

AUTOMATIC WEB NEWS CONTENT EXTRACTION

Gusti Lanang Putra Eka Prismana*

Universitas Negeri Surabaya e-mail: lanangprismana@unesa.ac.id *Correspondence: lanangprismana@unesa.ac.id

Submitted: 27 January 2022, Revised: 06 February 2022, Accepted: 18 February 2022 Abstract. The extraction of the main content of web pages is widely used in search engines, but a lot of irrelevant information, such as advertisements, navigation, and junk information, is included in web pages. Such irrelevant information reduces the efficiency of web content processing in content-based applications. This study aimed to extract web pages using DOM Tree in the rationality of segmentation results and efficiency based on the information entropy of nodes from the DOM Tree. The first step of this research was to classify web page tags and only processed tags that affected the structure of the page. The second step was to consider the content features and structural features of the DOM Tree node comprehensively. The next was to perform node fusion to obtain segmentation results. Segmentation testing was carried out with several web pages with different structures so that it showed that the proposed method accurately and quickly segmented and removed noise from web page content. After the DOM Tree was formed, the DOM Tree would be matched with the database to eliminate information noise using the Firefly Optimization algorithm. Then, testing and evaluating the Firefly Optimization method in effectiveness aspect were done to detect and eliminate web page noise and produce clear documents.

Keywords: DOM tree; web; news; extraction; firefly.

Online news is one of the big data sources. Information in the form of news articles is published every minute (Allen, Howland, Mobius, Rothschild, & Watts, 2020). There is a lot more information than the person doing the analysis, and this is a potential problem where a lot of data can be ignored (Newman & Cain, 2014). Search often used to obtain engines are information. Search engines use web spiders to surf the web and retrieve links that may contain the information sought, and present the information in the form of a collection of hyperlinks. Search engines are capable of retrieving information from the web but not from the unseen or hidden web, so this makes data extraction a very impractical task. The challenges faced by extractors include heterogeneous formats, changes in the structure of web pages, the introduction of more and more advanced technologies to improve UX, and others.

Extracting information from multiple sources has many problems such as finding useful information, extracting knowledge from large data sets, and studying individual users. Various methods and techniques have been developed (Abburu & Golla, 2015). Because the amount of information obtained on the web increases radically, the amount of redundant web content also grows at the same time. Therefore, updating the incoming data and retrieving useful information without duplicating data from the web, the web mining research community pays attention to an existing activity related to the information from the web quickly and efficiently (Dey & Jain, 2020).

The articles published on a website are mostly in the form of unstructured information because they usually contain main information or main content, advertisements, navigation, and other additional information. The amount of that information resulted in the difficulties of getting the main core information and finding relevant values and knowledge in the form of structured information, such as the form of a database. The mechanism for extracting a collection of texts to obtain facts in the form of events, entities, and relationships in the form of structured information as input to a database or ontology is called information extraction (Kara et al., 2012).

This study aimed to extract DOM Treebased web pages in the rationality of segmentation and efficiency results. This study used a method based on the information entropy of nodes from the DOM Tree. The first step carried out in this research was classifying web page tags and only processing tags that affected the structure of the page. The second step was considering the content features and structural features of the DOM Tree node comprehensively, calculating the information entropy of the nodes and the maximum text density of subnodes, and determining whether a node was a block page or independent. The third step was performing node fusion to obtain segmentation results. After getting the segmentation results, the web page noise will be removed to match the DOM Tree built with the database (Velloso & Dorneles, 2013). After the DOM Tree was formed, then the DOM Tree was matched with the database to eliminate information noise by

using the Firefly Optimization algorithm. Furthermore, testing and evaluating the Firefly Optimization method in effectiveness is done to detect and eliminate web page noise and produce clear documents (Yu & Jin, 2017). This research is expected to get a better approach for extracting data from semistructured documents, both based on structure and data using several optimization methods, techniques, and algorithms as well as finding a method for removing web page noise that often arises from web content extraction.

METHODS

A. System Architecture

The system architecture of the information extraction is an adaptation of the general architecture of the information extraction system and the general architecture of the system. The input of the information extraction system is in the form of unstructured natural language text from the source text on the web page (HTML text). The information extraction process according to the OBIE concept will involve an ontology as an extraction guide and produce output in the form of extracted information which is represented in the form of XML and annotated text. The ontology-guided extraction process will extract things such as classes, properties, and instances (Wimalasuriya & Dou, 2010).

The information extraction system architecture in this study was divided into three phases, namely the training phase, the development phase, and the evaluation phase. In the training phase, the system identifies patterns and lists of dictionaries (called semantic lexicon), which were learned using а bootstrapping approach. Previously, the corpus had to go through the preprocessing stage before the training process. The purpose of the training phase was to generate patterns and semantic lexicon. The development phase was a phase to identify and classify relevant information in a new collection of texts. The text used was not included in the corpus in the training process. The pattern was generated to get the extraction rules. In the development phase, the input text was passed to the OBIE system to produce an output. The last phase was the evaluation or testing phase. The architecture of the OBIE system in this study can be seen in Figure 1 below:



Figure 1. System Architecture

B. Preprocessing

The initial stage of information extraction to perform preprocessing on text input which aims to prepare the text into data that can be processed as input to the information extraction system. The parsing process was carried out on all text documents to identify all nouns or noun phrases (NP) and their

The parsing contexts. process in preprocessing consists of sentence detection, cutting sentences into (tokenization), tokens or words providing syntactic information (POS tagging), and cutting phrases (NP chunker). After parsing the text was complete, then the indexing and filtering process was carried out. The sentence indexing process was done by words/ NP detecting phrases in sentences, tokens on the left of NP, and tokens on the right of NP. After the indexing process was complete, the output (called a document-set) was stored in the database for processing.

C. DOM Tree

The Document Object Model (DOM) specification is an object-based interface developed by the World Wide Web Consortium (W3C) that constructs XML and HTML documents as tree structures in memory. Applications access XML data through an in-memory tree, which is a replication of how the data is structured. The DOM also allows users to dynamically traverse and update XML documents. It provides a model for the entire document, not just for a single HTML tag. The Document Object Model represents a web document as a tree. It is highly adaptable and can be used to renovate entire web pages. This is an explicit HTML document model. Some HTML tags do not include closing brackets. For some of these tags, the closing parenthesis is inferred by the following tag, for example, the tag is closed by the following tag. To analyze a web page, first, check the HTML document syntax because most HTML Web pages are not well-formed. After that, it passes the web page through an HTML parser which increases the markup and generates a DOM tree. Then the system breaks it down into several sub-trees according to the threshold value. Different websites have different layouts and serving styles, therefore the depth of the Web page tree varies according to the presentation style (<u>Kim & Lee</u>, 2017).

system must know the The maximum level of the DOM tree to select the best option. DOM tree-based method, as a segmentation method of high interest, is proposed based on the characteristics in which after parsing the DOM, HTML documents can form a tree structure that can accurately describe the hierarchical relationship between elements in a web page and is convenient for computer processing. After the web pages were parsed into a DOM tree, the algorithm grouped the web pages mainly by content features and DOM tree structure features. DOM is a common tool for representing web pages. In the DOM, the web page will be represented as a set of tags and a hierarchical relationship between the tags with the function of each tag, which allows the user to classify a message and an HTML tag (Sun, Song, & Liao, 2011).

