

How to use Data Miner to extract data from websites

The Data Miner tool helps you remove data from webpages and import them into beautifully formatted Excel / CSV files by looking through the text of the pages you have loaded.

If you're copying and pasting everything from web pages and putting them into a spreadsheet manually, you won't know what data scraping is (or web scraping). Maybe you know what they are but don't really care about the idea of learning to code just to save yourself a few hours of clicks.

Either way, there are plenty of code scraping tools that don't need code to help you in such situations, and the Chrome Data Miner extension is one of the most intuitive options. If you're lucky, the task you are performing will be included in the recipe book section of this tool. At that time, you do not even have to go through the steps to build your own formula.

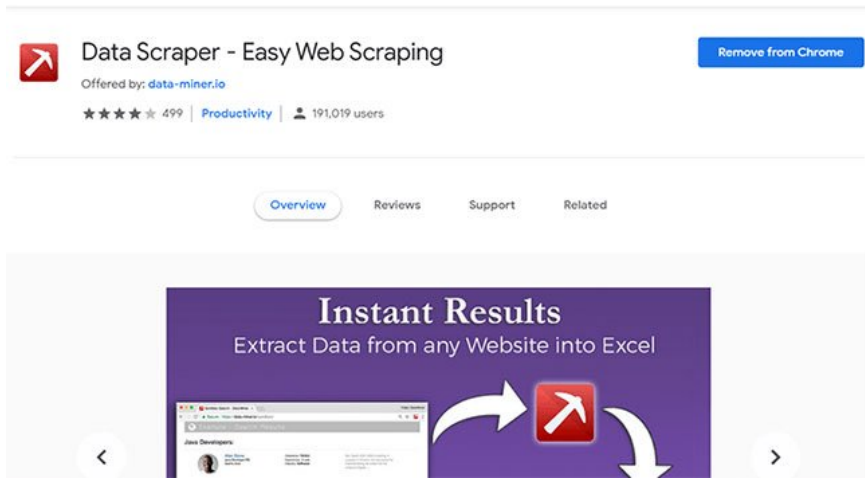
Instructions for using Data Miner to extract data from the web

1. How does Data Miner work?
2. 1. Set up Data Miner
3. 2. Load data
4. 3. Check the recipe
5. 4. Page type
6. 5. Create a row
7. 6. Divide the data into columns
8. 7. Tell the Data Miner how to get to the next page
9. 8. Tell the Data Miner where to click or scroll to load the data
10. 9. Save and run the formula
11. If there is a problem, is there an easier way?

How does Data Miner work?

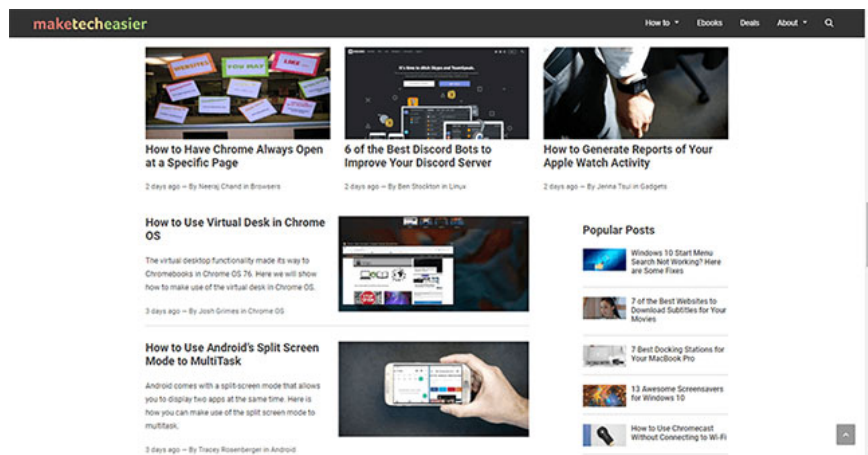
The Data Miner tool helps you remove data from webpages and import them into beautifully formatted Excel / CSV files by looking through the text of the pages you have loaded. That means you will at least need to be comfortable working with HTML to recognize some patterns, but not too broad knowledge. Advanced HTML and / or JavaScript skills will definitely help some tasks, but are not necessary for most things. You should also have at least basic spreadsheet skills so you can make sure your output is neat and organized.

