IMPLEMENTATION OF THE BFS ALGORITHM AND WEB SCRAPING TECHNIQUES

by Nurdin, Bustami, Muhammad Hutomi Marischa Elveny, Rahmad Syah

Submission date: 25-Jul-2021 03:31PM (UTC+0700)
Submission ID: 1623679306
File name: 5_6334531111774847675.pdf (433.77K)
Word count: 5319
Character count: 27868

Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPLEMENTATION OF THE BFS ALGORITHM AND WEB SCRAPING TECHNIQUES FOR ONLINE SHOP DETECTION IN INDONESIA

[1]NURDIN, [2]BUSTAMI, [3]MUHAMMAD HUTOMI, [4]MARISCHA ELVENY*, [5]RAHMAD SYAH

[1,2,3]Department of Informatics, Universitas Malikussaleh, Lhokseumawe, Aceh, Indonesia.

[4]Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia.

[5]Universitas Medan Area, Medan, Indonesia.

E-mail: [1]nurdin@unimal.ac.id, [2]busabiel@gmail.com, [4]marischaelveny@usu.ac.id*, [5]bayurahmadsyah45@gmail.com

ABSTRACT

This The number of online shops in Indonesia on the Shopee e-commerce web makes it difficult for consumers to detect the authenticity of online stores. This online shop detection system is an application to detect genuine online shops or fake online stores (dropship). This study aims to assist consumers in the process of searching for genuine online stores that sell the desired products quickly and automatically. The method used in this research is the Breadth First Search (BFS) algorithm and the Web Scraping technique which will be applied to the Shopee e-commerce web in Indonesia based on three parameters, namely delivery, store rating, and response rate. The results of this study indicate that the Breadth First Search algorithm with the Web Scraping technique can be used to complete the process of retrieving store data and product data in e-commerce and able to explore as well as check online stores automatically with good performance. The test results are based on factors such as precision, recall, F-Score, and Accuracy, with the results of 81% precision, 89% recall, 84.82% f-score, and 90% accuracy with 100 search data.
Keywords: *Online Shop, BFS Algorithm, Web Scraping, E-Commerce, Detection*

1. INTRODUCTION

Indonesia is a country with the fastest growth rate for e-commerce in the world. The rapid growth of e-commerce in Indonesia is due to the number of internet users in Indonesia that exceeds 100 million users. Based on statistics, the average internet user in Indonesia spends money on online shopping sites reaching Rp. 3,190,000 per person [1]. The cultural shift from conventional buying and selling to buying and selling online has a significant impact on the development of e-commerce in Indonesia. One of the factors for the growth of e-commerce in Indonesia is the existence of a marketplace. Shopee is one of the marketplaces in Indonesia. Shopee provides various facilities as well as various categories of items that are attractive to buy [2].

The original online store is an online store that sells most of its products directly, has a warehouse, and has stock products that are ready to be sold and sent to consumers. Meanwhile, online dropship shop (fake) is an online store that does not have a warehouse and does not have stock items that are ready to be sent directly to consumers. Dropship online shop does not mean an online shop that sells counterfeit goods, but an online shop that sells goods by taking pictures of other people's products to be promoted in their shop [3], [4].

There are several previous studies related to the research topic that the author did such as research conducted by [5] with the title On Multi-Thread Crawler Optimization for Scalable Text Searching with the results of his research comparing the BFS and DFS algorithms, where it was found that the BFS algorithm has advantages in time efficiency, simplicity, and flexibility in visiting pages on the Wikipedia web. Breadth First Search is also superior to Depth First Search in gathering searches for popular topics both at the global level (across the Web) and at the national level (domain.nl): Google Trends, WikiStats, and Queries collected from users of archives of Dutch historical newspapers [6]. Breadth First Search has good efficiency in terms of URL link browsing [7].

Web scraping is the process of retrieving a semi-structured document from the internet, generally take the form of web pages in a markup language such as HTML or XHTML, and analyze the

document for data retrieval. Also called web scraping applications (intelligent, automated, or autonomous agents) only focus on how to obtain data through data collection and extraction with varying data sizes [8]. Web scraping techniques can also be used in collecting promo data on e-commerce sites in providing appropriate promo information [9].

As for the problem in this study, because every year the number of online stores that sell their goods / products at Shopee continues to grow. The number of online stores contained in the Shopee marketplace will make it difficult for consumers to determine whether the online store that sells the goods that consumers want is the best online store from several aspects such as having stock in its warehouse or not, has the best selling price or not, and has a delivery service the best or not. This study aims to make it easier for consumers to choose the original online shop before buying products from the online shop at Shopee effectively and efficiently. Make it easier for consumers to check the authenticity of the original online store on Shopee e-commerce quickly and automatically according to the product keywords that consumers want.

