Comparison of E-commerce Products using web mining


* Department of Computer Engineering, Sinhgad Institute of Technology and Science, Savitribai Phule Pune University
** Department of Computer Engineering, Sinhgad Institute of Technology and Science, Savitribai Phule Pune University

Abstract- Web mining is an application data-mining technique used to extract information from web services. E-commerce websites nowadays have become one of the most important sources for buying all kinds of products. Many strategies have been developed by analyzing customer’s behavior so as to attract more business and participation of people. As there are many e-commerce websites available it becomes difficult for users to choose best deal for desired product amongst these websites. Comparison of E-commerce products using web mining enables users to analyze prices and get desired product at minimum price. Users can also select multiple products that belong to same category for comparing its features. To obtain best deals from e-commerce websites web crawlers and web scrapping techniques are used to fetch detailed information. This way, paper aims to provide solution for online customers to buy products at good deal and save their valuable time, effort and money.

Index Terms- web mining, e-commerce, web crawler, MongoDB, Django

I. INTRODUCTION

In the current era of online business, ecommerce have become a huge market for the people to buy goods online. Increasing use of smart devices and other mediums has paved the way for users to buy products almost from anywhere. This has increased involvement of online buyers evolving e-commerce business. These large numbers of ecommerce websites put users in turmoil to search and choose to buy a single product from multiple ecommerce websites. The proposed solution helps online users to grab best deal for their product from multiple ecommerce websites on single web interface. This will in turn save users time, money and efforts to find the same product prices on different ecommerce websites. Proposed system uses web scrapping technique to extract data from ecommerce web pages and also web crawler to links for products. This will also allow users to analyse prices and select products from same category for comparing its features. This system uses the following technologies:

1) Web Crawler:
The system deals with price comparison engine. The first thing required are to gather large amount of data from different e-commerce websites. It is not possible to manually collect the data from websites. Hence the best way is to create a web crawler that will navigate to these e-commerce websites. The fetched URL’s are send to scraper for scrapping process.

2) Web Scrapper:
Web Scrapping is used to extract HTML data from URL’s and use it for personal purpose. As this is price comparison website, data is scrapped from multiple e-commerce websites. In this system, Scrapping is done using python libraries like requests and beautifulsoup4. Beautifulsoup4 is a python library which is used for parsing html pages. Using these, product information from different e-commerce sites is scrapped and stored in database.

3) MongoDB:
MongoDB is classified as NoSQL database which is a document oriented database. As system deals with large amount of unstructured data, it is flexible to use mongodb as database. Data extracted from scraper is stored in MongoDB database.

4) Django Web Framework:
Django is a python web framework. Comparison of E-commerce products using web mining is product and price comparison website which is created using Django framework. Products that are been requested by user are queried in mongodb database. Products that are queried in mongodb database are then fetched and displayed on the website.

II. SYSTEM ARCHITECTURE

Figure 1 describes system architecture and its detailed working procedure. The front end system provides a graphical user interface (GUI) in the form of website where clients interact with the system whereas the backend consists of web crawling and scrapping techniques in order to extract product information from different e-commerce websites. The extracted information of e-commerce products is stored in MongoDB database. Client requests for desired product from main website and query is fired in local database. Product Information is displayed on main web page. Client can see prices of required product at one place present on different E-commerce firms. Another feature is provided on the website that compares products. User can add products of same the category to compare. User may also analyze the product for its details and specifications.

III. IMPLEMENTATION

Working of the proposed system is as follows: The backend system consists of two important techniques web crawling and web scrapping. Web scrapping is a technique that is used to extract information in the human readable format and display it on destination terminal. Before scrapping the output, Web Crawlers are responsible to navigate to the destination once the crawler reaches the correct page and matches up with the products, scrapping process starts. Crawler periodically fetches information from ecommerce websites so as to check for updates. If updates are available crawlers carries those updates and makes necessary changes in the database. Web scrapping essentially consists of two tasks: first is to load the desired web


www.ijsrp.org
Figure 1. System Architecture

The system architecture consists of several key components:

1. **Web Crawler**
   - Fetch URLs

2. **E-Commerce Websites**
   - Filter product info

3. **Web Scraper**
   - Fetch price

4. **Periodic Trigger**
   - Cron job to fetch price

5. **Client**
   - Search product
   - Request for price
   - Display price
   - Display result

