ISSN No:-2456-2165

Real Time Product Analysis

Pinky Rathod
Student, Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Ruchita Yeole
Student, Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Manisha Udasi
Student, Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Sandeep Utala
Professor, Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, India

Abstract:- Real-time product analysis enables real-time comparison of products from different e-commerce websites based on their prices. In today's era, customers shopping online need to check the desired product at different e-commerce websites to avail the product with the best prices and quality. The paper describes an overview of the website which increases the user experience. The website enables the online users to analyse the product specifications and compare prices of the specific product available on other e-shopping sites and buy their desired products saving a lot of user's time, efforts and money. Using web crawling and scraping techniques the web content of products on different sites is obtained.

Keywords:- *Web Crawler, Web Scraper, Beautifulsoup, Mongodb, Flask, Elasticsearch, E-Commerce.*

I. INTRODUCTION

Nowadays people are shifting from traditional shopping to online shopping as most people use smart devices like tablets, mobile phones, laptops etc. to access these e-commerce websites. With this increasing number of people shopping online, online shopping as a whole is rapidly growing. Also increasing the number of e-commerce websites has increased the involvement of online users to find the best products in the right deal. However, it takes a lot of people's time and effort to search for their desired product at multiple e-commerce websites, compare prices and get that product with best price and quality. It's difficult to get all the prices of a product at a single point for the decision making of end users. Thus, it doesn't satisfy the user's demand due to restrictions, or limitations of technologies. To overcome this problem, Realtime product analysis is a web-based project that will enable the users to compare the prices of a product from different ecommerce websites at a single place. Apart from this, it recommends buying the best products over a number of ecommerce websites. Thus, saving a lot of user's time, effort and money.

In Real-time product analysis the main task is to fetch the data of e-commerce sites and this Intelligent agent is used for crawling through to different websites to fetch URLs of different products. The intelligent agent is a web crawler or a Shop-Bot that is an automated program that continuously browses the www and fetches the URLs from different Ecommerce websites. Further, the web scraper scraps and retrieves the data from these fetched URL and stores it in the database.

Whenever a user searches for a particular product it queries the database and required results are displayed. The user can then compare prices from different e-commerce websites and when the best deal is selected, the user will be redirected to the original website to purchase the product.

II. TECHNOLOGIES USED

A. Beautiful Soup

Beautiful Soup is a Python library for extracting HTML and XML documents. It is used to create tree structures of HTML elements. It is a way of navigating, searching elements in the parse tree. The required content is reached by navigating through various elements. It makes it easy to scrape information from extracted HTML documents. It saves a lot of programmer's time of work.

B. MongoDB

MongoDB is a NoSQL database which is basically a document-oriented database. It is used for the systems having a large amount of unstructured data. It stores the data extracted from the web scraper. It can store various forms of data. Also, it has high availability, horizontal scalability, and flexibility.

C. Flask Web Framework

Flask is a web framework for python. It is used to build complex, database-driven websites. It is used as a backend framework where the frontend and backend are tightly coupled. It is easy to use and is used to build websites and web apps rapidly. It is used with MongoDB to have more control over the database.

D. Elasticsearch

Elasticsearch is a real-time, open-source distributed search engine based on Lucene. It is used to search all the kinds of documents and provides scalable search. It uses the concept of invert search for indexing. It does the full-text search over the documents. When a query is fired it looks up into the index table for required to data.

III. IMPLEMENTATION AND WORKING

A. E-commerce Websites

It is a comparison website that fetches the data of a product available at different e-commerce websites and enables the comparison of prices of the product. The website is made for clothing products and the comparison is limited to a few ecommerce websites. It brings all the details of the searched product at one place saving a lot of user time efforts and money. The user can compare the products and buy the best quality products at reasonable prices.

B. Web Crawler

A web crawler is a web spider also called an Internet bot that navigates to different e-commerce websites and extracts links and hyperlinks. It continuously browses www for web indexing. It collects the data from all the e-commerce websites and that data is further filtered and sent for scraping to the scraper. There is a scheduler associated with web crawler which gives periodic updates to the crawler about the changes in product details making the system real time.

C. Web Scraper

Web Scraping is used to extract HTML data from the URL sent by a web crawler. The data is scraped from multiple ecommerce websites and is stored into the MongoDB Database which is a NoSQL database used to handle a large amount of unstructured data. In this system, web scraping is done using python libraries like requests and beautifulsoup4.

D. MongoDB

MongoDB is a NoSQL database which is basically a document-oriented database. It is used for the systems having a large amount of unstructured data. It stores the data extracted from the web scraper. It can store various forms of data. It is a cross-platform, open source distributed database. Also, it has high availability, horizontal scalability, and flexibility.

E. Front End

Flask front end Framework is used to build complex web applications instantly. It is used as a backend framework where the frontend and backend are tightly coupled. It is used to build complex, database-driven websites. It is easy to use and is used to build websites and web apps rapidly. It is used with MongoDB to have more control over the database.

