ISSN NO: 2249-3034

WEB SCRAPPER SYSTEM FOR SMART ONLINE SHOPPING

PROF.KALAM NARREN
Assistant Professor, *Department of Computer Science & Engineering, Dnyanshree Institute of Engineering & Technology, Satara*

Dr.CHANDRA MOHAN
Students, Department of Computer Science & Engineering, Dnyanshree Institute of Engineering & Technology, Satara
rutujanalawade9185@gmail.com

ABSTRACT

In today's era, industries are based on different technologies that lead to living life more conveniently. E-commerce is one of the emerging industries which have a bright future and more competition with others. Nowadays most the people prefer online shopping through E-Commerce websites. But it is difficult to visit every website for comparing the deals or the price of a particular product. Even shoppers have a problem looking for the best deals and offers for a particular product which they are looking for. To solve these issues, we proposed a responsive web application that will help E-commerce merchants and shoppers collect product information, including a price list, from participating retailers and then it displays collective information and comparative result analysis in response to a shopper's or E-commerce merchant's search query.
KEYWORDS: E-commerce, Web Scraping, Web crawling, Machine learning, Big data.

1. INTRODUCTION

E-commerce is a mechanism of selling/purchasing products/services over the Internet. The customers are required to complete the transaction requirements by filling the transaction form with the required information, such as complete address, number of products, and the credit card number [1]. Following the successful completion of the transaction requirement, an email notification is sent to the customer.

Data scraping pertains to the process of extracting data from online files through computer scripts. Such extracted data exists in the form of tables and lists [2]. The interface between the script and the Internet for extracting data is basically a set of commands extracting data is basically a set of commands, i.e., an application programming interface (API). These APIs can be trained and used to extract data for search results across across a group of websites. Automating web searches and extracting data from multiple pages for search results [2].

We propose a scraping algorithm for detecting marketing trends in online shopping websites [3]. Specifically, our contribution is the pioneer work on using web scraping for extracting the best price of the target products from multiple websites rather than a single website. This work entails the products from the top online websites. When the user wants to buy a product, he/she can search the product in one consolidated website and the search results are pulled up from fabulous marketing websites in just one consolidated place.

ISSN NO: 2249-3034

2. BACKGROUND REVIEW

1. Web Scraping: Process of web scraping

- Identify the target website
- Collect URLs of the pages where you want to extract data from
- Make a request to these URLs to get the HTML of the page
- Use locators to find the data in the HTML
- Save the data in a JSON or CSV file or some other structured format

2. Web crawling:

A web crawler, which we generally call a "spider," is an artificial intelligence that browses the internet to index and searches for content by following links and exploring, like a person with too much time on their hands. In many projects, you first "crawl" the web or one specific website to discover URLs which then you pass on to your scraper.

3. Machine Learning:

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

4. Big Data:

The term "big data" refers to data that is so large, fast or complex that it's difficult or impossible to process using traditional methods. The act of accessing and storing large amounts of information for analytics has been around a long time.

3. PROPOSED SYSTEM

3.1 Problem Definition:

To develop a responsive web application using improved web scraper system for smart online shopping.

3.2 Objectives:

1. To develop a responsive web application for smart online shopping using web scraping.
2. To give E-Commerce trader a good opportunity to boost their sales, attract new customer and go ahead against their competition.
3. To build a model which will predict the best suitable product for the shopper by comparing different E-commerce sites.

3.3 Libraries:

1. **Flask:** Flask is a web framework. Flask provides tools, libraries and technologies that allow you to build a web application.
2. **Selenium:** Selenium provides support to multiple libraries such as Ruby, Python, Java, etc. as language bindings have been developed by Selenium developers to provide compatibility for multiple languages
3. **Beautiful soup:Beautiful Soup** is a pure Python library for extracting structured data from a **website**. It allows you to parse data from HTML and XML files.