6. **Main Website**
   - Fire query
   - Display result

7. **Database**

The process begins with fetching URLs to the web crawler, which then filters the product information from the e-commerce websites. This data is stored in a MongoDB database using pymongo connectivity. The front end consists of a main website, where the client searches for the required product in the search bar and queries are fired in the local database. The website is designed using Django, a Python-based web framework, and communication between the Python web framework and MongoDB is done using the MongoDB object-document mapper (Mongoengine). Required results are retrieved and displayed on the main website. Users can then compare the prices of available products from different e-commerce websites. A user is redirected to the original e-commerce website that offers the best deal. Another feature provided is that clients can compare products within the same category to differentiate specifications and choose accordingly.

**IV. RESULT**

Comparison of product prices from different e-commerce websites is displayed on a single web interface. The system allows users to analyze and compare product specifications for up to four products that belong to the same category. To achieve this result, web mining is used to fetch required product details, and web scraping is employed to extract information from different e-commerce websites. The system facilitates users to redirect to the original e-commerce website that offers the best deal. Following images showcase how product analysis and comparison for e-commerce sites are performed.

**V. CONCLUSION**

Comparison of e-commerce products using web mining creates a web-based system that helps users in decision-making while shopping online. This website will facilitate users to analyze prices present on various e-commerce shopping websites, allowing them to know the cheapest price of each product with a best deal. The website will also have the facility of comparing products with all their specifications, ensuring that users can choose from the same category. This will save buyers' efforts and time, ultimately bringing together strategies, best offers, and deals from leading online stores and helping buyers to shop online.

www.ijsrp.org
Image 1. Product details scrapped using web scraper

```
{
  "id": "ObjectID(574c815d2b5b8df33b1dc662)",
  "site": "flipkart",
  "title": "Samsung Galaxy On5(Gold, 8 GB)",
  "price": "Rs. 8,198",
  "Specifications": {
    "call_features": "Conference Call, Speed Dialling, Loudspeaker, Call Divert",
    "resolution": "HD, 1280 x 720 Pixels",
    "model": "Galaxy On5",
    "internal_memory": "8 GB",
    "memory": "1.5 GB RAM, 8 GB ROM",
    "sim_type": "Dual SIM, LTE + LTE (Dual Standby)",
    "processor_speed": "1.3 GHz Exynos 4412 Quad Core",
    "battery": "Yes",
    "expandable_memory": "microSD, upto 128 GB",
    "secondary_camera": "Yes, 8 MP",
    "screen_size": "5 inch",
    "primary_camera": "Yes, 8 MP"
  }
}
```

Image 2. Products are displayed on main web page as soon as client queries in search bar
Image 3. User can compare 2 or 3 products

Image 4. User can compare prices of a product that are present on different e-commerce sites
ACKNOWLEDGMENT

It gives us great opportunity in presenting the preliminary project paper on ‘Comparison of E-commerce products using web mining’. We sincerely express our gratitude to our internal guide Prof. (Ms.) P.N. Vengurlekar for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Prof. G.S. Navale, Head of Computer Engineering Department, for her indispensable support, suggestions. In the end our special thanks to Dr. S.N. Mali, Principal, who has given us valuable teaching and guidance which has inspired us to attain new goals.

REFERENCES


AUTHORS

First Author- Riya Shah, BE Computer Engineering, Sinhgad Institute of Technology and Science, shahriya1995@gmail.com
Second Author- Karishma Pathan, BE Computer Engineering, Sinhgad Institute of Technology and Science, karishma.ap1609@gmail.com
Third Author- Anand Masurkar, BE Computer Engineering, Sinhgad Institute of Technology and Science, anand.masurkar1819@gmail.com
Fourth Author- Shweta Rewatkar, BE Computer Engineering, Sinhgad Institute of Technology and Science, shweta.rewatkar2@gmail.com
Fifth Author- Prof. (Ms.) P.N. Vengurlekar, Assistant Professor, pooja.v.85@gmail.com
Correspondence Author-Riya Shah, shahriya1995@gmail.com, +91-8793387150