

Googling for Genealogy



Saturday March 10, 2012

Central New York Genealogical Society

<http://www.rootsweb.ancestry.com/~nycnygs/>

Dennis A. Hogan, <http://www.dennisAhogan.com>

This Complete Presentation is available online

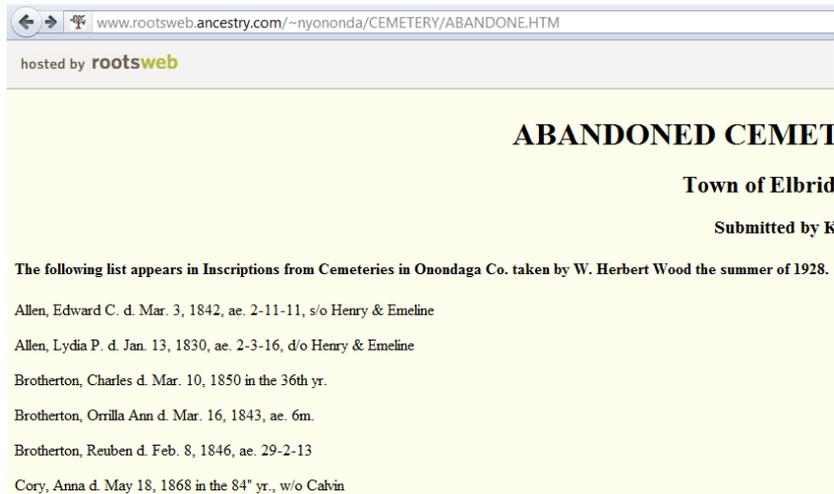
- At www.dennisAhogan.com, click on Lectures and Handouts tab
- Select a handout and save on your computer
- Then you can click on links to try out websites

How Search Engines Work

- Search Engines continually update their index
 - **Crawl** the web using software spiders/robots
 - Crawlers try to follow all links
 - Crawlers are stopped by search boxes/forms and login screens
 - **Index** each page crawled (each word gets a pointer to the web page where it was found)
 - **Analyze** each page for content value
- When a user submits a search string, search engines:
 - **Search** their index
 - **Rank** the results based on perceived relevancy
 - Google rank emphasizes the number of incoming links
 - **Display** the top results

How Search Engines Work

What We See



The screenshot shows a web browser window with the address bar containing the URL www.rootsweb.ancestry.com/~nyononda/CEMETER/ABANDONE.HTM. Below the address bar, it says "hosted by rootsweb". The main content of the page is centered and reads: "ABANDONED CEMETERY NEAR JORDAN Town of Elbridge Submitted by Kathy Crowell". Below this, a paragraph states: "The following list appears in Inscriptions from Cemeteries in Onondaga Co. taken by W. Herbert Wood the summer of 1928." This is followed by a list of names and dates: "Allen, Edward C. d. Mar. 3, 1842, ae. 2-11-11, s/o Henry & Emeline", "Allen, Lydia P. d. Jan. 13, 1830, ae. 2-3-16, d/o Henry & Emeline", "Brotherton, Charles d. Mar. 10, 1850 in the 36th yr.", "Brotherton, Orrilla Ann d. Mar. 16, 1843, ae. 6m.", "Brotherton, Reuben d. Feb. 8, 1846, ae. 29-2-13", and "Cory, Anna d. May 18, 1868 in the 84th yr., w/o Calvin".

What the Crawler Sees

Search billions of records on Ancestry.com ABANDONED CEMETERY NEAR JORDAN Town of Elbridge, New York Submitted by Kathy Crowell The following list appears in Inscriptions from Cemeteries in Onondaga Co. taken by W. Herbert Wood the summer of 1928. This cemetery was just off the hill road from Elbridge to Jordan near Jordan. Allen, Edward C. d. Mar. 3, 1842, ae. 2-11-11, s/o Henry & Emeline Allen, Lydia P. d. Jan. 13, 1830, ae. 2-3-16, d/o Henry & Emeline Brotherton, Charles d. Mar. 10, 1850 in the 36th yr. Brotherton, Orrilla Ann d. Mar. 16, 1843, ae. 6m. **Brotherton, Reuben d. Feb. 8, 1846, ae. 29-2-13** Cory, Anna

The Future of Genealogical Searching

- In the past year, Google has worked on a project to make search results better for genealogy.
- HTML5 is the next version of the programming language used by web developers. HTML5 can handle schemas – structures that organize knowledge.
- Google, FamilySearch, and other genealogy firms have worked together to develop a schema for genealogy (see historical-data.org).

