

Finding & Getting Data - Advanced Search with Google

In this video I'm going to show you how to find data on the web using different Google tools, and a great tool to start is Google.

So when you go to google.com there are certain commands that might make it easier for you to find files on the Web, especially spreadsheet files in this case. And what you do is just type "filetype" and then the extension of the spreadsheet file which is "xls" or "xlsx". These spreadsheet files usually, usually are made for humans so they're not machine readable most of the times but those are files that might be useful in your data journalism work stream because you might be able to clean those files and then prepare it to then make them for some kind of visualization or some kind of analysis. The important thing is that you can use Google as a search engine for files. And you do that when you type in "filetype" and then the extension of the file that you want. Not all of the extensions are compatible and I'm going to show you in a second the list of files that are compatible with this command.

But then if we type in here for example "education data" and then hit enter, Google will try to list the files that it thinks it's most useful for this query but those are not web pages. Those are actually files, so if you click on any of these links you're not gonna be directed to a web page but you're gonna be directed to a file. So, if you click on this one then the download window shows up and then you can download it and then open on your application of choice to process spreadsheets. I can also show you how to import Excel files to a Google Sheet.

The other thing that you might want to do using the file type command in Google that becomes more powerful is to use this site command, because the file type might point to, just a file type might point to multiple websites on the web but maybe you want to find which spreadsheet files are available on a given website, a specific website. Maybe it's a government website, maybe it's an official source but you want to see which files are available because the download section might be hidden or it might be hard to find where the files are scattered throughout and the search function on the website doesn't work.

On those cases, you can use the command site. So if you have a specific site in mind, so for example, World Bank, we might want to replace here so we use "filetype" and then the extension of the file here. We're using the extension for spreadsheets, then we use "site" and then we put "worldbank.org". So what's gonna happen now is that Google will look for spreadsheet files on the World Bank domain, worldbank.org. And then, if we select a topic that we might want to search about, for example "education" and hit enter then Google will give us a list of files about education or that has the term "education" in it. There are spreadsheet files but only on the domain worldbank.org. So that's how you might want to use Google to find files on the web sources or of databases or files that were prepared as spreadsheets for your investigation, and combined with the site command you make it even more powerful.

Let me show you how to find the list of files that are compatible with the file type command at Google. So if you go to the homepage, on the right lower corner you'll see the "Settings" option and then you go to "Advanced Search" and then there is a form where it helps you with all of the available commands to use at Google.

So, down here you find the file type command and then it says any format but there's actually a list of files that you want that you might want to try different from, you know, spreadsheet files. You could try PDA files or KML files, or even presentation files or documents. Right. So that's how you do it. You go to the right lower corner "settings", "advanced search" and then you go here to this form.