

Western Sullivan Public Library

Improving Google Search



Improving Google Search

The hugely popular search engine Google is a great tool for finding content on the World Wide Web. But often, searches can return millions of results. This course will focus on better understanding Google, and in so doing, learning how to search more effectively.

During this course we will look at:




- What is Google?
- Googlebot
- What does Googlebot see?
- PageRank
- AutoComplete
- Search Tips
- Searching for More than Websites
- Exercise



Improving Google Search

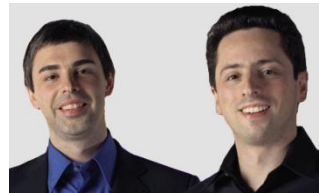
Western Sullivan Public Library
Public Computer Center

What is Google ?

Google is an internet search engine at the web address www.Google.com. It allows users to search through the content of billions of websites that comprise the World Wide Web by typing in key words or terms and delivers a list of results ranked in such a way to ensure that the best, most reliable, or relevant information is returned. But it is not the only search engine out there! Dozens of others exist, including popular search engines like , , and .

In 1990, when the World Wide Web was very young and consisted of very few websites, the first tool for searching the Internet was created. Called [Archie](#) it simply cataloged every document title on the web, and made it a searchable index. It did not look at the documents themselves, only the titles, since the documents were few enough that they could all be looked through manually.

In 1994, one of the first "full text" crawler-based search engine was introduced. **WebCrawler**[®] (still exists at webcrawler.com), unlike its predecessors, let users search for any word on a webpage. In the following years, many additional search engines debuted.



In 1998, two Stanford University students, Larry Page and Sergey Brin, created Google. It quickly became the dominant search engine, and eventually, the most visited website on the World Wide Web.

Its success was based largely on the fact that it was able to deliver the most relevant results to users, while other search engines often gave results that required much picking through in order to find the best information. This was made possible by Google's "PageRank" system.

Googlebot

Googlebot is Google's web-crawling bot (sometimes also called a "spider"), which "crawls" from website to website discovering new and updated pages to be added to Google's index of websites and searchable content.



Improving Google Search

Western Sullivan Public Library
Public Computer Center

Google uses a huge set of computers to crawl billions of pages on the web. Googlebot uses an algorithmic process: computer programs determine which sites to crawl, how often, and how many pages to fetch from each site.

The crawl process begins with a list of webpage URLs, generated from previous crawl processes. Usually, this would include the most popular websites on the internet. As Googlebot visits each of these websites it detects links to other websites on them and adds them to its list of pages to crawl. New sites and changes to existing sites are noted and used to update the Google index.

What does Googlebot See?

Googlebot can only see text in the website's code. As an example, here is a simple web page, displayed as the code that comprises it, as well as how it appears in a browser.

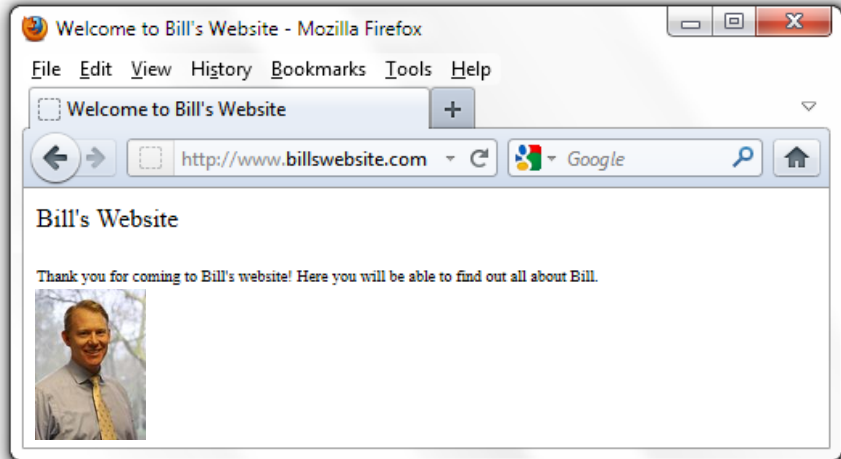
```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0
Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-
transitional.dtd">
<html lang="en" dir="ltr" class="client-nojs"
xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Welcome to Bill's Website</title>
</head>
<body>
<font size="3">Bill's Website</font>
<br><br>
<font size="1">Thank you for coming to Bill's website!
Here you will be able to find out all about
Bill.</font>

</body>
</html>
```



Improving Google Search

Western Sullivan Public Library
Public Computer Center



Notice how all of the text on the website is still there in the code, it is just in between bits of formatting code. Notice also how the image itself isn't visible in the code. As a result, Googlebot cannot see images!

PageRank

When someone performs a search on Google for "Maytag Washing Machine", Google will find hundreds of thousands of websites it has indexed with those terms. So how does it know which of the hundreds of thousands of results should be displayed at the top of the list?

