

WEIDJ: Development of a new algorithm for semi-structured web data extraction

Ily Amalina Ahmad Sabri, Mustafa Man

Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Terengganu, Malaysia

Article Info

Article history:

Received Mar 29, 2020

Revised Aug 9, 2020

Accepted Aug 29, 2020

Keywords:

Document object model
JavaScript object notation
Web data extraction
Wrapper extraction of image

ABSTRACT

In the era of industrial digitalization, people are increasingly investing in solutions that allow their process for data collection, data analysis and performance improvement. In this paper, advancing web scale knowledge extraction and alignment by integrating few sources by exploring different methods of aggregation and attention is considered in order focusing on image information. The main aim of data extraction with regards to semi-structured data is to retrieve beneficial information from the web. The data from web also known as deep web is retrievable but it requires request through form submission because it cannot be performed by any search engines. As the HTML documents start to grow larger, it has been found that the process of data extraction has been plagued with lengthy processing time. In this research work, we propose an improved model namely wrapper extraction of image using document object model (DOM) and JavaScript object notation data (JSON) (WEIDJ) in response to the promising results of mining in a higher volume of image from a various type of format. To observe the efficiency of WEIDJ, we compare the performance of data extraction by different level of page extraction with VIBS, MDR, DEPTA and VIDE. It has yielded the best results in Precision with 100, Recall with 97.93103 and F-measure with 98.9547.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ily Amalina Ahmad Sabri
Faculty of Ocean Engineering Technology and Informatics
Universiti Malaysia Terengganu
Kuala Nerus, Terengganu, Malaysia
Email: ilylina@umt.edu.my

1. INTRODUCTION

The numbers of devices and gadgets connection to the Internet is on the rise. This increase in internet's connection makes the web as the largest source of information worldwide. With the large amount of data residing in the web, and complemented by advanced technologies in database processing, it is therefore a seamless effort to gather, collect and process the data. As the consequence of the exponential data growth, it is most important for users to adopt advanced data analytics technologies for an efficient storage, retrieval and analysis of the data. The main aim is to usefully utilize this data, to learn about patterns and trends that can be used to make a positive impact on our lifestyle. However, the data itself doesn't produce these objectives, but rather it's solutions that arise from analyzing it and finding the answers we need. This accumulation of data in terms of volume, technology and techniques are often being discussed in relation to mine data from world wide web. Figure 1 shows the number of scholarly works over time by their publication type such as book, dissertation, journal article, report, conference proceeding and so forth via lens.org. From this graph, it can be easily seen the trend in this research field.

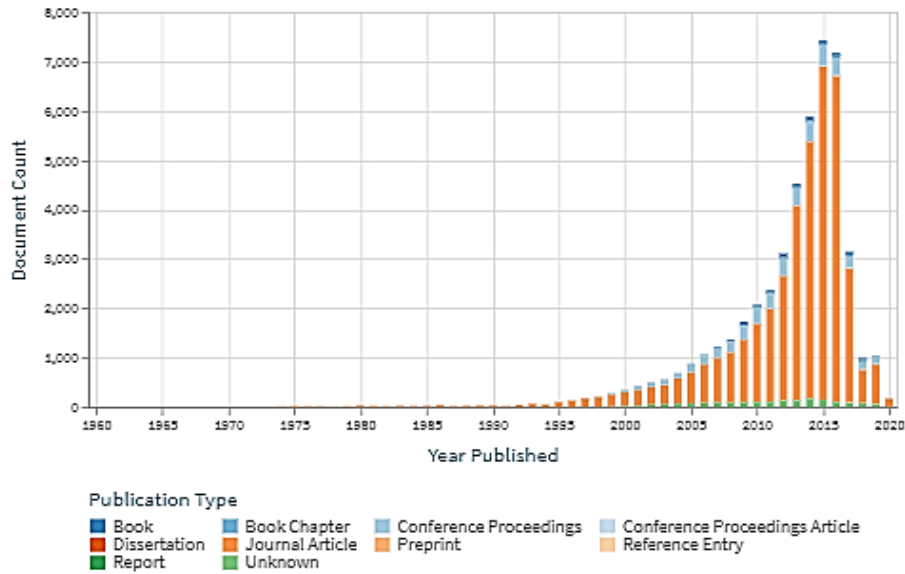


Figure 1. Number of scholarly works from 1970 till 2020

Mining data uncover new facts and relationships using useful patterns and techniques in order to give a solution for handling big data [1]. Data mining techniques are implemented to find useful patterns in large database such as MySQL and Oracle. It is the process that tries to discover patterns or techniques that can be applied in large dataset [2]. The main goal of data mining is to extract information from large dataset. Enough data and supported tools are important and need to complement each other's in dealing with large data set. It may be leveraging onto the implementation of the big data that provides great opportunities for various of fields such as e-commerce, industrial controls and smart medicals [3]. However, the characteristics of large volumes, large varieties, large velocities and large veracities of information need to be considered in order to handle the challenging for data mining [4]. Finally, the extracted information will be transformed into a structured way for further use. Web mining is the application of data mining techniques to discover potential information automatically from the web.