ISSN NO: 2249-3034

3.4 Methodology:

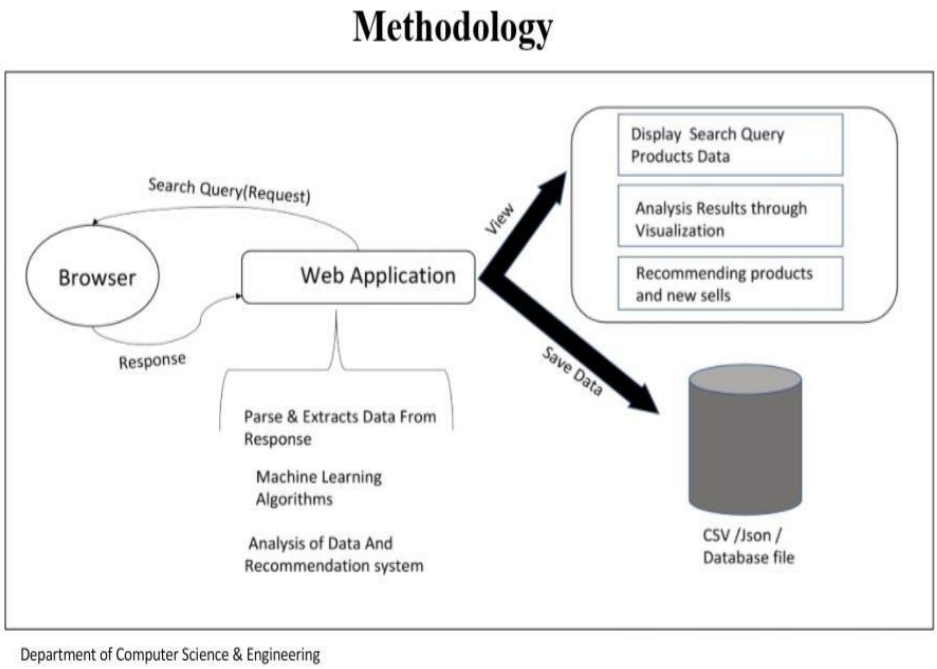

Fig1.1. Methodology

In this project we are developing dynamic website. We are using HTML, CSS and JavaScript Language for Frontend design and python language and some python libraries for the web scraping of e-commerce website of particular product.

1. **Web Application and Browser:**
   When customer is searching for a product on web application, such search query gives request to browser and then this site will fetch prices and gives you results using which customer can make the best purchase. At backend we are using web scraping library for scraping product price, name and image data from target e-commerce site.

2. **CSV/Database file:**
   Such Scrapped data will be stored in csv file or in dedicated database for further use like analysis of results and visualization

3. **Frontend/View:**
   So, at final stage, frontend will display product details from different website for particular search query, Some Analysis results in the form of visual reports for particular website, on previous search history of customer we are recommending them for new sales and offer through e-mail using machine learning algorithm or python.

ISSN NO: 2249-3034

4. RESULTS

Landing page of website where customer can use feature of websites after sign in.

