Practice Collecting Multiple URL Review Data

Digital Business Ecosystem Research Center





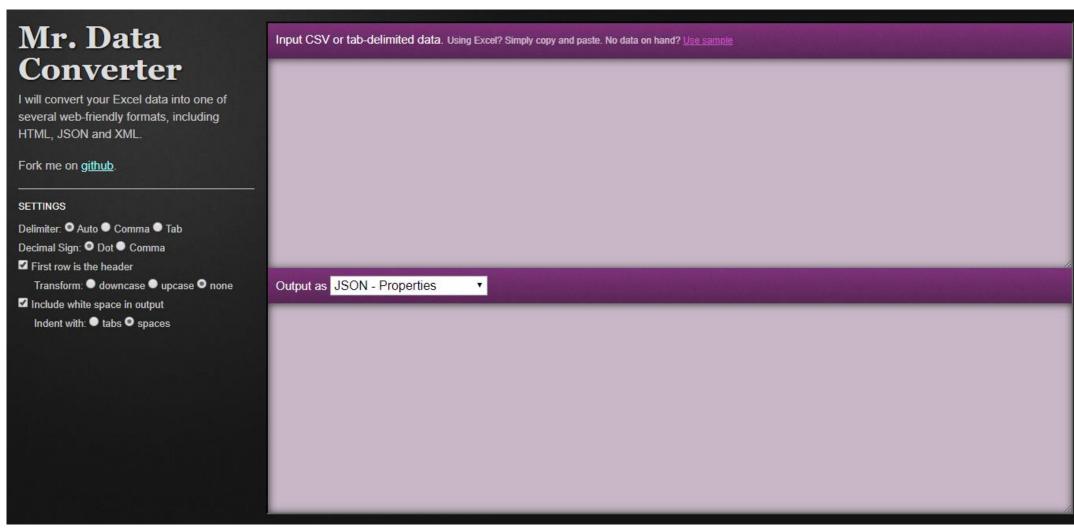


Make sure you have scraping data on product names, prices, and urls on the webpage.