In relation to Figure 2, web mining is divided into three categories; web content mining, web structure mining and web usage mining. Web content mining is all about discovering useful content on the world wide web (WWW) by using data integration and data extraction. Web structure mining places websites and web pages that contain in a network of connected websites by using hyperlinks. A hyperlink is an element in HTML documents that links an object such as text, image, and video. to another HTML document altogether. In other hand, web usage mining focuses on browsing behavior either using pattern track or personalize usages track. This paper focuses on web multimedia mining focuses on images.

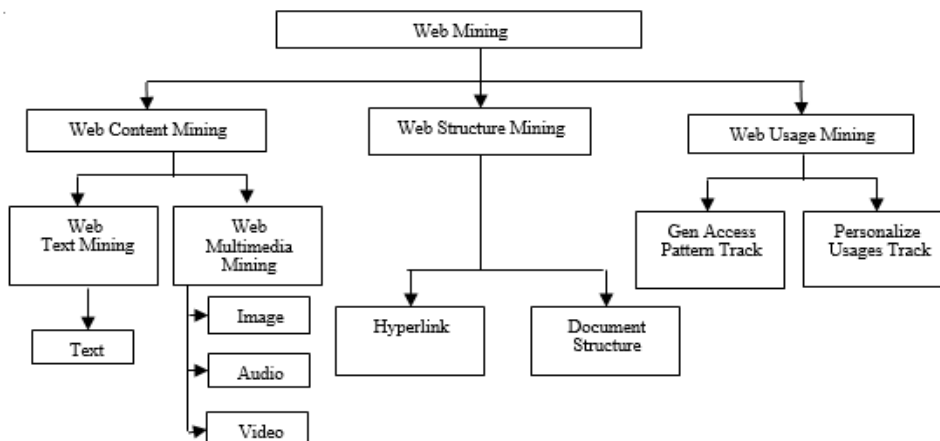


Figure 2. Web mining categories

Mining data or extracting data from web pages is a major feature for human to lead and get huge benefits. Websites are designed for various people and they are known as semi-structured data. The structure of each web page is different for each page. Thus, it is not easy to capture all the data in different structure [5] and many studies discuss about extracting data from websites and various methods have been developed. The large volume of images and their information requires new solutions to manage and analyze them. We have proposed wrapper extraction of image using document object model (DOM) and JavaScript object notation data (JSON) (WEIDJ) in order to address this concern. The main motivation for this research is image's extraction, mining of image's details and its storage in single multimedia database. In ideal scenario, if image need to be saved, it should be manually extracted. Extraction and saving of required files or images is important since these documents can be beneficial for further purpose. However, problems in loading times exist when the size of the images to be extracted are too big. Therefore, another solution must be developed to automatically extract the images to reduce the consumed time. A data extraction engine should be able to extract all the required from web page. The initial step in extracting data from a specific web page is to define the uniform resource locator (URL) of the web page, where the data is located.

2. DATA EXTRACTION

Data extraction is where data is been analysed and crawled through from data sources such as web or databases. It depends on specific patterns of user requirements. The goal of data extraction is to retrieve relevant information. It organizes data into usable and valuable resource so that we can use for further purposes. The extraction process may involve different data types. Prior to extraction processes, data needs to be well organized. If the data is in a structured format, it will be more applicable. There are three types of data; structured data, semi-structured data and unstructured data. There are many ways to deal with all these types of data. This research focused on the extraction of semi-structured data. There are three basic steps in data extraction process as shown in Figure 3.

The advantages of data extraction from semi-structured data is that it can be applied in various fields such as in education [6], advertisements [7], housing managements [8]. In former works, the discussed data extractions have been modelled using a single model or combination of several models for an optimum assessment [9] While web has developed into a large source of information, there are different data types of information that will be discussed in next section. This paper aims to advocate the potential of two-phase query paradigm for web mining. Our extensive experiments indicate by following criteria:

- Having an explicit target for the extraction process.
- Using a large set of information from several website which also has different structure.

This approach turns out to be highly effective in practice. In our view, these results hint that a fully automatic solution for querying the structured images and related information, non-hidden images refer to Figure 4 including aspects of the structure for each web and the redundancy of the images Figure 5.