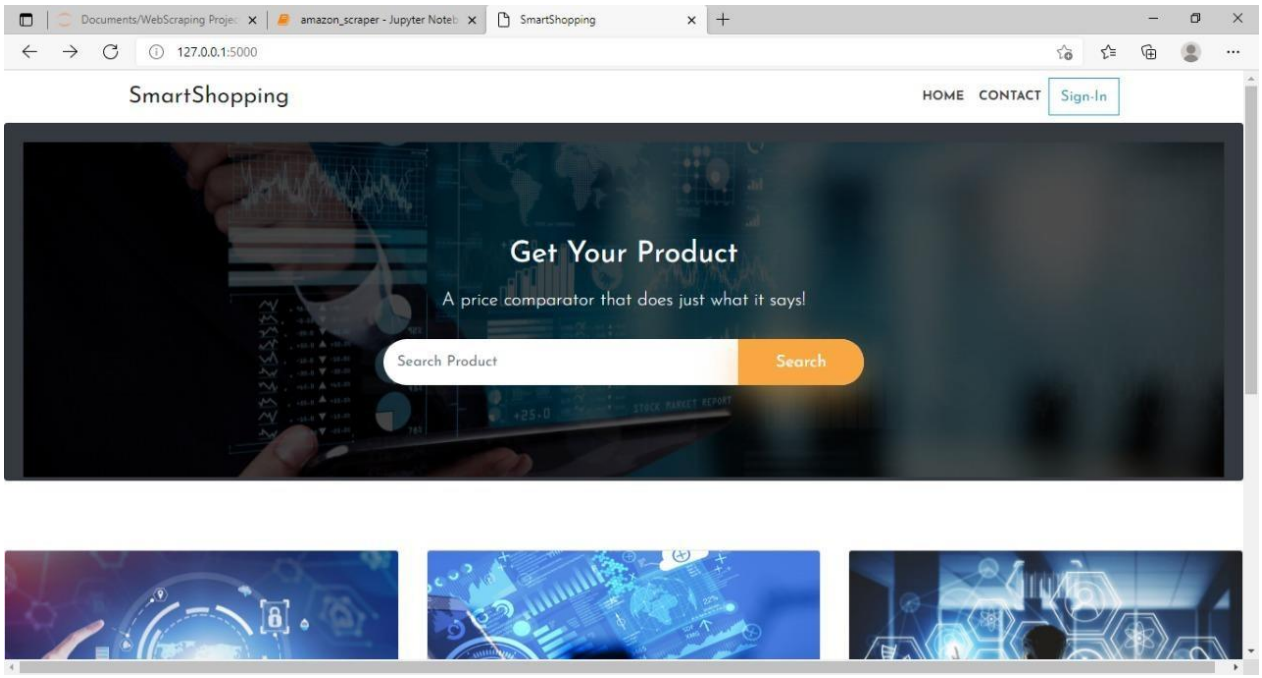

Fig 1.2. landing page

After Searching for particular products e-commerce website get opened in selenium web driver and parser will fetch query related data

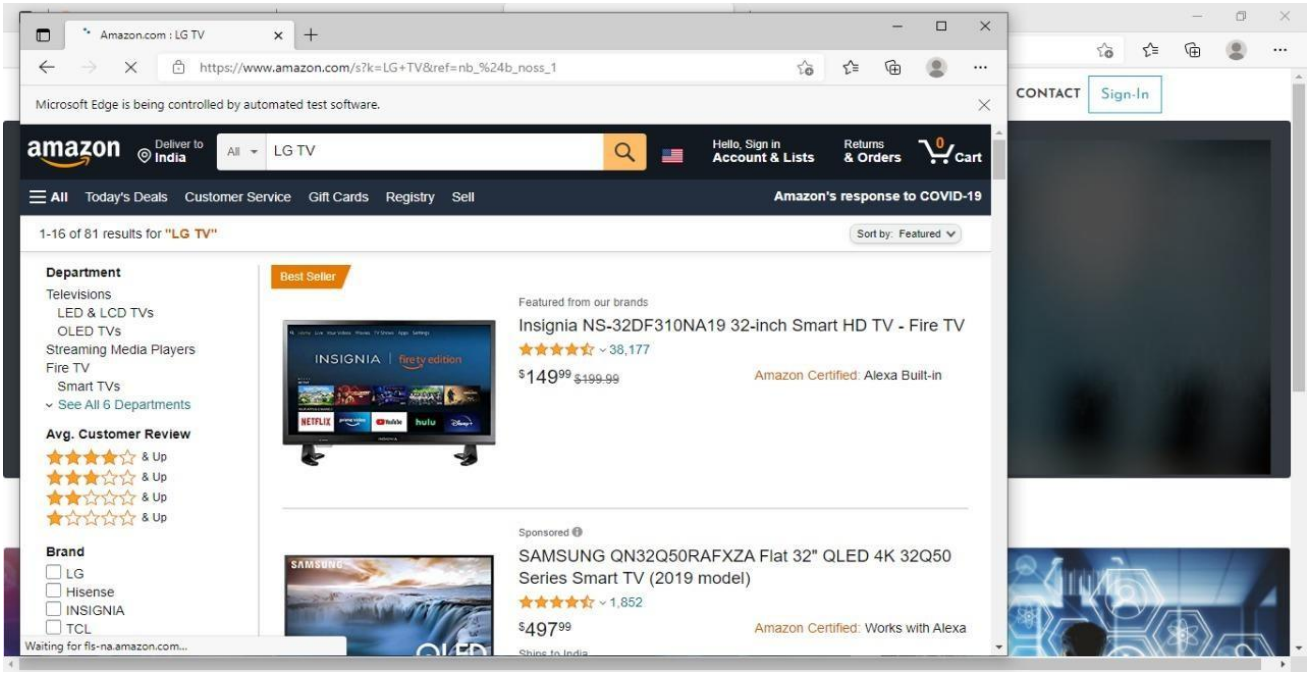

Fig1.3. Amazon Web scraping

Fetched data get displayed on our website which contains name of product, description, price, rating and review-count.