1. Set up Data Miner



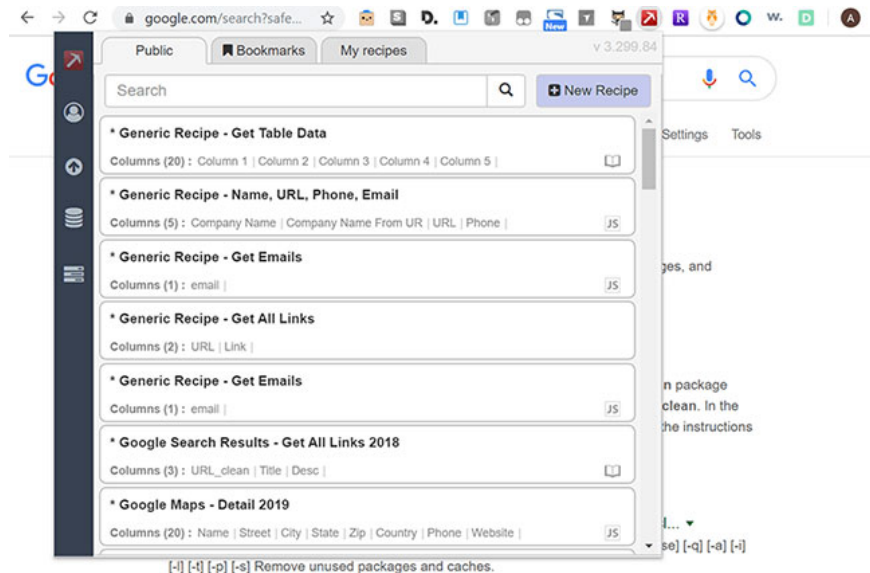
Use Chrome or another Chromium browser, then install the extension. The icon of the extension will appear on the toolbar. Clicking on it will take you to a page where you can set up an account. The free version gives you 500 scrapes (data extraction) per month, which is probably sufficient, unless this is what you do every day.

2. Load data



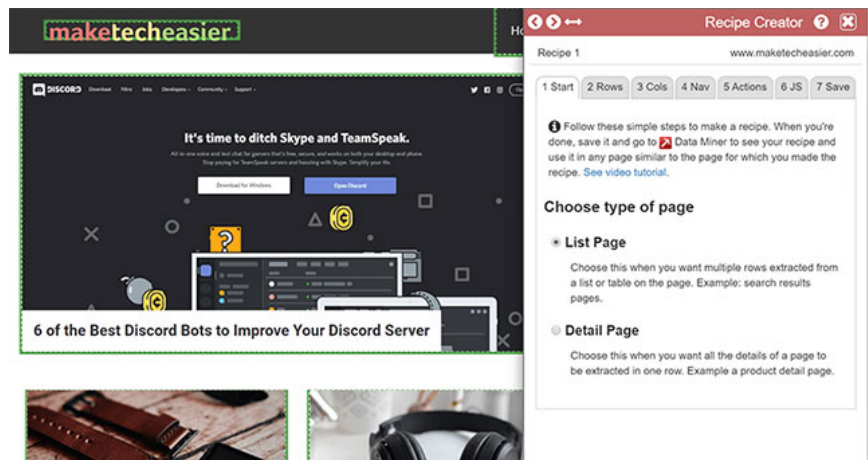
First, navigate to the page you want to extract the data from. If you have a lot of data pages or some data hidden behind buttons, that's fine - there are ways to deal with this situation. For now, you'll only need a representative form so the program knows what to look for.