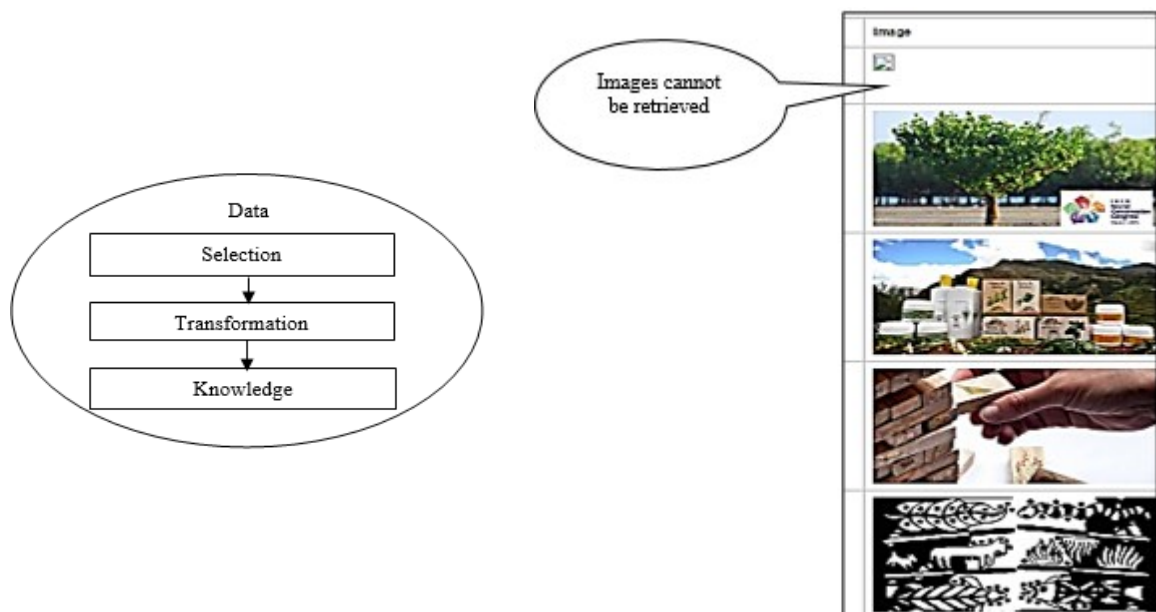


Figure 3. Data extraction process

Figure 4. Images cannot be retrieved








25	http://d1daee5goevto1.cloudfront.net/_skins/pandaorg3/img/logo_vida_silvestre-min.jpg	
26	http://d1daee5goevto1.cloudfront.net/_skins/pandaorg3/img/logo_vida_silvestre-min.jpg	
27	http://d2qz5f444n933g.cloudfront.net/img/carrera_la_hora_del_planeta_fundacion_vida_silvestre_28_03_28162.jpg	
28	http://d2qz5f444n933g.cloudfront.net/img/logo_vida_silvestre_22542.jpg	
29	http://d2qz5f444n933g.cloudfront.net/img/payunia_wcs_1_28142.png	
30	http://d2qz5f444n933g.cloudfront.net/img/logo_vida_silvestre_22542.jpg	
31	http://d2qz5f444n933g.cloudfront.net/img/logo_vida_silvestre_22542.jpg	

Figure 5. Redundancy of images

3 WEB DATA EXTRACTION (WDE)

The importance of web data extraction (WDE) depends on the fact that large amounts of data are continuously been generated, shared and utilized in every second. WDE techniques are implemented to reduce labor intensive tasks and play important roles in raising the accuracies in data extraction. Many factors should be considered in designing WDE including the techniques. One of the critical factors is the ability of the developed techniques in processing a large amount of data in a short time.

Web data extraction system is a software application that can extract data from web sources [10]. This application usually interacts with a web source and extracts the stored data. The extracted contents consist of elements in the HTML web pages and can be post-processed, converted to the most appropriate structured format and stored for further usage. Table 1 shows web data extraction tools that are using different techniques.

Table 1. Web data extraction tools

(Author, year)	Tools	Model
Fang, Xie [11]	STEM	Suffix Tree Based Method
Pouramini, Khaje Hassani [12]	Handle-based Wrapper	DOM Tree
Jiménez and Corchuelo [13]	TANGO	DOM
Chitra and Aysha Banu [14]	DWDE	Tag based Feature
Tripathy, Joshi [15]	VEDD	DOM Tree Breadth First Search (BFS)
Derouiche, Cautis [16]	ObjectRunner	
Liu, Pu [17]	XWRAP	DOM Tree
Chang and Kuo [18]	OLERA	
Liu, Grossman [19]	MDR	
Cai, Yu [20]	VIPS	DOM Tree Visual Cues
Crescenzi, Mecca [21]	Road Runner	-
Chang and Lui [22]	IEPAD	Pattern Discovery
Hsu and Dung [23]	SoftMealy	-
Hammer, Garcia-Molina [24]	TSIMMIS	Object Exchange Model (OEM)

4 RESEARCH METHOD

Prior to extraction processes, data needs to be well organized. If the data is in a structured format, it will be more applicable. There are three types of data; structured data, semi-structured data and unstructured data. There are many ways to deal with all these types of data. This research focused on the extraction of semi-structured data. There are three basic steps in data extraction process; selection, transformation and knowledge [25]. Web wrapper is a procedure which is implemented based on any of the specified algorithms. The goal is to seek and find data required by human users from the web sources, which includes unstructured or semi-structured data. The extracted data will be transformed into structured representation for further usage. Lately, the problems of extracting information from unknown sites, focusing on unstructured or

semi-structured data are getting much attention from the researchers. The works on WDE has lots of reviews. This section discusses about our proposed method, WEIDJ.

