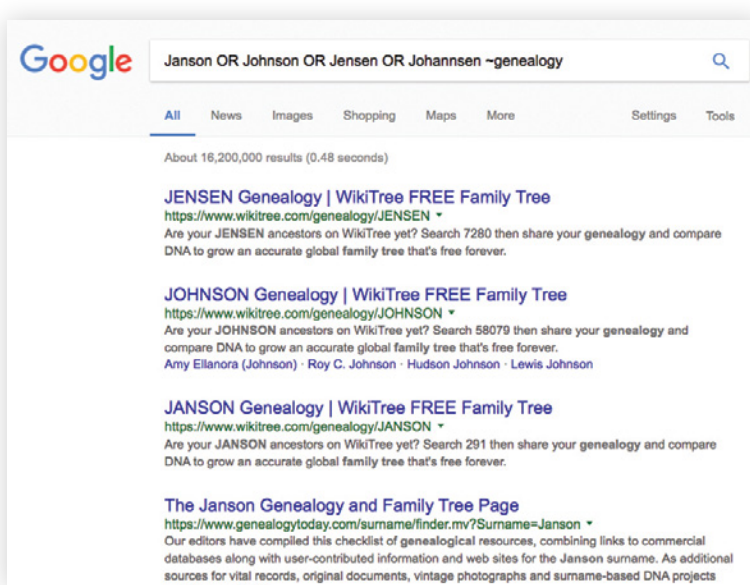


Search Methodologies for Genealogists



Jordan Jones is a writer and lecturer on genealogy and technology. He has been a genealogist since 1973, an editor and publisher since 1987, and a webmaster since 1995. He is the past president of the National Genealogical Society and can be reached at jordan@genealogymedia.com.

By developing a methodology for Internet searches, genealogists can find more relevant information to address brick walls and enrich family history with the personal details of their ancestors' lives and struggles. In this article, I suggest a methodology for Internet searches, describe the underpinnings of Internet navigation and search, provide samples of search syntax, and walk through a case study using search syntax to narrow results in a Google search.

Methodologies

Search engines are a treasure trove of information, and it is easy to approach them with a mindset of serendipity. I encourage a

methodology for search that will help genealogists know what they have searched and discover other avenues for search in the future.

- Most sites have **search tips or guidelines**. Read them!
- Look for **advanced search pages** on any site used to perform a search.
- **Experiment!** Make the search more specific to narrow the number of results. Make the search more general to widen the number of results.
- **Map out strategies** for more specific or varied searches. Plan more complex web searches like planning a trip to a major repository. This process includes sourcing searches as for any record search, including negative results. Consider Internet searches as part of a research plan and research report.
- **Automate search results** into an inbox with Google Alerts and similar tools.

Access: navigation and search

Genealogists trying to improve their searches on the Internet can benefit from a better understanding of how the web functions. There are two methods to access information on the web, navigation and search.

Websites provide click-through navigation paths to find information. Well-designed navigation paths provide an easy way to get to relevant information

simply by drilling down, click by click, into the content. The downside of navigation is that it can be cumbersome, especially for large quantities of information. An example of well-designed navigation is the Library of Congress's "Chronicling America: Historic American Newspapers," http://chroniclingamerica.loc.gov/#tab=tab_newspapers.

Search is helpful for discovering which website to navigate to, or when a navigation path to the information needed on a site is not obvious.

While web designers focus on access through both search and navigation, the web is a vastly accelerating collection of information. The number of public web pages quadrupled between 2005 (11.5 billion)¹ and 2017 (47 billion).² It takes powerful search skills to find relevant genealogical data.

Search is on the cusp of some very important changes. Artificial intelligence will allow generic search engines like Google and genealogy-specific search engines like Linkpendium, Ancestry, and MyHeritage to better understand what is being sought, by processing the user's natural language and the natural language meaning of the documents ingested. These are promising but still nascent developments. For the moment, search is still best done in a more rudimentary way.