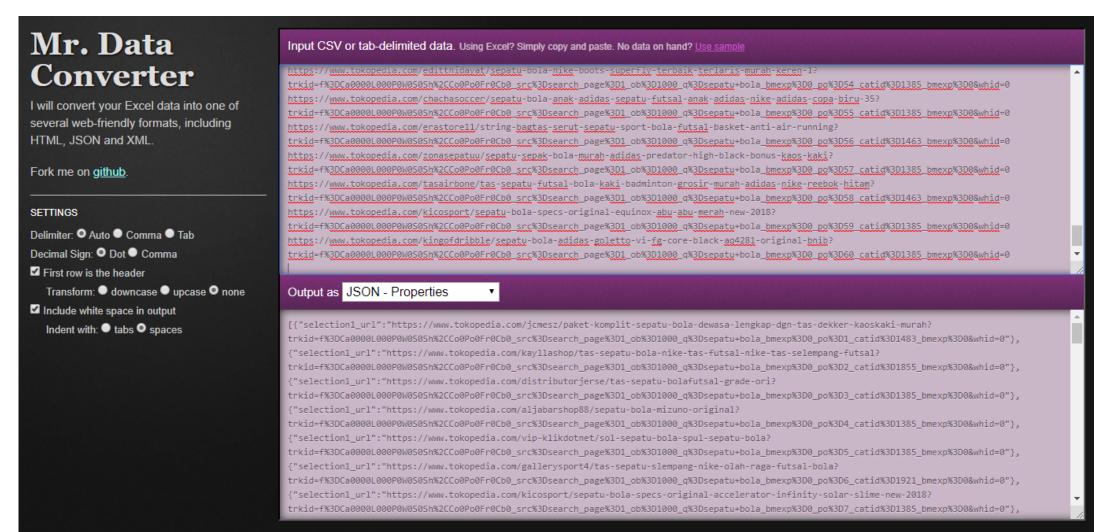
A	А	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	Q	R	S	Т	U
1	selection	selection	1selection	1 selectio	n1_selection	n2_url															
2	Paket Ko	n https://v	v Rp 115.00	(https://	www.tokop	edia.com/j	jcmesz/pa	ket-kompl	it-sepatu-l	oola-dewa	sa-lengkap	-dgn-tas-de	kker-kaos	skaki-mura	h?trkid=f%	63DCa0000	L000P0W0	sosh%2CC	o0Po0Fr0C	b0_src%3D	search
3	Tas sepat	tt https://v	v Rp 29.000	https://	www.tokop	edia.com/l	kayllashop	/tas-sepat	tu-bola-nik	e-tas-futs	al-nike-tas	selempang	-futsal?trl	kid=f%3DC	a0000L000	P0W0S0Sh	%2CCo0Po	0Fr0Cb0_s	rc%3Dsear	ch_page%3	3D1_ob
4	Tas Sepat	tı https://v	v Rp 25.000	https://	www.tokop	edia.com/	distributor	jerse/tas-	sepatu-bo	lafutsal-gr	ade-ori?trk	id=f%3DCa0	0000L000P	0W0S0Sh%	2CCo0Po0	Fr0Cb0_src	%3Dsearc	h_page%3	D1_ob%3D	1000_q%3[Osepati
5	Sepatu B	o https://v	v Rp 589.00	(https://	www.tokop	edia.com/a	aljabarsho	p88/sepat	u-bola-miz	uno-origi	nal?trkid=f9	63DCa0000L	.000P0W0	S0Sh%2CC	o0Po0Fr0C	b0_src%3D	search_pa	ge%3D1_c	b%3D1000	_q%3Dsep	atu+bo
6	Sol Sepat	tt https://v	v Rp 28.000	https://	www.tokop	edia.com/	vip-klikdot	tnet/sol-se	epatu-bola	-spul-sepa	tu-bola?trl	kid=f%3DCa	0000L000F	POW0S0Sh9	62CCo0Po	0Fr0Cb0_sr	c%3Dsearc	:h_page%3	3D1_ob%30	01000_q%3	Dsepat
7	tas sepat	u https://v	v Rp 29.800	https://	www.tokop	edia.com/	galleryspo	rt4/tas-se	patu-slem	oang-nike-	olah-raga-f	utsal-bola?	trkid=f%3	DCa0000L0	00P0W0S0	Sh%2CCo0	Po0Fr0Cb0	_src%3Ds	earch_page	%3D1_ob9	%3D100
					www.tokop																
					www.tokop																
					www.tokop																
					www.tokop														_	_	
	-			-	www.tokop			-	-												
					www.tokop																
					www.tokop																
					www.tokop																
				_	www.tokop																
					www.tokop																
					www.tokop																
					www.tokop																
					www.tokop			-			_								_		
	-			-	www.tokop								_								
					www.tokop							_									
23	SEPATITE	unttns://v	vi kn 775 <u>,00</u>	inttns://	www.tokon	edia com/i	niakiirniat	i/4/senati	ı-nola-adid	ias-x-techi	it-noots-m	urah/frkid=	:1%3DCa00	1001.000100	WOSOSh%2	COUPOUEr	OChO src%	RUSEarch	nage%3D1	ob%3D10	<u>/00_0%</u>

Go to Mr. Data Converter.