WEIDJ is developed to assist user in extracting semi-structured data from web page. A web page can be represented by a tree structure DOM. It converts and store a given web address of web page from a search engine into a DOM tree [26]. Recently, the extraction process is focused on image [27, 28]. When user input the uniform resource locator (URL) and the query is submitted to a search engine, the search engine will dynamically generated result page containing the result records. The results consist a link path for each element of image, image, size of image and time processing to load each image [29]. WEIDJ used alpha jet experiment (AJAX) technology to extract data from web sources. AJAX, is the abbreviation of Asynchronous JavaScript and XML, is a set of web development techniques that allows a web page to update portions of contents without having to refresh the page. AJAX represents a similar concept to the client-server development. During client-server phase, the amount of data transferred is very minimal over a terminal application by transferring only the necessary data back and forth. Similarly, with AJAX, only the necessary data is transferred back and forth between the client and the web server. This minimizes the network utilization and processing on the client. The time for extraction process has been reduced. Figure 6 shows an overview of WEIDJ using AJAX and JSON data.

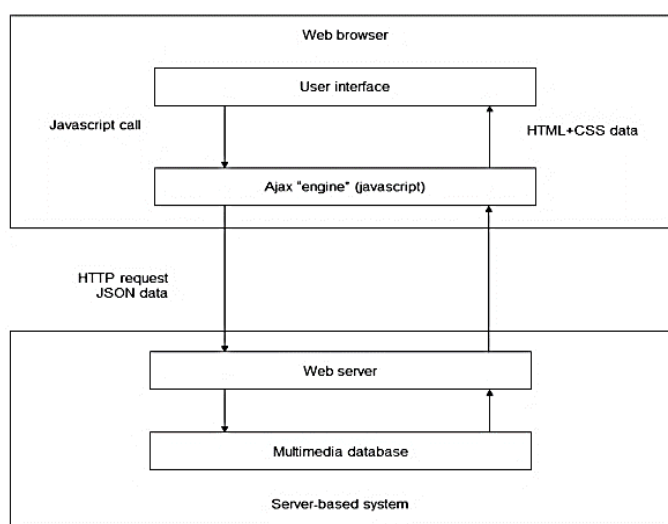


Figure 6. Overview of WEIDJ model

It can be difficult to properly create extraction rules describing required data. In this paper, we propose WEIDJ [30] model to extract images from a web page. The work described in this section uses a combination of both techniques, DOM and JSON [31]. In addition, we also do the checking of images by blocks in the HTML documents. It also focusses on arranging the extracted data in a tabular format. Lots of applications focuses on extracting information and then have it arranged accordingly [32, 33]. Every web page has their own structure includes main topic, related topics, additional information, advertisement, contact information, images, audio and video file. The problem that we want to solve is what is the best technique can be applied in order to extract images automatically [34, 35]. Mining information records in data regions plays important role in defining tags of semi-structured data. It is easy to extract data from data regions because it contains useful data. It is recognized as data area. A technique is requisite in order to mine data area. In the earlier stage, this model proposed DOM tree as based technique to mine data regions in web page.

4.1. WEIDJ algorithm

The industrial of big data completes the index function of big volumes of data especially in extracting image as the data of preference. There are many other researchers who work on data extraction from different sites in order to test the performances of extraction. In this research work, we retrieve images and their information from web sources to be analysed for further usage. A mediator tool call WEIDJ approach has been proposed. This tool aims to extract images according to uniform resource locator URL. The image's details will be mined and presented in a structured format before storing them into multimedia database. In this research, we propose a mediator tool call WEIDJ approach. This tool aims to extract images

according to uniform resource locator URL and mine image details then present images in structured format before storing them into multimedia database.

In WDE, a web-based method based on DOM is applied. DOM provides a structured way to describe documents. The HTML documents will be converted into DOM tree structure. Each element in the tree structure is known as node. The main task of data pre-processing in web data extraction includes pre-build the DOM tree of the web page. This wrapper will analyse the specific targets in the sources of Internet world, websites. First, it obtains the relative of URL from a website. Each URL may contains a few, hundreds or thousand images. Information will be extracted from from different levels of web pages such as single web, different sources of web pages and deep web. The extraction of information need to deal with page refinement to clean and extract useful information such as images, path of images, size of images and so forth in the rule of extraction. This wrapper is proposed to extract images from web. In this way, the processing of images will be converted into a form of computer processing; which is represented by the extraction of data in tabular format. This representation is important in order for providing research analysis of data extraction. Figure 7 describes the whole process of the realization of WDE.

