1-1-2009

Get More Out of Google Than It Gets Out of You

Maureen Cahill
University of Georgia School of Law Library, mcahill@uga.edu

Get More Out of Google

Than it Gets Out of You

Maureen A. Cahill J.D., M.L.I.S.
Student Services Librarian

The University of Georgia
School of Law
Alexander Campbell King Law Library
(http://www.law.uga.edu/library)
Athens, Georgia

CONTENTS

I. INTRODUCTION

"Every thing is on the Internet." That is the refrain we all hear. Heck, it's the mantra we all repeat–and live by. Whether we are trying to track down an expert witness, find some report issued by a federal agency, pinpoint an historical stock quote, or discover the perfect cabins on a secluded cay off the coast of Belize, we turn to the Internet. And, chances are excellent that we are among the 70% of U.S. Internet users who choose Google as our beginning point to find these nuggets online. The interface is clean and unobtrusive. The process is simple and immediate. The results are excellent? Maybe the results are just good? the best we can expect? better than the next search engine? The primary purpose of this paper is to help you consistently get better results from your Internet searches. As an extra, we will also explore many of the other services and applications that Google offers.

The key to effectively using the Internet as a resource is first to know when to use it at all; and, only then, how to use it. The same may be said for Google, or any search engine. First, master the intricacies of when to use it; then, work on details of how to make it work well. Many of the facts, figures, and documents that we all need or want *are* online. The best way to find a good portion of them is by utilizing a good search engine, such as Google. Nonetheless, not everything *is* on the Internet; and much of what is there cannot be found through a search engine. Consequently, although this is primarily a paper about searching the Internet and optimizing your use of Google, a fair portion of it will be devoted to explaining when not to turn to the Internet at all , and when not to use Google even when you search for online resources.

II. WHAT IS ONLINE? WHAT'S NOT?

Having a good notion of what to look for on the Internet is not a matter of memorizing lists of online documents. Online content changes rapidly. Much is added, and many things disappear, as well. Any list would have a short shelf life. However, it's relatively simple to develop a pretty accurate feel for the likelihood of anything being online. The first trick is to begin thinking about who puts things online, and why.

There are two really easy categories of online content providers, and their reasons for doing so are obvious. The first is government entities, which provide online content to save money and reach more people. The second is folks who want to sell you something (who do it because they want to sell lots of things to lots of people).

The federal government and a handful of states spear-headed the movement toward making government information available online beginning in earnest about ten years ago. Now all states as well as most cities and towns have a web presence. Legal and regulatory information is largely available, though there are still gaps. We can expect the range of online government information to continue to expand. At the same time, only the federal government is making any real push toward archiving older materials. In most instances, states and municipalities have only the most current information available. For instance, the Georgia Supreme Court's web site has the court's decisions only from 2008 and 2009. By contrast, the federal Government Printing Office (GPO) maintains the superceded U. S. Code back to 1994, the Code of Federal Regulations back to 1996, and all volumes of the Federal Register back to 1994. And, while GPO has an ambitious program underway to identify, scan, and post ever more historical and superceded Federal government information, we can probably expect most states and localities to continue maintaining only the most current information online.

If anything, the "people who want to sell you something category" is even easier to remember. Still, when you consider this category, stretch your thinking beyond personal goods and services and remember that information and services for your practice are also likely marketed online. You will certainly find expert witnesses, and surveyors, and process servers in far away states. But you'll also have luck if you look for medical illustration charts, and office management software, and detailed company profiles.

There are other categories of people and groups who make useful information freely available online. The most prolific among these are individual experts and scholars (who may just be trying to sell themselves or their ideas), professional associations (such as the ABA), interest groups (think Children's Defense Fund, AARP, National Rifle Association), and libraries and universities. By simply thinking about these varieties of groups and individuals, the information they produce, and the reasons that they produce it, you begin to have a very good notion of the kinds of content that are likely to be available online. For example, most groups and individuals that produce online content are interested in efficiently publicizing their current work or thought, or encouraging your immediate response. Consequently, their web sites probably include only fairly recent studies or papers or information. At the same time, libraries and universities work to preserve and archive information, so their web sites are more likely to contain older material.