The Future of Genealogical Searching

- At the RootsTech Conference in early February 2012, Google did a live demo of how searches will work in the future.
- At sites with genealogy content, crawlers will see:

Surname	Given Name	Event Type	Event Date
Brotherton	Reuben	Death	8 Feb 1848

- Advantages to using these schemas:
 - Announces to crawler the existence of genealogy content
 - Vendors will develop tools to take advantage

Resources

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

- *The Genealogists Google Toolbox* by Lisa Louise Cook
 - (Cook, Lisa Louise *The Genealogist's Google Toolbox*. San Ramon, CA:Genealogy Gems Publications, 2011.)
 - Also Google Earth for Genealogy Volume I & II (DVD) by Lisa Louise Cook

- *Google·pedia* by Michael Miller
 - (Miller, Michael *Google·pedia : the Ultimate Google Resource*, Third Edition. Indianapolis, IN: QUE, 2009.)

- <http://www.searchforancestors.com/archives/googlegenealogy.html>

Preamble

- Genealogy is NOT = Internet
 - 5% of the world's genealogy content is on the web
 - Randall J. Seaver, [Genea-Musings](#) blogger
 - 5% of the internet's genealogy content is indexed by Google – [mocavo.com](#) sited by Robert Gardner of Google
 - X% of Google's index shows up on Page 1 of results (For easy math let's use 1% - but that's high)
 - So 1% of the 5% of the 5% ($.01 \times .05 \times .05 = .000025$)
or about .0025 of 1% of the world's genealogy content can be accessed via Google

Preamble Cont'd

- Google is not the only search engine
 - Search syntax is not the same for every search engine
 - Anything said about the Internet is almost immediately obsolete
- There's some grey areas with Google

Googling for Genealogy

Part I - Google searches for genealogy

Part II – Other Google tools relevant for
genealogy

Part III – Other search engines relevant for
genealogy

Googling for Genealogy

Part I

Google searches for genealogy

Google Search Commands & Syntax

➤ **Stop Words**

- Depending on context, Google **may** ignore words such as *a, of, the, and, or, to, where, how*
- **the republic of ireland (128,000,000) vs republic ireland (120,000,000)**
- contents of 1st page of results are different
- **Take Away: Try with & without stop words.**

Google Search Commands & Syntax

➤ **Quotes “ ”**

- Exact contents of quotes must be found
- **“republic of ireland”** (5,320,000) does NOT yield the same results as **republic of ireland** (123,000,000)
- Exception to finding exact contents of quotes:
“hogan john” returns results such as:
 - Nick **Hogan, John** Mayer
 - Hulk **Hogan: John** Graziano
- **Take Away: Quotes are important & punctuation never counts**

Google Search Commands & Syntax

➤ **Case**

- Case insensitive even inside quotes (except Boolean commands AND and OR)
- **“republic of ireland”** (4,860,000) vs **“Republic of Ireland”** (4,860,000)
- **republic of ireland** (106,000,000) vs **Republic of Ireland** (106,000,000)

Google Search Commands & Syntax

➤ **AND vs and**

➤ **and** is usually ignored (stop word)

➤ **AND** between keywords indicates page must contain both keywords

Google Search Commands & Syntax

➤ Automatic AND

- **hogan john** (should) yield same results as **hogan AND john**
- **hogan john** (13,600,000) vs **hogan AND john** (14,100,000)
- 1st page of results has only 3 common items
- Daniel Lynch notes that for some unknown reason, the 1st statement above is not true.
- Take Away: Try with & without AND

Google Search Commands & Syntax

➤ **OR vs or**

➤ **or** is usually ignored (stop word)

➤ **OR** between keywords indicates page must contain either keyword (| = OR)

➤ **hogan OR hagan** (Particularly helpful with spelling variations)

Google Search Commands & Syntax

➤ **Order of keywords**

- First (leftmost) is considered most important
- Surname placename given
- **john hogan** (5,350,000) vs **hogan john** (13,400,000)
 - Is more better than less?
 - Why is **john** returning less results than **hogan**?
- **Take Away: Change the order to see different results**

Google Search Commands & Syntax

➤ **Word Variations – Stemming**

➤ Incorporates additional searches based on the stem of the search keyword.