D. Firefly Algorithm

This algorithm is inspired by nature which is based on a firefly's flash of light and mimics how fireflies interact with each other. In the firefly algorithm, some web documents or web pages are taken as input. The following are the steps in implementing the firefly algorithm as shown in the figure below.

Firefly Algorithm for Web Page Noise Removal Input: Multiple web pages Output: Extract relevant content from web pages Step 1: Access multiple web page Step 2: Read one by one page Step 3: Check web HTML tag Step 4: Consider the document with various tags Step 5: Objective function f(wi) w=(w1,w2,w3..) Step 6: Generate an preliminary population of fireflies Step 7: Formulate light intensity Step 8: Define absorbtion coefficient γ Step 9: While (t<Max_Generation) Step 10: For i=1:n Step 11: for j=1:n (n fireflies) Step 12: If(I>Ii) Step 13: Move firefly i to j Step 14: Calculate new solutions and update light intensity Step 15: End if Step 16: End for j Step 17: End for i Step 18: Identify the noisy information Step 19: Eliminate the noises

After reading the web page, the checked HTML tags were given in step 3 then consider the web document with various tags. In step 5, the objective function was calculated and generated the initial population of fireflies in step 6. The light intensity was formulated in step 7 and determines the absorption coefficient in step 8. In step 9, the maximum generation was evaluated based on the new solution of updated light intensity. In steps 18 and 19, noisy information is identified and eliminated (Bumbaca et al., 2011).

Finally, the main content was extracted. All information related to web pages was stored for efficient pattern retrieval using the Firefly technique. A database was created using an artificial neural network to store related data from web pages. Matching the constructed DOM tree with the database was to eliminate noisy information (Mangat, 2014). In the end, we can get the main content. The initialization of the objective function f (*wi*) is calculated using the light intensity I (*o*) which varies according to the inverse square law with the following formula:

$$I(o) = \frac{I_s}{I_o}$$
.....(1)

I(o) is the intensity at the source and r is the distance of the observer. The light intensity I varies with the square of the distance d. The absorption coefficient s calculated using the following formula:

The attraction of fireflies is proportional to the intensity of light perceived by other fireflies. The brightness observed by adjacent fireflies calculated using the formula below:

$$\beta\beta_0 e^{-yd^2}$$
..... (3)

The next step is initializing the firefly population. Firefly *i* is attracted to firefly *j* which is more attracted, its movement is evaluated using the following fouls:

$$x_i = x_i + \beta (x_j - x_i) + \alpha \varepsilon \dots (4)$$

The webpage noise removal funtto is calculated using the following formula:

$$fitnes = \alpha \frac{toT_{neg}}{T_{tot}} + \frac{\beta}{F} \dots$$
(5)

Ttot is represented as the total number of tags on a web page, meanwhile, *Tneg* is the negative tag on

a web page. Then, *F* is denoted as *F* and β is firefly attraction.

RESULT AND DISSCUSION

To verify the effect of the proposed method, the algorithm was implemented using the DOM method to analyze HTML tags which were then processed using an optimization algorithm to remove page noise using the firefly method. There were trials of several pages from different websites, such as Baidu Encyclopedia, Sina Blog, Tencent News, Blog Park, and other websites. This page has a clear distinction in content and structure and can illustrate the implementation of the algorithm well. This study proposed a DOM tree-based web page segmentation method that comprehensively considered the structural features and content features of web pages, used node information entropy to group nodes from the parsed DOM tree, and obtained the final segmentation results with node fusion in the form of news text content.

1. Description of Datasets

To carry out the experiment, datasets were collected from different web pages. These web pages contain meaningful content and also have noise such as advertising banners, copyright links, page icons, irrelevant navigation, and junk information included in the web pages (Kaur, 2014).

2. Performance Measure

In this study, valid blocks, invalid blocks, execution time, precision, recall, and F-Measure were considered as performance factors that were evaluated based on the number of negative tags and the total number of tags. Experiments were carried out to test and evaluate the proposed method, the effectiveness of detecting and removing noise to block web pages and produce clear documents. The validity and accuracy of the proposed algorithm were checked using recall, precision, and F-measures from the information retrieval field. The datasets used in the experiment consisted of several pages from different websites.

3. Scrapping Web Using DOM

The first step done in this research was to classify the web page tags and only process the tags that affected the page structure. The second step was to comprehensively consider the content features and structural features of the DOM tree node. This calculated the node information entropy and the maximum sub-node text density to determine whether or not a node was an independent page block. Next, the node fusion was carried out to obtain segmentation results. This research was applied to several pages from different websites for scraping news content on web pages. The following was a DOM standard used for initial website identification.



Figure 2. DOM Standard

The input in this process was the URL address or the HTML code of the web page to be extracted. The initial step was to convert HTML documents



Figure 3. DOM Process

There are several sections on a web page such as texts, images, lists, or tables. Therefore, the first step was to determine which part was the text of the HTML document. The texts were detected by making use of the string function. Texts in HTML match documents can be identified by the tag at the beginning and <\p> at the end of the table. The tag was used to detect reading in the form of a paragraph. If there are two or more paragraphs on a web page, the application stores the table at the second index of the variable in the form of an array. The next step was to create a DOM tree from the detected paragraphs. The DOM tree is composed of text-forming tags, namely tags . This tag is detected to determine the news content portion of the paragraph.

The third step was to extract or retrieve the data portion of the DOM tree. After the data section in the table was obtained, the next step was to save the data in the form of a CSV file. The selection of the form of the storage file is intended so that the extraction results can be used for subsequent needs, such as integration with data from other tables or to be stored in a database. Different websites have different layouts and serving styles, therefore, the depth of the Web page tree varies according to the presentation style. The system must know the maximum level of the DOM tree to select a good choice of threshold levels. That is why the system traverses the entire DOM tree to obtain the maximum DOM depth.