3. Check the recipe



Next, open the Data Miner and check the **'Public'** tab for available formulas. If you are on a popular website, someone else may have created the process to retrieve the data you are looking for. This will save you a lot of time. For example, websites like Google, Amazon and Twitter have lots of formulas available to help you instantly download links, prices, text and other data. You can check the formulas by clicking the **Run** button to see a preview of the worksheet created by the Data Miner. You can also adjust existing formulas to suit your needs, by clicking the **Edit** button .

4. Page type



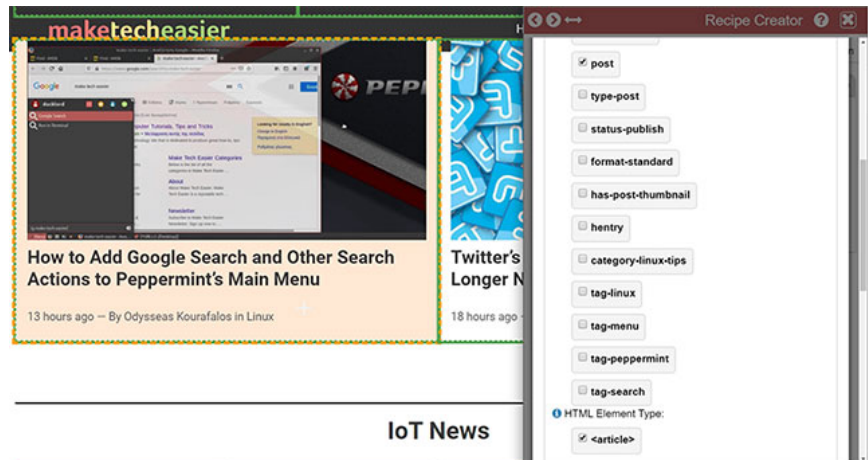
If none of the formulas is right for you, that's okay, you can create your own. Just click the **'New Recipe'** button to get started.

Your first choice will be either **'List Page'** or **'Detail Page'** .

1. Select **'ListPage'** if you are trying to retrieve multiple rows of data on a page. For example, you might want to download the link and page title of each search result or get the date and content of the post in the feed. This is probably the most common type and will be used as a model in this article. (The steps for the **'Detail Page'** are basically the same.)

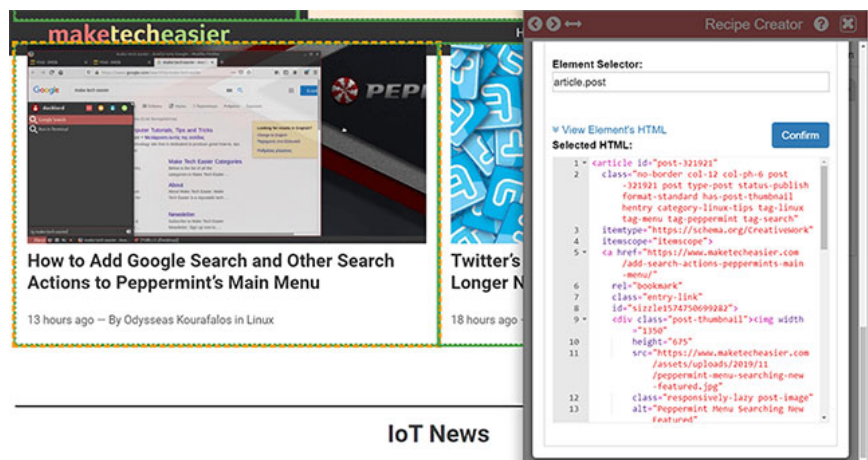
2. Select **'Detail Page'** if you have a lot of different information about something on the page - for example, the product page, where you need to get its price, description, link and rating, then put it all in. a row.