2. LITERATURE REVIEW

Previous research conducted by (Kapil, 2019) shows that the BFS algorithm is a simple algorithm that can be used for the purpose of crawling web pages http://www.cdlu.edu.in which is built with ASP.Net and the VB language. The task of a web crawler is to navigate the web and extract new pages for storage in the database [10]. While this author's research examines the Shopee e-commerce to retrieve page data with the PHP programming language and Java Script.

Research conducted by (Sun et al, 2019) compared the BFS and DFS algorithms, where it was found that the BFS algorithm has advantages in time efficiency, simplicity, and flexibility in visiting pages on the Wikipedia web. Browsing Wikipedia sites with BFS for 273.05155 seconds, while DFS 1000, 29163 seconds [5]. The difference with this author's research is the process of visiting Shopee's e-commerce web in retrieving data using the BFS algorithm and to find out the status of the online shop visited by the original or dropship based on the search keywords that the user enters into the search system.

Comparison between DFS, BFS, BEFS algorithms in searching for geotagged images. Where it is found that the BFS and DFS algorithms have better time efficiency than BEFS, but the BEFS algorithm is better at detecting geotagged images, namely 2% (201 out of 9832 images), BFS by 1.3% (52 out of 3898 images), and DFS 0.4% (13 of 3576 images) [11]. Research conducted by [7] is a comparison of the accuracy of browsing results where BFS is 42.33% and TF-IDF is 52.78% of the 100 crawled URL links. However, BFS has good efficiency for URL link browsing.

The crawler algorithm can search for relevant pages on web pages [12]. Web Crawler or spider is a computer program that searches the WWW sequentially and automatically. Crawlers which are sometimes called spiders, bots, or agents are software whose purpose is done for web crawling [13]. Of the 100,000 pages, BFS provides 28,797 pages of relevant pages with a precision rate of 28,797% [14]. Web crawlers are the main part of search engines, and the details are in their algorithms and architecture [16]. Meanwhile, the Treasure-Crawler's performance in performing web browsing in retrieving certain data with a precision and recall rate of almost 50% [15]. Web browsers can also be used to search and store data in order to keep it indexed to facilitate t searching by clients [16]. Apart from that, web crawlers are an important component of search engines, data mining, and other internet applications [14]. In search engines, crawlers are responsible for finding and downloading web pages [17].

Apart from this research, there are several studies that have been conducted by the author, including Searching the shortest route for distribution of LPG in Medan city using ant colony algorithm [18], Data driven optimization approach to fish resource supply chain planning [19], MILP model for integrated fish supply chain planning [20] and Robust optimization approach for agricultural commodity supply chain planning [21].

3. RESEARCH METHODOLOGY

3.1 Types and Sources of Data

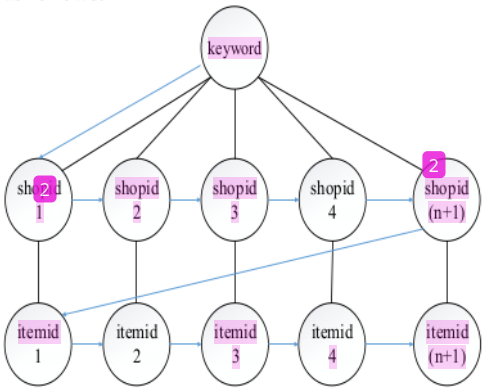
The type of primary data used in this research is online shop information data on the Shopee website in Indonesia, with the keyword "Iphone 7", the data collection process is carried out in real time. The data taken is a link to the entire online store listed based on the entered keywords and then taken online store information such as number of products sold, product ratings, total number of products sold, overall product rating, length of time joining, percentage of chats replied to, chat time replied, store appraisal.

Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Breadth First Search Algorithm

In this study, the algorithm used is the Breadth First Search algorithm, which is used as a search process as well as online shop link visits based on the entered keywords, then data is taken in the form of shopid and itemid which are used as nodes. The application of the Breadth First Search algorithm is as follows.



BFS: :keyword, shopid 1, shopid 2, shopid 3, shopid 4,.., shopid (n+1), itemid 1, itemid 2, itemid 3, itemid 4,..., itemid (n+1).

Figure 1: Breadth First Search Algorithm

In Figure 1 above, the node level 0 is the store search keyword based on product search, where each search keyword has a product that shopid will take as level 1 node and itemid as level 2 node of the BFS scheme in this study.

3.3 Web Scraping

Web scraping is the process of retrieving a semi-structured document from the internet, generally in the form of web pages. Web scraping has the following steps [22].

1. Create Scraping Template: The programmer learns the HTML document from the website which information will be taken for the HTML tag that encloses the information to be retrieved.
2. Explore Site Navigation: Program makers learn navigation techniques on websites which information will be taken to be imitated in the web scraper application that will be made.
3. Automate Navigation and Extraction: Based on the information obtained in steps 1 and 2 above, a web scraper application was created to automate information retrieval from the specified website.
4. Extracted Data and Package History: The information obtained from step 3 is stored in a table or database tables. How it works see Figure 2. (The Computer Advisor) [23].