		- 11 - 11 - 12 - 14 Feat	- A' A' ≡ : - A - ▲ - = :	5 2 5 5 6 Alexand	Www swep Ilinap Text Morge & Center + 1	Second Reg = % 9 1 1 Number	Constrand Formating -	Former Table Myleri	Call Call Signes -	Cons	$ \begin{array}{c} \sum & - & \sum \\ \square & - & \sum \\ \square & - & \sum \\ Sorth & Find h \\ \varphi & - & Sorth \\ S$	dear Idear	Sectory textvy
	A 11110	с н.	D.				6		н				
Ina	* Line Numbe *	Dest Li +		- and the second second	+ CSS		Aller	+	101			1	STREET, STREET, STREET, ST
-	605 615	270	270 Saat ini u	ntuk produk kendan	ian niaga htrol	body.themeot	createl.page-nead	d	и.мпар	div.container	.clearftx		div.row.col-
	609 609	222	222 "Saya sud	lah perintahkan ke #	ak Ongloy Minsl	body there-of	orisetif, page-read	d	и литар	div,container	clearful		dw/ow.col-
	625 634	203	203 Depatkan	update dan setiap h	ari dari sihtrel	body-thenie-of	oreatif.page-read	d	и литар	div.container	clearfix		div.row.cel-
	ເວລີເວວ	202	202 di sini nat	mpakrisa berperan s	ebagai ke html	body.therseot	been-egge/hitoend	d	w.sarap	div.container	clearfix		div.row.col-
	582 582	201	201 Sast ini s	udah ada beberapa p	roduk ke html	body.themect	preset if, page-read	d	e.urap	div.container	clearfix		div.row.col-
	621 635	362	352 "Kedus, k	its jugs mengember	gkan . Ki html	body.themeof	practif.page-read	d	и.лигар	div.container	clearfix		div.row.col-
	725 722	162	352 Jkie men	cari berila yang dek	et denger Brekom	er div.title.dearlis		d	s.jote_log	div.jinie_log	o_ing		div.tooltips
	554 553	234	154 Founder?	PT MAB Moeldoko d	alore acad htrel	body there-of	orestrf.page-read	0	клитар	div.container	clearfur		div/ow.col-
r	588 587	115	115 -1] menu	pakan salah satu per	usahaan ihtrol	body.thenie-of	ornotif, page-vead	đ	и нитар	div.container	dearful		dw.row.col-
	602 605	\$7	97 Moeldok	o mendeklarasikan p	endinian #508525	45 strong							
	565 564	\$7	87 Puluhant	truk pengangkut bat	abara tak hirsi	body.therseot	prectif.page-read	d	w.wrap	div.container	clearfix		div.row.col-
	595 587	65	65 Kebiataan	t Buruk Pengemudi	ung Bisa html	body.themeot	prectif.page-read	d	w.wrap	div.container	clearfix		div.row.col-
	620 629	37	37 Dus Baru I	PO Junigart90, Prem	um dari 5 html	body.themeot	omotif.page-read	d	v.nnip	div.container	,clearfix		div.row.col-
	34	35	35 MA8 Sias	kan Dua Produk Ban	, Truk Ushtreit	head		0	tie -				
	456 455	35	55 MAB Siap	Ran Dua Produk Ban	A Truk Lis #dshare	sitie							
1	\$36 \$35	\$5	55 MAB Siap	kan Dua Produk Ban	, Truk Lishtral	body.themeof	oreotif.page-read	d	и лигар	div.container	clearfix		div.row.mt3
	676 675	47	47 Teka-Teki	i Santuy Edisi Kata 9.	iku dan T #kompa	sç hé.info-banner-	ist-tide						
	655 655	45	-46 (Phone L)	2 Pro Max dan Vouch	er Hanya Kompa	ap hiunfo-banner-	ist-trole						
	605 607	42	42 Depatkan	Informal, Impiral	dan di ka html	body.themeot	prestil page-read	d	x.smip	div.container	clearfts		div.row.col-
	681 680	41	41 TTS - Teka	Tekt Santuy Edist Is	stah Kata Riompa	sc h4.info-binner-	ist-title						
	685 685	43	40 Teka Telo	Salitury Edisi Alat %	insportas titionpa	sg h4.info banner-	istatie :						
	100.004	54	26 Louis de	automose month	Inite accession	holu those	been, eeen litnen	- 4		discounting.	distrific		ALLERANS

Figure 4. Scrapping 1

Autofawe 1		G •				westapping - Ertel	P. See							🛦 ivean Alpana 😩			
File His	me	mart	Fagi	e Layout	Form	las Data Revie	w View Help								g shan	Cores	nenta
Frate Ch.	0	Abri 2 II		11 . A	A		5 Wap Test	General THE N. M. M.	-	Conditional	-	1	Ellivet -	Σ - Δ. Ο Ξ - Σου δ - Γου δ 8	4 Ideas	-	
- d Optoerst	a C	10 A	Font	State of		Kgrmi	n 6	Number	19	Pormating	Table = Styles	Style =	Cells	6/ * Filter * Select * Editing	10000	Servitivity	1
6	•		1	6 htte	>body	there otomotifica	e-read > div wrap :	- EN.CONTRAINER A	iearfix >	RA FORA COF-0	fiset-filled	dearth	is part wo st	div-parent > div.col dis	10-7.js-rea	diancie >	
a) A			c	D											t		
1804		1535	.0			RgmModal > div.kgm	Modal-block > div.kg	mildodal-footer	> pix.fpm	n-reworth-cf	sild(2)						
1805		1532	38	Lakukan Se	kiran	RegmModal > sity.kgm	Modal-block > div.kg	mModal-footer	> dra for	a-reworth-ch	Hid(2) Fall	button.	primary				
1806		1553	0			ntml>body.theme-a	terriotif.page-read	> soriptieth-child	0(30)								
1807		1534	0			htrel > body.thereea	tomotif.gage-read	birkorth-child	35)								
1808		1515	0			wkgmModal-complete											
1809		1536	0			RgmModal-complete	+ div.kgmModal-blo	ock .									
1870		2537	0			RightModal-complete	> div. kgmModal-blo	ck > div, ightMo	dal-head	er							
1811		1538	35	Lengkapi P	rofii	RightModal-complete	> div.kgret/iodal-blo	ick > div.kgmMo	dal-head	er > tiv.kgmb	Modal-ticle	Ki -					
1812		1519	0			RgmModal-complete	> div.kgmModal-blo	ck > div.kgmMo	dal-head	er > div.kgml	bi-labole						
1810		1540	.0			RgmModal-complete	> div.kgmModal-blo	sck > div.kgmMo	dol-head	er > civ.kgml	Hodal-Img						
1914		1541	. 0			mgmModal-complete	> giv.kgnModal-blo	ick > div/kgniMo	dal-head	er > thy, light	Andal-Ing	> ang					
1815		1542	0			ekgraModal-complete	> tiv.kgmModal-bio	ick > div.kgniMo	dal-body								
1816		1543	50	Segera len	plapi-	ekgmModal-complete	> div.kgmModal-blo	ck > div.kgmMo	sdal-body	3.p							
7817		1544	- 38	Hernihöer	korne.	RgmModal-complete	> div.kgmModal-bio	ck > div.kgmMo	dal-body	> p > strong							
1818		1345	0			RigmModal-complete	> sty.kgmModal-blo	ick > div.kgmMo	dal-feeta	f							
1819		1546	.0			rkgmModal-complete	> div.kgmModal-bio	ick > div.kgniMo	dal-foote	r> div.form-	rowinth-cl	10101					
1820		1547	20	Nanti Saja		skgmModal-btn-clism											
1821		1548	0			RgmModal-complete	+ div.kgmModal-bic	ck > div.kgmMd	del-foote	r > div.fotm-	row prits-ct	10:03					
1822		1349	15	Longkapi P	rofit	RigmModal-complete	> div.kgmModal-blo	ck > dw.kgritMo	dal-foots	r>dw.form-	rowmth-cl	Fiel21>	a button -prim	ary			
1823		1554	0			html>body.theme-o	tomotif.page-read	> soniptieth-shill	d(83)								
1824																	
1825	910	cess fin	ished with	th exit code	0												- 0
1854																	1.
	1.5	beet	Sheet2	Sheet2	-594	ea 🕘					4						
														COMPANY AND AND			

