

Exploiting Web Scraping for Education News Analysis Using Depth-First Search Algorithm

Endah Ratna Arumi¹, Pristi Sukmasetya²

^{1,2}Department of Informatics Engineering, Universitas Muhammadiyah Magelang, Indonesia

Article Info

Article history:

Received April 14, 2020

Revised June 24, 2020

Accepted June 27, 2020

Published July 15, 2020

Keywords:

Algorithm
Depth-first search
Education news
Online news
Web scraping

ABSTRACT

Online news is one source of data that is always up to date and provides information or factual data. The search engine is one of the features for users to be able to enter keywords based on the expected category quickly. The development of education in Indonesia makes it essential to discuss, in this study using unstructured data in online news with the keyword Education included as a parameter, and adding search methods in the field of Artificial Intelligence so that the data becomes more accurate. Data that used here was from online news, namely CNN Indonesia, Detikcom, and Liputan6. Using Python Programming with depth-first search method (DFS), when compared with the results data for relevant news. Web erosion using DFS will be very helpful in searching because this method can check the date data was sent and then track the destination URL. Of the three online media sites, Detikcom produces the highest monthly data yielding an average of 885 news about education. At the same time, Liputan6 has the least amount of data on average, 28 news per month, but the data obtained are very relevant compared to Detikcom and CNN Indonesia.

Corresponding Author:

Endah Ratna Arumi,
Department of Informatics Engineering,
Universitas Muhammadiyah Magelang,
Jl. Mayjend Bambang Soegeng Km.5 Mertoyudan, Magelang, Indonesia
Email: arumi@ummgl.ac.id

1. INTRODUCTION

The development of information technology, especially the web site, making social media and online news will continue to use in the era of big data, online news is one of the sources of the data used by the people of Indonesia in gathering information/news for the current. The wide range can be accessed using digital devices / mobile devices[1]–[4]. Online news provides information in the form of data or facts that can be obtained easily by providing a search engine feature by entering a keyword in accordance with the expected category [5]–[7]. Nowadays, information and news that spread to much online news, it is likely many lies will be accepted by the public [8], users should be more selective, and watch any news to know the truth of the news. News category of online news is very diverse, ranging from economic, political, cultural, sports, natural disasters, and the news is no less critical on education.

News of the use of information technology, especially the Internet in the field of education, by making use of online systems in developing students learning processes [9]. Educational development should be monitored to create more advanced countries and able to compete in the world. One effort to improve human resources in Indonesia in terms of education, the quality of teaching and learning process conducted by educators[10]. Therefore, the necessary information regarding the current state of education, educational

constraints, and related infrastructure in support of education. News of Education in several online news, but related news of education has not grouped in a data bank in the form of a web, so users should use the search feature to filter the news.

Analysis of text mining is a quick way to read and summarize textual information to collect sample data on the website or text documents, finding information that can be accessed via a web browser [11]–[13]. Extracting information from online news, web sites or web mining is a solution that can be used to collect information related to education issues [2]. Collecting data using surveys, interviews and questionnaires, often constrained by funding and validity of the data [14]. Web mining or scraping of the Web as an application of data mining techniques to the website to categorize the relevant information, the classification step, similar cases, and to detect web spam from several sources available on the Web, in the form of documents and web services [15]–[18]. The process in the same web mining with data mining, which only differentiates ways of collecting data. In data mining, processing data retrieved from the data that has been received and stored in the database, whereas, in web mining, data collection using a web search can automatically download a web page [19]. Automated data collection, extract information, and traversing a hyperlink on a web page called Web crawling or Robot [2] [7] [14] [20] [21].