How do search engines work?

To perform a search, a web search engine needs to do the following:

- **Gather**—By "crawling" through links on the web using software "spiders," a search engine gathers the information it will use to provide search results. Crawling is limited in two ways.
 - Login or navigation requirements: Some web pages are not accessible to spiders because of the technology used or because of login requirements. These sites, including many genealogy sites, are considered part of the "dark web" because they are not accessible in search engines.
 - Site owner requests for inclusion or exclusion of URLs: Webmasters use a file called robots.txt to request limitations to crawling. Conversely, a sitemap suggests pages to search, and can include pages not directly available to the search tool.

- **Organize**—The search engine "indexes" the sites it has crawled by creating and managing a series of relationships between index terms (words and phrases) and URLs (websites).
- **Cache**—Some web search applications (such as Google) store all the pages they crawl. This capability allows viewing of what was on the webpage at the time the search engine crawled it, in case the page changed or moved some time after being crawled and before the serving up of search results.
- **Rank**—Links delivered are ranked in terms of relevance (is the term in the title, the metadata, or some other prominent place on the website?), popularity (how many pages link to it?), authoritativeness, and other criteria. The details of ranking algorithms are tightly held secrets, often called "the secret sauce." Search Engine Optimization (SEO) is a field dedicated to improving site rank based on some understanding of page-ranking algorithms.
- **Search**—Understanding the previous operations will help get better search results.

Genealogical searching

So how do researchers access the desired genealogical information? Search can be divided into two modes, both useful to genealogists:

- **Keyword**—Every significant word is part of the search (Google, NewspaperArchive, Fold3).
- **Database**—Words are searched against particular fields in a database, such as "surname" or "state." This type of search is done in the major genealogy sites including Ancestry, Fold3, GenealogyBank, NewspaperArchive, FamilySearch, and SteveMorse.org.

It is important to keep in mind the kind of search being performed. A full-text search may not be able to distinguish a surname from any other word. One factor that makes Google search such a powerful tool is its ability to fine-tune searches in a number of ways. Despite the fact that Google has many of the downsides of simple full-text search, symbols can narrow the search down to more relevant results. Below are some search strategies that should be part of the search toolkit.

Websites cited in this article were viewed on 10 May 2017.

1. A. Gulli and A. Signorini, "The Indexable Web is More than 11.5 Billion Pages," <https://pdfs.semanticscholar.org/cd80/fccb92b64ebeb3459b65dd434f55fb916980.pdf>.

2. Maurice de Kunder, World Wide Web Size, <http://www.worldwidewebsize.com>.

The following examples relate to searches for Jane Graham (1811–1854). The text inside the square brackets is typed literally into a search engine to get the result.

Basic searches

- **Word**—Find pages that include “Jane” and “Graham.”
Google: [Jane Graham]. To ignore plurals and synonyms and search for a term exactly as typed, enclose the term in quotation marks. [Jane Graham “barn”]
- **Phrase**—Find pages that include “Jane Graham.”
Google: [“Jane Graham”] finds only pages with the exact phrase “Jane Graham” and ignores pages that have only phrases such as “Jane Eliza Graham.”
- **Proximity**—Find pages where “Jane” is near “Graham.”
Google: [Jane * Graham OR Graham * Jane] finds “Jane Eliza Graham,” “Jane ‘Liza’ Eliza Graham,” and “Graham, Miss Jane,” but not “Jane Graham” or “Graham, Jane.”
- **Boolean**—AND/OR/NOT
Google: [“Jane Graham” OR “Graham, Jane”] finds pages where Jane Graham is listed, whether or not she is listed surname first or given name first.

