

Comparison of E-commerce Products using web mining

Riya Shah*, Karishma Pathan**, Anand Masurkar**, Shweta Rewatkar**, Prof. (Ms.) P.N.Vengurlekar**

* 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 scraping 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 scrapper 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 scrapper 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 using an object relational mapper mongoengine. Admissions in reputed varsity. Now, here we enlist the proven steps to publish the research paper in a journal.

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. But 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 e-commerce 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

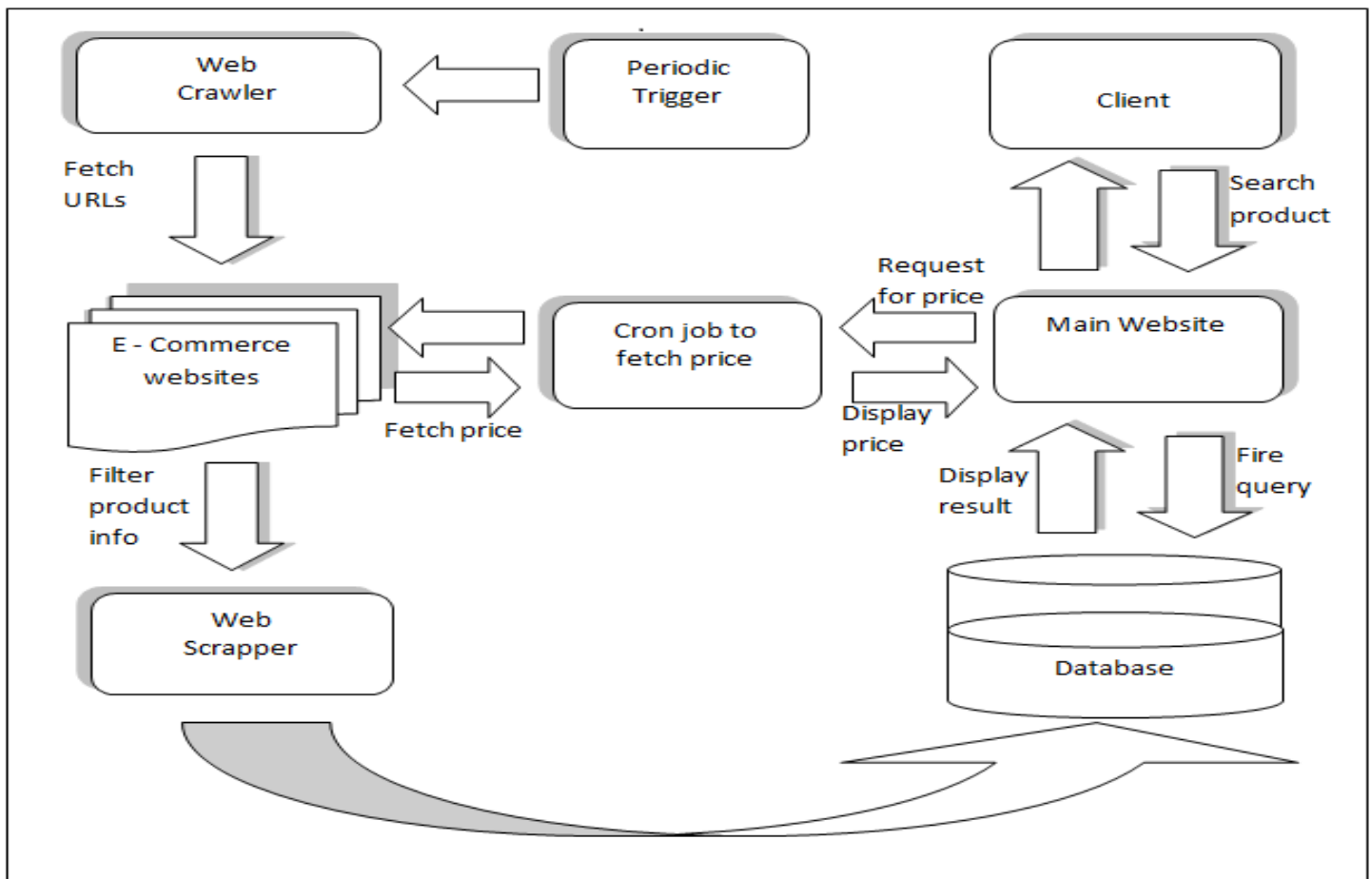


Figure 1. System Architecture

page and second is to parse HTML information of the page to locate intended information. In this system Scrapping is done using python as it provides rich set of libraries to address these tasks. “requests” is used to load the urls and “Beautifulsoup” library is used to parse the web page. After scrapping the products information from different e-commerce websites the data is stored in MongoDB database. Using pymongo connectivity data is scrapped and stored in database. The front end consists of Main website. The client searches for the required product in search bar and query is fired in local database i.e. MongoDB. The website is designed using Django web framework which is written in python. Communication is done between python web framework and MongoDB using Mongoengine which is a python object –document-mapper working with MongoDB. Required results are retrieved and displayed on Main website. The client can then compare prices of products that are available on e-commerce websites. A soon as client selects on best deal according to him ,he will be redirected to the original e-commerce website .Another feature provided is, Clients can compare products that belong to same category so as to differentiate specifications and choose accordingly.

IV. RESULT

Comparison of product prices from different ecommerce websites and result is displayed on single web interface. Also system allows user to analyze and compare product

specifications for maximum four products which lie under same category. To achieve this result web mining is done to fetch the required product details and concept of web crawler and web scraper is used to extract information of these products available on different ecommerce websites. System will allow users to redirect to original website of that specific product selected by the user as a best deal. Following images show how product analysis and comparison of e-commerce sites is done.

V. CONCLUSION

Comparison of E-commerce products using web mining is web based system which will help users in decision making while buying products online. This website will facilitate users to analyze prices that are present on different e-commerce shopping websites so that they get to know the cheapest price of product with best deal. The website will also have the facility of comparing products with all its specifications that belong to same category. This will surely save buyers efforts and valuable time. Ultimately, this will bring together strategies, best offers and deals from all leading online stores and will help buyers to shop online.