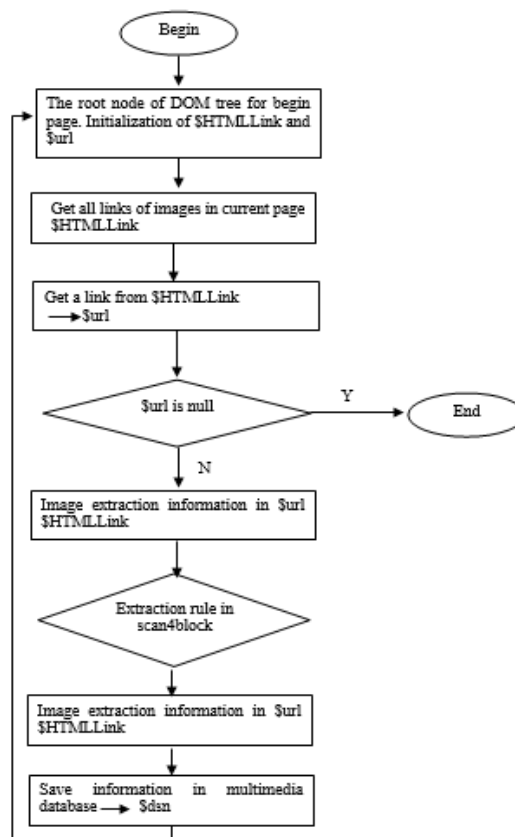


Figure 7. The process of the image's extraction

5. RESULTS AND ANALYSIS

In this research work, web data extraction experiments had been set up to compare the performance of WEIDJ with existing method. The software configuration that has been used in this experimentation can be referred in the previous work [35]. The findings of experiments tabulated in Figure 8 shows that when the amount of extracted images increase, the time of the two retrieval methods; DOM and WHDJ are increased but the time of WEIDJ on images extraction is significantly lower than other approaches. Five different websites from biodiversity field has been selected to test the performance of web data extraction as shown in Table 2. Each website has different data volume and different data size. For a web data extraction experiment, different data volume and data size are been tested by four different extraction methods.

This paper also selects the website of FangJia which is <http://sh.FangJia.com> as show in Figure 9. The reason why this website is selected as a guideline because there is a discussion of findings for image extraction that has been constructed [27]. Four typical data extraction algorithm VIBS, MDR, DEPTA and

VIDE were selected as comparing target. The experiments were conducted on the prototype system of the above algorithm. There are two types of performance measurement that have been conducted during this experiment. The first measurement is execution time analysis and second is precision, recall and F-measure.

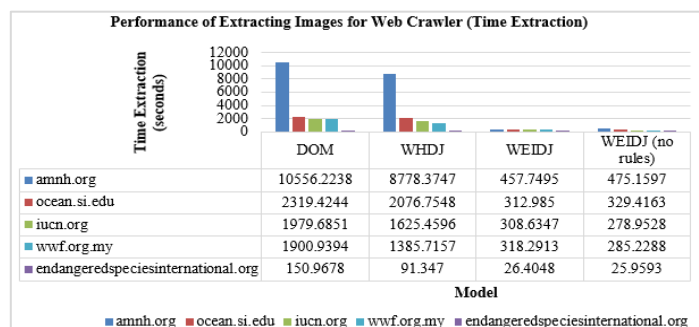


Figure 8. Performance of image extraction for deep web

Table 2. Characteristics of instant dataset

URL	Uniform Resource Locator (URL)	Domain
	General Biodiversity and Endangered Species Information	
1	http://www.amnh.org/	American Museum of Natural History (AMNH) Hall of Biodiversity
2	http://ocean.si.edu/	Ocean Portal: Smithsonian Institution
3	http://www.iucn.org/	International Union for Conservation of Nature
4	http://www.endangered-species-international.org	Endangered Species International
5	http://www.wwf.org.my	World Wide Fund for Nature

