

The Taming of Google

Maxwell Memorial Library's Technology Class
Thursday, September 27, 2018

0 Introduction

The rumors of the death of the reference librarian have been greatly exaggerated.

Just as electronic books have killed neither paper books nor the libraries that provide them, neither has Google made the specialized skills of a good reference librarian obsolete. As Neil Gaiman told us, “Google can bring you back, you know, a hundred thousand answers. A librarian can bring you back the right one.”

There are ways, though, to tame the beast. This month's tech program at Maxwell will introduce you to some of the ways you can tweak your online searches to get results that better answer your actual questions. You won't come out with *all* of a reference librarian's skills, but your Google searches will be more productive.

0.1 A Note about the typographic conventions in this handout

Because quotation marks are among the special characters I'll be talking about, I won't use them to set off text that you'll type into Google's search box. Instead, I'll use **bold face** for those exact strings. So in plain face text, you can interpret quotation marks (and other punctuation) normally, but all characters in bold face (including quotation marks and other punctuation) are characters you would type in.

I will also be using the computer-world convention of placing quotation marks only around text that is explicitly part of the string I mean. In particular, this means I will be violating the American typographic convention that requires that closing punctuation be placed inside the quotation marks whether that punctuation is part of the quotation or not. In other words, if I include closing punctuation inside the quotation marks, type it, and if I don't, don't. This convention isn't as important with Google (and other search engines) as it is in many parts of the computer world since search engines tend to downplay punctuation. However, I don't think they ignore it entirely as varying punctuation can at least affect the ballpark estimate of matching pages that Google gives you.

1 Search-box tricks

1.1 Site-specific searching

Very often, you will know what website has the information that you want or you will want to restrict your search to a type of institution (say only educational or governmental institutions for a scientific or medical question) to increase the quality of the hits you get. To do this use **site:**. For example,

- **Pertussis site:edu** returns hits on pertussis but only from institutions in the .edu top-level domain (so colleges and universities, mostly).
- **Pertussis site:gov** is similar, but the sites are governmental (.gov).
- **site:help.getpocket.com crash** gets pages about crashes from the Help site for the Pocket application (potentially useful in figuring out why the application crashes on me).
- **site:utexas.edu ladusaw** retrieves pages at the University of Texas at Austin mentioning Bill Ladusaw, a linguist who got his Ph.D. there.

1.2 Use double quotation marks for phrase searching

It doesn't seem to matter whether you use straight quotes or “educated” (curly) quotes, but they do need to be double quotes (") rather than single quotes (').

- **"Fred Jones"** or **“Fred Jones”** will both yield only results containing “Fred Jones” as a phrase.¹
- **'Fred Jones'** or **‘Fred Jones’** (using single quotes) will yield the same results as **Fred Jones** with no quotation marks at all. That is, pages that seem relevant to someone looking for pages about “Fred” and “Jones” but who isn't insisting that either term must be in the pages.

1.3 Use double quotes to require individual words

This tip is really just a special case of the previous one. The difference is that the phrases in question just consist of only 1 word.

- **“Fred Jones” “Shaggy”** will return pages, each of which contains both “Fred Jones” as a phrase and “Shaggy” as a word.

1.4 Use an asterisk within a quoted string as a wildcard

In a quoted string, you can use * as a place holder with the rough meaning of “any string of words whatsoever or even none at all”. This can be helpful if you almost know the exact string you want to search for but can't recall some of the words in the middle. It can also be useful for searching for pages that contain a certain pattern as a template of sorts.

¹ Remember, in keeping with my convention of plain face text having its usual interpretation, the hits Google returns may or may not have the quotation marks, but they will have Shaggy and Scooby's sidekick (or someone with the same name).

- **"In a hole in the * in a land"** gets pages with phrases that start with "In a hole in the", possibly follow with some undetermined text, and then "end" with "in the land".² For example, it should find pages with:
 - "In a hole in the *ground* in a land called the Shire".
 - "Not in a hole in the *ground but* in a land where once they dwelt-the Beorningas".

I've italicized the portions of the strings that the asterisk ended up standing in for. Note that content in the string coming before "in a hole" or after "in a land" gets a free ride since Google is only looking at whether a page matches your query and is not (here) filtering any of the matches out.

1.5 Use a minus sign to exclude words or phrases

This is a useful trick when you are looking for something that might be obscure relative to something else that dominates searches under the terms you have.

For example, you heard of some older event or program or something that went under the name "Python", but it was neither the Python computer programming language, the Monty Python acting troupe, nor the snake. Basically, the word "python" is all you have to go on.

When I searched Google with **python** just now, all but 2 hits on the first 3 pages (so 28 of 30 results) were directly about the Python programming language, and 1 of those 2 exceptions was a webcomic episode with a plot that featured the language.

By searching on **python -interpreter -machine -program -programming -programmer -developer -development -software**,³ though, I get hits for PYTHON ("a Cold War contingency plan of the British Government for the continuity of government in the event of nuclear war"), Python motorbikes, Python Products aquarium supplies, and the entry for python on wiktionary (which mentions the Python of Delphi, in case that was the thing I was actually looking for) all on the first page. There was still a lot about the programming language, but the trick did thin that a bit so that other uses could come through.

