

Genealogical Googling

SINCE YOU'RE HOLDING a copy of this magazine in your hands, I'd be willing to bet that you've used at least one — if not several — of the leading Internet search engines. In fact, if you're both a family history enthusiast and an Internet user, the odds are even greater that you need your daily fix of Google, Yahoo! or MSN more than your morning coffee. After all, compulsive searching is what we genealogists have been doing for decades — why should things be any different in this age of the Internet?

WHAT IS AN INTERNET SEARCH ENGINE?

Before we dissect the Internet's leading search engine as a tool for genealogists, let's consider the fundamentals. At its most basic level, the job of a search engine (Google or any other) is simply to compare a string of characters provided by the user with those stored on one or more computers. Sounds simple, right? But what happens when the volume of content is scattered across the globe on billions of individual pages? How can one possibly find those few hidden pages that contain precisely the content they have been hoping for?

In essence, you're searching for a needle among billions of "haystacks" scattered worldwide. Google, in turn, becomes the high-powered magnet that helps you extract your needle with relative ease (sorry... no pun intended). We'll take things step-by-step; you may even want to sit by your computer as you read this article a second time. By spending less than 30 minutes practicing the tips outlined in this article, you'll be able to obtain great results faster than a census enumerator can misspell your ancestor's surname!

KEYWORD BASICS

A single word entered into a search box is generally referred to as a "keyword". Multiple words are collectively referred to as a "keyword phrase" or simply as "keywords". This is an important concept for two reasons — first because keywords are what Google will use as a means to determine the relevancy of results returned in response to your query and second because these are

also the words that advertisers will buy to have their sponsored listing (advertisement) appear on the page

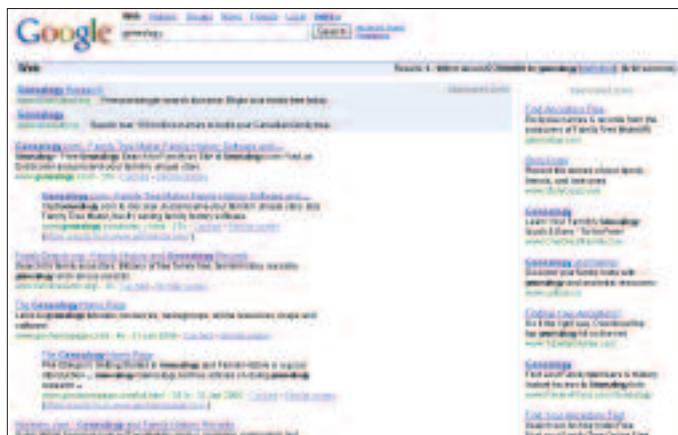
above or adjacent to your search results.

A single keyword is the most basic element of any search. All words are processed as lower case, so it makes no difference how you type them. Genealogy, GENEALOGY, genealogy and gEnEaLoGy will all yield the same results on Google. Submission of this single-word search with no other commands or characters will — in most cases — yield the greatest number of results, many with little true relevance to what you may actually be seeking.

With more than 16 billion pages currently indexed by Google, an average single-word search will result in more than 25 million pages that have some relevance to the term being queried. Popular terms will yield far more potential results, while more obscure terms can result in fewer than 100. If you are fortunate enough to be researching a fairly uncommon surname, you may have good luck with a single-word search. On the other hand, if you are conducting research for a common surname — including names with other popular meanings (e.g. Brown, Ford, White, etc.), you'll need to look beyond a single-word search. A typical online search most often involves a combination of two or more words and will likely yield more meaningful results for the user when compared to a single-word search.

IGNORING COMMON WORDS

In an effort to improve speed and provide more meaningful results, Google ignores commonly used words, unless instructed otherwise. Words such as 'a',



Google is one of the few sites on the Internet whose goal is to help you leave their site as quickly as possible.

Take a close look at the familiar Google results page shown above for the search term 'genealogy'.

How can one possibly find those few hidden pages that contain precisely the content they have been hoping for?



Dan Lynch is a technology marketing consultant and professional genealogist based in Connecticut.

A frequent lecturer, he is an APG member and also serves as vice president for the Connecticut Society of Genealogists. He can be found online at www.danlynch.net

Genealogical Googling

Using the author's common Irish surname, you can see how 39 million results can be reduced to 58,300 just by using additional words.

'of', 'the', 'or', 'and', 'to', 'how' and others will be ignored when processing your search. Example: The phrase "land of the free" will be processed the same as the two-word query "land free", both yielding about 148 million results.

In addition to being common words, both AND and OR are also reserved for use by Google as special operators, performing pre-defined functions when included in a search. They will only be treated as operators when entered in all upper-case letters. In addition, Google uses a technique called an "implied" or "automatic AND" — meaning they will interpret multiple word queries as if the word AND appears between each entry. This concept is discussed in greater detail later in this article.