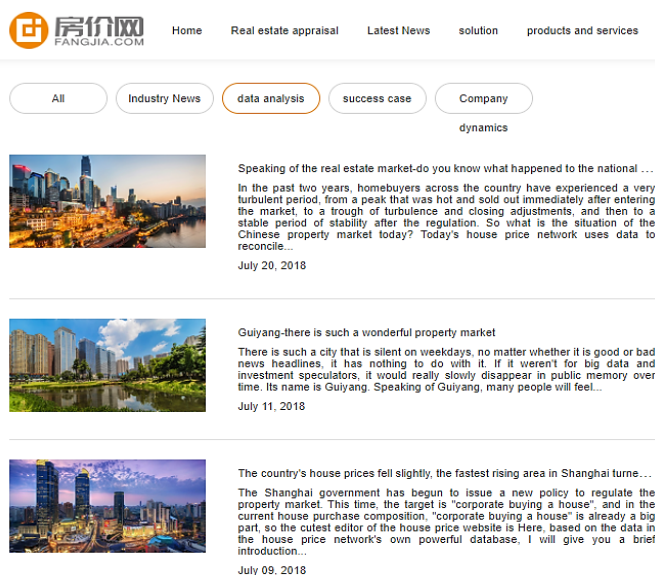


Figure 9. Structured pages from FangJia.com

5.1. Time extraction analysis

In this experimental work, 40 pages from the same website FangJia (<https://fangjia.fang.com/bj/>) have been selected randomly. Then, the extraction time will be calculated from the beginning of the extracted page to the next page. Figure 10 shows the sample output for extracting 40 pages by corresponding page. The duration of the extraction process is shown in details in Table 3. From the performance analysis, in the preliminary for 5 and 10 pages VIBS it is excellent in extracting images but when the HTML documents become larger, we found that WEIDJ clearly outperforms existing tools.

5.2. Precision, recall and F-measure

According to [27], the interference of web page noise to data extraction is important to be considered besides efficiency and accuracy of different deep web page heterogeneity. This issue motivates us

to improvise existing algorithm on noisy information. So, besides focusing on the performance of time extraction for extracting information, we also want to extract the significant information of image and remove the noisy information. Table 4 shows the result of the experimental evaluation for WEIDJ using FangJia webpage as tested URL. Figure 11 shows the comparison of the five algorithm of the experiments. Our model, WEIDJ has proven that its ability to extract data as accurate as VIBS. This accuracy in extraction is achieved because of two factors that we include in this research, which are noises filtration and the use of JSON which helps to transform the data faster.

$$Precision = \frac{Dataretrieved}{Dataretrieved+Datafalse} \tag{1}$$

$$Recall = \frac{Dataretrieved}{Totalofimage} \tag{2}$$

$$Fmeasure = 2 \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

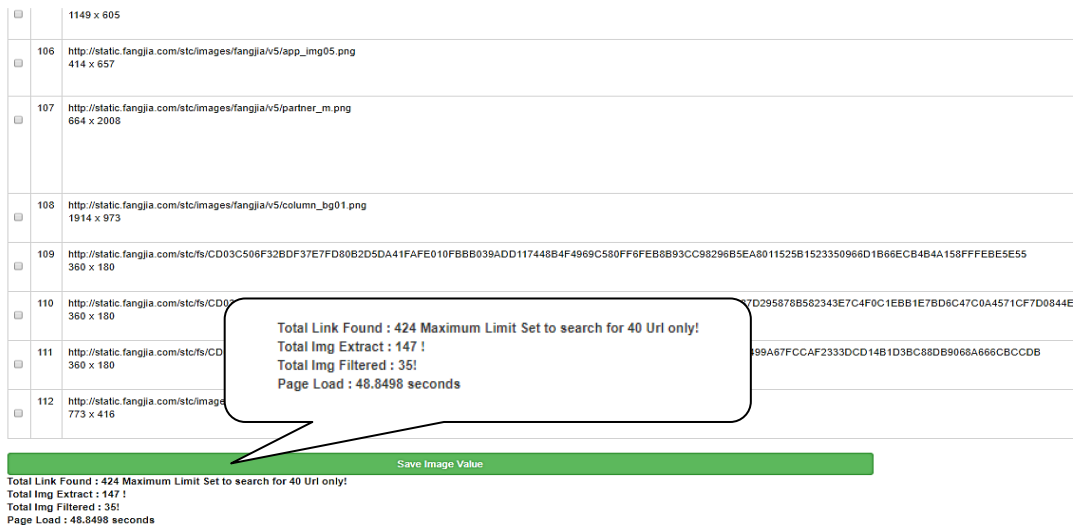


Figure 10. Extracting 40 pages

Table 3. The performance of data extraction

Method	Time Extraction							
	5 pages	10 pages	15 pages	20 pages	25 pages	30 pages	35 pages	40 pages
WEIDJ	12.6972	18.639	22.18	29.1468	29.5079	35.2651	37.977	48.8498
VIBS	7.25	12.7	23.74	30	35.01	44.37	49.76	62.69
MDR	19.29	40.11	61.18	83.78	101.07	122.63	148.33	164.16
DEPTA	20.98	43.79	66.66	90.63	114.04	135.72	153.55	180.71
VIDE	53.13	94.37	144.33	195.23	246.29	291.08	341.18	389.52

