in Dynamical & Control Systems, Vol. 10, 10-Special Issue, Aug 2018.

AUTHORS PROFILE:

Ms. N Greeshma is studying her B.Tech in CSE at Institute of Aeronautical Engineering, Hyderabad. She is interested in areas like Data Science and Artificial Intelligence. She also completed her internship in IOT applications using Data Science advanced python techniques.

Mr. C. Raghavendra Graduated in B.Tech (CSE) from Jawaharlal Nehru Technological University, Hyderabad in 2008. He received Master's Degree in M.Tech (CSE) from Bharath INSTITUTE OF Higher Education and Research (BIHER), Chennai, in 2011. He pursuing a Ph.D. in Computer Science & Engineering from BIHER, Chennai. Presently, he is working as Assistant Professor of CSE Dept., Institute of Aeronautical Engineering, Hyderabad. He has published more than 20 Publications in various International Journals and conferences. His research interests are Image processing & Security, Big data and Databases. He is a life member of ISTA, CSI, IAENG, CSTA professional societies.

Dr. K Rajendra Prasad is professor and head in the department of CSE. He received B.Tech degree in CSE from JNT University, Hyderabad; M.Tech degree in CSE from VTU, Belgaum; Ph.D. in CSE from JNT University, Anantapur. He has 14 years of indispensable experience in teaching. Previously he served as a member in Board of Studies where he was actively involved in TEQIP work. He has published 15 papers in various international conferences and 8 papers in reputed journals. He is a member of CSI, IEEE professional societies. His major research interests are Data Mining, Pattern Recognition and Speech Processing.