Figure 2: How Web Scraping Works
(Source: The Computer Advisor, 2019)

3.4 System Design Stages

The stages of the overall system design process are as follows:

1. Input search keywords, input number of searches: is the process of entering search keywords, as well as the number of searches to be performed.
2. Scraping all shopid, scraping all itemid: is a process to retrieve all shopid and itemid results from keyword-based searches.
3. Create roaming queue: the process of creating a roaming queue for implementing the Breadth First Search algorithm.
4. Breadth First Search Browsing: is the process of browsing shopid and itemid queues using the Breadth First Search algorithm.
5. Scraping shop and product information: is the process of retrieving store and product information based on search keywords.
6. Checking the original shop or dropship: is a process to check whether the online store is genuine or dropship with predetermined criteria.
7. Empty queue ?: is a process to check whether the browsing queue is still there or not. If it's still there, the Breadth First

Search will be explored again. If not, then go to the next process.

8. Original or dropship store information output: is the process of displaying whether the online store is genuine or dropship.

4. RESULT AND DISCUSSION

4.1 Queue Formation

At this stage, it is the formation of a queue for node visits using the Breadth First Search algorithm. At this stage the system being built will automatically build a queue based on the search keywords that the user enters. The search keyword will be the source node / root node. Each node will store a unique code and will be used for link visits later. In the first layer of forming the queue tree, it will be used to accommodate shopid nodes which are unique to visitors from online stores, and in the second layer of the queue formation tree will accommodate itemid nodes which are used as nodes for product visits from online stores that are queued up. In this study, the authors conducted a search input for 100 online stores to detect whether the store's status was genuine or fake using the Breadth First Search algorithm. Further explanation regarding the formation of queue trees in this study can be seen in Figure 3.

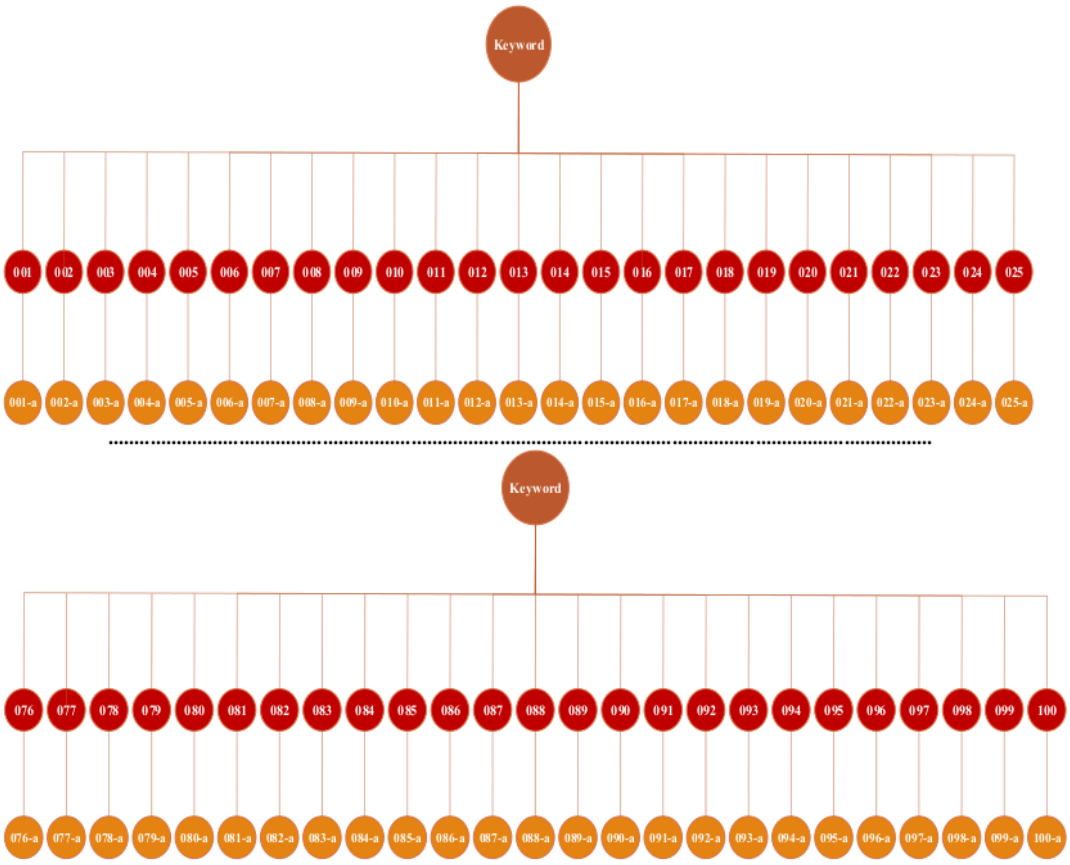


Figure 3: Forming Queue Tree

Each node in the first layer accommodates the shopid and the second layer accommodates the item id. The explanation of Figure 3, which can be seen in table 1.