5. Create a row



Click the **'Find'** button and move the mouse until the yellow check box includes all the data you need for the final worksheet entry. For example, if you download the search results, you will need to highlight an area large enough to include the title, URL and description, each of which can be placed in separate columns in the next step. To make a selection, press the **Shift** key. Don't worry if you accidentally let go. Data Miner saves all processes even when you navigate away from the page.

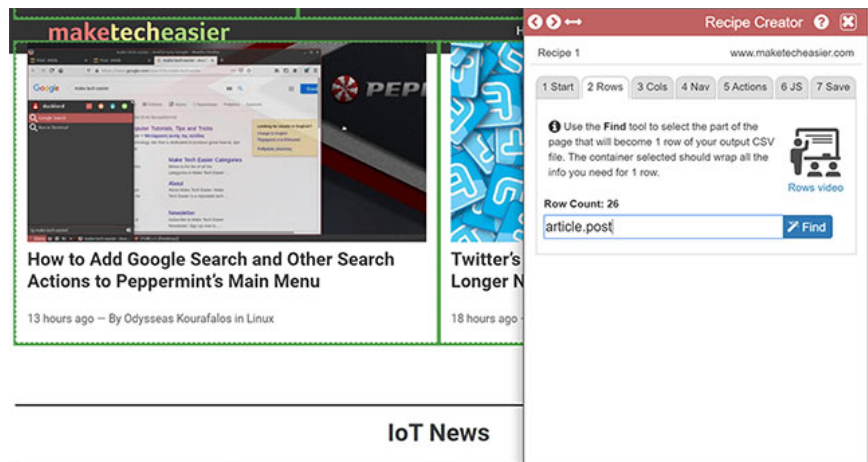
You will then want to select at least one of the boxes under **'Element's Classes'** or **'HTML Element Type'**. Ideally, you should see the selection cover all elements on the page that are of the same type as the one you selected.



If you find that the selector doesn't include everything you need, try selecting just one of the elements and clicking **'Select Parent'**. This will make the box bigger and be able to capture everything you need. If not, you may need to dig into the HTML a bit and identify the classes and element types you need. When in doubt, click **'Select Parent'** until the box is as big as possible without including multiple list entries, as this will give you

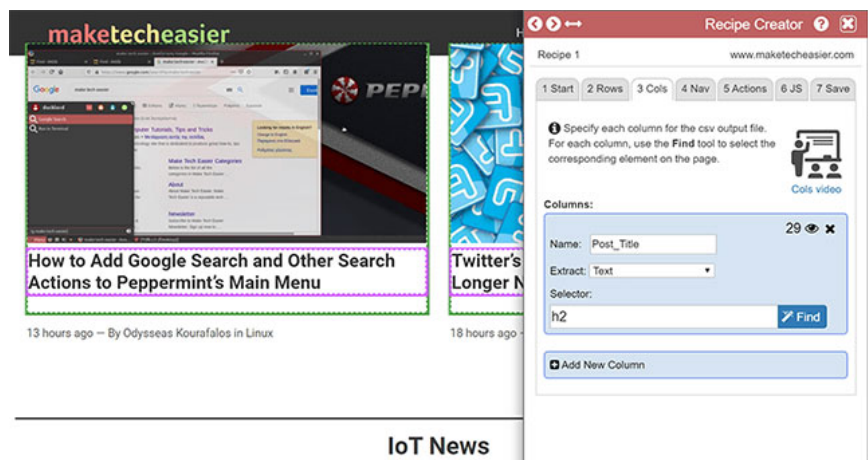
more flexibility when selecting columns.

Data Miner gives you the **'View Element's HTML'** option at the bottom and also allows you to enter custom selectors. If you want to get all the links on a page with class **'product'**, you can simply enter a `.product`. This is where some basic HTML / CSS knowledge will really help.