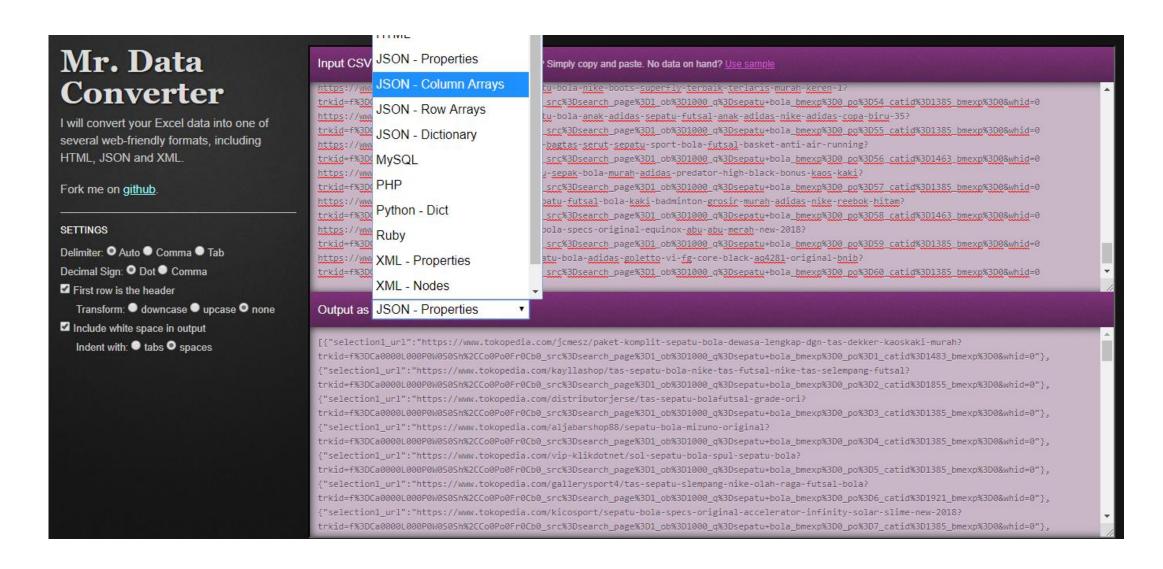
https://shancarter.github.io/mr-data-converter/



In the first box enter all of your links. Make sure to type in a heading name at the top of the column such as "urls".

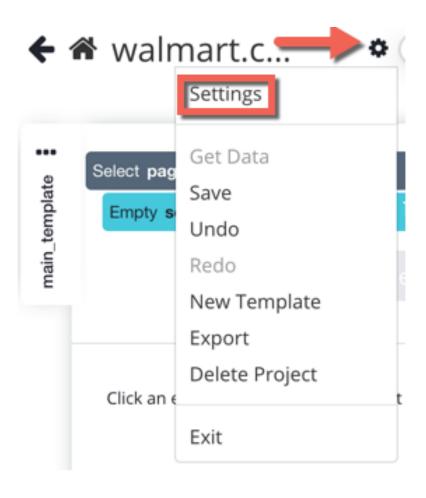


From the dropdown select "JSON - column arrays".



Open your project using any page as the URL, such as the homepage for your website.

Go to the "Settings" menu in the project.



Paste JSON from Mr. Data Converter into the "Starting Value" textbox, you can format it as follows:

- The list name "urls" can be renamed to anything you want such as "links" or something more descriptive like "shoes" or "brands".
- You can enter as many links in the structure as you want. We have 3 in the example below, but you can keep adding more links inside quotation marks and separated by a comma.

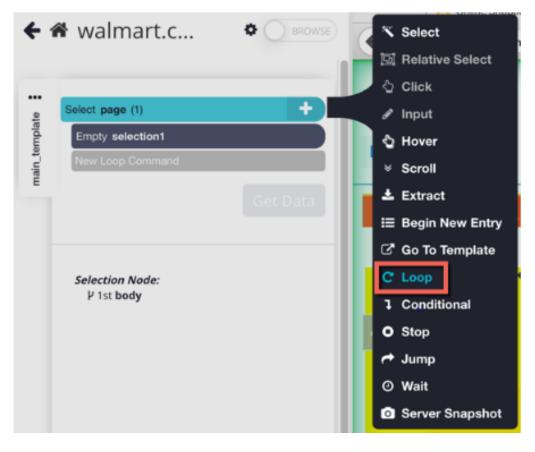
```
{
  "urls": [
  "https://www.walmart.ca/en/ip/hp-stream-14-cb110ca-14-inch-laptop-white-intel-celer
  "https://www.walmart.ca/en/ip/hp-17-by0002ca-173-laptop-natural-silver-and-ash-silv
  "https://www.walmart.ca/en/ip/acer-aspire-3-156-laptop-amd-e2-9000-amd-radeon-r2-gr
  ]
}
```

Click on "Back to commands" to go back to the project view.