To do this, Google uses a system to rate which websites are most reliable, relevant, or popular, in which it assigns every website a "PageRank" from 0-10. So if something is searched for, and there are two results, one with a PageRank of 10 and one with a PageRank of 7, the site with the PageRank of 10 will be displayed first, then the one with the PageRank of 7.

PageRank is determined by the links to a web site. Each time somebody adds a link to a web site, Google interprets this as a vote for that site. The more links a site has, the more votes.

But Google also looks a little deeper than just sheer volume of links and analyses the importance





Improving Google Search

Western Sullivan Public Library
Public Computer Center

of the web site that is linking to the other site. Sites that Google determines are important are those with a higher PageRank. So a link to a site from another site with a PageRank of 6 is better than a link from a site with a PageRank of 3. In fact, 1 link from a site with a PageRank of 6 is better than 10 links from sites with a PageRank of 3.



When Google is determining how important the link to a site is, it also checks how many other links are on the web page. So if a site with a PageRank of 6 has 1000 links on a page Google will determine that the site's 'vote' for any one of those web sites is only worth 1/1000 of the PageRank 6 value. If there were only 3 other links on that page, the 'vote' would be interpreted by Google as much more important.

Search Tips

Every word matters. Generally, all the words you put in the query will be used. Search is always case insensitive. A search for [new york times] is the same as a search for [New York Times].

Punctuation is usually ignored, including @#\$%^&*()[]\ and other special characters.

Think about how the page you are looking for will be written. A search engine is not human. It is a program that matches the words you give to pages on the web.

Use the words that are most likely to appear on the page. For example, instead of saying [my head hurts], say [headache], because that's the term a medical page will use. The query [in what country are bats considered an omen of good luck?] is very clear to a person, but the document that gives the answer may not have those words. Instead, use the query [bats are considered good luck in] or even just [bats good luck], because that is probably what the right page will say.

Describe what you need with as few terms as possible. The goal of each word in a query is to focus it further. Since all words are used, each additional word limits the results. If you limit too much, you will miss a lot of useful information. The main advantage to starting with fewer keywords is that, if you don't get



Improving Google Search

Western Sullivan Public Library
Public Computer Center

what you need, the results will likely give you a good indication of what additional words are needed to refine your results on the next search. For example, [weather cancun] is a simple way to find the weather, and it is likely to give better results than the longer [weather report for cancun mexico].

Choose descriptive words. The more unique the word is the more likely you are to get relevant results. Words that are not very descriptive, like 'document,' 'website,' 'company,' or 'info,' are usually not needed. Keep in mind, however, that even if the word has the correct meaning but it is not the one most people use, it may not match the pages you need. For example, [celebrity ringtones] is more descriptive and specific than [celebrity sounds].

Phrase search ("")

By putting double quotes around a set of words, you are telling Google to consider the exact words in that exact order without any change. Google already uses the order and the fact that the words are together as a very strong signal and will stray from it only for a good reason, so quotes are usually unnecessary. By insisting on phrase search you might be missing good results accidentally. For example, a search for ["Alexander Bell"] (with quotes) will miss the pages that refer to Alexander G. Bell or Alexander Graham Bell.

Search single word exactly as is ("")

Google employs synonyms automatically, so that it finds pages that mention, for example, childcare for the query [child care] (with a space), or California history for the query [ca history]. But sometimes Google helps out a little too much and gives you a synonym, when you don't really want it. By putting double quotes around a single word, you are telling Google to match that word precisely as you typed it.

Search within a specific website (site:)

Google allows you to specify that your search results must come from a given website. For example, the query [iraq site:nytimes.com] will return pages about Iraq but only from nytimes.com. The simpler queries [iraq nytimes.com] or [iraq New York Times] will usually be just as good, though they might return results from other sites that mention the New York Times. You can also specify a whole class of sites, for example [iraq site:.gov] will return results only from a .gov domain and [iraq site:.iq] will return results only from Iraqi sites.

Terms that must be matched (+)

Using a plus [+] symbol before a word ensures that both words appear in the results. For example, in some instances [George Washington's dog] might return



Improving Google Search

Western Sullivan Public Library
Public Computer Center

results that don't mention a dog, but [George Washington +dog] will ensure that all results have the word dog in them.

Terms you want to exclude (-)

Attaching a minus sign immediately before a word indicates that you do not want pages that contain this word to appear in your results. The minus sign should appear immediately before the word and should be preceded with a space. For example, in the query [anti-virus software], the minus sign is used as a hyphen and will not be interpreted as an exclusion symbol; whereas the query [anti-virus -software] will search for the words 'anti-virus' but exclude references to software because of the space after the word "virus". You can exclude as many words as you want by using the - sign in front of all of them, for example [jaguar -cars -football -os]. The - sign can be used to exclude more than just words. For example, place a hyphen before the 'site:' operator (without a space) to exclude a specific site from your search results.