Advanced searches

- **Synonym or “like”**—Find pages with words like the search term or phrase. This search is more useful outside genealogy, but it has some similarity to a Soundex search.
Google: A search for [~genealogy] returns results about genealogy and family history.
- **Wildcards**—Some sites allow wildcards (*_?) to replace one or more characters. Check the site’s guidelines. These symbols are very handy for checking occurrences of variably-spelled surnames.
Ancestry.com: ? replaces one character; * replaces zero to five characters. Names must contain at least three non-wildcard characters.
Google: * replaces one or two whole words. There is no wild card for less than a word on Google. For multiple surname spelling searches on Google, use the Boolean OR operator [Johnson OR Jensen OR Johannsen].

- **Exclude word**—Find pages that don’t include a particular word.
Google: [Jane Graham -Eliza]
- **Exclude phrase**—Find pages that don’t include a particular phrase.
Google: [Jane -“Eliza Graham”]
- **Include specific site**—Find pages in a specific site, such as usgenweb.org, where a search term appears.
Google: [site:www.usgenweb.com Graham]
- **Exclude specific site**—Find pages where a search term appears, but exclude a specific site.
Google: [-site:www.usgenweb.com Graham]
- **Location**—City, county, state, or other locale can narrow search results drastically.
Google: [“Jane Graham” “Monroe County”]. Major genealogy sites, Flickr, and others provide fields to narrow searches by location.
- **Record type**—Birth certificate, obituary, newspaper.
Google: [“Jane Graham” newspaper]. Major genealogy sites provide methods for limiting searches to particular record types.
- **Individual databases**—Databases for particular records or kinds of information.
Google: Image Search, Map Search. Major genealogy sites, Flickr, and others provide tools for limiting searches to specific databases or tag clouds, or groups of related user-contribution tags or descriptors.

Special Google search features

- **Search types**—News, Videos, Images, Personal.
Google’s search results page allows focus on these different facets of results. The Personal results are based on the user’s Google account (if logged in) and recent searches.
- **Searches limited by time**—Search recent pages. After running a search, click “Tools” on the right to see a new pull-down menu that says “Any time.” Choose a time period to search based on when the content was updated.
- **Searches focused on content type**—Narrow the search to images (including images of documents), videos, blogs, voice (for voicemail in Google Voice), news (including recent obituaries), books, social media. Search starting with a picture, not text, using Google Goggles on a smart phone or Images search on Google’s main page.

- **What words were used when**—Google Ngram Viewer (<https://books.google.com/ngrams>) tracks the currency of words in the millions of volumes in Google Books. It's handy to understand when words became popular or passé.
- **Shopping**—Google Shopping can be used for quick comparison pricing for that next scanner.
- **@**—Search social media, such as @Twitter or @Facebook.
- **Related sites**—Find results on sites that have similar content to a specified site. For example, [related:Ancestry.com] finds results on sites that Google determines have similar content to Ancestry.com.
- **Numerical range**—Find results that include any numbers within a range. This search works well with years and any other types of numbers. The syntax consists of the starting number followed by two periods and the ending number of the range: [1811..1854].

