IOSR Journal of Mobile Computing & Application (IOSR-JMCA)
e- ISSN: 2394-0050, P-ISSN: 2394-0042.Volume 7, Issue 6 (Nov - Dec 2020), PP 09-15
www.iosrjournals.org

Improving Search Engine Performance for e-Commerce Transactions using the REST-based API

Iwara Arikpo[*,1], Emmanuel Echa[2]
[1]*Department of Computer Science, University of Calabar, Nigeria*
[2]*Directorate of ICT, Cross River University of Technology, Calabar, Nigeria*
Corresponding Author: Iwara Arikpo

Abstract: *This paper examined the difficulty in getting suitable and frequently updated prices for goods while shopping on e-commerce websites with the aid of price comparison websites. The study proposed to develop a price comparison search engine for e-commerce websites that will be faster, and possess a dynamic database frequently updated with prices of goods. This process is online-based, and the prices on the labels of displayed results of goods are hyper-linked to websites of different merchants who have these goods on their websites up for sale. Five price comparison websites were used as case study. The Representational State Transfer (REST) API was used to carry out search and time efficiency analysis of two goods across the five price comparison websites. This analysis was also carried out on the enhanced price comparison website as well with its own distinctive properties that differentiated its queryTime and totalTime from the other five. The resultant data showed that the crawler functions deployed in this study performed 0.6 seconds quicker than those already available in the five price comparison websites which were used as case study.*

Key Words: *website, search engine, price comparison, web crawler, shopbots, indexer, information retrieval*

I. Introduction

The Internet is a vast collection of billions of web pages containing terabytes of information stored on thousands of servers using Hypertext MarkupLanguage (HTML) (Gupta and Johari, 2009). To locate businesses and their products, it was customary to use directories, newspapers, and advertisements, while family and friends were also a relevant source of information.

Through the Internet, people can easily access large amount of data. Collecting and managing the massive amounts of data from a rapidly growing network of devices and sensors, processing that data, and then sharing it with other connected entities has become a complex task over the years, considering the yearly accumulation of data on the Internet (Oracle Corporation, 2014). Internet browsers constitute a first tool to ease the navigation experience, with Microsoft shipping the Windows Operating System offeringthe Microsoft Edge (formerly Internet Explorer), and other software companies doing the same. Thanks to the browsers, users move easily across documents and reach a large amount of information in just a few mouseclicks. Search technologies have beeninstrumentalin facilitatingthenavigation of the Internet, because they help users locate and aggregate content closely related to their areas of interest (Bughin, 2011). Search engines like Google, Yahoo, Bing, among others, constitute one type of search technologies. These search tools are often offered at no cost to users, but are financedthroughthe selling of advertisements that are related to the search queries. Since users tend to find these advertisements to be highly relevant to their interests, advertisers will pay well to place them (Varian, 2006). The search engines help the buyers to locate and purchase their products of interest.By taking advantage of the keywords the user provides while searching for information, search engines can pick and deliver targeted advertisements to the most interested audience. In this way, search engines become a relatively precise channel through which producers and retailers can reach consumers, thereby increasing their search(ing) engine optimization value. E-commerce search engines function to provide an efficient matching platform for the supply and the demand side, and to facilitate transactions aimed at generating revenue. Besides, they ensure user satisfaction by presenting the most relevant items that a potential customer is searching for (Goswami et al., 2018)

The Problem

With the growing adoption of the Internet and online storefronts, the question of how to find information of interest with regards to shopping on the Internet has become a daunting task for a lot of people. In the past, the Internet could be thought of as just a repository of static information, and search engines merely offered Internet users basic information retrieval. However, there is a need for more specific web search capabilities,as the web has evolved into a bustling marketplace where online business transactions have become the ideal way of buying and selling.It is hoped that, at the end,an ideal search engine will reduce searching costs in terms of time spent and Internet for consumers transacting on these platforms (Chai & Jones, 2005).

With this in mind, this study focuses on developing an optimized price comparison search engine which is expected to improve the relevance of products search queries on e-commerce websites. The returned query will work with the concept of cheapest products taking the higher places on the list followed by the higher in price and then hyper-link the accepted search query to the main e-commerce websites of the selling merchants.