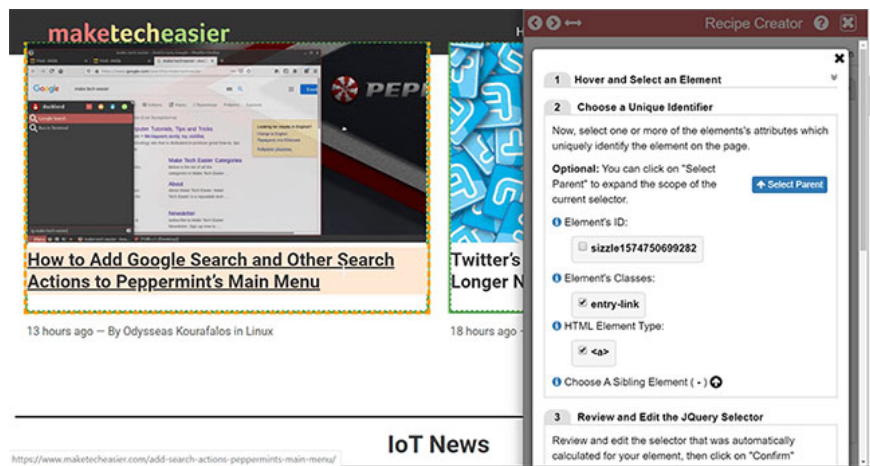


When you return to the main row menu, you'll see the **'Row Count'** option with the number of items your formula will create in a spreadsheet. If it does not include everything, you will need to double check your item selection.

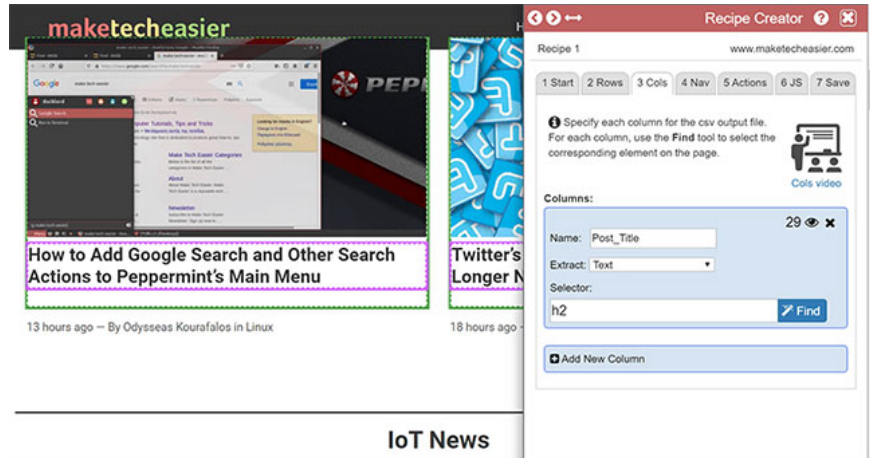
6. Divide the data into columns



Once you've selected all the data for your rows, it's time to make everything look good by dividing it into different types of columns. Each selection made here must be a subsection of the box you selected for your row.