Research about the web scraping has been done, one of the study is to analyze the market in Indonesia using IP manipulation so that the user agent used is not hindered by the destination site [22]. Scraping Web can be used to collect data on natural disasters, in a study conducted source data retrieved from online news, then enter a keyword as a parameter, and extracted into Excel format (.csv) [5]. Web scraping declared valid if it does not steal data, and manipulate the data, even with the scraping of the Web provides reciprocal symbiosis with increased traffic from the scraping of the source link [23]. Web crawling application may also be combined with in-depth search methods to find specific news because the results are more accurate and valid [6]. Crawler robots can be used in libraries, whose job is to find book content that users often look for before the library makes a book purchase [24]. Web crawlers can be integrated with REST-based web services, to support the use of the cross-platform system, and then serves as a library for Python that provides an API to access the data contained in Google Play Store [25]. Currently, web scraping/web mining can combine with several methods in Artificial Intelligence to speed up the search data is valid and relevant to the keywords.

Data collection in this study uses web scraping, and web crawlers with Python programming language, combined with one of the methods applied in Artificial Intelligence [16], to identify and select a reliable source of data from web sites and online news so that the data obtained is accurate and in accordance with the purpose of collecting information related to education [26]. Web scraping need to pay attention to the quality of the data, the accuracy of the data, especially when dealing with the World Wide Web, where a lot of unstructured data, so a lot of data cleansing and data necessary fail-safe program scraper made. The source of data was taken from the online news, i.e. www.cnnindonesia.com, www.detik.com, www.liputan6.com. Based on data that are collected from web crawlers, there is a link from the search data and unstructured online news. In this research, a depth-first search method is used to be a searching method. Then, this study used unstructured data from online news with the keyword Education included as a parameter and added search methods in the field of Artificial Intelligence so that the data becomes more accurate. The addition of this method is expected to increase the accuracy and obtain the relevant data, and results will be stored in a NoSQL database.

2. METHOD

This study using a combination of Python Programming web scraping and depth-first search method. We can see the detail in Figure 1. In this study, the source of data collection was taken from the online news, i.e. www.cnnindonesia.com, www.detik.com, www.liputan6.com. Web scraping process used to input a search using the keywords "Education", with data taken in the form of dates, headlines, links to photos, news links, categories, so that specific categories can group them.

Figure 1 is an explanation of the first and in-depth search method. The method used in the search process starting from the root node of the date to be traced, down to the news URL node, afterwards into categories, and finally to the final node or news content page that can be determined. This process will continue until the last node is executed; it will do a reverse search or reverse to search for the next node [6]. When the process of backtracking is not found, it will be repeated to other new branches to get to the endpoint search by categories that have been determined [27].

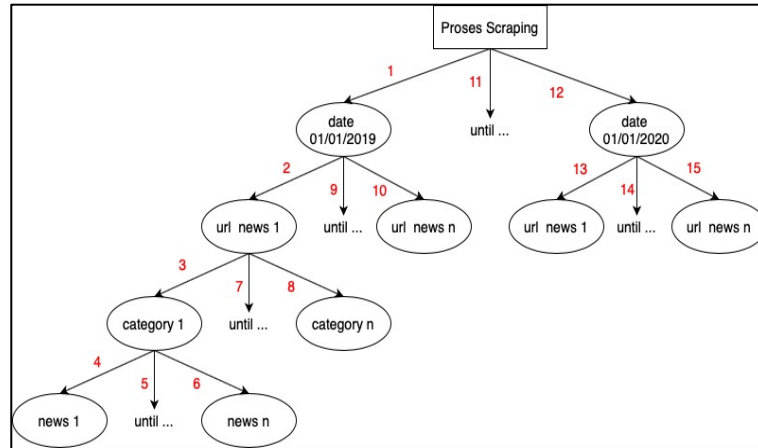


Figure 1. Web Scraping using Depth first search [6]

