ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 5, Issue, 11, pp. 5923-5926, November, 2015

International Journal of
DEVELOPMENT RESEARCH

Full Length Research Article

PRODUCT ANALYSIS AND COMPARISON USING WEB MINING

***Shweta Rewatkar, Anand Masurkar, RiyaShah, Karishma Pathan and Pro. Vengurlekar, P. N.**

Sinhgad Institute of Technology and Science-411041, Savitribai Phule Pune University, India

ARTICLE INFO

Article History:
Received 11th August, 2015
Received in revised form
19th September, 2015
Accepted 06th October, 2015
Published online 30th November, 2015

Key Words:
Product comparison,
Product analysis,
Web scrapper,
Web crawler,
Web mining,
e-shopping,.
MongoDB.

ABSTRACT

Web mining – the application of data mining techniques is used to extract knowledge from web documents. Web has become the root for emerging online business or shopping websites. Increasing number of ecommerce websites has put users in turmoil to search for shopping sites to buy a specific product in interest with best price and quality. The paper describes overview of the web system (website) which increases the user experience .The website enables the online users to analyze the product specifications and compare prices of the specific product available on other e-shopping sites. These web content of products on e-shopping sites is assembled using the technique called web scraping along with web crawlers.

INTRODUCTION

Today, the Internet has become an essential part of daily life of end users. The companies realize that Internet can be a shopping channel to reach existing and potential customers. People nowadays give priority to on-line shopping as most people use smart devices like tablets, mobile phones, laptops etc. to access these e-commerce websites. With this consensus online shopping as a whole is rapidly growing. Increasing number of ecommerce websites has increased the involvement of online users to find the best products in a right deal. Although online shopping continues to grow, there are few issues like comparing prices of buying items shown on different e-shopping websites. Hence, it's difficult to get all prices at a single point for the decision making of end users. To overcome these issues, product analysis and comparison using web mining is a web-based project (website) that will enable online users to analyze prices that are present on e-shopping websites at one place. In this project, the main task is to fetch the data of ecommerce sites. For obtaining this information we need to use the web crawler and web scrapper. Web crawler continuously browses the www typically for the purpose of fetching URL's of different ecommerce websites.

****Corresponding author: Shweta Rewatkar,***
Sinhgad Institute of Technology and Science-411041, Savitribai Phule Pune University, India.

It constantly gives the updates about the change in prices of the different products; the upcoming offers etc. on the various ecommerce websites. Further, the web scrapper scrapes and retrieves the data from these fetched URL's and stores it in the database. Whenever user searches for a particular product it queries the database and required results are displayed. User can then compare prices from different e-commerce websites and when a best deal is selected, the user will be re-directed to original website to purchase the product.

Scope

The proposed system deals with analysis and comparison of products from ecommerce websites into a single web application. The website will allow users to browse, search, compare and analyze the product details. The website will have constraints on what it can do. The project uses limited number of renowned ecommerce websites and products due to lack of hardware availability.

Definitions

Web Scraper

It is a technique of extracting information from different websites. Web scrapper extract multiple types of data –text,

images, tables etc. Web scraper can handle dynamic data to be fetched from e-commerce websites.

Web crawler

A web crawler is a web spider also called as Internet bot which systematically browses WWW typically for the purpose of web indexing. Web crawler navigate through the e-commerce sites to extract links and hyperlinks available on web pages.

Beautiful Soup

It is python library for extracting data out of HTML and XML files. It works with the parser to navigate, search and modify the parse tree .It supports HTML parser which has been included in python's standard library and also third party python parsers.

Product Perspective

For customer or any online user, they do not require any more specific hardware or software requirements. Website will interact with the user and display product price from different ecommerce website as soon as user queries in the search bar. The website will also display product specification along with the compare option of products within the same category. The following are the perspective of the project:

Multiple browsers support

The website must be available to all online users and must work correctly on browsers like Google Chrome, Internet Explorer and Mozilla Firefox.

User Privacy

The users are allowed to create their own account at front end in order to access and track their interested product availability and price drop over a period of time on e-commerce sites.