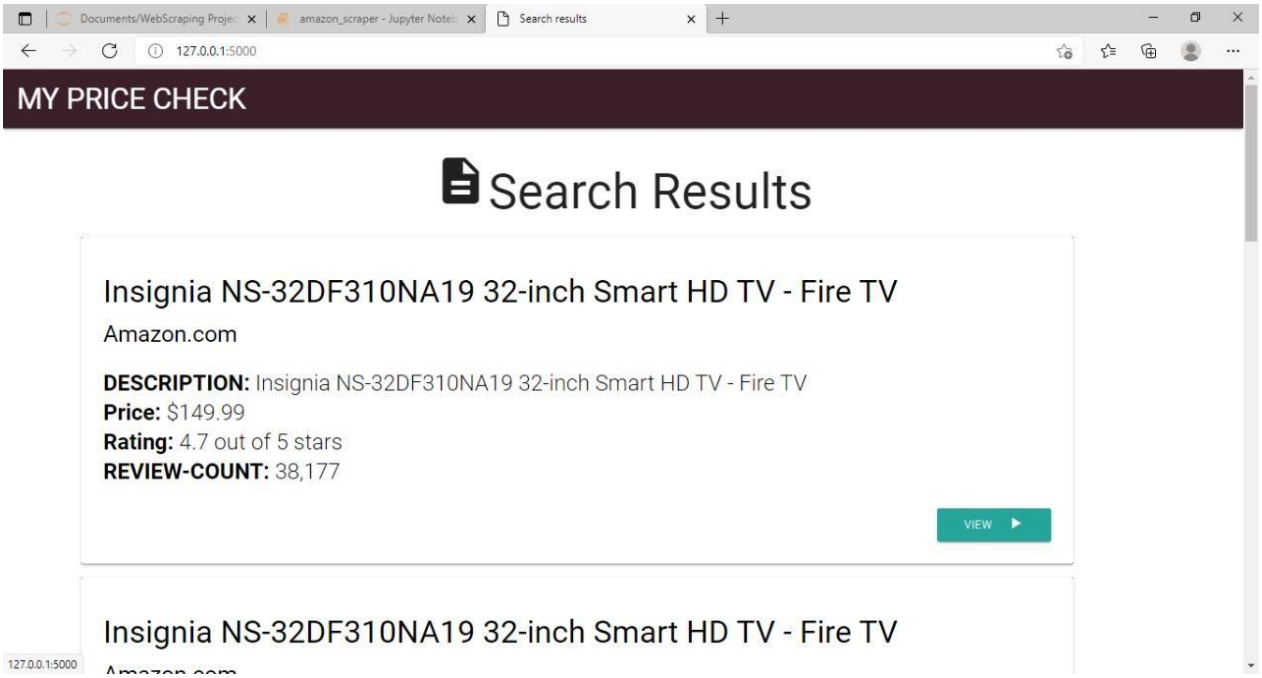

Fig 1.4. Result

5. CONCLUSION AND FUTURE SCOPE

In proposed system "Improved web scraper system for smart online shopping.", With this study, we have compiled a list of the most common issues which can arise when scraping web pages: various anti-scraping techniques, the prevalence of dynamically loaded content, as well as having to consider the legal and ethical aspects of scraping. The designed framework incorporates a Scrapy framework for web crawling and scraping.

As of future work, we aim to enable the proposed frame-work to suggest relevant and non-relevant items based on a factor k. In future we can extend this project Web scraping tool can make market research of the particular product/services and enormous benefits to offer in the marketing field.

6. REFERENCES

[1] **2019 International Conference on Computing, Mathematics and Engineering** Technologies – iCoMET 2019, Exploiting Filtering approach with Web Scraping for Smart Online Shopping, Penny Wise: A wise Tool for Online Shopping

[2]2017 International Conference on Sustainable Information Engineering and Technology(SIET), Increased Information Retrieval Capabilities on e-Commerce Websites using Scraping Techniques

[3] 2019 International Conference on Computer Communication and Informatics (ICCCI -2019), Jan. 23 – 25, 2019, Coimbatore, INDIA. A Review on Web Scraping and its Application

[4] Blogs: 25 Best Price Comparison Websites for Your Online Products in 2021, Web Scraping using selenium with python

[5] 2019 E-commerce network with price comparator sites, 5-7 June 2019, Ceske Budejovice, Czech Republic. This paper uses network analyzes and simulation methods to model network dynamics to address these issues.