Figure 5. Scrapping 2

	• • • •		ment imaging - face	a (2) Sa	and-						🔥 Inter Alpiera			
rile He	treast one	Page Layout Per	malas Data m	leview View He	ilp:							15 Shi	e Com	ments
Nere S	CMD4 B I U	• 11 • A' A' • ⊡ • • • ▲ •	= = 3 € • 5 = 3 = 5 = 5	fig Wag Test Marge & Center	- 500001	• (*2.8	Conditional Formatting	Farred an	Byle -	State - State - Information	Σ - 27 Ξ - 27 φ - 5018 fm φ - 5018 fm	0 d8 st=	and the second s	
Childrand	a	Park 1	2 24	primeral	16 Not	ider Ta		Styles		CH01	Idorg	Man.	Secularly.	
E28	* - X	J. A Meet>b	ody.thomeskole.pa	et-read > tity when >:	dis container-	clearftx > div.	ew.ool.offse	t-fluid.cle	artin.(5-g	unt-wp-stick	- parent > div col-b	00-T11-read-	article >	
		-												
And a state of the			<i>v</i>											_
1	(r)	120			* #no	ot								
1	634	234 Dikutin dari situ	in a start of the	falah menakul perte	most dera hir	ni > bods ther	ne-okola pa	no-read in	div.wrap	> div.contain	er clearf is h div row	Cost-offset-fit	uid dearfre h	-parts
4	650	209 Dapatkan updat	te dan setiap hari dar	Konsas.com.Mattb	ergalaung chtre	ni > body they	ne skola pa	ge-read >	diw wirap	> ewcentain	er.clearfix > div.row	col-attset-fi	ad deaths B	1-gant
5	630	166 Tisp menutis ju	ga berhak mendapat	perlindungan, memel	iluk agama, htm	ni > body.then	ne-skola pa	po-read »	div.orap	> piv.contain	er,clearfix > div.rov	.col-offset-fi	uid dearfis ja	-piant
1	79.6	test transportanting	and to deb man of the	the conference day and										
		The first state and the second state of the	COMPANY AND A COMPANY AND A	POLITICA CLEAR PROFILM BRID	shan Anda; Re	<pre>icomendaci>i</pre>	div.10e.dea	rftin > dên Jê	10: 10gs	NR18 1020-	nght> invitor lo	100 Img > (h)	A 100 R Jas > 50	pan I
1	638	152 Contoh mengak	ui pertamaan deraja	t antar secama manus	ishan Anda, Bie sia yaitu tidhtr	komenilasi > ni > body.thei	div 1/De diea na-skola pa	rhis > dis (i ge-med >.	<pre>ble_logs div.wtsp</pre>	> div.contain	er.clearfix > div.rov	igo_ing>dv .col-offset-fil	v tee Rugs > sp uid clearfix, js	pan i-giant
1	638	152 Contoh mengak 147 Sikap Ini henda	kui pertamaan deraja knya ditumbuhkan d	t antar sesama manus alam lingkungan perpi	shan Anda, Reg sia yaitu tidhte sulan masshte	Nomendasi > ni > body.thei ni > body.thei	div 1/0e diea roa-skola pa rea-skola pa	rhis > dis pi go-read >- go-read >-	div.unap div.unap	> div.contain > div.contain	-right > div_tote_k er.clearfix > div_rov er.clearfix > div_rov	igo_ing>dk .col-offset-fi .col-offset-fi	n too Kugs > sp uid clearfix, js uid clearfix, js	pan L-giart L-giart
7	638 637 632	152 Contoh mengak 147 Sikep Ini hende 148 Sikep ini berart	cui percansan deraja knya ditumbuhkan d Lantarmanutia saling	t antar sesama manus alam lingkungan pergi menghornati, mengi	shan Anda, Imp sia paitu tidhtr pulan masyhtr fuargai, serthtr	komenilasi > ni > body.thet ni > body.thet ni > body.thet	div 1/de diea nae-skola pa nae-skola pa nae-skola pa	rha>dia ja go-read > go-read > go-read >	He_logi div.wrap div.wrap div.wrap	> div.contain > div.contain > div.contain > div.contain	-right > div_boxlo er.clearfix > div_rov er.clearfix > div_rov er.clearfix > div_rov	igo_ing>dk .col-offset-fi .col-offset-fi .col-offset-fi	n too Mars > sp uid clearfix, js uid clearfix, js uid clearfix, js	pan L-giart L-giart L-giart
2 8 8	638 617 613 615	152 Contoh mengak 147 Sikap ini hende 146 Sikap ini berart 138 Meteroir deri si	kui peruamaan deroja Anya ditumbuhhan d Lantarmanusia soling tuo, sikep mengekui	ar preserver of the pro- t antar sesama manus alam lingkungan perg- persemaan derajat an	shan Anda, Rig sia paita tidhtr padan masyhtr (sargai, sarthtr rtarmanusi htr	komenilasi > ni > body.ther ni > body.ther ni > body.ther ni > body.ther	div 1/Ge diea ma-skola pa ma-skola pa ma-skola pa ma-skola pa	rha>dea ja go-read > go-read > go-read > go-read >	<pre>kie_logs div.unap div.unap div.unap div.unap div.unap</pre>	> div.contain > div.contain > div.contain > div.contain > div.contain	- ngH > div_botek er_clearfis > div_rov er_clearfis > div_rov er_clearfis > div_rov er_clearfis > div_rov	igo_ing>dk .col-offuet-fi .col-offuet-fi .col-offuet-fi .col-offuet-fi	n too Rugs > sp uid die arfin je uid die arfin je uid die arfin je uid die arfin je	oan L-giart L-giart L-giart
0	638 637 632 635 643	152 Contoh mengak 147 Sikap ini hende 148 Sikap ini berart 138 Meteroir dari si 113 Tidak memberi	kui pertamaan denija kinya ditumbuhkan d Lantarmanusia saling fusi, sikap mengakui perlakukan atau sika	t antar sesama manus alam lingkungan perg persemaan derajat an o yang berbeda kepad	shan Anda, Reg sia yaitu tidhtr jaulan masyhtr fuargai, saethtr tiarmanusii htr da teman, i htr	komenilasi > ni > body.ther ni > body.ther ni > body.ther ni > body.ther ni > body.ther	div 1/3e diea nie-skola pa nie-skola pa nie-skola pa nie-skola pa nie-skola pa	rfis > dis pi go-read > go-read > go-read > go-read > go-read >	dis serap dis serap dis serap dis serap dis serap	> div.contain > div.contain > div.contain > div.contain > div.contain > div.contain	right > div_polelo encleartis > div_row encleartis > div_row encleartis > div_row encleartis > div_row encleartis > div_row	igo_ing>dk .col-offset-fi .col-offset-fi .col-offset-fi .col-offset-fi	A too Rigs > sp uid ziearfin js uid ziearfin js uid ziearfin js uid ziearfin js uid ziearfin js	i-giant i-giant i-giant i-giant i-giant i-giant