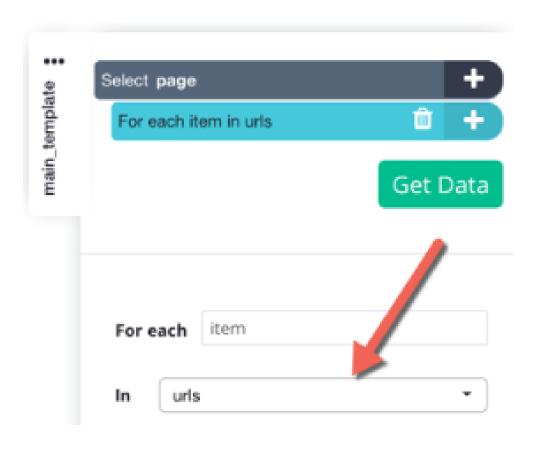
Click on the + button on the right side of the "Select page" command.

From the tool box, choose the "Loop" tool. It is inside "Advanced".

The loop tool iterates over a list. It is good for repeating commands multiple times.

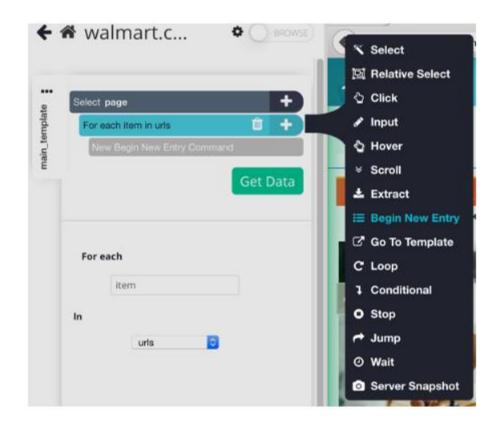


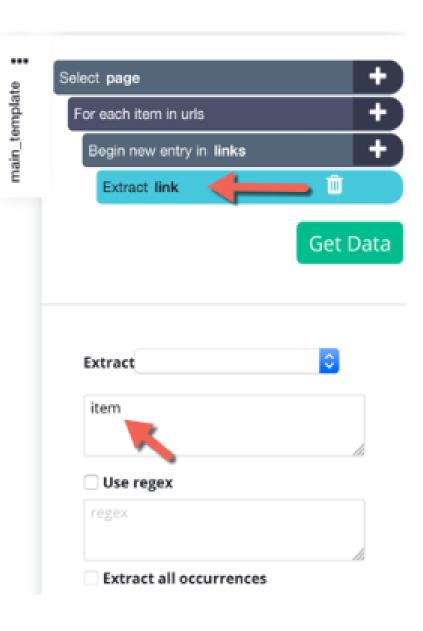
When added, your Loop command should default to looping through your "urls" list.



Click on the + button on the right side of the "For each item in urls" command.

From the tool box choose the "Begin New Entry" tool. Now the results of each one of the urls will go into a separate row in CSV and a separate object in JSON.

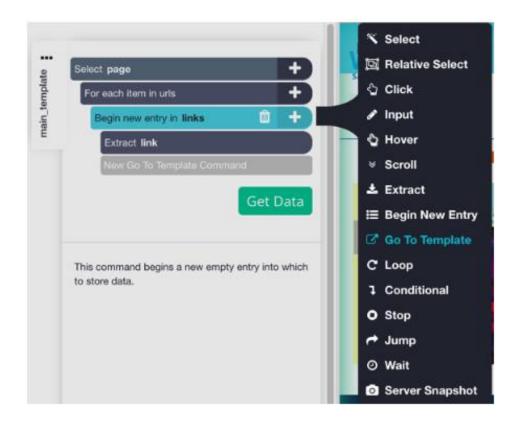




Rename the "list1" command to something else like "links". Make sure not to name the Begin new entry command the same as the list that holds your urls.