Table 1: Queue Formation

Node	Layer	Data	Information
Keyword	Layer 0	Iphone 7	Root Node
001	Layer 1	374936	1st node (shop id 1)
002	Layer 1	2299609	2nd node(shop id 2)
003	Layer 1	3110500	3rd node (shop id 3)
004	Layer 1	3572585	4th node (shop id 4)
005	Layer 1	3954177	5th node (shop id 5)
006	Layer 1	4121774	6th node (shop id 6)
007	Layer 1	4304336	7th node (shop id 7)
008	Layer 1	7298199	8th node (shop id 8)
009	Layer 1	9900293	8th node (shop id 9)
010	Layer 1	10124325	10th node (shop id 10)
011	Layer 1	10198806	11th node (shop id 11)
012	Layer 1	10878607	12th node (shop id 12)
013	Layer 1	11211469	13th node (shop id 13)
014	Layer 1	11911115	14th node (shop id 14)
015	Layer 1	12319692	15th node (shop id 15)
016	Layer 1	13056258	16th node (shop id 16)
017	Layer 1	13637076	17th node (shop id 17)
018	Layer 1	14351274	18th node (shop id 18)
019	Layer 1	14553130	19th node (shop id 19)
020	Layer 1	14671128	20th node (shop id 20)
.....	.....	.....	.....
100	Layer 1	126491982	99th node (shop id 99)

4.2 Node Visiting

The process of visiting nodes is in accordance with the Breadth First Search algorithm, namely with FIFO (First In First Out) where the first node is entered into the queue, it will be issued [25]. At this point, each node in the queue is visited one by one. After the visit, the data scraping process will be carried out on that page. The data taken is store data such as store name, number of products, store location, store rating, shop owner, good rating, normal rating, bad rating, following, follower, response rate, and delivery. And product data such as product names, product prices, products sold, product status, product ratings, ratings with photos, ratings with text, number of reviewers, 1 star, 2 star, 3 star, 4 star, 5 star, and links product image. However, the writer's focus is on several attributes in store data such as store name, store location, type of delivery, store rating, and response rate. This data will be used at the online store detection stage later. Store data can be seen in table 2.

Table 2: Online Shop Data

Store Name	Location Store	Delivery	Store Rating	Response Rate
X1	DKI Jakarta, Id	Reguler	5.0	83
X2	Banten, Id	Reguler	4.7	98
X3	Jawa Timur, Id	Reguler, Reguler (Cargo), Instant, Next Day	4.7	99
X4	DKI Jakarta, Id	Reguler, Same Day, Instant, Next Day	4.4	100
X5	DKI Jakarta, Id	Reguler, Same Day, Instant	4.8	98
X6	DKI Jakarta, Id	Reguler	4.5	98
X7	DKI Jakarta, Id	Reguler, Instant, Next Day	4.8	96
X8	Banten, Id	Reguler	4.9	100
X9	Jawa Barat, Id	Reguler, Same Day, Instant	4.4	68
X10	Jawa Barat, Id	Reguler, Instant	4.7	97
X11	Jawa Tengah, Id	Reguler	4.8	80
X12	DKI Jakarta, Id	Reguler, Same Day, Instant	4.8	100
X13	DKI Jakarta, Id	Reguler, Instant	4.9	92
X14	Jawa Tengah, Id	Reguler	5.0	100
X15	Jawa Tengah, Id	Reguler	5.0	82
X16	DKI	Reguler	4.7	70

	Jakarta, Id			
X17	Jawa Barat, Id	Reguler, Next Day	4.8	66
X18	Jawa Timur, Id	Reguler	4.7	97
.....	.....	.....	.....	.....
X100	DKI Jakarta, Id	Reguler, Same Day, Instant	4.9	99

4.3 Online Shop Detection Parameters

Online shop detection parameter is carried out based on the data obtained in the search process or node visits. The detection of whether the online shop is original or dropship is based on several criteria in the shop as in table 3.

Table 3: Criteria for Online Stores

Parameter	Online Store	
	Original	**Dropship**
Delivery	There are Same Day Service, Regular, Next Day, Instant.	Only Shopee Logistics shipments are available
Store Rating	≥ 4	< 4
Response Rate	≥ 90 %	< 90 %

4.4 Test Result

After testing the shop data based on the criteria and parameters for the detection of the online shop described above, the status of the online store is shown in table 4