Now this example was somewhat contrived, but you will find times where your search for an obscure sense of a term returns a lot of noise from a culturally prominent sense, and that's where using this trick can help.

2 I have placed scare quotes around "end" since, of course, the text in the page could continue without any break, but "in a land" is the end of the text we're focussing on.

3 As you've probably guessed, I did not start with this string. I started with just **python** and gradually added terms I wanted to exclude because those terms seemed to occur in many of the hits about the programming language. Had there still been only pages about snakes or British comedy, I was prepared to add **-snake -monty**.

1.6 Use OR to join words when only 1 of them needs to be present

When you don't restrict Google, it seems to be pretty good about finding pages related to your search even if they don't contain the exact text you searched on. So, for example, if you search for **venetian blind**, some pages will come up that don't contain the word “venetian” but are still relevant to a search for blinds with horizontal slats.

However, Google's ability in this regard is limited. Moreover, you might have some terms in mind that are related in the domain of knowledge you're exploring but not in general. Or you might want to make sure that the top results talk about at least some of the terms but are ranked independently of which terms they use.

This is where using the conjunction **OR** (it is important that it is in all-caps) comes in. The **OR** tells Google that you want results that relate to any of the terms being connected by **ORs**, so your results will be a mix of the results that you'd get by searching each term by itself.

- **Platonism OR essentialism OR idealism** gives links to pages talking about each of the 3 philosophical outlooks (as well as extraneous pages on both *Essentialism: The Disciplined Pursuit of Less* by Greg McKeown⁴ and idealism in the sense of fighting for one's ideals).

1.7 Ranges of numbers or units

Searching for text with individual numbers (more precisely, numerals) works the same as for any other characters or words, but what if you want a range of numbers? For instance, pages talking about newborns between 9 and 11 pounds? To specify a range of numbers, type 2 periods between the beginning and end numbers. You can also include a unit of measure:

- **Newborn 9..11 lb** for pages with information about newborns between 9 and 11 pounds;
- **Clocks \$12..\$22** for clocks where a range of \$12 to \$22 is relevant;⁵
- **Roads 12..75 mi** for roads having some association with 12;
- **Poland 1700..1800** for pages that mention both Poland and some number between 1700 and 1800.⁶

4 This shows us a place where excluding terms by using the minus sign might be helpful. If pages about the book were too frequent and that made it hard to get at the pages about philosophy, we could add **-McKeown** to our search string.

5 Note that the **\$** has to be on both numbers.

6 The way I've formulated the explanation in the Poland example (“pages that mention both x and y”) is actually all you're promised in a Google search. The explanations for the other 3 examples would be false advertising if expressed as guarantees, but they might well express what a person using those searches hoped to find.

2 Options below the search box

In addition to the tricks shown in the previous section that help you tweak what you type in the search-box, Google has some controls below the box that let you filter the results.

Starting from the left-hand side, you see links labeled “All” (this is probably underlined or otherwise highlighted for you), “Images”, “Books”, “News”, and “Videos”. These filter your search in the way you might expect: clicking on “Images” finds pages with images that are related to your search and so on. “More” shows a drop-down menu with 4 more filters of the same sort: “Maps”, “Shopping”, “Flights”, and “Finance”.

On the right-hand side, you see “Settings” and “Tools”. Tools are ways to further refine your search, and the particular tools offered will depend on which left-hand side filter you’ve chosen, so under “All” the tools let you specify a time period for when the page was published and whether to do a verbatim search (the same as using double quotes). The tools for “Images” are “Size”, “Color”, “Usage rights”, “Type”, “Time” (that is, when the page was published), and whether to display sizes on the thumbnails.

The Settings menu is more about setting your preferences for general searches on Google rather than for tweaking individual searches. There are, however, 3 options here that can be useful beyond the “set it and forget it” of settings:

1. “Search help” brings you to a page where you can search for Google’s pages on using Google itself.
2. “History” lets you look at the searches you’ve conducted with Google (if you’ve enabled Google to save your search history). This can be useful if you want to revisit a search that you made previously, but it does mean giving some privacy up to Google, and there are ways to accomplish much the same thing within the browser on your machine.
3. “Advanced search” is treated in the next section.

3 Advanced search

If you go to https://www.google.com/advanced_search or go to <https://www.google.com/> and select “Advanced search” under the Settings drop-down menu just beneath the right-hand side of the search box, you’ll see several tools for making your searches more precise. The fields under the first heading (“Find pages with...”) should look familiar since they appeared in the first section. The tools under the second heading (“Then narrow your results by...”) give you a convenient means to narrow results. The options and the explanations Google provides appear in the table below.

Option	Use this to...
language	Find pages in the language you select.
region	Find pages published in a particular region.
last update	Find pages updated within the time you specify.
site or domain ⁷	Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov
terms appearing	Search for terms in the whole page, page title, or web address, or links to the page you're looking for.
SafeSearch	Tell SafeSearch whether to filter sexually explicit content.
file type	Find pages in the format you prefer.
usage rights	Find pages you are free to use yourself.

⁷ This does exactly the same thing as the trick in subsection 1.1.