Dependencies

- Client: We have assumed that all of the computer systems are in proper working condition. User is capable of operating these system's basic functions including to power on the system, open either Internet Explorer or Mozilla Firefox, login and navigate the browser to the address the websites online.
- Web Server: We have assumed proper back end working of web server and database system with internet connection that allows to perform all communications with clients.
- The project completely depends on the python interpreter. We assume that the interpreter is bug free and any bugs that may occur in the interpreter may ultimately affect the system in a negative way.

Design and Implementation Constraint

The following points specify the constraints within which the system must operate.

- The computer must be equipped with browsers like Internet Explorer, Google Chrome, and Firefox.
- Response time for loading searched product must be as low as possible.
- As it a web based product i.e. website there are no memory requirements.
- The language with which user will interact is constraint to one language only that is English.

Software and Hardware Interface

Software Interface

Front End

The front end will have the JAVA stand-alone application along with other web designing tools like CSS, PHP, JavaScript etc. which will provide user's User Friendly Interface in order to make decisions and review on any product item displayed on the destination web page.

Back End

The back end will consist of web crawler and web scraper along with Mongo DB database .Web crawler runs continuously and extract all links and hyperlinks of products available on e-shopping site .Web scraper uses these extracted links of web crawler and scrape required product information and store in Mongo DB database for further processing of user requests.

Product Functions

Provide Product details

The website shall show detail description of selected products. The website shall provide browsing options to see product details. The product details includes the availability of the product on the e-commerce sites , its ratings , as well as display prices of selected product present on multiple e-commerce websites.

Provide Search Facility

The website shall provide a search bar for end user to enter product name of specific interest. The website shall display product information or specification based on users search.The website shall notify if searched product is not available or not found on any e-commerce websites or on the destination site.

Provide Comparison Facility

The website shall provide a compare option where end user will select products of same category for comparison. Detail description shall be compared of selected products in tabular format. This comparison strategy will help customers to decide the product of interest with maximum benefits.

Provide Purchase Redirect Facility

The user shall analyze the prices of selected product that are of different E-commerce sites and shall be redirected to selected ecommerce website to purchase the product.

Allow admin to update products

Administrator shall have the access right to initiate update process to check for recent product description, prices, availability of product on e-commerce websites.

Use Characteristics

The user is simply anyone who can access website through internet. It is assumed that user is familiar enough with computer to operate browser and is capable of browsing to, from and within websites.

Expected Result

Comparison of products is done of multiple e-commerce websites. Using web crawlers, url's are fetched from the e-commerce websites. From these url's, product information will be scrapped using scrapper and will be stored in database i.e. MongoDB. All product details will be present in database, a simple query from user will fetch data from the database to compare particular product. Results will be displayed on main website where user will analyze prices of that product.