II. Literature Review

There is a vast amount of literature in the area of price comparison search engines. However, the literature relevant to this study shows that, in recent years e-commerce has grown into an enormous business environment. This has resulted in a competitive business atmospherewhere marketers, both in the physical and online markets worldwide have a global face to brand themselves and sell their goods and services.

Brown and Goolsbee (2002) find that online price comparison in retail insurance markets has led to a substantial reduction in prices for consumers. Also, in many sectors, intense price competition in online markets has put downward pressure on prices in the conventional retail elements of that sector. Browsing from home to shop for an item from a store's website to check prices and window-shop is easier than physically visiting the store. 'Robots-software' that simultaneously querymany stores for price information aids cheap, easy and productive price comparison, for goods like gadgets, electronics, consumer durables and tickets.

Comparison sites or 'Shopbots' are electronic intermediaries that assist buyers when they search for product and price information on the Internet (Anderson, 2011). Shopbots have been functioning on the Internet since the late 1990s. Compared to other more traditional price comparison websites, most Shopbots stand alone, do not sell items themselves, but instead they rather gather and compare prices, products, and other important information from third-party merchants and present them on the search interface to the consumers in a readable way. In the same vein, consumers can easily compare offers presented by the search robots and have direct access to displayed links to the vendors' websites. These displayed links allow a buyer to click on the linked address and navigate directly to the site of the seller that offers the best deal on the product.

According to Serenko(2009), Shopbots are platforms through which buyers and sellers can establish contact with one another. In this sense, comparison sites essentially play an intermediation role. Furthermore, approaches have been considered to improve the technology behind price comparison websites, and how goods and services matched with their prices can be collected from merchants. Merchants who intend to list their products on the website can then supply their own lists of products and prices, and these are matched against the items of other merchants in the database. The technology behind this is made possible through a mixture of information extraction and fuzzy logic (an extension of Boolean logic that is based on the mathematical theory of fuzzy sets, which is a generalization of the classical set theory. It provides a very valuableflexibility for reasoning by introducing the notion of degree in the verification of a condition. By this, it enables a condition to be in a state other than true or false [1 or 0]; thus, taking cognisance inaccuracies and uncertainties(Denoncourt, 2013)).

Price comparison websites in the same context can also collect data from several merchant websites through a data feed file. Merchants make information available through a set of electronic formats. This data is then fed into the applications plugin's feeds of the comparison website and data is imported successfully. In the same vein, other third-party businesses are providing compiled data feeds so that comparison websites do not have to import from many different merchants (Chai &Jones, 2005).

III. Materials And Methods

System Design

The system design methodology was based on the object-oriented analysis and design methodology (OOAD)using the unified modelling language (UML). The system architecture for the price comparison search engine is shown inFigure 1. Other logical and physical design components for the system are also presented.

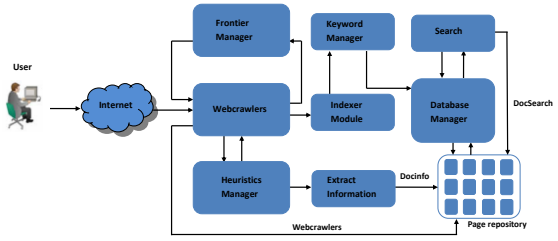

Figure 1: Adapted and modified architectural design for the system (Elwin &Rick, 2005).