Another means to help you understand what is likely to be online is to consider the kinds of information that lend themselves most readily to online presentation, or those that best exploit the talents of computers. Online presentation is perfect for constantly created or quickly changing information (like news and weather forecasts) since there is

basically no lag time between writing the information down and publishing it online.

Directories make use of database technology; and calculators of various sorts rely on a computer's ease with computation.

Having said all of this, let me still encourage you to perform a quick Internet search whenever you are presented with a research problem. There are countless exceptions to the general rules for what you should expect to find online. What I do urge you to do is know enough about what to expect to find online that you are quick to abandon an unsuccessful search for material that is unlikely to be on the Internet.

III. WHEN IS A SEARCH ENGINE THE WRONG TOOL FOR FINDING ONLINE INFORMATION?

To answer this question, you must understand the basics of how a search engine works. The first detail to digest is that search engines do not search the Internet. They search a database that they have created of Internet pages. Google uses web crawlers (essentially software robots) to traverse the web and capture pages. These crawlers begin with a list of known web addresses, capture those pages, scan them for links, and then turn the pages over to an automated indexing system. Then the crawlers go on to follow the links from the first batch of pages and repeat the process. Since there are trillions of pages on the web, the index that Google or any search engine searches never includes all of the web, nor does it necessarily include the most current version of the page at a specific address. Even though Google recently announced that its index now includes over one trillion pages, that is still only a fraction of the entire Internet.

In addition, each search engine stores only a set amount of data from each page included in its index. The latest experimental results I saw recorded Yahoo! including

around 200 Kb; Google capturing just over 500 Kb; and MSN indexing over 1000 Kb. These are larger amounts of storage than in the recent past (101 Kb was the cut off for many years); but this still means that there are some partial pages included in each of these indexes. In other words, if the information you want is after the magic cut off point, it won't show up in a search.

If a web site is very large (including hundreds or even thousands of pages) it is likely that all search engine indexes are going to include only a portion of the pages on the site. Often this happens because some pages can only be reached by following many links from the beginning page of the site. Web crawlers do not readily find links that drill that deeply into the site. Sometimes the only way to find material buried deeply in a web site, is to go directly to the site.

Finally, search engines most often index only static web pages (those pages that exist on servers in a definite form). However, much of the most useful information on the web resides in databases. When a database available to web users receives a query, it generates an Internet page on the fly which contains the answer to the query. This 'answer page' exists only in response to the query. It does not live on a server waiting to be found and linked to. The same is true of information generated by the many web calculators and translators. The answer to your question is generated only when you pose the question. You can easily find a wealth of useful information by using web accessible databases, calculators and translators. However, the only help that a search engine can offer you to do this is to help you find the database, calculator, or translator.

IV. WHEN I TRULY NEED GOOGLE, HOW CAN I SEARCH LIKE AN EXPERT?

6

There are a few simple rules for any online search. Know what you are looking for as well as you can. Think of details. Describe it well in your search. Think of synonyms. Use important, unique words. Try and understand what kind of information you are seeking. Is it a simple fact, or a quotation, or a government document? Ask yourself, 'who or what group might produce the information?'. If your first stab at a search returns nothing useful in the first 30 or so results, try a new search. If you are still unsuccessful, try a different search engine. The various search engines do index different pages. Finally, try and think about where the information might be. Is it likely to be in a database? Search for a likely database. Then search the database. Was it produced by a specific group? Go to the group's web site, and search there. Or, you can use Google to search an individual web site (site: www.sec.gov).

After you go through that exercise, the easiest way to get good results from Google is to understand a bit more about how it works. The search engine automatically assumes that you want to find all of the words you enter in a search. What Google sees when you type 'red clay conference' is 'red AND clay AND conference'.

Google will accept (use) up to thirty two words in a search. In many searches, seven or eight words are much better than two. For instance, I bank at First American Bank and Trust in Athens. Everyone in town simply calls it "First American". However, if I enter that search into Google, my bank is not listed in the first ten results. If my search is 'first american bank', my bank just barely makes it to the first page of results. The ideal search is 'first american bank trust athens georgia'. With all of those words, my bank is the very first result.

Google pays attention to the order in which you enter terms. More emphasis is placed on earlier terms. Pages that contain words together in the same order as your

query are given a higher rank. If I search for my bank by entering 'american bank first', the result that I want is no longer on the first page of results.