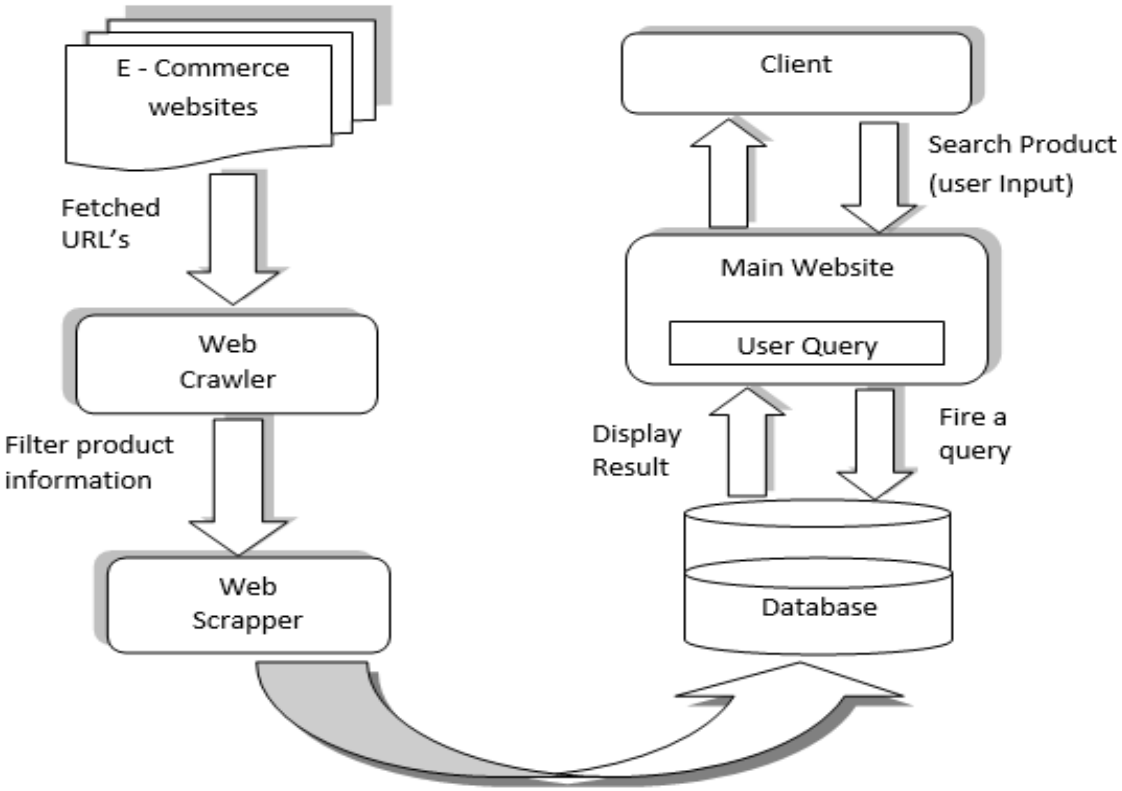

Figure 1. System Architecture

System Architecture

In our proposed system Figure 1, web crawler after defined time period fetches URLs of different ecommerce websites. Web Crawler runs in the back end fetches URLs and also checks for any update of products on ecommerce sites. It also displays new prices if changed on that particular ecommerce website. Filtered URLs are then passed to the scrapper who in turn scrapes the required data from the multiple ecommerce sites and stores in the database. Mongo DB database is used. Mongo DB is an unstructured database and it stores data in the form of documents. A search bar is provided to the user for searching for a particular product. Based on the input from user, required details of the product are then displayed to the user. When user queries for a product, a query is fired to the database. The result of the required product is then displayed on the website. Another feature is provided on the website that compares products. User has to add products of same the category to compare. User may also analyze the product for its details and specification.

Products that belong to same category can be compared which will help buyers to choose the best product have best features available in market.

Conclusion

Product Analysis and Comparison 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.

REFERENCES

Arvind Kumar Sharma1, P.C. Gupta, 2012. "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.

Jos´e Ignacio Fern´andez-Villamor, Jacobo Blasco-Garc´ıa, Carlos ´A. Iglesias, Mercedes Garijo "A Semantic Scrapping Model for Web Resources" Spain.

Kali Pradeep, I., I. Bhagyasri , P. Praneetha, 2013. "E-Commerce With Backbone Of Data Mining", Proceeding of the *International Journal of Engineering Research & Technology (IJERT),* Vol. 2 Issue 7, July.

Li Mei, Feng Cheng, 2010. "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.

Rahul Dhawani, Mrudav Shukla, Priyanka Puvar, Bhagirath Prajapati, 2015. "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.

Rajendra Nath, Khyati Chopra, 2013. "Web Crawlers: Taxonomy, Issues & Challenges" Proceedings of the *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 4, April pp. 944-948.

Rama Mohan, B., A. Govardhan, 2013. "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, 201.

Richard, K. Lomotey, Ralph Deters, 2013. "RSenter: Tool for Topics and Terms Extraction from Unstructured Data Debris", Proceeding of the IEEE *International Congress on Big Data*.

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

Shridevi Swami, Pujashree Vidap, 2013. "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.

****