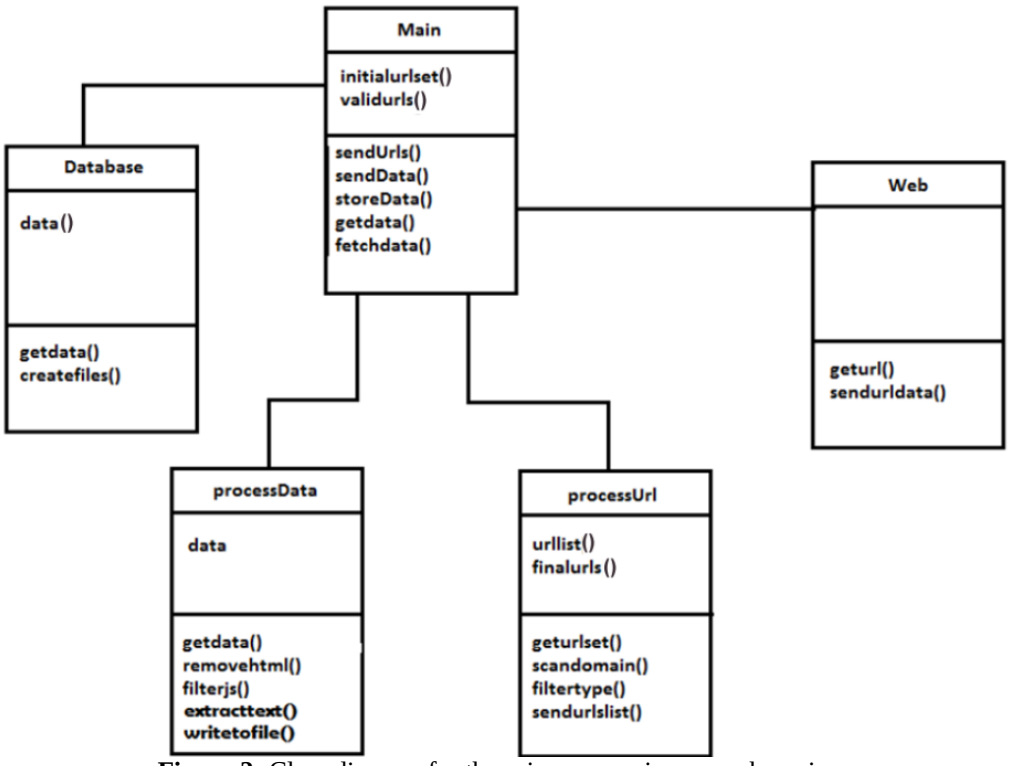

Figure 2: Class diagram for the price comparison search engine

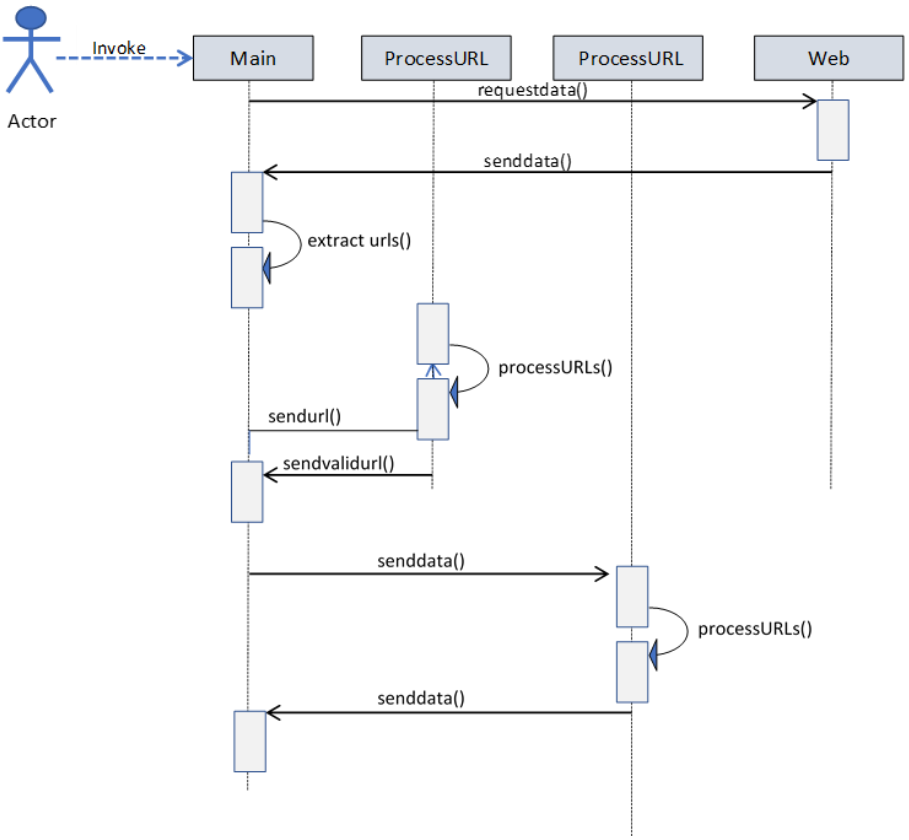

Figure 3: Sequence diagram for the price comparison search engine

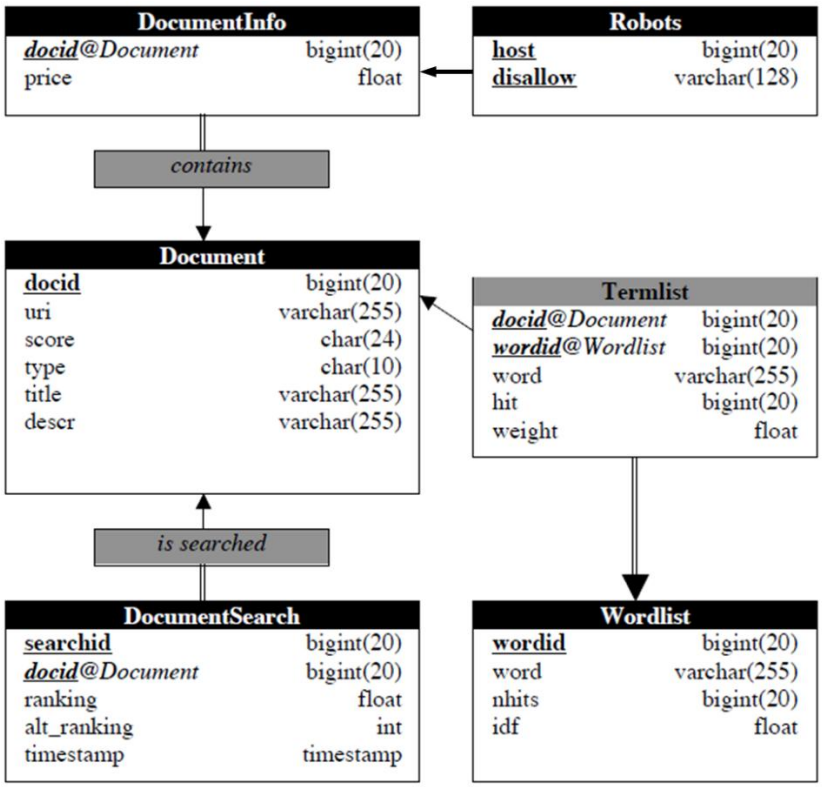

Figure 4: Database ER diagram for the price comparison search engine

Algorithm for the Crawler Class

The algorithm used for the crawler is shown below:

```
For(each url in urllist)
{
If (MIME type == html or txt)
{
If(url == given domain)
{
Extract data();
Extract urls();
Cleaner.script()
Extract text();
Extract metadata();
Save to disk();
}
}
}
```

The algorithm precisely describes the working of the Crawler. All the URLs are stored in a list called URL list. The system begins by taking each URL from the URL list iteratively and checking if the URL has a MIME type that will hold data or not. If the MIME type proves to be anything else rather than HTML or txt type, the URL is ignored and proceeds to the next URL. If the URL goes through, it is then checked again to see if it belongs to the domain specified by the user. This helps the crawler to be domain-specific. If the crawler is domain-specific, it will be able to download the most relevant web pages in an early stage of the crawl; which is then supplied to the Data Extractor (Filipowski, 2014). The Data Extractor extracts data from the URL and also extracts all the URLs in the data, and then stores them in the URL list. The data extracted is now filtered to remove the JavaScript and then the HTML tags are also removed. The required text is then extracted from this data and the metadata appropriately. The residual data is saved to the disk as appropriate.

Using REST-based API for Analysis

The search analysis of this study was carried out on five price comparison websites used as case study. These websites are **ngpricehunter.com**, **pricecheck.com.ng**, **bestprices.ng**, **priceinfo.com.ng** and **expedite.com.ng**. The Representational State Transfer (REST) Application Program Interface (API) was deployed to carry out search analysis of twoparticular products across the five price comparison websites (Rodriguez, et al., 2016).

