(REVIEW ARTICLE)

Distributed crawler using python and Redis

Sonia Bajaj, Shlok Bharuka, Danish Sheikh, Mechaiver Chherke and Rutika Barde

School of Engineering & Technology Department of Computer Science and Engineering, G H Raisoni University, Saikheda, India.

Abstract

An online platform known as a price comparison website serves as an intermediary between consumers and sellers. It presents various lists of prices for products selected by the user, enabling them to make informed decisions and save money. Additionally, it fosters price awareness in consumers, making them less susceptible to deceitful retailer advertisements. In essence, the website also serves as a promotional tool for retailers, supermarkets, hypermarkets, and shops to reach out to customers. In today's economic environment, where living costs are high, and time is limited, a price comparison website can significantly assist consumers with their shopping and household needs.

Keywords: Distributed-crawler; Scrapy; Python; Data Scraping; Generic Structure; Crawler; Web scraping

1. Introduction

A price comparison website is an online platform that acts as a bridge between consumers and sellers. It allows users to view a range of prices for a particular product, thus helping them make an informed decision about which product to choose in order to save money. This website not only benefits the consumers but also provides an excellent opportunity for retailers, supermarkets, hypermarkets, and shops to reach out to a broader audience.

One of the most significant advantages of using a price comparison website is that it allows consumers to compare prices and features of products from various sellers. This makes it easier for them to identify the best deals and avoid paying higher prices for the same product. Additionally, it promotes price awareness among consumers, making them less likely to fall prey to deceitful advertisements that claim to offer the lowest price, but in reality, do not.

Moreover, a price comparison website plays a crucial role in promoting the retailers, supermarkets, hypermarkets, and shops to the customers. It allows these businesses to showcase their products and services to a vast audience and provides them with a platform to compete with other businesses in the same industry. With the rise of e-commerce, having an online presence has become critical for businesses to remain competitive, and price comparison websites offer them an excellent opportunity to expand their reach and increase their sales.

In the current economic situation, where the cost of living is continually increasing, and time is limited, a price comparison website is an excellent resource for consumers. It provides them with a convenient and efficient way to shop for the products they need without having to spend hours visiting different shops or websites. With a few clicks, consumers can compare prices and features of products from different sellers and make an informed decision about which one to purchase.

In conclusion, a price comparison website is an essential tool for both consumers and businesses. It offers consumers the opportunity to save money and promotes price awareness, while also providing businesses with a platform to

---

* Corresponding author: Shlok Bharuka

showcase their products and services and expand their reach. With the ever-increasing demand for online shopping and the rise of e-commerce, price comparison websites have become an integral part of the retail industry and are expected to grow in popularity in the coming years..

2. Material and methods

2.1. Stage 1: Project Planning

In this stage, the project team identifies the problem related to the project and determines its significance. They outline the scope and objectives of the study while ensuring that the project is feasible within the given timeframe.

The team conducts literature reviews to identify potential solutions and the necessary tools and system types required for development. This stage is critical to ensure that the project is well-planned and that the team has a clear understanding of what they need to achieve.

2.2. Stage 2: Data Gathering and Analysis

The second stage involves conducting studies on energy and electricity consumption. The team reads and gains an understanding of comparison sites and how they help solve shopping problems.

They research existing comparison sites to analyze competitors and gain insights on what works and what doesn't. The team also studies consumer behavior related to home groceries products, to get a better understanding of the target audience.

2.3. Stage 3: Research on any Existing Similar Systems

The third stage involves studying existing similar systems to understand how they work.

The team analyzes the concepts and computation methods applied in similar systems and evaluates how the system helps solve the problem.

This stage is important to ensure that the team learns from existing systems and builds upon them rather than starting from scratch.

2.4. Stage 4: Drafting the Main Components of the System

In this stage, the team identifies the main components that will make up the system.

They require visitors to register with their basic information to use the website's services. Registered users will be automatically subscribed to the website's newsletter.