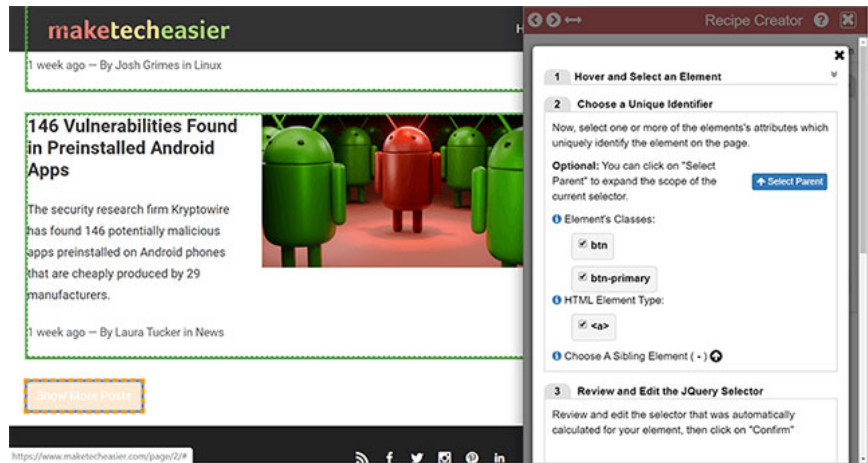


To create a column, just enter a name for it and use the **Find** button to select what you want to extract, just as you did for the rows. The most common data likely will be text, URL or image URL. Getting the URL by hovering over the text links can be a bit tricky. You may have to click on '**Select Parent**' until you reach the level in which **Element Type** is, which is the HTML tag for the link.

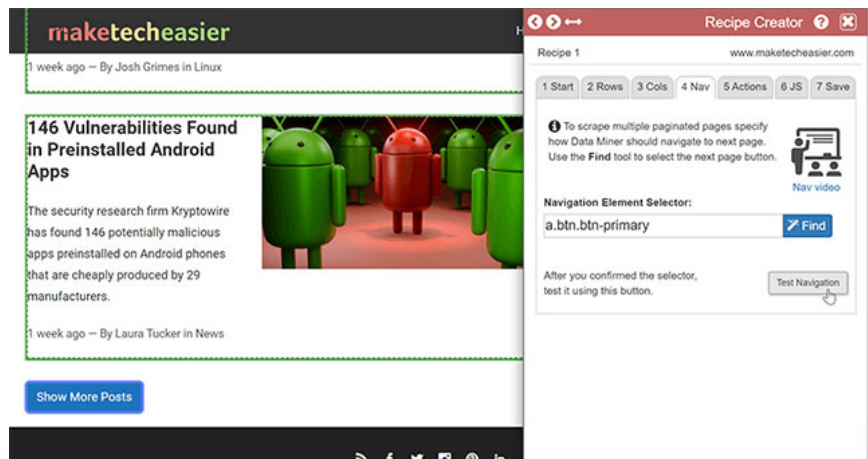


To ensure you have the right type of data in your column, simply click the eye icon to the right of each column name, next to the number indicating how many columns you have selected. This will show you a preview of each row for that column. If something goes wrong, go back to correct the tags and categories you have chosen to identify the rows. Don't be afraid to open the HTML viewer and check out the patterns related to the data you are trying to get.

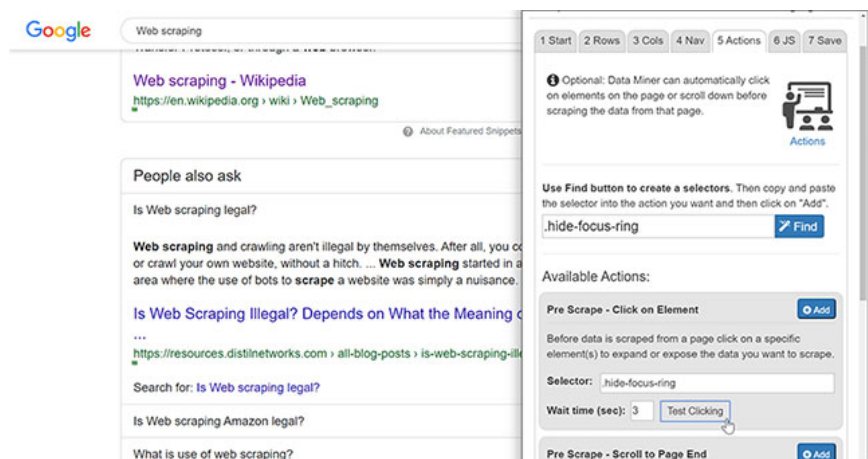
7. Tell the Data Miner how to get to the next page



If you have multiple data pages to extract, you probably don't want to click through each page and run your formula again. To solve that problem, just tell the Data Miner where to find the navigation button it needs to click to go to the next page. Be careful not to tell it to click on something like 'Page 2', because then it will only go to **Page 2**. Again, make sure you are selecting an element and use the **Test Navigation** button to make sure it works.