Note that my searches don't include capitalization. The search engine does not distinguish between capital and lower case letters. In fact, the only time that Google calls for capital letters (or even notices) is when you want to us 'OR' as a boolean term. In that instance, the full word must be capitalized.

Google often ignores common words in searches. Articles such as 'the' and 'an' and prepositions such as 'from' and 'to', as well as conjunctions and most adverbs are all generally ignored.. The search engine will alert you that it has ignored terms in your search when it returns results. If any of these "stop words" are crucial to your search and Google is ignoring them, you can force the search engine to consider all of your terms by placing a + in front of each word (i.e., ' +but +I +know +it +when +I +see +it').   Of course, there is a much easier and more effective way to get Google to return a quote such as the one above: simply include the entire phrase in quotation marks. Using phrases in quotes whenever possible is perhaps the single most effective search strategy that you can employ.   You can also force Google to exclude a word from your results by placing a - in front of the word in your search.

The advanced search form in Google lets you simply fill in terms that you want specifically included or excluded or treated as a phrase. It also allows you to easily choose the language, file type, and date of last update of the pages in your results. In addition, you can limit your results to a single site or to a domain (such as: .org or .edu) .

There are other easy search tools available in the Google basic search. If you place a tilde (~) in front of a word in your search, Google will search not only for the word itself, but also for its synonyms. The search engine will look for a range of

numbers. Just enter the two numbers marking the limits of the range with two periods

between them (1986..1999). If you use the phrase "what is" in front of a search term,

Google will return a definition or explanation. (what is krok) Type the phrase "current

time" with a city or country, and that's what you'll get. (current time beijing)

Everyone has had the experience of remembering most, but not all of a quotation.

Google can help. The asterisk (*) serves as a wild card in a search. ("twas * and the *

* did * and * in the wabe") Believe it or not, the very first result for that search is the

beginning of the poem Jabberwocky.

Google recognizes ticker symbols. Just enter the symbol in the search box and

your first result will be an up to the minute stock quotation (not that you want to see that

right now). The search box can be used as a calculator ((6 * 7)/3). (The answer is 14.)

It's also a tool for unit conversions. (65F in C) (It's 18.333 degrees)

Your can search for local businesses, restaurants, or movies very easily. Just

enter the type of business or restaurant and a zip code. (plumber 30605 or thai food

30605, even movies 30605–you'll get listings and show times) Check the weather at the

beach (weather tybee island); or get a map of soperton (map soperton georgia).

V. FINALLY, A WORD ABOUT WHAT GOOGLE IS GETTING OUT OF YOU

Google is a two way street. All search engines are. What they are getting from

you is information. About you.

One of the ways that search engine technology has improved over the past few

years is that search results are now tailored to the user. This is accomplished by the

various search engines placing software (cookies) on each computer from which a search

is executed. These cookies keep a history of searches. Over time, the search engine gets

enough information about earlier searches that it can first offer up local results or results

related to previous searches. In reality, no matter what exact words I enter in Google from my computer, if my search is similar to my bank's name, I get my bank.

If you are uncomfortable with cookies, you can set your Internet browser's security to refuse them, to delete them each time you log off, or to ask you about whether or not to maintain each cookie it downloads. You can also manually go into your browser and clear cookies. There is a tradeoff for doing this. Without cookies, your search results are not tailored to you. Many of the conveniences of sites remembering your log in or auto filling some forms are lost.

Search engines, including Google, keep search information not just on individuals' computers in the form of cookies but also on their own servers. Every search engine company does give assurances that the information they maintain on their servers does not link search information to any individual user. However, they do all keep search information. Many of them keep it for many years. They use statistics about search activity to sell advertising. We know that in some countries, they share search statistics with the government.

Privacy groups offer advice about how to protect personal information. In general, they suggest you always delete cookies after you use a public computer. To minimize the accumulation of personal information about yourself in a single place, many groups suggest that you not use the same company for search and for email or other services. A few of the many good web sites that explain privacy issues and offer clear advice about protecting yourself are: http://www.aclu.org/privacy/internet/index.html; http://www.eff.org/wp/effs-top-12-ways-protect-your-online-privacy; and http://privacy.org/ .