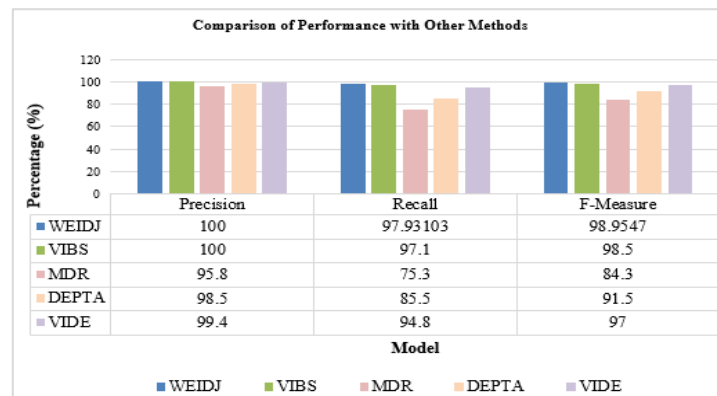


Figure 11. Comparison performance existing method

Table 4. Result of the experimental evaluation for WEIDJ

Total img	Data retrieved	Data (False)	Unknown Data	Precision	Recall	F1
145	142	0	3	100	97.93103	98.9547

6. CONCLUSION

All the World Wide Web has become a vast information store that is growing at a rapid rate, either in number of sites or in volume of useful information. WDE is time consuming when the html documents becomes larger. Single DOM did not perform very well in extracting multimedia data such as image if the volume of data become increased. However, when another technique JavaScript object notation is implemented in enhanced model namely as wrapper hybrid DOM and JSON (WHDJ), the time execution in extracting image and its information has been reduced to 50% greater than DOM technique. Even the time execution has improved but the limitation of this model is the redundancy of similar filename in images extraction. Complementary to this, we intend to combine both approaches and apply visual segmentation to get the best performance and extract the constructive images. This wrapper has been developed based on proposed model, WEIDJ. The findings result of time execution of WEIDJ is greater (90%) than existing tools should be interpreted because of the page level of extractions which is deep web, used in the analysis of experimentation for the execution time. In this study, the benchmark of dataset (FangJia) and biodiversity websites were heterogeneous with respect to image, path of images, size of images and execution time. Beside the execution time is focused as main guideline, the experimentation of images extraction would have improved the validity of significant information by removing noisy information of images. In future studies, it is recommended that the selection of dataset involves variety of fields which includes social networks or other platform. This is because the structure of website have been developed in different technologies.

ACKNOWLEDGEMENTS

I sincerely thank all those who helped me in completing this task especially Biasiswa Universiti Malaysia Terengganu (BUMT).

REFERENCES

- [1] R. Suresh, *et al.*, "Data mining and text mining - A Survey," *2017 International Conference on Computation of Power, Energy Information and Communication*. 2017, pp. 412-419, March 2017.
- [2] A. Apaolaza, M. Vigo, "Assisted Pattern Mining for Discovering Interactive Behaviours on The Web," *International Journal of Human-Computer Studies*, vol. 130, pp.196-208, October 2019.
- [3] Q. Zhang, *et al.*, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146-157, July 2018.
- [4] C. S. Saravana Kumar and R. Santhosh, "Effective information retrieval and feature minimization technique for semantic web data," *Computers & Electrical Engineering*, vol.81, January 2020.
- [5] V. Kayser and E. Shala, "Scenario development using web mining for outlining technology futures," *Technological Forecasting and Social Change*, vol. 156, July 2020.
- [6] K. Williams, *et al.*, "Scholarly big data information extraction and integration in the citeseer χ digital library," *Data Engineering Workshops (ICDEW)*, March 2014.
- [7] M. S. Pera, *et al.*, "Web-based closed-domain data extraction on online advertisements," *Information Systems*, vol. 38, no. 2, pp. 183-197, April 2013.
- [8] Dewaelheyns, V., I. Loris, and T. Steenberghen, "Web data extraction systems versus research collaboration in sustainable planning for housing: smart government takes it all," *REAL CORP 2016 Proceeding*, 2016.
- [9] N. V. Kamanwar, S. Ka, "Web data extraction techniques: A Review," *Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, March 2016.
- [10] Laender, A. H., *et al.*, "A brief survey of web data extraction tools. ACM sigmod record," *DBLP*, vol.31, no. 2, pp.84-93, June 2002.
- [11] Fang, Y. X., *et al.*, "STEM: A suffix tree-based method for web data records extraction," *Knowledge and Information Systems*, vol. 55, no. 10, pp.1-27, May 2017.
- [12] A. Pouramini, *et al.*, "Data extraction using content-based handles," *Journal of AI and Data Mining*, January 2018.
- [13] P. Jiménez and R. Corchuelo, "On learning web information extraction rules with TANGO," *Information Systems*, vol. 62, pp. 74-103, December 2016.
- [14] M. Chitra, B. Aysha Banu, "Deep web data extraction based on url and domain classification," *ISAACA JOURNAL*, pp. 1-4, July 2015.
- [15] A. K. Tripathy, *et al.*, "Vedd-A visual wrapper for extraction of data using DOM tree," *Communication, Information & Computing Technology (ICCICT)*, October 2012.
- [16] N. Derouiche, *et al.*, "Automatic extraction of structured web data with domain knowledge," *IEEE 28th International Conference on Data Engineering*, April 2012.
- [17] L. Liu, *et al.*, "XWRAP: An XML-enabled wrapper construction system for web information sources," *Data Engineering, 2000. Proceedings. 16th International Conference on*. 2000, February 2000.

