

Flipkart Web Scraper

1. Introduction

The Flipkart Product Scraper is a Python-based tool that extracts detailed product information from Flipkart, an e-commerce website. Using the BeautifulSoup library, this scraper collects and organizes product data such as name, price, ratings, availability, and other relevant details. This project is designed to help users analyse and compare products, track price trends, and gather data for further analysis or research. The source code and sample csv file submitted uses data for “Bluetooth Headset” search on Flipkart.

2. Technologies Used

Python – Programming language used for writing the scraper.

BeautifulSoup – Python library used for parsing HTML and extracting data.

Requests – A library used for sending HTTP requests to fetch web pages.

Pandas (optional) – For storing and manipulating the data in a CSV format.

3. Features

- Scrapes product details including title, price, rating, product description, and product replacement policy.
 - Supports scraping multiple pages for fetching products in bulk.
 - Option to export data in CSV, JSON, or Excel format.
 - Handles dynamic content loading to get accurate information from Flipkart.
 - Ability to filter and scrape products based on categories or keywords.
-

4. How It Works

1. **HTTP Request:** The scraper sends a GET request to the Flipkart product page.
 2. **Parsing HTML:** Using BeautifulSoup, the HTML content of the page is parsed.
 3. **Pagination:** Scraper extracts different links for different web pages from a single search page and goes to each page to extract information.
 4. **Data Extraction:** Specific tags, classes, and IDs are identified to extract key product information such as name, price, and rating.
 5. **Data Export:** The extracted data is stored in a structured format such as CSV or JSON for easy analysis.
-

6. Setup to use the source code

1. **Installed libraries required in python:**

pip install bs4

pip install requests

pip install pandas

2. **Which product information do you want to extract?**

- I. Open the Flipkart website on your browser
 - II. Search the desired product
 - III. Copy URL from the top of the webpage
 - IV. Paste it into the "url" variable of the source code
 - V. Execute the python script.
 - VI. Open the CSV file named "flipkart_data.csv" present in the same directory as the source code
 - VII. You have your required information :)
-

Possible Future Enhancements

- **Prediction of Price Trends:** Implementing **machine learning** to predict price trends through the data extracted. Comparing the prices, and demand of a particular brand.
 - **Scraping reviews:** Can scrape the reviews and use specific keywords to extract main highlighted points by the buyers.
 - Making the scraper more **robust** against CAPTCHA or anti-scraping measures.
-

That's it.