MORE WORDS, FEWER RESULTS

Consider the example of a consumer researching an item for potential purchase — in this case, searching for a

	Search Term	Google Results
1	camera	257,000,000
2	digital camera	118,000,000
3	underwater digital camera	5,340,000
4	underwater digital camera lens	2,360,000
5	underwater digital camera zoom lens	1,390,000
6	Nikon underwater digital camera zoom lens	914,000

zoom lens for their Nikon underwater digital camera. Submitting a single-word search for 'camera'

on Google would yield more than a quarter-billion potential results. This is far more than anyone can evaluate and while the top 10 results (those appearing on Google's first results page) will deal with cameras, they may not provide the depth of specific content need-

	Search Term	Google Results
1	lynch	39,100,000
2	lynch family	9,340,000
3	lynch family genealogy	4,910,000
4	connecticut lynch family genealogy	1,450,000
5	waterbury connecticut lynch family genealogy	58,300

ed for making an informed product decision. Extending this example, if the consumer

conducts a new search for "Nikon underwater digital camera zoom lens", they would filter out more than 99.5 percent of the original results, leaving only those most directly relevant to their actual needs. Evaluating the first few sites, one does quickly find information specific to the product in question.

SEARCHING FOR GENEALOGICAL GOLD

The concept of using several carefully

selected keywords applies nicely to family history research. You can combine one or more surnames with the names of cities, towns, counties or villages to pinpoint the exact family you are searching for. This technique can be your first step in quickly sorting through mountains of results to find your nugget of genealogical gold. Be careful not to apply too many filters too soon — you don't want to accidentally filter out possible results that may hold valuable clues.

Using the author's common Irish surname, you can see how 39 million results can be reduced to 58,300 just by using additional words. As we continue, you'll see how other simple commands can be used in conjunction with one another to refine these results even further — leading to a very manageable number of highly relevant results.

AND, OR AND OTHER OPERATORS

In the previous examples, you can easily see how the addition of each word reduced the volume of search results. Google's use of the 'automatic AND' was the invisible filter being used on our behalf to reduce results to a more meaningful group.

In the third Lynch example, Google searched for pages including lynch AND family AND genealogy, seeking the presence of all three terms somewhere on the page to be included in the search results. Those pages matching three terms out of three would be considered most relevant, placing them high on the results list. Some pages might match just two of three terms, but still be included toward the bottom of the results list.

The plus sign (+) can be used in place of the AND operator, placing it directly in front of the term. When submitting a query with the absolute requirement that a specific term be included in resulting pages, use the plus symbol. (Example: lynch +genealogy could be used instead of lynch AND genealogy)

In a similar fashion, a minus sign (-) placed immediately before a word can be used to exclude results containing that word. This is especially helpful for genealogists when surnames or place names are words with alternate meanings. The city of Waterbury, located in New Haven County, Connecticut, is also the name of a town in Vermont. Waterbury is also the brand name of

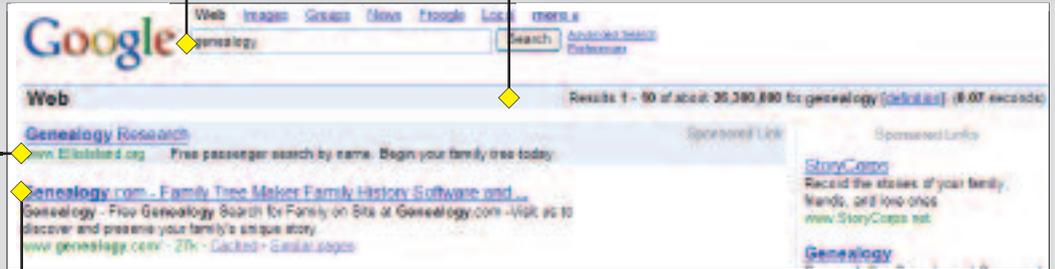
ANATOMY OF THE GOOGLE WEB RESULTS PAGE

Learning how to submit more intelligent queries is just half the equation when using Google. The other half lies in understanding how results are presented by Google so that you can quickly scan the list and select one or more for closer inspection. The Google Web results page may appear crowded, but once you understand the basic layout, you will find it quite easy to use.

Search Box – The Google search box appears at the top of each results page adjacent to the Google logo. This box will contain the query most recently submitted, allowing further refinements if necessary. Above the box are small text links providing access to other Google tools and services.

Result Statistics – A single light-blue line appears horizontally across the page containing the approximate number of web pages matching the search criteria, along with the time it took to respond. Individual search terms are also repeated and a [definition] link will sometimes appear allowing you to obtain information about the word.

Sponsored Links – Google sells text-based advertising units on their results pages to generate revenue. Depending upon the individual words contained in your search, you may see text advertisements called ‘Sponsored Links’ along the top and/or the right side of the page. The advertiser pays Google for the placement on that page only if you click on their link and are delivered to their web page. Since the ads are based on specific keywords you provided, you may sometimes find the products or services featured helpful in your family history pursuits.



Search Results – These are what you’re after! Each Google results page contains text listings, as many as 10 per page, for web pages that match your search criteria. Results are presented from most relevant to least relevant using Google’s proprietary algorithm for analyzing the contents of each page. Each individual result summary can contain more than a half-dozen link options. You should become familiar with this standard format.

Next & Previous – At the bottom of each page, immediately following the last result summary, you may see Next and Previous links. Depending upon the volume of search results, you may see numbered links. Each number corresponds to a page which will include as many as 10 additional results for your review.