Table 4: Online Shop Status

Store Name	Online Shop Status Based		
	Delivery	**Store Rating**	**Response Rate**
X1	Original	Original	Dropship
X2	Original	Original	Original
X3	Original	Original	Original
X4	Original	Original	Original
X5	Original	Original	Original
X6	Original	Original	Original
X7	Original	Original	Original
X8	Dropship	Original	Original
X9	Original	Original	Original
X10	Original	Original	Dropship
X11	Original	Original	Original
X12	Original	Original	Dropship
X13	Original	Original	Original
X14	Original	Original	Dropship
X15	Original	Original	Original
X16	Original	Original	Dropship
X17	Original	Original	Dropship
X18	Original	Original	Dropship
X19	Original	Original	Original
X20	Original	Original	Original
X21	Original	Original	Original
X22	Original	Original	Original
X23	Original	Original	Original
X24	Original	Original	Dropship
X25	Original	Original	Original
X26	Original	Original	Original
X27	Original	Original	Original
X28	Original	Original	Dropship
X29	Original	Original	Dropship
X30	Dropship	Original	Original
X31	Original	Original	Original
X32	Original	Original	Original
X33	Original	Original	Dropship
X34	Original	Original	Original
X35	Original	Original	Original
X36	Original	Original	Original
X37	Original	Original	Original
X38	Original	Original	Original
X39	Original	Original	Dropship
X40	Original	Original	Original
X41	Original	Original	Original
X42	Original	Original	Original
X43	Original	Original	Dropship
X44	Original	Original	Original
X45	Original	Dropship	Original
X46	Original	Original	Original
X47	Original	Original	Original
X48	Dropship	Original	Dropship
X49	Original	Original	Original
X50	Original	Original	Dropship

X51	Original	Original	Original
X52	Original	Original	Original
X53	Original	Original	Original
X54	Original	Original	Original
X55	Original	Original	Original
X56	Original	Original	Original
X57	Original	Original	Original
X58	Original	Original	Dropship
X59	Original	Original	Original
X60	Original	Original	Original
X61	Original	Original	Dropship
X62	Original	Original	Original
X63	Original	Original	Original
X64	Original	Original	Dropship
X65	Original	Original	Dropship
X66	Original	Original	Original
X67	Original	Original	Original
X68	Original	Original	Dropship
X69	Original	Original	Original
X70	Original	Original	Dropship
X71	Original	Original	Original
X72	Original	Original	Original
X73	Original	Original	Original
X74	Original	Original	Dropship
X75	Original	Original	Original
X76	Original	Original	Original
X77	Original	Original	Dropship
X78	Original	Original	Dropship
X79	Original	Original	Dropship
X80	Original	Original	Original
X81	Original	Original	Original
X82	Original	Original	Dropship
X83	Original	Original	Dropship
X84	Original	Original	Original
X85	Original	Original	Original
X86	Original	Original	Original
X87	Original	Original	Dropship
X88	Original	Original	Original
X89	Original	Original	Dropship
X90	Original	Original	Original
X91	Original	Original	Original
X92	Original	Original	Original
X93	Original	Original	Original
X94	Original	Original	Original
X95	Dropship	Original	Original
X96	Original	Original	Original
X97	Original	Original	Dropship
X98	Original	Original	Original
X99	Original	Dropship	Original
X100	Original	Original	Original

Based on data table 4 online shop status, a recapitulation can be made based on the parameters

Table 5: Online Shop Detection Results

Parameter	Number of Online Shop Detection	
	Original	Dropship
Delivery	98	2
Store Rating	98	2
Response Rate	70	30

From the above test data, the authors get data that the application can detect online stores well, from the keyword "Iphone 7" with many searches of 100 different online store data. The time it takes for the system to visit each node, from the first node to the 200 node by implementing the Breadth First Search algorithm, takes about 161.6 seconds. With the number of detections based on delivery, there were 98 genuine online shops and 2 dropship, based on store ratings, 98 genuine online shops were obtained and 2 dropship, and based on the response rate, 70 genuine online shops and 30 dropship.

The following is a graph of the test results of the detection of genuine or fake online stores based on the criteria or parameters of the online shop in Figure 4.

Figure 4: Online Shop Status Diagram

Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.5 Breadth First Search Algorithm Performance

From research conducted by applying the Breadth First Search algorithm in the detection of genuine or fake online stores in e-commerce by searching 100 online stores. Based on observations, the Breadth First Search algorithm has a pretty good performance in searching for relevant data.

Table 6: Relevant Visits Pages

Algorithm	Search	Number of Relevant Pages
Breadth First Search (BFS)	100	81

The performance testing of the Breadth First Search algorithm in this study uses several factors such as precision, recall, F-Score, and Accuracy [12].

1. Precision
   It is the ratio of positive true predictions compared to the overall positive predicted results (all pages retrieved) [20].
   Precision (P) = (TP) / (TP + FP)
   = (81) / (81 + 19)
   = 81
2. Recall
   It is the ratio of true positive predictions (relevant pages) to the overall true positive data (pages retrieved) [24].
   Recall (R) = (TP) / (TP + FN)
   = (81) / (81 + 10)
   = 89
3. F1 Score
   F1 Score is a weighted average comparison of precision and recall [25].
   F1-Score = 2 * (R * P) / (R + P)
   = 2 * (89 * 81) / (89 + 81)
   = 84.82
4. Accuracy
   Is the ratio of true predictions (positive and negative) to the overall data.
   Accuracy= (Number of Correct Predictions) / (Total of predictions)
   = (81 + 9) / (100)
   = 90

Based on the results of the performance test above, it can be concluded that the performance of the Breadth First Search algorithm in online shop detection with 81% precision, 89% recall, 84.82% f-score, and 90% accuracy.