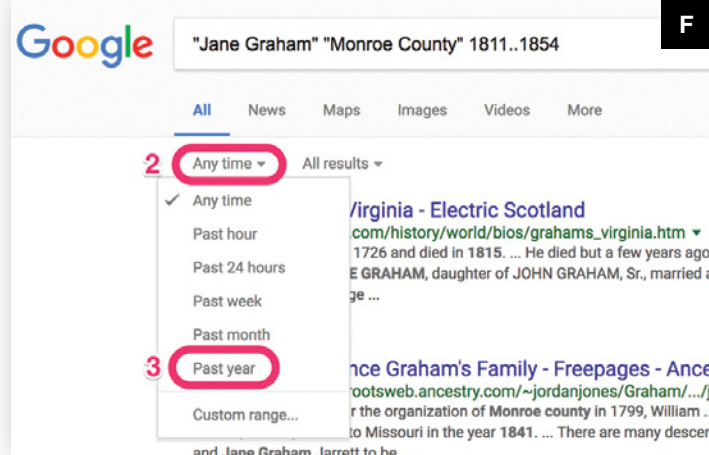
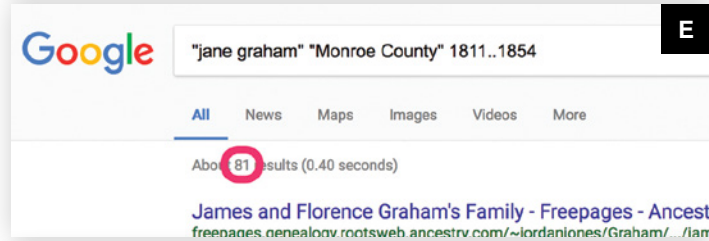
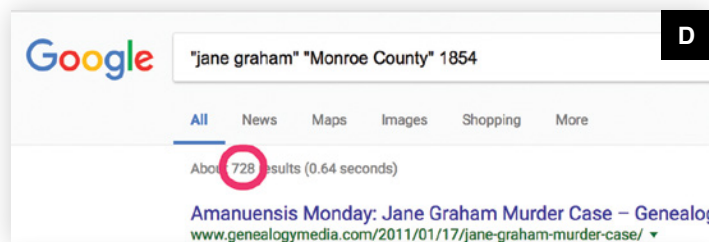
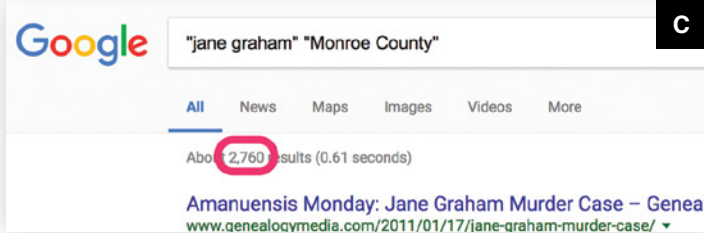
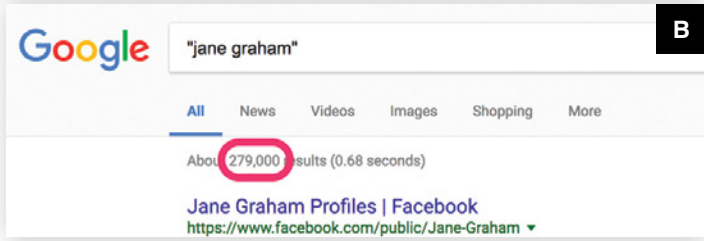
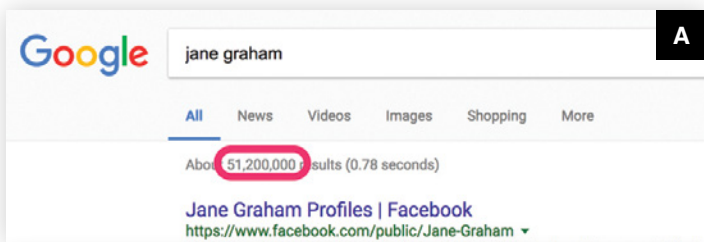
Case study: Jane Graham

Using these strategies, including searching by phrase and searching by location, I was able to narrow down results for my ancestor Jane Graham from 51.2 million to seven, or by a factor of 7.3 million.³

- A simple search for [Jane Graham] yielded 52.2 million results. **A**
- Making this a phrasal search ["Jane Graham"] reduced this to 279,000 results. **B**
- Adding a place to the query, and making the place a phrase ["Jane Graham" "Monroe County"] reduced results to 2,760. **C**
- Adding a year ["Jane Graham" "Monroe County" 1854] reduced the results list further, to 728. **D**
- Making the date a date range ["Jane Graham" "Monroe County" 1811..1854] further reduced results to 81. **E**

At this point, I had whittled the millions of results down to something manageable. But let's assume I know the publication is recent. If I narrow the search to pages published in the past year, I get down to seven results. **F**

3. Google, <https://www.google.com>. Various searches.



Conclusion

It is easy to use search tools casually, even haphazardly. I recommend that genealogists build their own methodology to plan and document searches. Use these search recommendations and advanced tools to find relevant data that can be the next major discovery.