8. Tell the Data Miner where to click or scroll to load the data

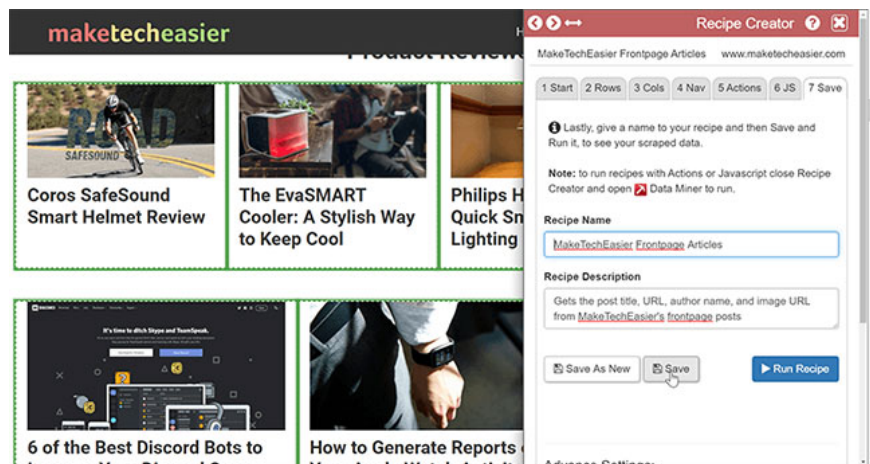


Some pages do not load data until you click on something or scroll down. Fortunately, Data Miner can do these things too! Use the **Find** tool at the top to select the element you need to manipulate, then place the selector in the appropriate box and check to make sure it works.

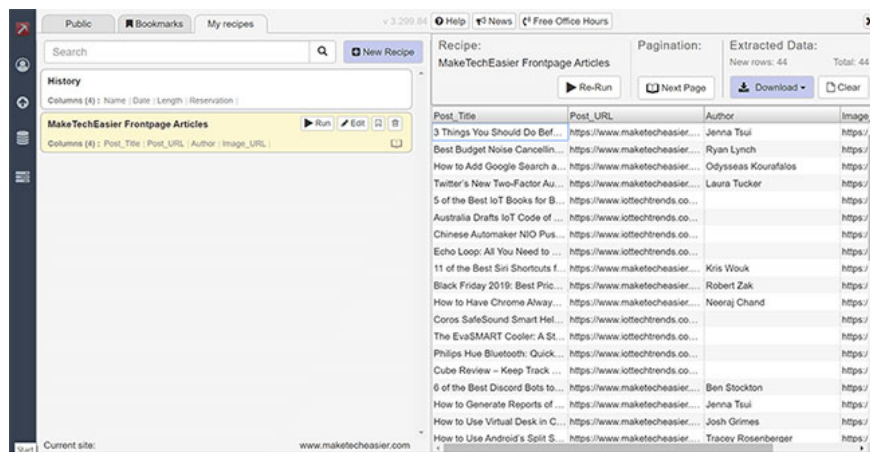
Finding out exactly which selector will trigger an element or infinite scroll bar can be difficult, but basic HTML knowledge, as well as some trial and error will be helpful here. Most of the things you'll need to manipulate here are JavaScript based, but the Data Miner only needs to know the CSS selector associated with the action to activate it, so you don't have to mess around with any code. in most cases.

The next step also allows you to add custom JS to do whatever you want, but that's quite advanced and far exceeds what is needed to extract basic data.