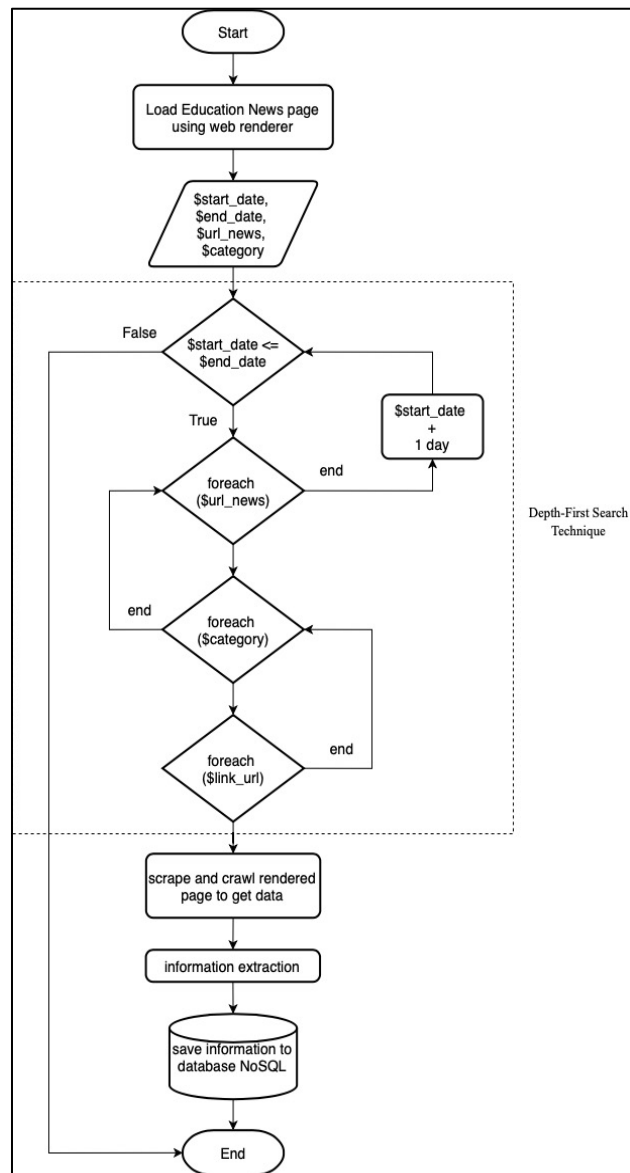


Figure 2. Flowchart of Web Scraping for Education News using Depth-First Search

Figure 2 shows the flowchart that describes the data searching technique based on keywords, looking fit of the data using a depth-first search, store data, and analyze results. This study will begin with the process of loading the URL in online news intended for the keyword "education". After that, the proceed with depth-first search starts by taking the start date, and the expiration date of the news, the URL news, and a category, and will be repeated until it is found news that matches the search. Search result URL will continue to scratch and crawl data in accordance with keywords. Web scraping and web crawling have different tasks even though they are used on web pages. In the web scraping, users can take the raw data in a web browser, perform text mining, building a data set is then stored and analyzed to obtain data in real-time.

Web crawlers can move between websites automatically for web scraping, and was assigned to make a copy of the relevant, seek information, and collect data on the web site. The process of extracting information then performed, and after scraping and crawling process data, news data will be exported to an Excel file format (.csv), and stored in a NoSQL database.

3. RESULTS AND DISCUSSION

This study implemented in Python Programming web scraping. An adding a search method by Depth First Search. The search method will be used as an alternative to obtaining relevant results in accordance with the three objectives of online news. Web scraping process is done by a number of attempts to produce valid data.

Python programming is used as a means of annihilation with a written script method Depth First Search in the search process data. The initial step user enters a keyword and date for the news; these measures will continue until the data on that date is found, then perform the sorting/extraction of data to choose the news that is relevant to the desired search. Results of this web scraping is the document/data compiled and stored using Excel format (.CSV).

After the data extraction process is complete, the process of comparing the data acquisition with three online news followed, namely Indonesia CNN, Detikcom, and Liputan6. Retrieving data with web scraping in this study had a span of one year, beginning in January 2019 - December 2019. The data were collected in accordance with relevant keywords, i.e. about education news. The results of the three-web scraping of online news will be stored and analyzed all the results obtained, to ensure that the news by keyword searchable.