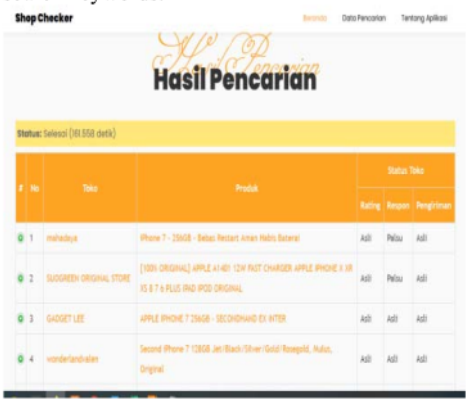
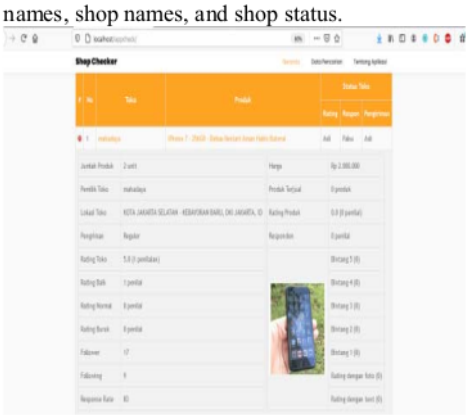
4.6 System Testing

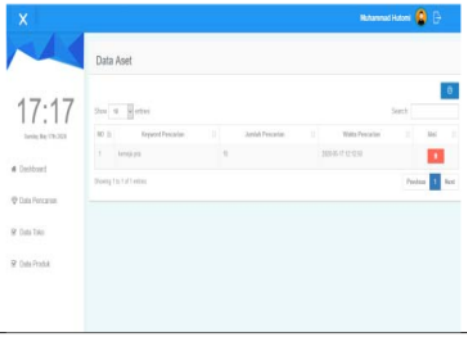
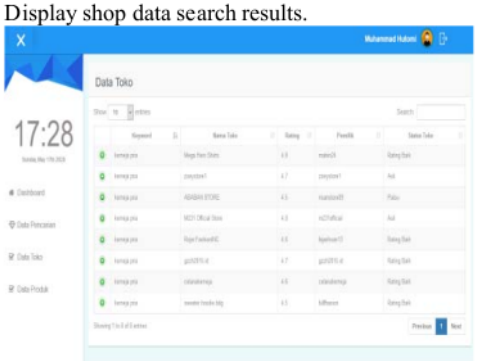
Testing the system used in this study is to use Black Box Testing. Testing with Black Box Testing is done by providing a number of inputs to the program. The input is then processed according to its functional requirements to see whether the application program can produce the desired output and is also in accordance with the basic functions of the program. The system testing can be seen in table 7.

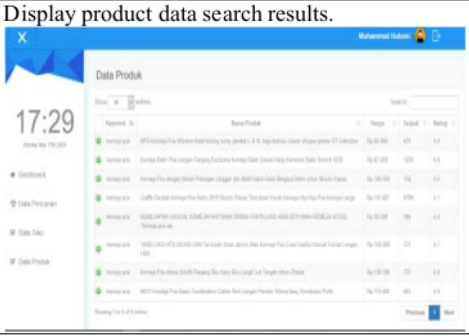
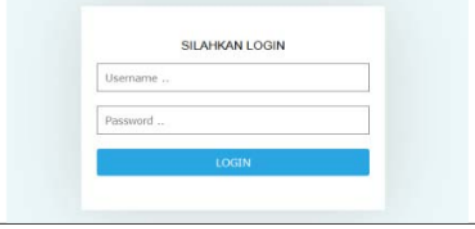
Journal of Theoretical and Applied Information Technology
30th June 2021. Vol.99. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 7: Black Box Testing

Scenario	Test Case	Hypothesis	Result
Online shop checking form	Enter the keyword and number of searches then click the check store button.	Perform shop and product searches based on search keywords. 	Valid
Search result detail form	Click the plus button on the table	Displays detailed search results for product names, shop names, and shop status. 	Valid
Search data form	Click the search data menu button	Displays the results of store and product search data. 	Valid

Admin login form	Enter your username and password, then click the Login button.	Displays the admin dashboard menu. 	Valid
Admin search data form	Click the search data menu button	Displays search keyword data that has been done before. 	Valid
Form clear search data	Click the delete button on the table.	Delete the selected search keyword data. 	Valid
Store data form	Click the shop data menu button.	Display shop data search results. 	Valid

Product data form	Click the product data button	Display product data search results. 	Valid
Logout form	Click the logout button	Exit admin panel and display admin login form. 	Valid