2 5 8 0 1 2	638 637 633 635 643 643	152 Contoh mengak 147 Sikap ini honda 148 Sikap ini berart 138 Melanair dari si 113 Tidak memberi 108 Menurut Hadi R	cui pertamaan denja knya ditumbuhkan d i antamanusia saling fus, silap mengekui perlekukan atau sila lianto dalam jumal, i	proprotector con pro t antar secama manual alam Ingkungan perg- menghomasti, mengi persemaan derajat an piyang berbeda kepad semua orang memiliki	than Anda, Brg cia yaitu tidhtr aulan masshir hargai, sarthtr riarmanusi htr da teman, Fitr i dengat dahtr	komenilasi > ni > body ther ni > body ther	div 1/3e diea mo-skola pa mo-skola pa mo-skola pa mo-skola pa mo-skola pa mo-skola pa	rfis > dis (i go-read > go-read > go-read > go-read > go-read > go-read >	dia sinap dia sinap dia sinap dia sinap dia sinap dia sinap	hjtole_logo > biv.contain > biv.contain > biv.contain > biv.contain > biv.contain > biv.contain > biv.contain	right > div_boa_c er.clearfix > div_rou er.clearfix > div_rou er.clearfix > div_rou er.clearfix > div_rou er.clearfix > div_rou er.clearfix > div_rou	igo_ing>ch icol-attust-fi icol-attust-fi icol-attust-fi icol-attust-fi icol-attust-fi icol-attust-fi	x toolitys > sp uid zieartik, js uid cleartik, js uid cleartik, js uid cleartik, js uid cleartik, js uid cleartik, js	oan I-giant I-giant I-giant I-giant I-giant
2 9 1 2 2 3	638 637 633 635 643 638 638 638	152 Contoh mengak 147 Sikap ini hende 148 Sikap ini berari 138 Meterari dari si 133 Tidak memberi 168 Menurut Hadi II 82 - merupakan - 8	cui pertamaan denja knya ditumbuhkan d i antamanusia saking fusi, sitaip mengekai perlekukan atau sika kanto dalam jumal , hanji sila kedua Pano	per protectos can pro trantar secama manus alam lingkungan pergi persomaan derajat ar o yang berbeda kepad semua orang memiliki osilo adalah kemanusi	man Anda, Brg cia paita tidhtr pudan masshtr hargal, carthtr tiornanusi htr da teman, kitr i dengat da htr laan sang a htr	komenilasi > ni > body ther ni > body ther	div 1/3e diea no-skola pa no-skola pa no-skola pa no-skola pa no-skola pa no-skola pa no-skola pa	rfus > dis pr go-read > go-read > go-read > go-read > go-read > go-read > go-read >	dia sarap dia sarap dia sarap dia sarap dia sarap dia sarap dia sarap dia sarap	h(tote_logo > biv.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain	right > div_duke_k er.clearfis > div_rou er.clearfis > div_rou er.clearfis > div_rou er.clearfis > div_rou er.clearfis > div_rou er.clearfis > div_rou er.clearfis > div_rou	igo_ing>ch icol-ottuet-fi icol-ottuet-fi icol-ottuet-fi icol-ottuet-fi icol-ottuet-fi icol-ottuet-fi icol-ottuet-fi	A too Muss > sp uid ziearthe, je uid clearthe, je uid clearthe, je uid clearthe, je uid clearthe, je uid clearthe, je uid clearthe, je	pan Ligiant Ligiant Ligiant Ligiant Ligiant Ligiant
2 5 1 1 2 5 4	638 637 633 635 643 638 659 659 619	153 Contoh mengai 147 Sikap ini hende 148 Sikap ini berart 138 Meterair dari si 133 Tidak memberi 108 Menurut Hadi B 82 meteratari b 61 implementari b	ua pertamaan detaja keya ditambuhkan d Lantamanusia soling tuo, sikap mengakai pertakukan atau sika kanto dalam jumal, hunyi sila kedua Paro silai Komarusiaan ya	per protection can per al am Englungan perg- persamaan derajat am o yang berbeda kepad semaa drang memifiki osola adlatik kemanab ng Adil dan Beradab d	mon Andia, Brg cia paita tidhtr padan matshtr hargal, carthtr tiornanusi htr da teman, i-ter i denijat da htr laan sang a htr di Lingkung htr	konsenilasi > ni > body ther ni > body ther	div 135e diea nie - skola pa nie - skola pa	rfus > dis (1 ge=read > ge=read > ge=read > ge=read > ge=read > ge=read > ge=read > ge=read >	dia unap dia unap dia unap dia unap dia unap dia unap dia unap dia unap dia unap	h(tote_logo > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain > div.contain	ngR > div pole_k er clearfix > div rou er clearfix > div rou	igo_ing>ch col-other.fi col-other.fi col-other.fi col-other.fi col-other.fi col-other.fi col-other.fi col-other.fi	x too ittos > sp uid cleartix, jo uid cleartix, jo	949 1-51470 1-51470 1-51470 1-51470 1-51470 1-51470 1-51470
2 5 1 2 3 4 5	638 637 633 635 648 638 679 679 639 645	153 Gontoh mengai 147 Sikap ini hende 148 Sikap ini herart 139 Meteroir dari si 153 Tidak nemberi 160 Menurut Hadi R 82 - metepatan it 16 inglementaan h 16 inglementaan h	oui perumaan denja keya citumbohian d Lantamanutis coling fus, sikap mengekul periakukan atu, sika kanto citam jumal, n kenyi sika kedua kan kiai kenanutian ya tar sesama menusia	per protection can pro a solar inglivingan perg- persamaan derajat am p yang berbeda kopad semaa orang memiliki sola adalah kemanusi ng Akil dan Beradala bergan menjunjung 10	man Anda, Bre sia paita tidhte padan masshte (hargal, saethte tlarmanusi hte da teman, khen da teman, khen da teman khen da teman da hte ban sang a hte ti Lingkung hte inggi persa hte	Nomentas) > ni > body ther ni > body ther	div 1/13e clea ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa	rfus > dis (1 go-read > go-read > go-read > go-read > go-read > go-read > go-read > go-read > go-read > go-read >	He_logi div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap	hjble_bjole_bjole_bjole_bjole_bjole_bjole > blv.contain > blv.contain > blv.contain > blv.contain > blv.contain > blv.contain > blv.contain > blv.contain > blv.contain	nght > liw juse_v enclearfu > div.row enclearfu > div.row	go_ing > ch col-offset-fl col-offset-fl col-offset-fl col-offset-fl col-offset-fl col-offset-fl col-offset-fl col-offset-fl col-offset-fl col-offset-fl	v toolitos > sp uid clearte, je uid clearte, je	ue 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant
2 8 0 1 2 3 4 5 6	638 637 633 643 643 643 643 643 643 644	153 Gontoh mengai 147 Sikap ini hende 148 Sikap ini herart 138 Welansir dari si 133 Tidak neehben 140 Wenunut Hadi 8 82 menyakan - 8 63 implementau h 78 Menghagai ora 67 Wenghagai ora	oui pertaman denja knya ditumbuhkan d Lantarmanusia oling fuo, sikap mengekau perlakukan atbu sika lanto dalam jumal, n langi sibabedua Pano vilai Komanusian ya har sesama manonya lar sesama mangan tidak	per protecter of two per protecter of two per and two per per per sense of two per se	than Anda, Bro sia patu tidhtr pagar mesyhtr (harga, serthtr rlarnanusi htr da temar, khtr da temar, khtr da temar, khtr di Lingkang, htr di Lingkang, htr higgi pirsa htr n sersena - htr	Nonentias) > hi > body, then hi > body, then	div 1/16 clea ma -skola pa ma -skola pa	rfis > dis () go-read > go-read >	He_logi div sinap div sinap	hjtvie_logo > div.contain > div.contain	nght > liw june_k er clearfu > div rou er clearfu > div rou	go_ing > ch icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi icol-attest-fi	A too ittas > sp uid. cleartin, jo uid. cleartin, jo	ue I-plant I-plant I-plant I-plant I-plant I-plant I-plant I-plant