The REST-based API architecture involves reading a web page which contains contents that make up Extensible Markup Language (XML) file. In Figure 5, a request is made through the REST API, the URL for the price comparison website provided as a request is being processed. As shown (for example, ngpricehunter.com), the result displayed for a search for Blackberry z10 on the konga.com website responds with the currentPage on which Blackberry z10 is found, the queryTime for processing the request, totalTime taken to respond to the search. Furthermore, product description is provided with the use of unique product identifier (sku) and "name".

```
#request:
http://api.ngpricehunter.com/v1/products(Description=blackberryz10*|sku=7619002)?show=sku,name
&pageSize=15&page=5&apiKey=YourAPIKey&format=json

#response:
{
  "currentPage": 5,
  "queryTime": "0.10",
  "totalTime": "0.45",

"/v1/products(Description=\"blackberryz10*\"|sku=7619002)?show=sku,name&page=5&format=json&api
Key=YourAPIKey",
  "products": [
    {
      "sku": 7619002,
      "name": "blackberryz10"
    },
    {
   ...
```

Figure 5: REST-based API result for ngpricehunter.com

IV. Results & Discussion

Evaluation of the REST-based API Results

Each price comparison website has its own distinctive properties that differentiates its queryTime and totalTime from the others. To make a better comparison, a well evaluated process is summarized in Table 1.

Table1: Results from REST-based API (i)

Comparison website(s)	Product name	Product sku	QueryTime (secs)	TotalTime (secs)
ngpricehunter.com	Blackberry z10	7619002	0.10	0.52
pricecheck.com.ng	Blackberry z10	1761045	0.11	0.51
bestprices.ng	Blackberry z10	1752654	0.12	0.50
priceinfo.com.ng	Blackberry z10	1729354	0.10	0.54
expedite.com.ng	Blackberry z10	1752291	0.10	0.55

The result for the price comparison search engine is presented in Table 2;and shows some differences with those displayed withthe five comparison search engines displayed in Table 1. To make a comparison of the result of this study with those of the five other comparison search engines, the queryTime and totalTime of all the searches carried out must be matched, as shown in Table 3.

Table2: Results from REST-based API (ii)

Comparison website(s)	Product name	Product sku	QueryTime (secs)	TotalTime (secs)
pricecomparisonsearchegine	Blackberry z10	53791	0.08	0.37
pricecomparisonsearchegine	LG TV 32"	53834	0.07	0.35

Table3: Results from REST-based API (iii)

Comparison website(s)	Product name	Product sku	QueryTime (secs)	TotalTime (secs)
ngpricehunter.com	Blackberry z10	7619002	0.10	0.52
pricecheck.com.ng	Blackberry z10	1761045	0.11	0.51

bestprices.ng	Blackberry z10	1752654	0.12	0.50
priceinfo.com.ng	Blackberry z10	1729354	0.10	0.54
expedite.com.ng	Blackberry z10	1752291	0.10	0.55
Pricecomparisonsearchegine	Blackberry z10	53791	0.08	0.37
Pricecomparisonsearchegine	LG TV 32"	53834	0.07	0.35

From Table 3, using the 'pricecomparisonsearchegine', the queryTime and totalTime time of the search result for Blackberry z10 and LG TV 32" is 0.08, 0.37 and 0.07, 0.35 respectively. With these results, it can beestablished that, search queries carried out on our price comparison search engine with a dynamic database will return results in a shorter time, compared to the five price comparison search engines considered in this study.

In the Search result interfaceof Figure 6, it is evident that, the real-life test results are shown to display the results of an item searched from the home page. This result has displayed the online shops that have the searched item within their products line, and also a direct link to their individual websites. It should be noted that, this result is still within the Price Comparison Search Engine. Every result and price displayed in hierarchical arrangement is drawn from the implementation crawler system and the dynamic database.