Date	Headlines	Photo	News links	Category
Agustus	VIDEO: Kontroversi Wacana 'Impor' Rektor Asing	https://akcdn.detik.net.id/visual/2019/08/01/c02db4e8-e196-40e0-a751	https://www.cnindonesia.com/tv/20190801194945-409-417542-video	TV
Agustus	Agung Hercules: Dari Atlet Pra PON ke Dunia Selebritas	https://akcdn.detik.net.id/visual/2019/08/01/3e3e366d-58f4-4015-a388	https://www.cnindonesia.com/olahraga/20190801173519-178-417484	Olahraga
Agustus	Rekrut Rektor Asing, Moeldoko Sebut Jokowi Mau Ada Kompetisi	https://akcdn.detik.net.id/visual/2019/07/19/931392ad-44c9-4916-8e8b	https://www.cnindonesia.com/nasional/20190801150823-20-417435	Nasional
Agustus	Guru Besar UI: Salah Jika Menristekdikti Impor Rektor Asing	https://akcdn.detik.net.id/visual/2015/02/12/d69661da-9629-41c6-8d3d	https://www.cnindonesia.com/nasional/20190801115842-20-417338	Nasional
Agustus	DPPI Minta Menristekdikti Kaji Ulang Wacana Impor Rektor Asing	https://akcdn.detik.net.id/visual/2019/06/26/99e62e7c-ac76-4260-9b3d	https://www.cnindonesia.com/nasional/20190801075750-20-417278	Nasional
Agustus	Selidiki Kematian, Polisi Datangi Rumah Paskibra Tangsel	https://akcdn.detik.net.id/visual/2018/08/24/ca0b6319-3d8e-4930-997f	https://www.cnindonesia.com/nasional/20190802202019-12-417941	Nasional
Agustus	Mendikbud Ajukan Penambahan 180 Ribu CPNS	https://akcdn.detik.net.id/visual/2019/01/08/7a1e8a11-f95e-41c1-8fac	https://www.cnindonesia.com/nasional/20190802181441-20-417893	Nasional
Agustus	130 Perguruan Tinggi Swasta Ditutup Sepanjang 2015-2019	https://akcdn.detik.net.id/visual/2019/06/05/bc422dbd-1a32-49b3-bde4	https://www.cnindonesia.com/nasional/20190802172238-20-417874	Nasional
Agustus	Polemik Rektor Asing dan Salah Diagnosis Menteri Nasir	https://akcdn.detik.net.id/visual/2017/11/27/785821d6-2c26-4b6d-98a8	https://www.cnindonesia.com/nasional/20190802150101-20-417824	Nasional
Agustus	Menristekdikti Targetkan Impor Rektor Asing Berjalan di 2020	https://akcdn.detik.net.id/visual/2015/06/30/cb610149-7b20-4781-bf7c	https://www.cnindonesia.com/nasional/20190802114916-20-417691	Nasional
Agustus	Tim Investigasi Cek Kesehatan Siswa SMA Taruna Palembang	https://akcdn.detik.net.id/visual/2018/10/01/023d11a9-5fe1-4a6c-880a	https://www.cnindonesia.com/nasional/20190801135448-12-417394	Nasional
Agustus	Catatan Guru Besar UI atas Wacana Impor Rektor oleh Dikti	https://akcdn.detik.net.id/visual/2019/01/05/6832c7f4-93dc-4e6d-8488	https://www.cnindonesia.com/nasional/20190802172456-20-417873	Nasional
Agustus	Buntut MOS, SMA Taruna Indonesia Palembang Dibekukan 1 Tahun	https://akcdn.detik.net.id/visual/2019/07/14/d9614de3-66fa-4070-bead	https://www.cnindonesia.com/nasional/20190805195417-20-418674	Nasional