7	638 637 633 643 643 669 643 643 643 644 389	133 Contoh mengai 143 Sikap Ini handa 145 Sikap Ini handa 138 Welanar dari sa 139 Menanar dari sa 130 Menanar dari sa 130 Menghangai ang 60 Contoh sikap m	oui pertamaan denja knya citumbuhkan d i antarmanusia saling perlakukan atuu sika ianto silakan jurnal , ianto silakan jurnal , ianto silakan jurnal , ianto silakan jurnal , iang lain Kernanusiaan ya tar sebania manusia ng lain dengan tidak emgelui petsamaan	pan protector una para ta nara seasana manua alam lingkungan peng- menghunnaki, menghunnaki, menghunnaki peranaman denjat am peranaman denjat an peranaman denjat di serua orang memilika sela adulah kemanusi sela adulah kemanusi peranaman mempunyang hi melakukan perbasaran	Wan Anda, Bro via yakan masyhte hargan, serihte tharmanusi hte da teman, Piter a dengut da hte da teman, Siter da teman, Site da teman, Site da teman, Site da teman, Site si dengut da hte manusia hte manusia hte	Nonentias) > Ni > body, then Ni > body, then	div 1/16 clea ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa ma-skola pa	rfis > dis () go-read > go-read >	HE_Op dis unap dis unap	hjtvie_logo > div.contain > div.contain	nght > liw (tota_tk) er.clearfia > div.rou er.clearfia > div.rou	go_ing > ch col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi col-offset-fi	A too Itas > sp uid. clearths, jo uid. clearths, jo	ue I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant
2 5 6 7 7 8 8 8 8 8	638 637 633 643 643 643 659 629 643 644 644 644 644 644	153 Contoh menggia 147 Sikap Ini hende 168 Sikap Ini hende 158 Melanan dari si 158 Melanan dari si 159 Melanan kindi 82 - menyakan, 8 83 Implementaki 178 Melaphagai ora 67 Melaphagai ora 68 Melahagai ora 86 Contoh sikap mela 18 Melabagai ora 18 Melabagai ora 18 Melabagai ora	oui pertaman denja knya ditumbuhkan d i antamanusis obing tus, sikap mengikul kianto dalam jumal, i nanyi sila kedua Paro Vilai Kenanyusiaan ya tar sesama menusia ng lain dengan tidak emgakui persaman i an yang kesudahan at	per protecter terretor tata per tantar seasan manua alam Ingkungan peng- menghormadi, mengi persamaan denajat am o yang berbedu kepad sertua orang memilika sertua orang memilika sertua orang memilika setua data kemanos dengan menjunyung ti melakukan perbatana dengat antar secara ditimpa mu	Wan Anda, Bro via parta tidhtr paolan masyhtr thargai, serthtr tharnanusi htr da teman, Fittr a dengat da htr San yang a htr i Lingkung, htr inggi persen htr manusia htr usibah, htr	Nonventiagl > Ni > body then Ni > body then	div 1/16 diea me skola pa, me skola pa,	rha > dis (1 po-read > po-read >	He_logi dix unap dix unap	hjitite_logo > div.contain > div.contain	ngit > lingit > lingi	go_ing>chi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi col-attest-fi	A book as so point of the article is used	ue I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant
0 1 2 3 4 5 6 7 8 8 9 9	638 637 633 643 643 643 643 643 643 644 645 644 645 645 642	153 Contoh menggia 147 Sikap Ini hende 148 Sikap Ini berah 158 Welman dari si 158 Welman dari si 158 Welman Hadi B 82 menyakan t 168 Implementah 169 Implementah 16 Implementah 16 Menghangai ora 16 Contoh silaga 16 Kenghangai ora 18 Menolong berai 18 Menolong berai 18 Menolong berai 18 Menolong berai 18 Menolong berai 18 Menolong beraith	ou perumun denja keye (Numbohan denja Lantamanula saling tur, sikap mengelau perlakukan atau sika kanto delam juani , ang sika denjan tidak engatu persamanusian ya an yang besutahan at ai dan kengian tidak	proprotection was per- proprotection and a series in a series of the energy company and the provide the series of the provide the series of the provide the series of the series of the series of the series of the series of the series of the series of the series of the series of the series of the series of	Wan Anda, Bre via pata tidher (hargar, sechte tharnanusi hin da teman, i hin higgi persa hin a secteta - hin manzala hin usbah, hen	Nonsentral > No body then > body then	div 1/16 diea ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa, ma-dkola pa,	rth > chi ji go-read > go-read >	He_logi dix unap dix unap	hjitite_bigo Silv.contain	- right > like, init_, i er scharfta > div.rou er scharfta > div.rou	go_ing > ch col-offset fi col-offset fi	A booktast > spo uid chearths, jo uid chearths, jo	ue I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant
5 5 6 7 6 6 7 6 6 0	638 637 633 648 638 649 649 649 644 589 644 589 644 589 644 589 644 589 644 589 644 589 644 589 644 589 644 589 644 586 644 645 645 645 645 645 645 645 645 64	123 Control meruppi 123 Step Ini hende 125 Step Ini hende 125 Melensin densin 126 Melensin densin 126 Indexensin densi 127 Melensin held im 126 Implementation 127 Melensination 127 Melensination 128 Melensination 128 Melensination 128 Melensination 128 Melensination 128 Melensination 128 Melensination 128 Melensination 128 Melensination 128 Melensination 129 Melensination 129 Melensination 120 Melensin	uu pertamuan denja knyo (Uhurkonhan d antaramaratis kalip pertakukan atku olas lanto dialam juma). unji ola kelua Anoo sitai komananisan ya ata sesama manoja i ang lain dengan dengan diala ata yang kesuathan at ak dan konsejaban ya	per per cereiro sono per per cereiro sono per per cereiro sono manane deraste in personamente deraste a por para permentifica de la para de la persoana come per ementifica con persoa de la persoana de	Wan Anda, Bee ila patta tidhte plaqain mashle hargai, sertito tiarmanusi hite da temau kite da temau kite da temau kite ili Ungkung hite inggi perse hite manusia hite sebah, ter her ken ken ken ken ken ken ken ken	Nonsentral > No body ther N > body	divitife diea ma-skola pa, ma-skola pa,	rtu > du ji go-read > go-read >	He_logs dis unap dis unap	h) bitlogo > div.contain > div.contain	right > liky, toit er charffis > divrou er charffis > divrou	go_ing>chiefterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil col-atterfil	A booklast > sp uid clearths (s uid clearths (s) uid clearths (s) uid clearths (s) uid clearths (s) uid clearths (s) uid clearths (s)	ue I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant
5 6 7 4 5 6 7 6 8 8 8 8 9 9 1	638 637 633 648 648 649 649 649 649 649 649 649 649 649 649	123 Contoh menggib 123 Step Ini hendel 124 Step Ini hendel 125 Melanar danisi 126 Melanar danisi 126 Menunut Hadi II 82 - menupakan ya 126 Inglisenanta II 78 Hadup nukun ani 127 Melangkangai ora 128 Menulang benu 128 Menulang benu 128 Menulang benu 128 Menulang benu 128 Kenulang benu 128 Kenulang benu 128 Contoh Silap M	nú perumun denja knya čitardonkan d i artannanusis koling periakuna njao slaka lianto dalam jumah, hanji Slakaba jumah, ng lain dengan tidak engetu persenan ng lain dengan tidak engetu persenan ng kan dengan tidak dan konsej ban jum	proprior de centre de la proprior de centre de la proprior de la centre de la proprior de la centre de la cen	Wan Anda, Ber via parta tidhte parta tidhte pargai, serthte tharmanusi hte da temput dahte da temput dahte da temput dahte da temput dahte da temput dahte da temput dahte ti Lingkung hte sama tid ti Lingkung hte ti Lingkung hte t	Nonentari > Nody the > body then i > body then	divi tifte diea ma-skola pa rea-skola pa	rtu > du ji go-ruad > go-ruad >	He_logi div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap	httis_logo > div.contain > div.contain	right > thy total er charfin > div zou er charfin > div zou	go_ing2-ch col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi col-chter.hi	A booklastis kup uid clearthe jo uid clearthe jo	44 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant 1-glant
2 8 8 9 0 1 1 2 2 3 4 4 5 6 6 7 8 9 0 0 1 1 2 2 3 4 4 5 6 0 0 1 1 2 2 3 4 4 5 5 0 0 0 0 0 1 1 2 3 5 4 5 1 5 1 5 1 5 1 5 1 5 1 5 1 5 1 5 1	638 637 633 643 648 649 649 649 643 644 645 644 642 4 45 5340	25. Constructions menuging (25.7 Strategin in Terrary 25.8 Construction) in Terrary 25.9 Webmark Handlin (26.1 Webmark Handlin (26.1 Wengthang) on (26.1 Wengthang)	nia percaman denja innya ditumbohan denja artamananis koling fato, sitep mengelar into dalam jumal, i artan dalam jumal, i arta biskema konsultan ya dar sesama manusian ya dar sesama biskema dalam dar sesama dan yang kesadahan da dan konsighan ya haramanan hengaku hestorman	properties de terres dans properties de terres de terres antes a resuma manues alam integringen perge personna de terres promitiés de terres de	Way Anda, Boy ila parta tid htr parta tid htr pargal, senthtr hargal, senthtr hargal, senthtr i dengar da htr a senaul, htr i dengar da htr blan parg à htr di Ungkung htr di Ungkung htr hungi pirsai htr n sensesa htr manzala htr htr htr htr htr htr htr htr	homeniai) > body their ii > body their harsotle ii > body their harsotle ii > body their ii > body their	citor 1196 cites cites rese-social pag- rese-social pag- lo	rfu > du () go-read > go-read >	ne_lop div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap div.wrap	Little Deputy > divicentian > divicentian	right > they to be exclusifie > divisors exclusifie > divisors	go_ing>chickson colortheath colortheath colortheath colortheath colortheath colortheath colortheath colortheath colortheath colortheath colortheath colortheath colortheath	A hostical > sp uid clearfu, ju uid clearfu, p uid clearfu, p uid clearfu, p uid clearfu, p	ue I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant I-glant
5 9 9 1 1 2 2 3 4 4 5 6 6 7 7 8 9 9 0 1 1 2 2 3 4 4 5 6 6 7 7 8 9 0 0 1 1 1 2 2 3 4 4 5 5 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	638 637 633 643 643 643 643 644 644 345 644 644 644 644 644 645 642 4 345 645	253 Control Intercepts 267 Storgent Intercepts 268 Sidap init Interact 259 Weiterin denia 269 Weiterin denia 260 Weiterin denia 261 Interception 261 Weiteringstein 262 October Sidap 263 Weiteringstein 263 Weiteringstein 253 Weinghormatil 263 Control Sidap M 263 Control Sidap M 263 Control Sidap M 263 Control Sidap M	nia pertaman denja hinya citumbohan denja artamananis oding tras, sitapi menghala pertakuna tato sika lianto dalam jumali, hanyi Silakebua Paro ata secama nensoba tai kemanutan ya da secama nensoba na jain dengan tidak mengkup persenaan anyong kecu catan at ak dan teorojaban ya dengkala (hospanaa) dengkala (hospanaa) dengkala (hospanaa) mengkala (hospanaa)	per para de contra una para la ministra resultan manual al ministra de la para persaman dera jale an persaman dera jale an persama da la para de la para la para de la dela dera de la dela dera para de la dela dera de la dela dera para de la dela dera de la dela dera para de la dela	way and a low and the second state of the seco	townenius > to > body then ii > body then	char II de José de la pa ma - seola pa le	rfu > du i go - read > go - read >	He_logi div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap div.unap	Little Digge Di	right > liky, to it exclusifie > divrou er clearfie > divrou	ego_ing> ching> chi col-attuar th col-attuar th	Notitizai > spisal disentia i spisal disentia i spisal disentia i spisal disentia pi sold disentia pi sol	nan Inglant Inglant Inglant Inglant Inglant Inglant Inglant Inglant Inglant Inglant Inglant