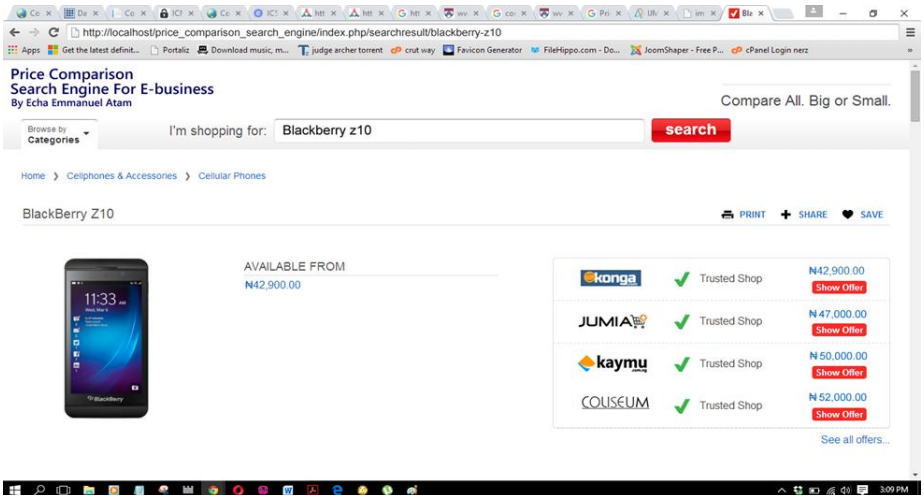

Figure 6: Screenshot of a search result interface

V. Conclusion

The researchers can conclude from the study that, there is some level of convention and structure among commercial websites for crawler functions for the reasonable extractionof information. It is fulfilling to realise that, the functions developed in this study have succeeded in improving search engine speed in e-commerce transactions. However, this study was limited to five price comparison websites. The diverse and dynamic nature of the wider web implies that, further research is necessary to attain a highly optimized price comparison search engine.

VI. Future Research

Although the price comparison search engine developed in this study has achieved its purpose,there is room for improvement. Of particular note, is the need to increase the number of web crawlers in future research, in order to improve thedistribution of the workload amongthe crawlers.

References

[1]. Anderson, S. P. (2011). *Advertising and the Internet, Chapter 11 in Handbook of the DigitalEconomy, Volume* 1 (M. Peitz and J. Waldfogel, Eds.), forthcoming.

[2]. Brown, J. and Goolsbee, A. (2002).Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry, *Journal of Political Economy*, Vol. 110, No. 3, pp. 481-507.

[3]. Bughin, J., Corb, L., Manyika, J., Nottebohm, O., Chui, M., Barbat., B. M., Said, R. (2011). *The impact of Internet technologies: Search*, McKinsey&Company: High Tech Practice, July 2011.

[4]. Chai E. and Jones R. (2005). Automated Price Comparison Shopping Search Engine PriceHunter. *Senior Design Project CSE 401*. University of Pennsylvania, Philadelphia. USA. Vol. 3, No. 2, pp. 3-6

[5]. Denoncourt, F. (2013). Introduction to fuzzy logic, *Massachusetts Institute of Technology*, USA.

[6]. Filipowski, K. (2014). Comparison of Scheduling Algorithms for Domain Specific Web Crawler. *European Network Intelligence Conference (ENIC)*, Wroclaw, Poland, 2014, pp. 69-74.

[7]. Goswami, A., Zhai, C., & Mohapatra, P. (2018).Towards Optimization of E-Commerce Search and Discovery, *The 2018 SIGIR Workshop On eCommerce*, June 2018

[8]. Gupta, P., Johari, K. (2009).Implementation of Web Crawler. In 2nd International Conference on Emerging Trends inEngineering and Technology, pp.838-843.
[9]. Oracle Corporation. (2014). The Internet of Things: Manage the Complexity, Seize the Opportunity, *Oracle Corporation*, USA.
[10]. Rodriguez, C., Baez, M., Daniel, F., Casati, F., Trabucco, J. C., Canali, L., &Percannella, G. (2016).REST APIs: A Large-Scale Analysis of Compliance with Principles and Best Practices. A. Bozzon et al. (Eds.): ICWE 2016, LNCS 9671, *Springer International Publishing Switzerland*, pp. 21–39, 2016.
[11]. Serenko, A. and Hayes, J. (2009). Investigating the functionality and performance of online shopping bots for electronic commerce: A follow-up study. *International Journal of Electronic Business*, 8(1): 1-15.
[12]. Varian, H. R. (2006). The Economics of Internet Search, *2007 Angelo Costa Lecture inRome*.