Image 3. User can compare 2 or 3 products

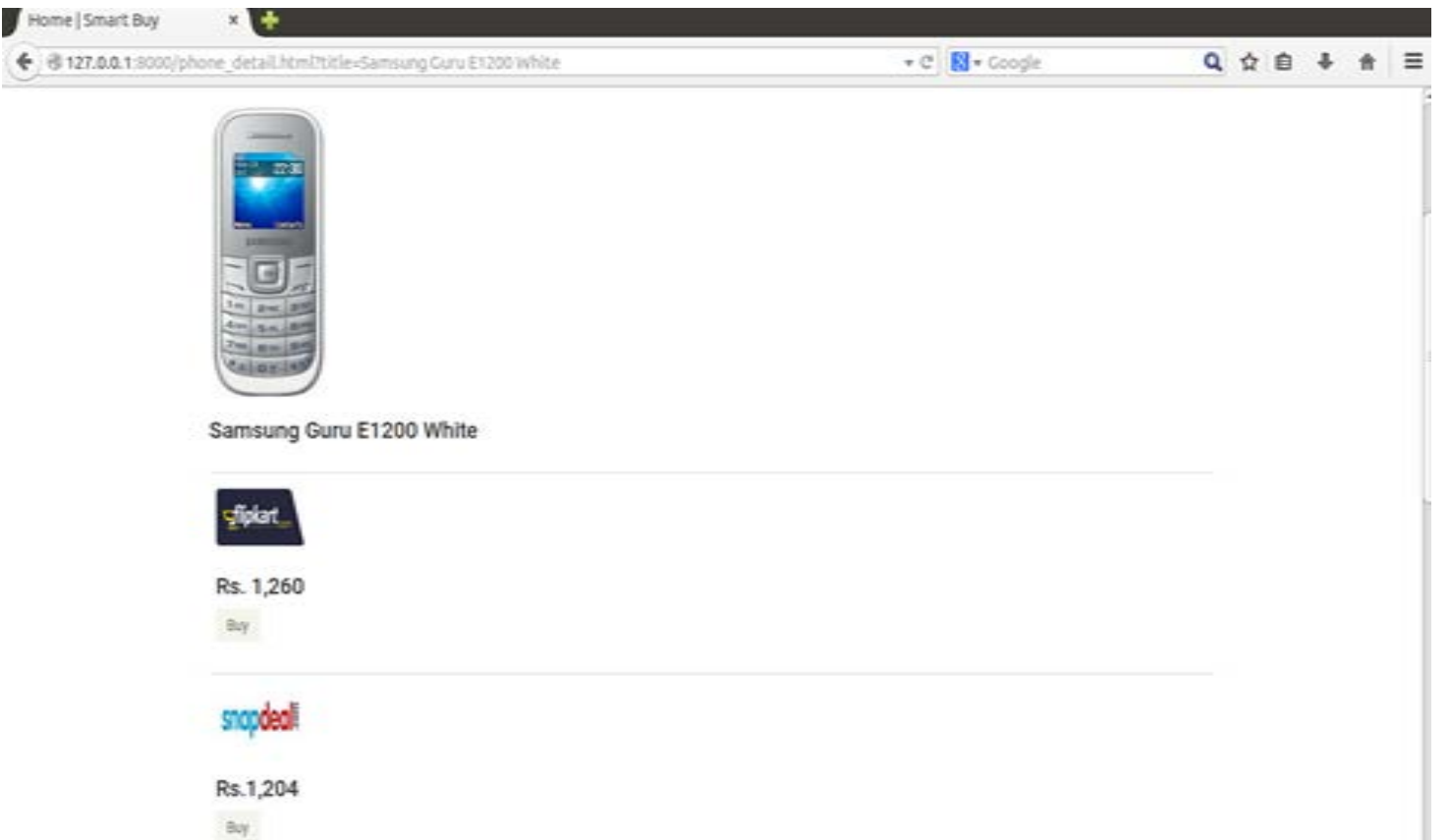


Image 4. User can compare prices of a product that are present on different e-commerce sites

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

- [1] Li Mei, Feng Cheng, "Overview of WEB Mining Technology and Its Application in E-commerce", Proceedings of the 2010 IEEE 2nd International Conference on Computer Engineering and Technology, Vol. 7, 2010.
- [2] Shridevi Swami, Pujashree Vidap, "Web Scraping Framework based on Combining Tag and Value Similarity" Proceedings of the IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.
- [3] Dr. Rajendra Nath, Khyati Chopra, "Web Crawlers: Taxonomy, Issues & Challenges" Proceedings of the International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013, pp. 944-948.
- [4] Jos'e Ignacio Fern'andez-Villamor, Jacobo Blasco-Garc'ia, Carlos 'A. Iglesias, Mercedes Garijo "A Semantic Scrapping Model for Web Resources" Spain.
- [5] Richard K. Lomotey, Ralph Deters, "RSenter: Tool for Topics and Terms Extraction from Unstructured Data Debris", Proceeding of the IEEE International Congress on Big Data, 2013.
- [6] I.Kali Pradeep, I. Bhagyasri, P. Praneetha, "E-Commerce With Backbone Of Data Mining", Proceeding of the International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 7, July - 2013.

[7] Arvind Kumar Sharma, P.C. Gupta, "Study & Analysis of Web Content Mining Tools to Improve Techniques of Web Data Mining" Proceeding of the International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 1, Issue 8, October 2012.

[8] Rahul Dhawani, Mrudav Shukla, Priyanka Puvar, Bhagirath Prajapati, "A Novel Approach to Web Scraping Technology" Proceeding of the International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 5, MAY 2015.

[9] Rupali Arora, Rinkle Rani Aggarwal, "Modeling and Querying Data in MongoDB" Proceeding of the International Journal of Scientific & Engineering Research, Volume 4, Issue 7, July-2013.

[10] B Rama Mohan, A. Govardhan, "Online Aggregation Using MapReduce in MongoDB", Proceeding of the International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9, September, 2013

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