Figure 6. Scrapping 3

tateler		19 - 12 - 17 men kaupping - Grad 🖉 States	×
File	Home Inte	nt Rage Layout Formulas Data Review View Help @Share @Commer	da
Paste Q	n v Calibei	Image: Second	
Clabeard	19	Feet 15 Alignment 15 Number 15 Styles Colls Editing Lides Sensitivity	~
<u>^1</u>	• 1 0	A V X	*
	8	C D E P O H I J K L M N O P Q R S T	U 🖻
1	x		-11
2	447	291 Sebanusk Methodealtext > 0.0th-child(9)	
4	445	249 Tak cuma #definited a canth-child 100	
3	432	245 Porosaha Rieta (ktaltert > o n/t)- (hild (2)	
6	433	29 Perusahui Metikaleta iteet > pinth-child(d)	
7	409	185 Secerti di #detikdeta/tect > grith-child (15)	
8	435	182 la pun per #detikdeta/tert > grith-child(7)	
2	457	120 PF service #detailtext > purch-child [15]	
10	133	165 <iframe html="" w=""> head > noscript</iframe>	
11	460	130 Pernimpir #detikdetalitext > p.nth-child(16)	
12	425	137 Penusaha #detikdeta/text > pr/th-child(2)	
18	449	135 Ada seper #detikdetaitext > pixth-child(11)	
14	458	121 Tak cuma edetikdetaiteet > prith-child [14]	
15	564	78 PPN Naik #otherArticles > article_list-content_tem.column-3rth-child[3] > div.media.media-image-radius.block-link > div.media_text > div.me	
16	522	76 Skim Gels RecessContent > article.list-contentitem.column-conth-child(3) > h3.list-contentUtile > a	
17	563	74 18 Foto Te Notherstrticles > article.list-content_item.column-limit-child[2] > div.media.media-image-radius.block-link > div.media_text > div.media_ttRl > a media_link	
10	425	72 Pangeran html > body > div.container > div.grid-row.content_bg > div.column-li > article.detail > div.detail_media > figure.detail_media-image > figurption.detail_media-captio	
19	446	71 Putera Ma #detikdeta/text> table Jinksis/ponth-child(8) > tbody > tr > td > div./ihatig > a	
30	531	71. Digwyw P. RcrossContent > article.itst-contentitem.column.exrth-child(8) > h3.itst-contentttile > a	
21	472	67 Otanja ki kaoupi title	
22	574	65 Asus ROG RotherArticles > article.list-content_tem.column-Jorth-child[4] > div.media.media-image-radius.block-link > div.media_text > div.media_title > amedia_title > amedia_title > and a link	
	443	B Accessed and the statement of the stat	
	Sheeti	Shert2 Shert3 Shert4 ④	
Finady			100%

Figure 7. Scrapping 4

4. Optimization of Firefly Algorithm Method

All information related to the web pages was stored for efficient pattern retrieval using the Firefly technique. A database is created using an artificial neural network to store related data from web pages. Matching the created DOM tree with the database is to remove the noise information to get the main content (Liu, 2017). The initialization of the objective function f *(wi)* is calculated using the light intensity I(o) which varies according to the inverse square law. The formula used I as follows

$$I(o) = \frac{I_s}{I_o}$$

I(o) is the intensity at the source and r is the observer's distance. The light intensity I varies with the square of the distance d. The absorption coefficient γ is calculated using the following formula:

$$II_0 e^{-yd^2}$$

The steps of the firefly algorithm can be described as follows:

- a) Step 1: Access several web pages
- b) Step 2: Read every web page, one by one
- c) Step 3: Check web HTML tags
- d) Step 4: Consider documents with multiple tags

CONCLUSIONS

Based on the study having been conducted entitled *Automatic Web News Content Extraction*, there are several things that can be concluded as follows:

- a. Automatic Web News Content Extraction Algorithm can be used in optimal extraction of main web page content (news) and results in a better approach.
- b. The use of different datasets has a varying effect on the performance of the indicated precision, recall, and fmeasure parameters. This depends on the page reference level of similarity with the extracted page. The more

- e) Step 5: Objective function of *f* (*wi*)
 w = (*w*1, *w*2, *w*3 ..)
- f) Step 6: Produce an initial population of fireflies
- g) Step 7: Formulate light intensity
- h) Step 8: Determine the absorption coefficient $\boldsymbol{\gamma}$
- i) Step 9: Meanwhile, (t < Max_ Generation)
- *j*) Step 10: For *i* = 1: *n*
- k) Step 11: For *j* = 1: *n* (*n* of fireflies)
- l) Step 12: If (*lj* > *li*)
- m) Step 13: Move the fireflies to j
- n) Step 14: Calculate the new solutions and renewing light intensity
- o) Step 15: End if
- p) Step 16: End for j
- q) Step 17: End for *i*
- r) Step 18: Identify noisy information
- s) Step 19: Eliminate noise
- t) Step 20: End it

similar, the more stable the performance shown.

- c. The use of raw and valid datasets also has varying effects on precision, recall, and f-measure performance. It depends on the validation process of the tag structure of the website page.
- d. The researcher suggests that extracting web page content can be tried using other methods that will produce a better approach for data extraction.

REFERENCES

Abburu, Sunitha, & Golla, Suresh Babu. (2015). Satellite image classification methods and techniques: A review. *International Journal of Computer Applications*, 9(8).

- Allen, Jennifer, Howland, Baird, Mobius, Markus, Rothschild, David, & Watts, Duncan J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), 3539. 10.1126/sciadv.aay3539
- Bumbaca, Daniela, Wong, Anne, Drake, Elizabeth, Reyes II, Arthur E., Lin, Benjamin C., Stephan, Jean Philippe, Desnoyers, Luc, Shen, Ben Quan, & Dennis, Mark S. (2011). Highly specific off-target binding identified and eliminated during the humanization of an antibody against FGF receptor 4. *MAbs*, *3*(4), 376–386. Taylor & Francis. <u>https://doi.org/10.4161/mabs.3.4.1578</u> <u>6</u>
- Dey, Arnab, & Jain, Sudhanshu. (2020). Automatic skimming of web pages on a single click efficiently. 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 596–602. IEEE. 10.1109/ICOEI48184.2020.9143003
- Kara, Soner, Alan, Özgür, Sabuncu, Orkunt, Akpınar, Samet, Cicekli, Nihan K., & Alpaslan, Ferda N. (2012). An ontologybased retrieval system using semantic indexing. *Information Systems*, 3(4), 294–305. https://doi.org/10.1016/j.is.2011.09.004
- Kim, Yeongsu, & Lee, Seungwoo. (2017). SVM-based web content mining with leaf classification unit from DOM-tree. 2017 9th International Conference on Knowledge and Smart Technology (KST), 359–364. IEEE. 10.1109/KST.2017.7886134
- Newman, George E., & Cain, Daylian M. (2014). Tainted altruism: When doing

some good is evaluated as worse than doing no good at all. *Psychological Science*, *25*(3), 648–655. <u>https://doi.org/10.1177/095679761350</u> <u>4785</u>

- Sun, Fei, Song, Dandan, & Liao, Lejian. (2011). Dom-based content extraction via text density. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, 245–254. https://doi.org/10.1145/2009916.2009 952
- Velloso, Roberto Panerai, & Dorneles, Carina F. (2013). Automatic web page segmentation and noise removal for structured extraction using tag path sequences. *Journal of Information and Data Management*, 4(3), 173.
- Wimalasuriya, Daya C., & Dou, Dejing. (2010). Ontology-based information extraction: An introduction and a survey of current approaches. *Journal of Information Science*, Vol. 3, pp. 306–323. Sage Publications Sage UK: London, England. https://doi.org/10.1177/016555150936

https://doi.org/10.1177/016555150936 0123

Yu, Xin, & Jin, Zhengping. (2017). Web content information extraction based on DOM tree and statistical information. 2017 IEEE 17th Conference International on Communication Technology (ICCT), 1308–1311. IEEE. 10.1109/ICCT.2017.8359846

© 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/).