The begin new entry command should have a unique name.

Click on the + button on the right side of the "links" command and choose the "Go To Template" command from the tool box. The Go To Template command will let you specify which url you want to go to and which type of page you want to open.



On the pop up window, choose the "Go to URL" option instead of the "Stay on the Same Page" option.

In the text box type in "item" without quotation marks, assuming you didn't name it something else.

Go To Template			×						
		n either execute the template ns, so you can combine them							
Stay on the Same Pag	е								
O Go to URL: item									
● Go to Existing Template main_template Go to Template from Another Project									
Create New Template	template name								
		Go to Existing Temp	late						

In the "Create New Template" text box type in the name of a new template you want to open for each link - such as "results".

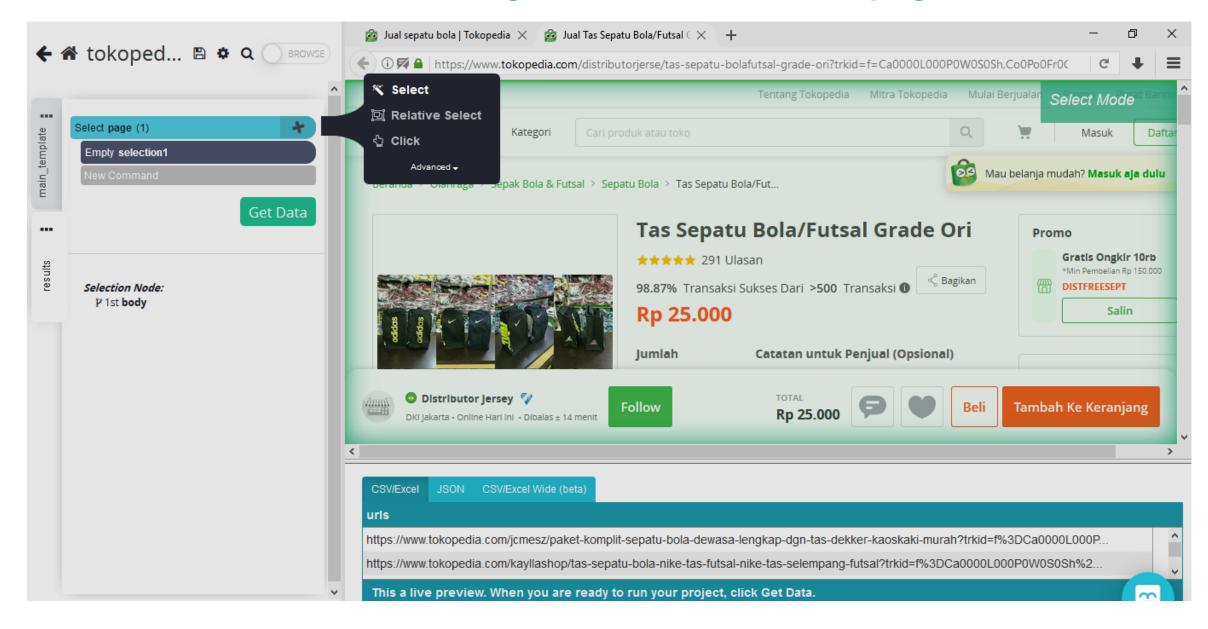
Click on "Create New Template". You should now be taken to the first url in your JSON list and a new template should be created for you.