Agustus	VIDEO: Kontroversi Wacana Rektor Impor	https://akcdn.detik.net.id/visual/2019/08/06/d63e430-dfa5-4e52-97b6	https://www.cnindonesia.com/tv/20190806201708-410-419023	TV
Agustus	Kasus Kematian Paskibraka di Tangsel, Polisi Periksa Pelatih	https://akcdn.detik.net.id/visual/2014/10/28/a753352e-b3dd-4819-b29	https://www.cnindonesia.com/nasional/20190806192102-12-419002	Nasional
Agustus	Terima Sanksi, SMA Taruna Palembang Akan Coret Semimiliter	https://akcdn.detik.net.id/visual/2019/07/14/d9614de3-66fa-4070-bead	https://www.cnindonesia.com/nasional/20190806190524-20-418999	Nasional
Agustus	VIDEO: Capaska Latihan Bersama Paspampres	https://akcdn.detik.net.id/visual/2019/08/05/91f47168-6155-4ee9-892e	https://www.cnindonesia.com/tv/20190806120504-403-418815	TV
Agustus	Timbang Impor Rektor, UGM Pih Dosen Asing Buat Dongkrak PTN	https://akcdn.detik.net.id/visual/2017/05/03/7a8207c9-3988-4a00-899b	https://www.cnindonesia.com/nasional/20190807173826-20-419273	Nasional
Agustus	Peraih Nobel Toni Morrison Meninggal Dunia	https://akcdn.detik.net.id/visual/2019/08/07/5f0833af-bb80-4f6c-93e0	https://www.cnindonesia.com/hiburan/20190807104206-241-419108	Hiburan
Agustus	VIDEO: Jam Belajar Siswa Dikurangi Akibat Asap Karhulia	https://akcdn.detik.net.id/visual/2019/08/08/c3446648-8f39-4d9f-a38a	https://www.cnindonesia.com/tv/20190808143708-407-419551	TV
Agustus	Usut Paskibra Tangsel Tewas, Polisi Selidiki Proses Latihan	https://akcdn.detik.net.id/visual/2019/08/01/829e20d0-238e-4e82-a72f	https://www.cnindonesia.com/nasional/20190809134815-12-419836	Nasional
Agustus	1.124 Titik Panas di Kalbar, Jam Sekolah Akan Dimundurkan	https://akcdn.detik.net.id/visual/2019/08/07/a8ca5219-b039-41ea-b704	https://www.cnindonesia.com/nasional/20190812095132-20-420430	Nasional
Agustus	Jadi Generasi #AntiNgaret Mulai dari Diri Sendiri	https://akcdn.detik.net.id/visual/2019/08/02/d90cd4b-7714-479e-9988	https://www.cnindonesia.com/gaya-hidup/20190806173134-284-4189	Gaya Hidup
Agustus	Tempat Penitipan Anak Hangus Terbakar, 5 Bocah Tewas	https://akcdn.detik.net.id/visual/2015/01/07/27a55028-8e71-46c4-9f5e	https://www.cnindonesia.com/internasional/20190812091903-134-420	Internasional
Agustus	Siswa SMA Penemu Obat Kanker Dapat Penghargaan Ikon Pancasila	https://akcdn.detik.net.id/visual/2019/08/13/1e9585f1-5e1c-4a5e-90f3	https://www.cnindonesia.com/nasional/20190813143041-20-420884	Nasional
Agustus	VIDEO: Capaska tetap Kuat Meski Rindu Keluarga	https://akcdn.detik.net.id/visual/2019/08/13/50b7842-3432-4249-9c34	https://www.cnindonesia.com/tv/20190813123034-403-420830	TV
Agustus	Mengenal Gen Alpha, Generasi Setelah Milenial dan Gen Z	https://akcdn.detik.net.id/visual/2018/10/25/122b5a9f-0f08-4949-8fde	https://www.cnindonesia.com/gaya-hidup/20190812144950-284-4205	Gaya Hidup