5. CONCLUSIONS

The online store detection application that was built can detect online stores properly, based on shipping parameters, store ratings, and response rates. The Breadth First Search algorithm is a simple algorithm that can be used for the process of retrieving store data and product data on Shopee e-commerce in Indonesia with the help of the Web Scraping technique. The results of the research on the keyword "Iphone 7" with the number of searches of 100 online shops resulted in online shop detection based on delivery parameters detecting 98 genuine online shops and 2 fake online shops, based on the shop rating parameters detecting 98 genuine online shops and 2 fake online shops, and based on parameters response rate detects 70 genuine online shops and 30 fake online shops. The detection results of an online shop do not necessarily categorize an online shop as genuine or fake as a whole, but it is detected based on parameters, so it is possible that in one parameter it can be said to be a genuine online store category, and in one other parameter it is categorized as a fake online shop. This application has a weakness that is very dependent on internet connection speed, because it will affect the number of online shops detected as well as the visiting time of each node.

REFRENCES:

[1] H. Widowati, 2019, "Indonesia Jadi Negara dengan Pertumbuhan E-Commerce Tercepat di Dunia, Databoks" Available at https://databoks.katadata.co.id/datapublish/2019/04/25/indonesia-jadi-negara-dengan-pertumbuhan-e-commerce-tercepat-di-dunia

[2] Hermawan, 2019. "Apa itu Shopee? Keunggulan Apa Saja yang dimiliki Shopee?" Available at https://www.nesabamedia.com/apa-itu-shopee/

[3] Yuli, 2017, "Cara Membedakan Online Shop Asli dan Dropship di Shopee," Available at: https://blog.bajuyuli.com/2017/10/toko-online-asli-Dropship-shopee.html

[4] N. Nurdin, M. Hutomi, M. Qamal, and B. Bustami, "Sistem Pengecekan Toko Online Asli atau Dropship pada Shopee Menggunakan Algoritma Breadth First Search,"*Jurnal RESTI*, vol.4, no.6, 2020, pp.1117-1123.

[5] G. Sun, H. Xiang, and S. Li, "On Multi-Thread Crawler Optimization for Scalable Text Searching," *J. Big Data*, vol. 1, no. 2, 2019, pp. 89–106.

[6] T. Samar, M. C. Traub, J. van Ossenbruggen, and A. P. de Vries, "Comparing topic coverage in breadth-first and depth-first crawls using anchor texts," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9819 LNCS, 2016, pp. 133–146.

[7] A. M. Tanvir, Y. Kim, and M. Chung, "Design and Implementation of an Efficient Web Crawling Using Neural Network," 2020.

[8] R. C. Wijaya, J. Andjarwirawan, and H. N. Palit, "Aplikasi Pencarian Produk Jual Mobile Devices dari Berbagai Situs E-commerce," *J. Infra*, vol.4, no.1, 2016.

[9] F. R. Wibowo, D. S. Rusdianto, and A. Arwan, "Pengembangan Sistem Pengumpulan Promo E-Commerce Berbasis Website Dengan Menerapkan Teknik Web Scraping Dalam Proses Pengambilan Data Promo," *J. Pengemb. Teknol. Inf. dan Ilmu Komputer*., vol. 3, no. 3, 2019, pp. 2887–2893.

[10] M.Y. Kapil, "Design of a Novel Interface for a Web Crawler," *International Journal of Electrronics Engineering*, vol. 11, no. 1, 2019, pp. 952-958.

[11] N. Fazal, K. Q. Nguyen, and P. Fränti, "Efficiency of web crawling for geotagged image retrieval," *Webology*, vol. 16, no. 1, 2019, pp. 16-39.

[12] S. Vijayarani, and M. E. Suganya, "Web Crawling Algorithms–A Comparative Study," *International Journal for Science and Advance Research in Technology*, vol. 2, no. 2, 2016, pp. 351-357.

[13] R. Kumar, J. Anurag, and A. Chetan, "Survey Of Web Crawling Algorithms," *Advances in Vision Computing: An International Journal*, vol. 1, no. 2/3, 2014, pp. 1-8.

[14] S. Kumar, and D. A. Kumar, "A Comparative Study of Scheduling Algorithms for Web Crawling Using Vb. Net Technology," *International Journal of Research in Science And Technology*, vol. 1, no. 2, 2011.

[15] A. Seyfi, A. Patel, and J. C. Júnior, "Empirical evaluation of the link and content-based focused Treasure-Crawler," *Computer Standards & Interfaces*, vol. 44, 2016, pp. 54-62.

[16] F. J. M. Shamrat, Z. Tasnim, A. S. Rahman, N. I. Nobel, and S. A. Hossain, "An Effective Implementation Of Web Crawling Technology To Retrieve Data From The World Wide Web (WWW)," *International Journal Of Scientific & Technology Research*, vol. 9, no. 1, 2020, pp. 1252-1256.

[17] M. A. Golshani, V. Derhami, and A. ZarehBidoki, "A Novel Crawling Algorithm For Web Pages," In Asia Information Retrieval Symposium, 2011, pp. 263-272.