Go To Template	×
Make ParseHub start executing a template. You can either execute the template on the same page or go to a URL. URLs are expressions, so you can combine them out of multiple parts.	
○ Stay on the Same Page	
Go to URL: item	
☐ Go to Existing Template main_template 😊	
Go to Template from Another Project	
Create New Template results	
Create New Template	

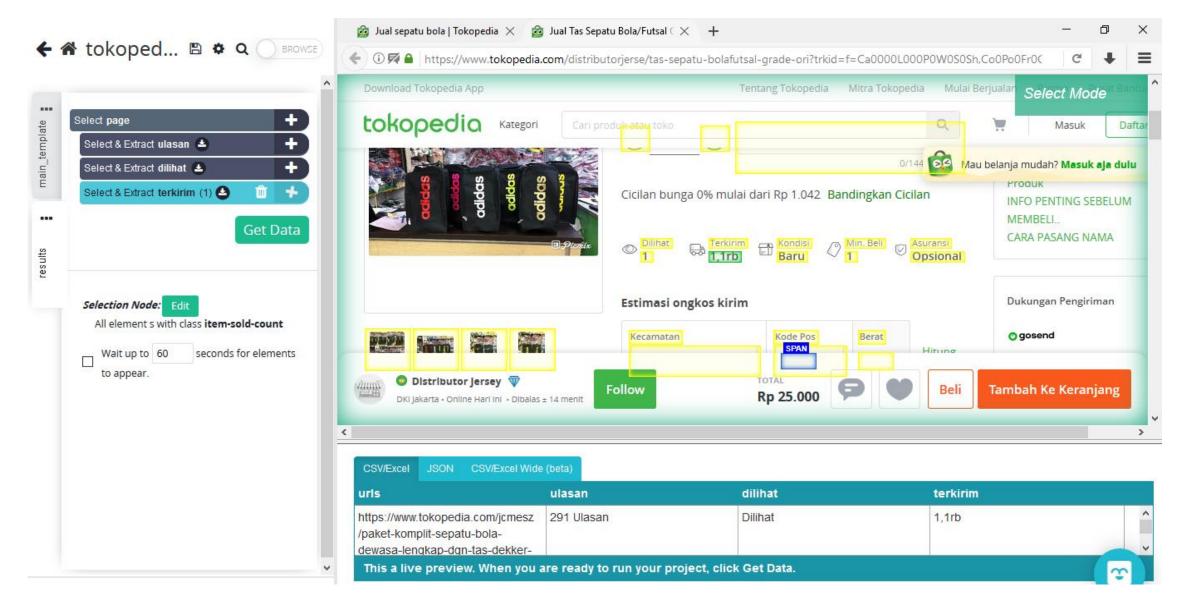
On this new template continue making commands that will be applied to each of the urls in your list in turn.



Click on the + button on the right side of the "select page (1)" command.