Figure 3. Result of Web Scraping in CNN Indonesia news

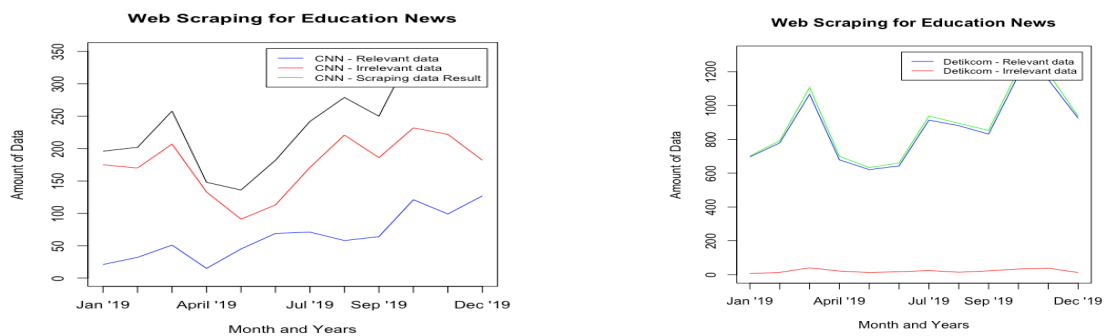
Figure 3 results of the scaling Web in Indonesia CNN web site, data is retrieved varied, relevant education news, but there is some news unrelated to education news. CNN Indonesia consists of five categories, namely, lifestyle, tv, national, international, technology, and economics. On average CNN Indonesia will send 239 news every month, with the keyword education news, all the news can be included in all categories.

of the data scraping results obtained, the next step is the comparison of data scraping the relevant and valid data less relevant even unrelated to the keywords that are searched, is evidenced in Table 1. In Table 1 are shown all data scraping of three online news, from January 2019 until December 2019.

Table 1. Comparison web scraping with relevant data

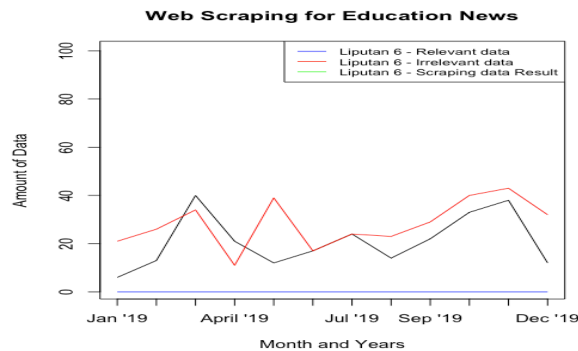
Month	Number of Scraping Data Results			Number of Relevant Data Results					
	CNN Indonesia	Detik com	Liputan 6	Relevant Data			Irrelevant Data		
				CNN Indonesia	Detik com	Liputan 6	CNN Indonesia	Detik com	Liputan 6
January	196	702	21	21	696	21	175	6	0
February	202	792	26	32	779	26	170	13	0
March	258	1107	34	51	1067	34	207	40	0
April	148	700	11	15	679	11	133	21	0
May	136	633	39	45	621	39	91	12	0
June	182	660	17	69	643	17	113	17	0
July	242	938	24	71	914	24	171	24	0
August	279	895	23	58	881	23	221	14	0
September	250	853	29	64	831	29	186	22	0
October	353	1213	40	121	1180	40	232	33	0
November	321	1192	43	99	1154	43	222	38	0
December	309	937	32	127	925	32	182	12	0
Total	2.876	10.622	339	773	10.370	339	2.103	252	0

Table 1 presents a recap of the results of a web scraping produced by CNN Indonesia, Detikcom, and Liputan6. Based on tests conducted, the results can be seen most news posting them on the website Detikcom, in the first year, there were 10,622 observation posts news about education news, with the amount of data relevant to as many as 10,370. Figure 6 displays the comparison between the data scraping with data relevant to the site Detikcom data discrepancies are not much different, which means that the results of the keyword search feature pertinent to the news category.