[18] Nurdin, Taufiq, and Fajriana. "Searching the shortest route for distribution of LPG in Medan City using ant colony algorithm". IOP Conf. Series: Materials Science and Engineering, 725 (2020) 012121, 2020.

[19] Nurdin, M. Zarlis, Tulus and S. Efendi, "Data Driven Optimization Approach to Fish Resources Supply Chain Planning in Aceh Province", *IOP Conf. Series: Journal of Physics: Conf. Series* 1255 (2019) 012081, 2019.

[20] Nurdin, M. Zarlis, Tulus and S. Efendi, "Mixed Integer Linear Programming Model for Integrated Fish Supply Chain Planning" *Journal of Theoretical and Applied Information Technology*., Vol.98, No.12, 2020, pp. 2017-2028.

[21] Nurdin, Bustami, and Maryana, "Robust Optimization Approach for Agricultural Commodity Supply Chain Planingg" *Journal of Theoretical and Applied Information Technology*., Vol.99, No.2, 2021, pp. 304-315.

[22] A. Josi, L. A. Abdillah, and Suryayusra, "Penerapan Teknik Web Scraping Pada Mesin Pencari Artikel Ilmiah," *Jurnal Sistem Invormasi (JSI)*, vol. 5, no. 2, 2014, pp. 159-164.

[23] The Computer Advisor. Web site scraper the most effective tool for web data extraction, [Online] (Updated 09 Juni 2014) Available at: http://www.thecomputeradvisor.net/web-site-scraper-the-most-effective-tool-for-web-data-extraction/

[24] S. Jaiganesh, P. Babu, and K. N. Satheesh, "Comparative Study Of Various Web Search Algorithms For The Improvement Of Web Crawler," *International Journal of Engineering Research & Technology,* vol. 2, no. 4, 2013, pp. 1020-1025.

[25] S. Saranya, B. S. E. Zoraida, and P. V. Paul, "A Study on Competent Crawling Algorithm (CCA) for Web Search to Enhance Efficiency Information Retrieval," *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, 2015, pp. 9-16.

IMPLEMENTATION OF THE BFS ALGORITHM AND WEB SCRAPING TECHNIQUES

ORIGINALITY REPORT

15%	8%	7%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	4%
2	jurnal.iaii.or.id Internet Source	3%
3	Submitted to President University Student Paper	1%
4	Rahmat Shaumi. "Indonesian-Aceh Application Translation Design Based on Android", International Journal Education and Computer Studies (IJECS), 2021 Publication	1%
5	I Gede Susrama Mas Diyasa, Agus Prayogi, Intan Yuniar Purbasari, Ariyono Setiawan, Sugiarto, Prismahardi Aji Riantoko. "Data Classification of Patient Characteristics Based on Nutritional Treatment Using the K-Nearest Neighbors Algorithm", 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), 2021 Publication	1%

6	Submitted to Universitas Brawijaya Student Paper	1 %
7	Irfan Mahendra, SW Sulistianto, Astriana Mulyani, Agus Wiyatno, Oki Rosanto. "Assessing E-Commerce Success from a Millennial Perspective in Indonesia", Journal of Physics: Conference Series, 2020 Publication	1 %
8	Puteri Prameswari, Isti Surjandari, Enrico Laoh. "Opinion mining from online reviews in Bali tourist area", 2017 3rd International Conference on Science in Information Technology (ICSITech), 2017 Publication	<1 %
9	ir.cwi.nl Internet Source	<1 %
10	www.koreascience.or.kr Internet Source	<1 %
11	Submitted to University of Northumbria at Newcastle Student Paper	<1 %
12	Moh Abdul Latief, Alhadi Bustamam, Titin Siswantining. "Performance Evaluation XGBoost in Handling Missing Value on Classification of Hepatocellular Carcinoma Gene Expression Data", 2020 4th International	<1 %

Conference on Informatics and Computational Sciences (ICICoS), 2020
Publication

13	www.mi-co.com Internet Source	<1%
14	Goldy V Nivaan, Andi W. R. Emanuel. "Analytic Predictive of Hepatitis using The Regression Logic Algorithm", 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2020 Publication	<1%
15	Nurdin, Taufiq, Fajriana. "Searching the shortest route for distribution of LPG in Medan city using ant colony algorithm", IOP Conference Series: Materials Science and Engineering, 2020 Publication	<1%
16	id.123dok.com Internet Source	<1%
17	S. Saranya, B.S.E. Zoraida, P. Victor Paul. "Chapter 2 A Study on Competent Crawling Algorithm (CCA) for Web Search to Enhance Efficiency of Information Retrieval", Springer Science and Business Media LLC, 2015 Publication	<1%

18	Submitted to School of Business and Management ITB Student Paper	<1%
19	Arvind K Sharma, Vandana Shrivastava, Harvir Singh. "Experimental performance analysis of web crawlers using single and Multi-Threaded web crawling and indexing algorithm for the application of smart web contents", Materials Today: Proceedings, 2020 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		