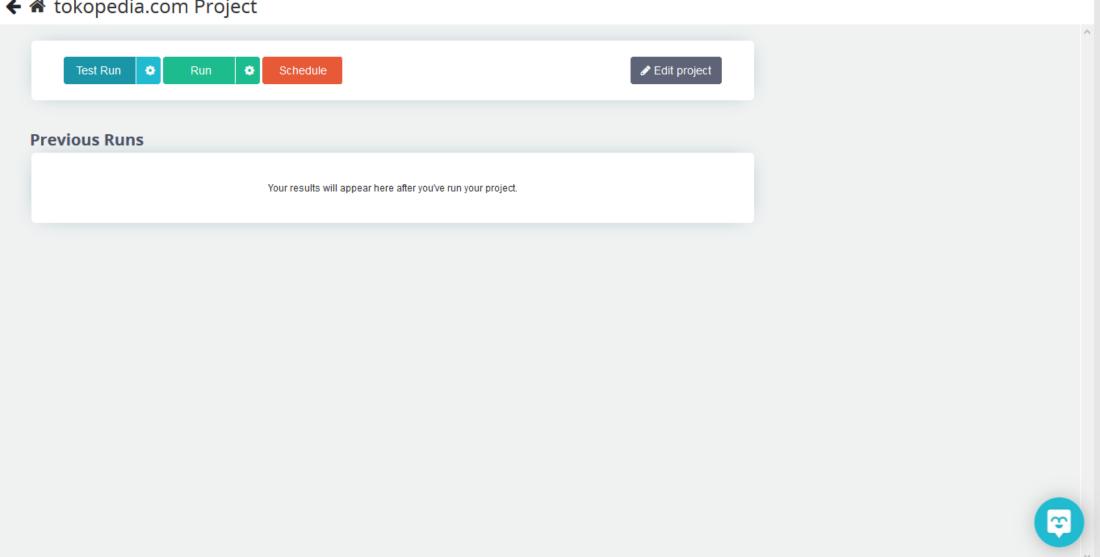


Select the data you want to retrieve and this time the number of reviews, the number of views, and the number sent. After that click "Get Data".

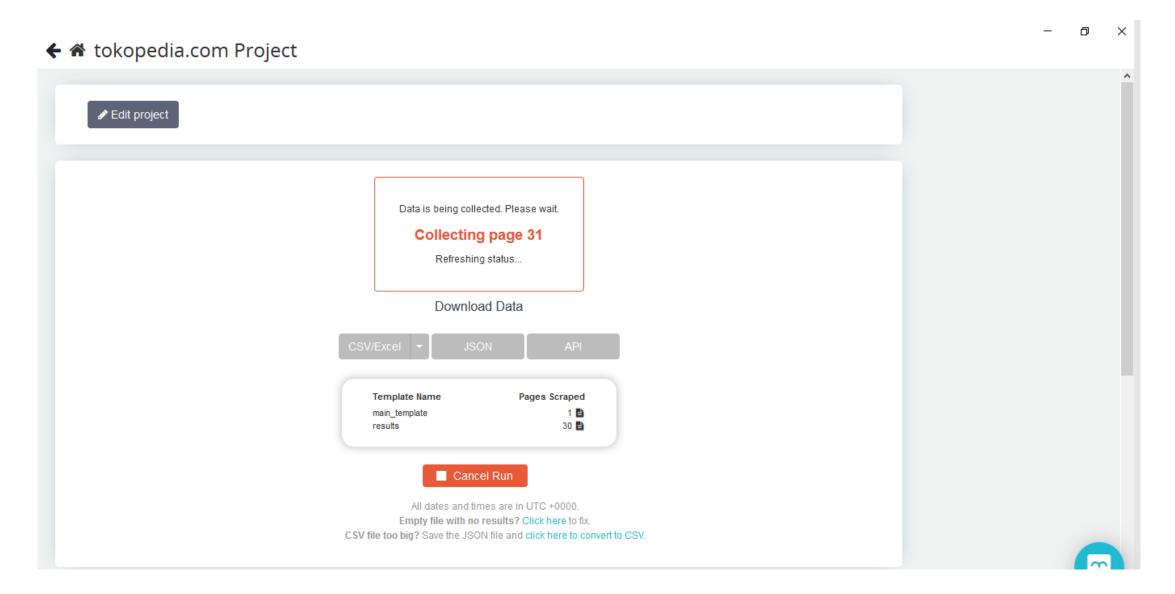


Click "Run"