This stage is important to ensure that the website's visitors can fully access the services provided by the website.

2.5. Stage 5: Develop System Architecture

The final stage involves developing the architecture of the system to create a clear understanding of how it will operate. The team ensures that the system is user-friendly and provides efficient and effective services. They also test the system to identify and resolve any potential issues before launching it. This stage is crucial to ensure that the system is fully functional and that users have a seamless experience.

3. Results

The results are shown when the crawler is given an input including the url of the site to be crawled using redis.
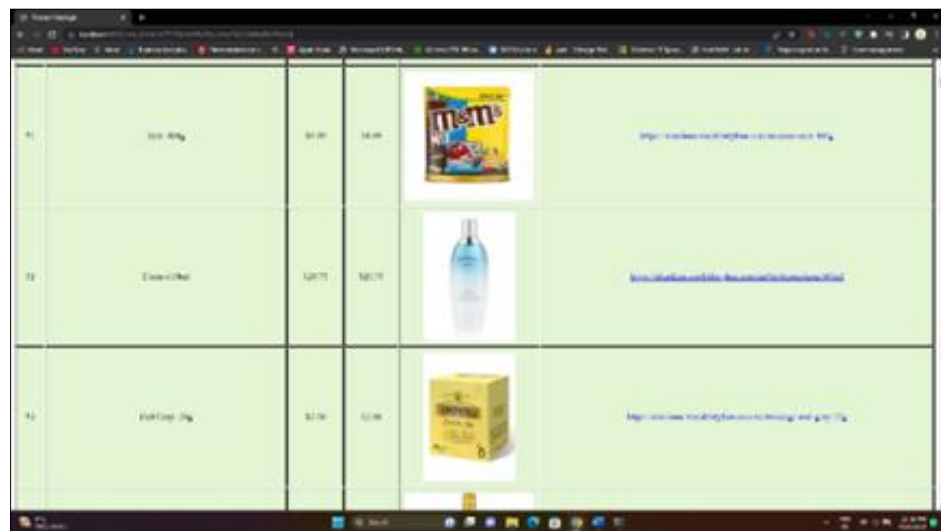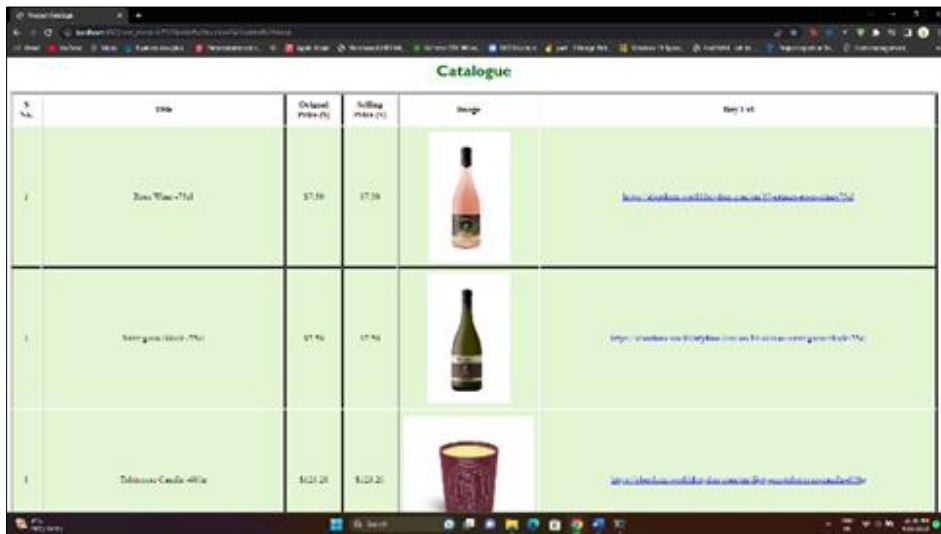
Then we get the data output in json format as a file. We use this file as the local database to present the products in a tabular format on the basic website we created.