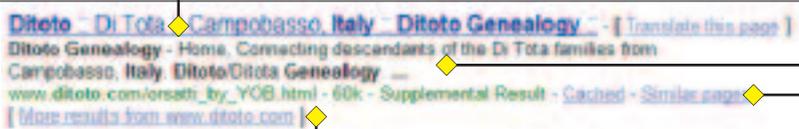
A CLOSE-UP OF AN INDIVIDUAL GOOGLE RESULTS SUMMARY

While the summaries vary slightly depending upon the contents of the page, the format is consistent.

Line 1 – Contains the Page Title as an underlined dark blue link (which turns purple once you have visited the site). For pages containing foreign language content, an optional link may appear enabling an automatic translation of that page.

Lines 2/3 – Either two or three lines, this black text contains a very brief summary of content from the site page. For both Page Title and content summary, any matching keywords will appear in bold-face type.

Line 4 – Appearing in green text, this line displays the web address (also called a URL for Uniform Resource Locator) of the page, followed by the file size of that page (60k in this example). Also appearing on this line are two important links in light blue. The first link provides access to a ‘Cached’ version of this page. A cached page is a recent snapshot of that page, usually less than 60 days old, created by Google. This can be helpful when the live version of that page has been removed or is temporarily unavailable. A ‘Similar pages’ link may also appear at the end of this line. Clicking this link will execute a new search using a special related: operator. The results will display pages with content closely matching that of the page where the link appeared.



Line 5 – The last line in the example shown above appears on the occasion when Google has determined that several pages from the same website may be relevant to your query. The first result is typically followed by a second indented summary, followed by the link [More results from www.sitenam.com]. Clicking this link will execute a search using a special operator instructing Google to restrict the search to only that site.

Genealogical Googling

many popular antique clocks and watches frequently featured for sale on the Internet. Submitting the search **waterbury -vermont -clock -watch** would look for pages including the term 'waterbury', but would then filter out those containing the words 'vermont', 'clock' and 'watch'.

Using OR (entered as upper case) can also be helpful for genealogists by providing additional flexibility in filtering search results. Submitting the search: "lynch family genealogy" waterbury OR bridgeport tells Google you want 'Lynch family genealogy' to appear in that exact order AND (implied) either the cities of Waterbury or Bridgeport. This query returns four results — all worthy of consideration.

These operators can be used individually or in combination with one another. By submitting a few practice searches, you will become more familiar with using these and other search filters.

USING QUOTATION MARKS

When evaluating results, you may sometimes be frustrated to find a top-ranked site has little to do with your desired outcome. The page may reference the words Lynch and family and genealogy, but in an unrelated context. If you want to view just those results where the terms appear together, then resubmit the query using quotations — "lynch family genealogy" — and you'll reduce the 4.9 million results to just 2,550. Note that Google will interpret anything inside quotation marks as a request for an exact ordered match. When placed inside quotations, the search "Patrick Lynch" will yield different results from "Lynch Patrick" and you are encouraged to try both and compare the results.

Earlier in this article, you learned the word 'and' would be ignored as a common word if entered in lower-case letters (e.g. internet and genealogy will yield substantially similar results to those resulting from a search for internet genealogy). If, however, you need Google to factor this as part of a search, you can place it inside quotations to essentially override the filter used for common words.

As an example, searching for "bed and breakfast" (with quotes around the entire search string) will yield 13.7 million results, compared to 35.8 million for bed AND breakfast.

SYNONYMS

In addition to the symbols + and - to add or remove terms from a search, Google can process synonyms for one or more search terms when preceded by the "~" symbol. This symbol is called a tilde and is the upper-case character to the immediate left of the number 1 on your computer keyboard. In the two examples below, note how expanding the second search to include synonyms for the word genealogy extends the pool of results seven-fold.

Example:

Search: lynch +connecticut genealogy
Results: 226,000 results

Search: lynch +connecticut ~genealogy
Results: 1,610,000 results (including terms such as roots, family tree, and others in place of genealogy)

SEARCHING BEYOND GOOGLE WEB

For many Google users, the default 'Web' search is the

only portion of the Google service they use. Other Google features use the same search conditions described earlier, so once you know how to navigate Google Web, you're ready to explore the other powerful features of the Google service. Those of particular interest to genealogists include Google Images and Google Language Tools.



GOOGLE IMAGES

Google Images uses the input specified in your query to find pictures or graphic images appearing on web pages matching your search criteria. One quick search by a genealogist will turn up a wide array of vintage portraits, headstones, homesteads and many other fantastic images. The results page displays graphics in a grid pattern, four images across and five images down, for a total of up to 20 thumbnail images per page.



Clicking on the thumbnail image will bring you to the page where the image is displayed. A small text link appears at the top of the page — "See full-size image" — providing direct access to a full-sized version of that image. As you explore the power of this Google service, you should use the search tips discussed earlier to more quickly filter through potential images.

Now that you're on your way to becoming a Google power user, explore the other text links found above the search box. The better you become at mastering these search techniques, the more quickly you'll uncover clues to advance your family story.