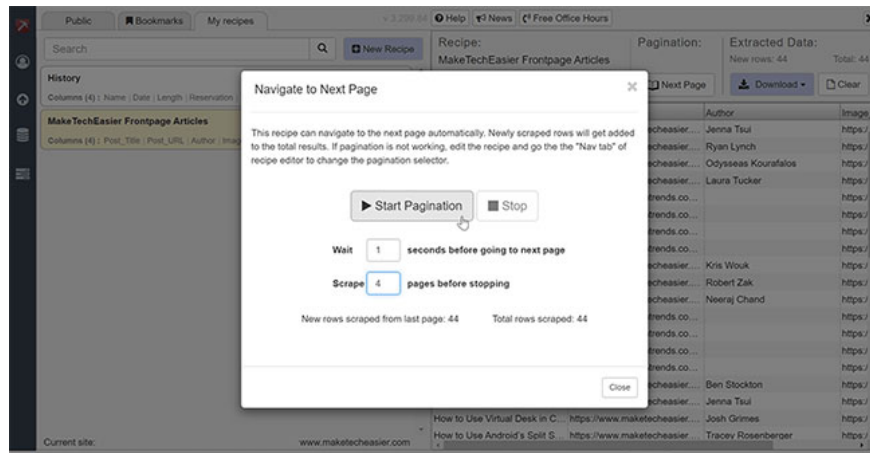
9. Save and run the formula



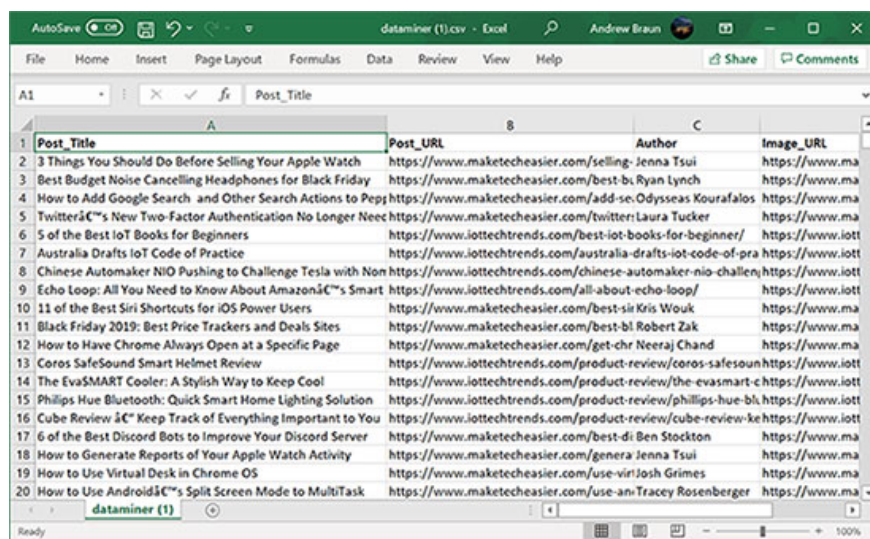
Congratulations! Now it's time to see if all combined properly. Run the formula on the page where you are logged in, then check the preview to see if the rows and columns are doing everything as intended. If not, you can go back and edit the formula.



If everything works as usual, you can use the 'Next Page' button to tell the Data Miner how many pages to crawl and how fast it is performing (Doing so quickly can cause the system to flag you. is bot).



After you have all the data you need, you can choose the file format you want to use for download.



If there is a problem, is there an easier way?

If the Data Miner program is not right for you, there are many other data scraping tools available such as ParseHub, Scraper, Octoparse, Import.io, VisualScrapper, etc. Some of these tools may have an online interface. This is a bit more intuitive and automated, but you'll still need to know at least a little bit about HTML and how websites are organized.

What makes Data Miner particularly good for beginners is the library of recipes from the community, capable of helping you avoid even the slightest 'encounter' with the code. That, combined with the generous monthly free scrape package, makes Data Miner a very good tool for almost any need.

Hope you are successful.

You finished reading the article "[How to use Data Miner to extract data from websites](#)" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.