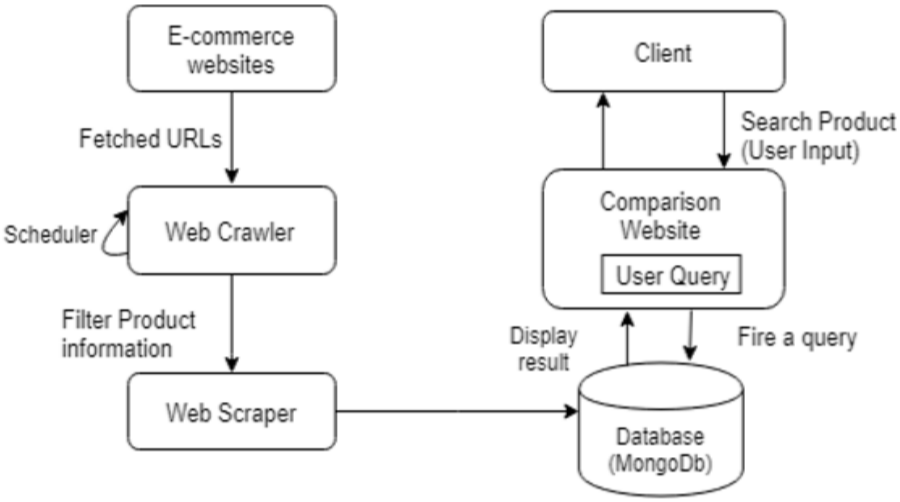

Fig 1:- System Architecture

F. Working

Figure 4.1 describes the system architecture and its working procedure. There is a web crawler that runs on the back-end of the website whereas the frontend technology provides a graphical user interface (GUI) for the users to communicate with the system. The web crawler continuously browses different e-commerce websites and fetches the urls from those websites. There is a scheduler associated with web crawler which gives periodic updates to the crawler about the changes in product details making the system real time.

Then the web scraper is used which scraps the information from HTML tags like , <p>, etc. This scrapped information is stored in the MongoDB database which is a NoSQL database used to handle a large amount of

data. It has high performance and scalability of key-value storage method and rich data processing functions of a traditional relational database. After that database is replicated to Elasticsearch for faster indexing and searching.

The user can search for any product using the search bar provided. Whenever the user searches for the products in the search bar, a query is fired to the Elasticsearch database to retrieve the required results. The details of the product fetched from all the e-commerce websites are displayed at one place and the user can thus compare and analyse the products and buy the products of their own wish. For optimized searching search optimized algorithms are used. After the user decides to buy the desired product and clicks on the buy button the user is redirected to the original website to purchase the product.

IV. CONCLUSION

Real-time product analysis is a web-based system that helps the users in decision making while buying products online. It is a comparison engine that does price comparison of the products. It grants the power to in the hands of the user to analyse the prices of a product available at different ecommerce websites and buy the product at cheapest price with the best deal saving a lot of user's effort, time and money.

Eventually it brings all the characteristics of a product at one place making it easy for the customers to take decisions. As far as product comparisons are concerned, manufacturers, companies and businesses also tend to benefit from them. The beauty of this project is that it can be customized for a specific business segment. It is a cost-effective way to advertise an online business and expand its recognition in the crowded.

ACKNOWLEDGMENT

We would like to take the opportunity to express our sincere thanks to our guide Prof.Mr.Sandeep Utala, Department of Information Technology, VESIT, Chembur for his invaluable support and guidance throughout our project research work. Without his kind guidance & support, this was not possible.

We are thankful to Dr.Mrs.Shalu Chopra, HOD, Department of Information Technology, VESIT, Chembur and all my B.E. teachers for providing advice and valuable guidance.

We are also thankful to all the references below who helped us sharpen our ideas about the topic. We also thank the people who helped us with their constructive criticisms.

REFERENCES

[1]. "A Dive into Web Scraper World" Deepak Kumar Mahto, Member, IEEE, Lisha Singh, Member, IEEE, 2015.
[2]. Yunhua Gu, Member, IEEE, Shu Shen, Member, IEEE, Jin Wang, Member, IEEE, Jeong-Uk kim, Member, IEEE, (2015). "Application of NoSQL Database MongoDB".
[3]. Xue-mengLi, Member, IEEE, Yong-yi Wang, Member, IEEE, "Design and Implementation of an Indexing Method Based on Fields for Elasticsearch",2015.
[4]. "Mining E-Commerce Data from E-Shop Websites" Andrea Horch, Member, IEEE, Holger Kett, Member, IEEE and Anette Weisbecker, Member, IEEE, 2015.
[5]. "Comparison of E-commerce Products using web mining", Department of Computer Engineering, Sinhgad Institute of Technology and Science, Savitribai Phule Pune University
[6]. Darshita Kalyani, Dr. Devarshi Mehta Gujarat, India. (June 2017) "Paper on Searching and Indexing Using Elasticsearch".
[7]. "Implementation of Web Crawler", Pooja gupta, Member, IEEE, Kalpana Johari, Member, IEEE.