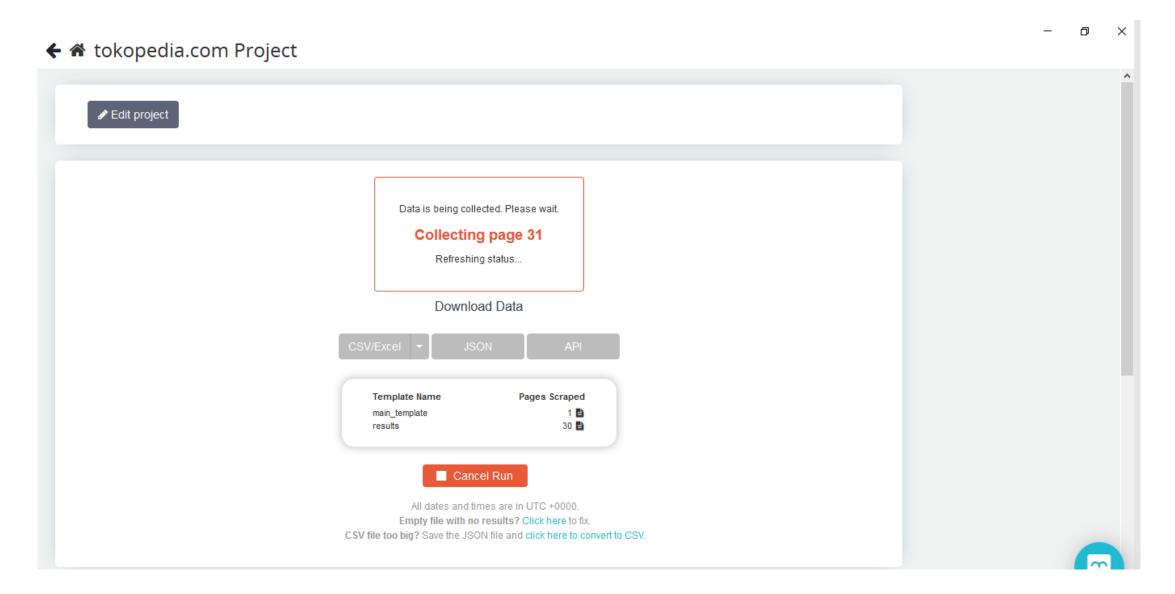
← ☆ tokopedia.com Project



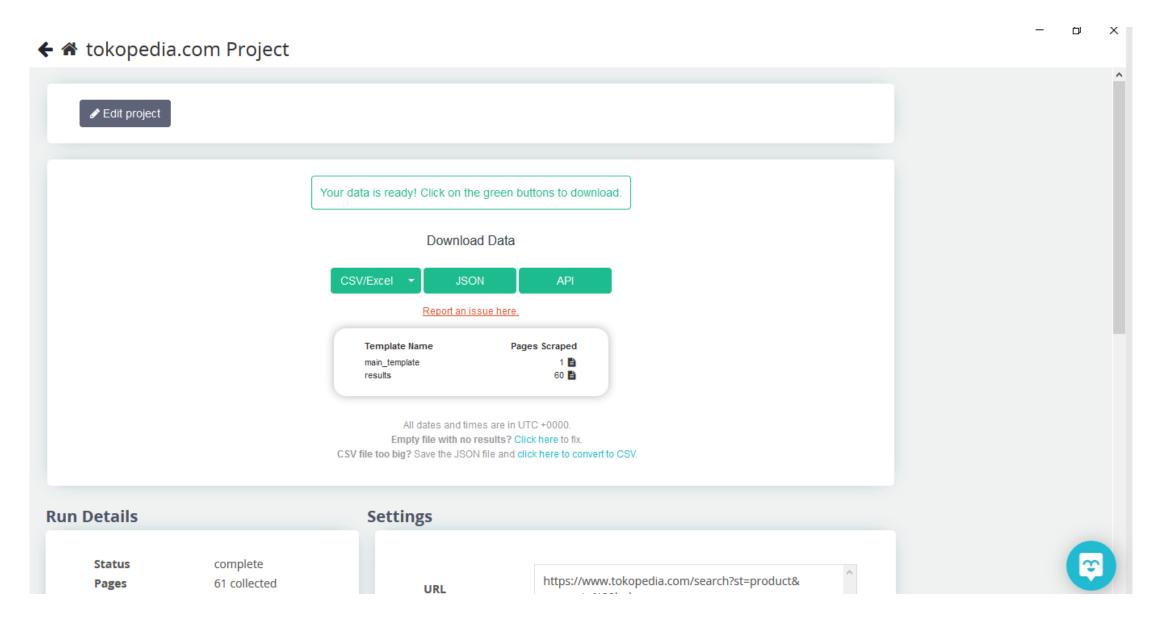
Please wait for a while until the data is ready to be downloaded



Please wait for a while until the data is ready to be downloaded



Please download data according to data type requirements.





THANK YOU