Fill in the blanks (*)

The *, or wildcard, is a little-known feature that can be very powerful. If you include * within a query, it tells Google to try to treat the star as a placeholder for any unknown term(s) and then find the best matches. For example, the search [Google *] will give you results about many of Google's products (go to next page and next page -- we have many products). The query [Obama voted * on the * bill] will give you stories about different votes on different bills. Note that the * operator works only on whole words, not parts of words.

The OR operator

Google's default behavior is to consider all the words in a search. If you want to specifically allow either one of several words, you can use the OR operator (note that you have to type 'OR' in ALL CAPS). For example, [San Francisco Giants 2004 OR 2005] will give you results about either one of these years, whereas [San Francisco Giants 2004 2005] (without the OR) will show pages that include both years on the same page. The symbol | can be substituted for OR. (The AND operator, by the way, is the default, so it is not needed.)

The Link operator

To search for web pages that link to a URL, use the "link:" operator. For example, to find pages that link to www.google.com, use [link:google.com]. You can also search for links to specific pages or directories, e.g. [link:google.com/webmasters].



Improving Google Search

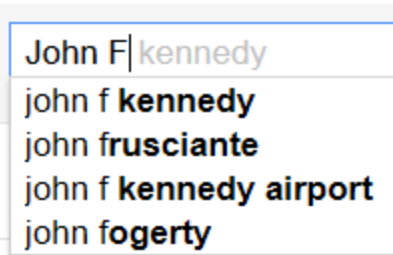
Western Sullivan Public Library
Public Computer Center

Geographic Search

To limit a search to a specific area, simply include "near: [location]" in the query, along with whatever is being searched for. [Location] can be a city and state, a zip code, or an actual address. For example, typing [Pizza near:Monticello, NY] will return a listing of pizza places within the general vicinity of Monticello, New York.

AutoComplete

Google's AutoComplete is the text that will appear beside or under whatever you are typing. It is usually a guess at what you are trying to search for, based on common or popular searches, in order to save you time. If the



AutoComplete does indeed come up with what you meant to type, you can simply click on the AutoComplete text to perform the search. For example, typing in "John F" might bring up the AutoComplete "John F. Kennedy". If this is what you're looking for, click the AutoComplete text rather than continuing to type out the rest of the name.

AutoComplete can be especially useful by showing popular ways of phrasing search terms that are often most effective. For example, if searching for the televised broadcast President Kennedy gave during the Cuban Missile Crisis, typing in [John F Kennedy Cuba] will list, among other things, [john f kennedy cuban missile crisis speech], clearly a popular and effective way to search for that information.

In addition, Google now performs a search based on whatever information you've entered before you're finished. This means that if you set out to search for "George Washington", as soon as you type the "G", search results will appear for the letter "G", then a search for "Ge", and so on. When you get to the "O", you'll see AutoComplete returns "George Washington".

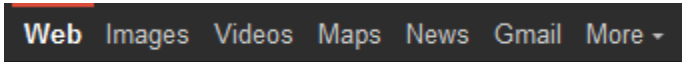
Unfortunately, AutoComplete can be a pest sometimes also. It may be focused on "guessing" what you want to type and for a few seconds or longer, you won't be able to type. The other nuisance is that you cannot turn AutoComplete off!



Improving Google Search

Western Sullivan Public Library
Public Computer Center

Searching for more than Websites



While Google is best known for searching websites, it has also indexed several other types of content on the World Wide Web. Searches can be limited to different content types via the links along the top of the Google website.

Clicking on images and performing a search will yield a list of images. Searching for images of John F. Kennedy will return the following as opposed to documents about him.



All results
By subject

On the left side of the page, you'll see several options for narrowing down your search, including size, color scheme, and more.

Any size
Large
Medium
Icon
Larger than...
Exactly...

Any color
Full color
Black and white

As mentioned earlier, Google cannot really "see" images, so the results are context-based. If it recognizes that a picture is on a website, it will index it by



Improving Google Search

Western Sullivan Public Library
Public Computer Center

words on the same website. So if a website for the Alamo features a picture of one of the Alamo's tour guides, the tour guide's picture might appear when searching for "The Alamo".

Other types of content that can be searched on Google include "Video", "Maps", and "News", but there is much more as well!

Exercise

Either on your own, or in groups, attempt to use some of the search skills you've learned to find answers to the following questions while making note of what tips and techniques you used to find the answers.

1. How much did a gallon of milk cost in 1940?
2. Where is the closest McDonald's to Watonga, Oklahoma?
3. Fill in the blank from George Washington's Farewell Address:
"Interwoven as is the love of liberty with every _____ of your hearts".
4. How many instances are there of the word "snack" on the library's website (www.wsplonline.org)?
5. In what movie involving a dog did Chevy Chase play a part?
6. What is a tom-tom (not to be confused with the gps device)?