For starting the website we need to first start a localhost on a port and then access this server using a browser and finally we are able to see the data we extracted in a visualized form.

Starting the local host:

Figure 1 Screenshots of the catalogue table

4. Discussion

Our objective is to simplify the process of data collection using our crawler. This is achieved by allowing changes to be made solely to the YAML file and inputting data into Redis, rather than having to modify the code itself. We have developed the crawler in a generic manner to eliminate the need for creating new crawlers for each website. By implementing this approach, we have made data extraction simpler and more efficient for our users.

5. Conclusion

This document serves as a comprehensive guide for the project, containing all the pertinent details related to its various aspects. The abstract, background, and objectives provide a broad overview of the project, while the methodology section delves into the specifics of the approach chosen for completing the project. The methodology outlines the research techniques, tools, and technologies used in the project, as well as the timeline and budget constraints.powe

In order to develop a successful and functional price comparison website, it is crucial to refer to previous research papers related to the project. The literature review section presents the research that has been previously conducted by other researchers on this topic. It is imperative to thoroughly study this research in order to gain a better understanding of the topic and to identify the most effective methods for data retrieval. Since different websites utilize their own formulas to gather prices, it is essential to find the correct method for retrieving the desired data according to user preferences. The website aims to provide users with valuable information that will aid in making informed purchasing decisions. By offering a platform for comparing prices, the website addresses the concerns of busy individuals who wish to compare prices before buying products. One of the primary objectives of the price comparison website is to simplify the process of comparing prices. To achieve this goal, the website has been designed in a user-friendly manner that is accessible to all. This is achieved by providing a simple and intuitive interface that allows users

to quickly and easily search for the products they are interested in. Moreover, the website provides an extensive range of products and prices, allowing users to make informed decisions based on their specific needs and budget constraints.

The price comparison website also serves as a platform for sellers to promote their products and advertise any promotions or sales that may be going on. This provides sellers with an opportunity to increase their reach and attract more customers to their products. In addition, the website provides detailed information about each product, including specifications, features, and reviews, which allows users to make well-informed purchasing decisions.

In conclusion, the price comparison website is an effective tool that simplifies the process of comparing prices for busy individuals. The website provides users with valuable information to aid in making informed purchasing decisions while also serving as a platform for sellers to showcase their products and advertise any promotions or sales. Thorough research, a user-friendly interface, and comprehensive product information are all crucial components of a successful price comparison website.

Compliance with ethical standards

Disclosure of conflict of interest

We, declare that I have no financial or personal interests that could be perceived as influencing the research conducted. However, it is important to note that I am a current Student of G H Raisoni University, Saikheda, India, which may be perceived as having a potential conflict of interest.

References

[1] Berger, P. D., & Nasr, N. I. (2019). Customer lifetime value: Marketing models and applications. Journal of Interactive Marketing.

[2] Brynjolfsson, E., Dick, A., & Smith, M. D. (2021). Search and Product Differentiation at an Internet Shopbot.

[3] Burdett, K., & Judd, K. L. (2019). Equilibrium Price Dispersion. Econometrica (pre-2019)

[4] Chevalier, J., & Goolsbee, A. (2021). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. Quantitative Marketing and Economics.

[5] Choi, D. H., Kim, C. M., Kim, S.-I., & Kim, S. H. (2019). Customer Loyalty and Disloyalty in Internet

[6] Retail Stores: Its Antecedents and its Effect on Customer Price Sensitivity. International Journal of Management (2021).

[7] Clay, K., Krishnan, R., & Wolff, E. (2018). Prices and price dispersion on the Web: Evidence from the online book industry.

[8] Dawkins, P. M., & Reichheld, F. F. (2021). Customer Retention as a Competitive Weapon. Directors and Boards.

[9] Degeratu, A. M., Rangaswamy, A., & Wu, J. (2021). Consumer choice behavior in online and traditional

[10] Supermarkets: The effects of brand name, price, and other search attributes. International Journal of Research in Marketing,(2021).