Resources

Ancestry Help. <https://support.ancestry.com/s>.

Click "Search & Records."

Cooke, Lisa Louise. *The Genealogist's Google Toolbox*. Second edition. San Ramon, CA: Genealogy Gems Publications, 2013.

Fold3 Advanced Search Page. <https://www.fold3.com/s.php>. Click "Advanced."

GenealogyBank Advanced Techniques.

<http://www.genealogybank.com/information/help/advanced-techniques>.

Google Advanced Search. https://www.google.com/advanced_search.

Google Alerts. <https://www.google.com/alerts>.

Google Books Ngram Viewer. <https://books.google.com/ngrams>.

Google Search Help. <https://support.google.com/websearch>.

Google Search Preferences. <https://www.google.com/preferences>.

Linkpendium Family Discoverer. <http://www.linkpendium.com/family-discoverer>.

Lynch, Daniel M. *Google Your Family Tree: Unlock the Hidden Power of Google*. Provo, UT: FamilyLink.com, Inc., 2008.

NewspaperArchive Advanced Search. <https://newspaperarchive.com/advancedsearch>.

One-Step Webpages by Stephen P. Morse. <https://stevemorse.org>.

Wikipedia. "Robots exclusion standard." https://en.wikipedia.org/wiki/Robots_exclusion_standard.

Wikipedia. "Sitemaps." <https://en.wikipedia.org/wiki/Sitemaps>.

Wikipedia. "Web search engine." https://en.wikipedia.org/wiki/Web_search_engine. 🌱

THE REPRINT COMPANY PUBLISHERS

Reprint editions of historical and genealogical books pertaining to Virginia, North Carolina, South Carolina, Tennessee, Georgia, Alabama, Mississippi, and Louisiana. Visit our website, www.reprintcompany.com, to view our complete catalogue of books and sign up for email notifications of new books coming back in print, special offers, and other information.

A Representative Sampling of Our Titles

Brewer, Willis: *Alabama: Her History, Resources, War Record and Public Men, From 1540 to 1872* (1872)

Claiborne, J.F.H.: *Mississippi As A Province, Territory and State with Biographical Sketches of Eminent Citizens* (1880)

Draper, Lyman C.: *King's Mountain and Its Heroes* (1881)

Hawks, Francis L.: *History of North Carolina*, 2 vols. (1857-58)

Hewatt, Alexander: *An Historical Account of the Rise and Progress of the Colonies of South Carolina and Georgia*, 2 vols. (1779)

Jones, Charles Colcock Jr.: *Antiquities of the Southern Indians, Particularly of the Georgia Tribes* (1873)

Jones, Charles Colcock Jr.: *The History of Georgia*, 2 vols. (1883)

Landrum, John B.O.: *Colonial and Revolutionary History of Upper South Carolina* (1897)

Logan, John H.: *A History of the Upper Country of South Carolina from the Earliest Periods to the Close of the War of Independence*, 2 vols., (1859, 1910)

Pickett, Albert James: *History of Alabama and Incidentally of Georgia and Mississippi from the Earliest Period* (1900)

Ramsay, David: *History of South Carolina from Its First Settlement in 1670 to the Year 1808*, 2 vols. (1809)

Rowland, Dunbar: *Military History of Mississippi, 1803-1898* (1908)

South Carolina Historical Society: *The Historical Writings of Henry A.M. Smith*, Vols. I-III (*The Baronies of South Carolina, Cities and Towns of Early South Carolina, Rivers and Regions of Early South Carolina*); *South Carolina Genealogies*, Vols. I-V



www.reprintcompany.com

864-579-4433 • Post Office Box 5401 • Spartanburg SC 29304