- [18] C. H. Chang and S. C. Kuo, "OLERA: semisupervised web-data extraction with visual support," *IEEE Intelligent Systems*, vol.19, no.6, pp. 56-64, December 2004.
- [19] B. Liu, *et al.*, "Mining data records in web pages," *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003.
- [20] D. Cai, *et al.*, "VIPS: A Vision-Based Page Segmentation Algorithm," *Microsoft technical report*, 2003.
- [21] V. Crescenzi, *et al.*, "Roadrunner: towards automatic data extraction from large websites," *VLDB*, September 2001.
- [22] Chang, C.-H. and S.-C. Lui, "TEPAD: information extraction based on pattern discovery," *Proceedings of the 10th International Conference on World Wide Web*, January 2001.
- [23] C. N. Hsu, M. T. Dung, "Generating finite-state transducers for semi-structured data extraction from the web," *Information Systems*, vol. 8, no. 8, pp. 521-538, December 1998.
- [24] J. Hammer, *et al.*, "Extracting semistructured information from the web," September 2002.
- [25] A. Gupta, Anand Shankar S., and C. Manjunath, "A comparative study on data extraction and its processes," *International Journal of Applied Engineering Research*, vol. 12, no. 18, pp. 7194-7201, 2017.
- [26] Ily Amalina A. S., and M. Man, "Multiple types of semi-structured data extraction using wrapper for extraction of image using DOM (WEID)," *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, 2018.
- [27] J. Liu, *et al.*, "Deep web data extraction based on visual information processing," *Journal of Ambient Intelligence and Humanized Computing*, October 2017.
- [28] A. Bhardwaj and V. Mangat, "An improvised algorithm for relevant content extraction from web pages," *Journal of Emerging Technologies in Web Intelligence*, vol. 6, no.2, May 2014.
- [29] Ily Amalina A. S., and M. Man, "The proposed algorithm for semi-structured data integration: case study of setiu wetland data set," *Journal of Telecommunication Electronic and Computer Engineering*, vol. 9, no. 3-3, pp. 79-84, 2017.
- [30] Ily Amalina A. S., and M. Man, "WEIDJ: An improvised algorithm for image extraction from web pages," *The 8th International Conference on Information Technology*, May 2017.
- [31] Ily Amalina A. S., and M. Man, "Improving performance of DOM in semi-structured data extraction using WEIDJ model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 3, pp. 752-763, 2018.
- [32] J. Wang and F. H. Lochovsky, "Data extraction and label assignment for web databases," *Proceedings of the 12th International Conference on World Wide Web*, January 2003.
- [33] Y. Zhai and B. Liu, "Web data extraction based on partial tree alignment," *Proceedings of the 14th International Conference on World Wide Web*, January 2005.
- [34] Ily Amalina A. S., and M. Man, "The proposed algorithm for semi-structured data integration: case study of setiu wetland data set," *Journal of Telecommunication Electronic and Computer Engineering*, vol. 9, no.3-3, pp. 79-84, 2017.
- [35] Ily Amalina A. S., M. Man, *et al.*, "Web data extraction approach for deep web using WEIDJ," *Procedia Computer Science*, vol. 163, pp. 417-426, 2019.

BIOGRAPHIES OF AUTHORS



Ily Amalina Ahmad Sabri was born on 20th March 1985 in Kuala Terengganu, Terengganu. Her primary education started at Sekolah Kebangsaan Rusila (1992-1997) and she accomplished her high school at Kolej Sains Pendidikan Islam Negeri Terengganu (KOSPINT) in 2003. She received her Diploma Information Technology from Polytechnic of Sultan Mizan Zainal Abidin (PSMZA). After that, she enrolled to Universiti Malaysia Terengganu to further her degree studies in Software Engineering, which was obtained in 2009. She continued her master degree in Master of Science (Computer Science) in the same university and graduated in 2014. During masters's degree, her research was in Decision Support System, focusing on Fuzzy AHP in decision making for tourism destination. Now, she has completed her Doctor of Philosophy (Computer Science), also in Universiti Malaysia Terengganu. Her current area of interest is Data Mining focusing on Web Data Extraction.



Mustafa Man is an Associate Professor in School of Informatics and Applied Mathematics and also as a Deputy Director at Research Management Innovation Centre (RMIC), UMT. He started his PhD studies in July 2009 and finished his studies in Computer Science from UTM in 2012. He has received Computer Science Diploma, Computer Science Degree, Master Degree from UPM. In 2012, he has been awarded a "MIec MOS Prestigious Awards" for his PhD by MIMOS Berhad. His research is focused on the development of multiple types of databases integration model and also in Augmented Reality (AR), android based, and IT related into across domain platform.