(a) Graph of data Scraping result CNN Indonesia

(b) Graph of data Scraping result Detikcom



(c) Graph of data Scraping result Liputan6

Figure 6. Graphs of data scraping results with three online news

Figure 6 graphs the results of the data erasure with three online news, the results of the second-largest amount of data on CNN Indonesia, the scraping of the entire record 2,876 news reporting, with relevant data as much as 773. From the result, it can be seen that every month the news about the news post more than 300 education news. Based on Figure 6, there is more than half the data be erased using the keyword search feature on the CNN website Indonesia is not optimal, many irrelevant posts with results taken. Results of data extraction Liputan6 within one year, there was news about education news that a record 339, extraction of all data relevant to the keywords searched. Although the post every month a little bit, but Liputan6 search feature on the site, displaying data corresponding to a keyword input from the user. Based on the three online news that has been studied, the results revealed a significant difference between the amount of data scraping and also the amount of relevant data.

4. CONCLUSION

Based on the implementation and testing of web scraping done with Python Programming with Depth First Search method, it can be concluded that the feature by entering a specific keyword search will affect the result of scraping. Depth First Search will significantly affect the process of searching in web scraping because this method will perform keyword and selecting post similarity date then find the destination URL. From scraping web for three online news, it can be concluded that the results obtained are very diverse, Detikcom produces result data erasure most, Liputan6 sites has a number of deletion of data is the smallest. Still, the data collected as a whole is very relevant than Detikcom and CNN Indonesia. But, CNN Indonesia has a lot of irrelevant data.

5. REFERENCES

- [1] A. Setiawan, E. U. Artha, E. R. Arumi, Sunarni, A. Primadewi, and S. Nugroho, "Task Analysis of Facebook users on Frequently used Menus," *J. Phys. Conf. Ser.*, vol. 1179, no. 1, 2019.
- [2] R. Hanifah and I. S. Nurhasanah, "Implementasi Web Crawling Untuk Mengumpulkan Web Crawling Implementation for Collecting," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, pp. 531–536, 2018.
- [3] I. Kim and G. Pant, "Predicting web site audience demographics using content and design cues," *Inf. Manag.*, vol. 56, no. 5, pp. 718–730, 2019.
- [4] P. Liu, X. Xia, and A. Li, "Tweeting the financial market: Media effect in the era of Big Data," *Pacific Basin Financ. J.*, vol. 51, no. May 2017, pp. 267–290, 2018.
- [5] I. P. Sonya, "Analisis Web Scraping untuk Data Bencana Alam dengan Menggunakan Teknik Breadth-First Search Terhadap 3 Media Online," *J. Ilm. Inform. Komput. Univ. Gunadarma*, vol. 21, no. 3, pp. 69–77, 2016.
- [6] A. Miqdad *et al.*, "Penerapan focused crawling pada situs berita online," 2016.
- [7] M. Kumar, A. Bindal, R. Gautam, and R. Bhatia, "Keyword query based focused Web crawler," *Procedia Comput. Sci.*, vol. 125, pp. 584–590, 2018.
- [8] F. A. Suharno and L. Listiyoko, "Aplikasi Berbasis Web dengan Metode Crawling sebagai Cara Pengumpulan Data untuk Mengambil Keputusan," in *Seminar Nasional Rekayasa Teknologi Informasi*, 2018, no. November, pp. 105–109.
- [9] E. R. Arumi, Sunarni, and P. Nuraini, "PENINGKATAN MINAT KONSELING DAN SELF DISCLOSURE SISWA MELALUI APLIKASI E-KONSELING DI SMP MUHAMMADIYAH PUJOTOMO," in *Seminar Nasional Hasil Penelitian dan Pengabdian Pada Masyarakat IV Tahun 2019*, 2019, pp. 574–579.
- [10] E. R. Arumi and U. Yudatama, "Pemanfaatan Curriculum Vitae dan Sasaran Kinerja Pegawai untuk Penilaian Kinerja Dosen Menggunakan AHP," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 3, pp. 170–176, Nov. 2017.
- [11] J. S. Jones-Diette, R. S. Dean, M. Cobb, and M. L. Brennan, "Validation of text-mining and content analysis techniques using data collected from veterinary practice management software systems in the UK," *Prev. Vet. Med.*, vol. 167, no. July 2018, pp. 61–67, 2019.
- [12] A. Saverimoutou, B. Mathieu, and S. Vaton, "A 6-month analysis of factors impacting web browsing quality for QoE prediction," *Comput. Networks*, vol. 164, p. 106905, 2019.
- [13] X. Xie, Y. Fu, H. Jin, Y. Zhao, and W. Cao, "A novel text mining approach for scholar information extraction from web content in Chinese," *Futur. Gener. Comput. Syst.*, no. xxxx, 2019.
- [14] D. Peng, T. Li, Y. Wang, and C. L. Philip Chen, "Research on information collection method of shipping job hunting based on web crawler," *8th Int. Conf. Inf. Sci. Technol. ICIST 2018*, pp. 57–62, 2018.
- [15] K. Sellamy *et al.*, "Web mining techniques and applications: Literature review and a proposal approach to improve performance of employment for young graduate in Morocco," *2018 Int. Conf. Intell. Syst. Comput. Vision, ISCV 2018*, vol. 2018-May, pp. 1–5, 2018.
- [16] S. H. Hong, S. K. Lee, and J. H. Yu, "Automated management of green building material information using web crawling and ontology," *Autom. Constr.*, vol. 102, no. March, pp. 230–244, 2019.
- [17] F. Asdaghi and A. Soleimani, "An effective feature selection method for web spam detection," *Knowledge-Based Syst.*, vol. 166, pp. 198–206, 2019.
- [18] T. Okuhara, H. Ishikawa, M. Okada, M. Kato, and T. Kiuchi, "Contents of Japanese pro- and anti-HPV vaccination websites: A text mining analysis," *Patient Educ. Couns.*, vol. 101, no. 3, pp. 406–413, 2018.
- [19] G. Deepak and J. S. Priyadarshini, "Personalized and Enhanced Hybridized Semantic Algorithm for web image retrieval incorporating ontology classification, strategic query expansion, and content-based analysis," *Comput.*

