

Project Report: Infinite Analytics Company - Amazon Web Scrapping for Product Prices

Introduction

In this project, we aimed to automate the process of web scraping product prices from Amazon's website. The objective was to create a tool that takes a list of product names as input, navigates to Amazon's search page, extracts the prices of the top 4 results, and provides the most accurate product price based on fuzzy matching, handling spelling mistakes, and ensuring the lowest price in case of multiple matches. The project was developed using VS Code and leveraged Python libraries such as Selenium, Pandas, and FuzzyWuzzy.

Project Scope

The primary focus of this project was to build an efficient web scraping tool for Amazon. Additionally, we expanded the data extraction capabilities to include liquor websites and Mondelez products, showcasing Infinite Analytics' AI platform's versatility in handling various product categories.

1. Design a web scraping tool to fetch product prices from Amazon.
2. Automate the process of searching for products based on a provided list.
3. Extract prices from the top 4 search results.
4. Implement fuzzy matching to handle variations in product names.
5. Handle spelling mistakes and random word order in the input.
6. Ensure the tool finds the most accurate product price and the lowest price in case of multiple matches.
7. Create a log file to track the product names, matched results, and extracted prices.

Implementation Details

Automated Browser Navigation with Selenium

We utilized Selenium with Chrome driver to automate the browser navigation. The tool took each product name from the provided list, searched for it on Amazon's website, and extracted the prices of the top 4 results.

Fuzzy Matching and Handling Variations

To handle variations in product names, spelling mistakes, and random word order, we implemented a fuzzy matching algorithm using the FuzzyWuzzy library. This ensured accurate matching and extraction of prices even when the product names were slightly different.

Error Handling for Top Results

Amazon sometimes displays different products as the top results, especially when there are multiple versions or editions of a product. In such cases, the code re-ran for these products to find the correct one and its corresponding price, for a maximum of three attempts.

Price Comparison and Extraction

The tool compared the fuzzy match scores of the product names in the search results and extracted the price of the product with the highest match. In case multiple products had the same match, the tool extracted the price of the product with the lowest price among them.

Log File for Tracking

We implemented a log file to keep track of the product names, their matched results, and the extracted prices. This allowed us to monitor the tool's performance and analyze any discrepancies in the results.

Conclusion

The Amazon web scraping project successfully demonstrated an automated and efficient approach to extracting product prices from Amazon's website. By leveraging the power of Selenium, Pandas, and FuzzyWuzzy, the tool achieved accurate and reliable results, even in the presence of variations in product names and search results.

The implementation of fuzzy matching and error handling ensured that the tool obtained the most accurate product price from the search results. The log file provided a transparent view of the scraping process, allowing easy tracking of the product prices over time.

The web scraping tool developed in this project empowers companies to access valuable consumer insights, monitor price trends, and make data-driven decisions in a competitive market. This cost-effective and scalable solution offers

an efficient way to collect and analyze pricing data from Amazon, providing companies with a valuable edge in understanding consumer behavior and market trends.

The successful completion of this project showcases the potential benefits of web scraping in democratizing consumer insights for businesses. With the ability to automate the extraction of product prices and handle variations in product names, companies can save valuable time and resources, enabling them to focus on strategic decision-making and enhancing customer experiences.

As the world of e-commerce and consumer preferences continue to evolve, data-driven approaches like web scraping become indispensable tools for companies seeking to stay ahead in a highly competitive landscape. By harnessing the power of web scraping technologies, businesses can gain deeper insights into pricing dynamics, monitor competitor activity, and adapt their strategies to meet the ever-changing demands of consumers.

In conclusion, the Amazon web scraping project represents a valuable contribution to data-driven decision-making and consumer insights, exemplifying the capabilities of Infinite Analytics in providing efficient and effective solutions to drive business success.