- Electr. Eng.*, vol. 72, pp. 14–25, 2018.
- [20] N. Hosseini, F. Fakhari, B. Kiani, and S. Eslami, “Enhancing the security of patients’ portals and websites by detecting malicious web crawlers using machine learning techniques,” *Int. J. Med. Inform.*, vol. 132, no. March, 2019.
- [21] J. Cheng, X. Zhao, J. Liu, and Y. Zhang, “Automated Test Generation Based on Colored Petri Net and Improved Depth First Search for Train Control System,” *2019 Chinese Control Conf.*, pp. 6761–6765, 2019.
- [22] E. H. Fernando, H. Sagala, A. E. Budiman, I. N. Husada, and H. Toba, “Ekstraksi dan Analisis Produk di Marketplace Secara Otomatis dengan Memanfaatkan Teknologi Web Crawling,” vol. 5, pp. 350–359, 2019.
- [23] A. Josi, L. A. Abdillah, and Suryayusra, “Penerapan teknik web scraping pada mesin pencari artikel ilmiah,” 2014.
- [24] N. Nafi’iyah and E. Sulistiono, “Pemanfaatan robot crawler pada pembuatan toko buku online,” *JOUTICA-PRESS*, pp. 12–16, 2016.
- [25] L. B. Ilmawan, “MEMBANGUN WEB CRAWLER BERBASIS WEB SERVICE UNTUK DATA CRAWLING PADA WEBSITE GOOGLE PLAY STORE,” *Ilk. J. Ilm.*, vol. 10, pp. 215–224, 2018.
- [26] R. Gunawan, A. Rahmatulloh, I. Darmawan, and F. Firdaus, “Comparison of Web Scraping Techniques : Regular Expression, HTML DOM and Xpath,” in *International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018) Comparison*, 2019, vol. 2, no. IcoIESE 2018, pp. 283–287.
- [27] A. S. Hidayatullah and C. Setianingsih, “REALIZATION OF DEPTH FIRST SEARCH ALGORITHM ON LINE MAZE SOLVER ROBOT,” in *The 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC) REALIZATION*, 2017, pp. 247–251.