➤ Singular/plural, other tenses

➤ **running** yields results including **running, run, runs**

Google Search Commands & Syntax

- **To eliminate stemming**, put quotes around word
 - “running”
- **Plus Sign** with preceding space is now NOT functional
 - Former functionality: Find exact keyword excluding variations (eliminate stemming)
 - **+running**

Google Search Commands & Syntax

- **Minus Sign** with preceding space
 - Removes pages from results that contain the concatenated word
 - **hogan –hulk –heroes** ** Particularly helpful for surnames with common meanings **
 - **hogan –“hulk hogan” –“hogan’s heroes”**

Google Search Commands & Syntax

- **Wildcard** (*, **, ***, etc.) within quotes
 - **“john hogan”** will NOT find John F Hogan
 - **“john * hogan”** finds John F Hogan, John F. Hogan, John Frances Hogan, John “JF” Hogan but **not** John Hogan or John J F Hogan (or Hogan, John)
 - **ONLY** whole word wildcard is available in Google (* = 1 word). Partial word wildcard is available in some other search engines (* = 0 to n characters)
 - **“john ** hogan”** finds John J. “Jack” Hogan and John Smith, George Hogan
 - **“john *** hogan”** finds
John and Gertrude (Bovee) Hogan
John Smith fought George Hogan

Google Search Commands & Syntax

- Searching for names
 - **“john hogan” OR “john * hogan” OR “hogan john”** [best practice]
 - **“john hogan” OR “john * hogan” OR “john ** hogan” OR “john *** hogan” OR “hogan john”**
 - Probably should add a placename/spouse/year range

Google Search Commands & Syntax

➤ **Number Range**

➤ (..) can be used for a range of any type of numbers, including a year range

➤ **“john hogan” born 1850..1870**

➤ **“john * hogan was born” 1850..1870**

➤ NOT “john * hogan” birthdate 1850..1870

➤ **“john hogan” “mary o’toole” married 1850..1870**

Google Search Commands & Syntax

- Some searches for BMD info
 - **“john * hogan was born” rochester OR canandaigua**
 - **“john * hogan married” rochester OR canandaigua**
 - **“john * hogan died” rochester OR canandaigua**
 - **“john * hogan was buried” rochester OR canandaigua**

Google Search Commands & Syntax

➤ **Tilde (~) symbol**

- Tilde preceding a keyword finds similar words (synonyms in a Google sense)
 - **Hogan ~genealogy** searches not just for genealogy, also words/strings like ancestry, family history, roots, family tree and others
 - Note that when searching only on **Hogan**, Google has no idea that you are interested in genealogy. Your results will include a lot of “noise.”
- How to determine the Google synonym for a word
 - **~cemetery -cemetery**
 - Then “subtract” bolded keywords from the results. Repeat process until no more results.
- **Best Practice: try your search with and without ~genealogy**

Google Search Commands & Syntax

- Some other search suggestions
 - **“john * hogan” ~obituary**
 - **“john * hogan” ~cemetery**
 - **“john * hogan” ~will**
 - **“john hogan” OR “john * hogan” OR “hogan john” ~genealogy**

Google Search Commands & Syntax

➤ State searches

➤ As in Michigan vs. MI vs. Mich.

➤ Prior to the 2-letter state abbreviations (which began ~1963), longer abbreviations were used.

See http://en.wikipedia.org/wiki/List_of_U.S._state_abbreviations

➤ Take Away (see next slide for results of my unscientific study): use tilda and full state name ~**michigan** which seems the most inclusive. Safest technique would be **michigan OR mi OR “mich.”**

Google Search Commands & Syntax

words bolded on results pages:

~mi mi, michigan, major cities in the state

~michigan michigan, mi, mich

~mich mich, mich., michigan, mi, umich

mi mi, michigan

michigan michigan

Mich mich, mich., michigan

Google Search Commands & Syntax

➤ **Compound clauses**

➤ **(John OR “John J” OR “John James”) AND (Hogan OR Hagan)**

➤ Up to 32 keywords can be in a search string

Google Search Commands & Syntax

- **site:** followed by any domain name
 - Enables a google search of that site only.
 - To search all RootsWeb pages for military, **military site:rootsweb.ancestry.com**
 - Another example, **site:www.facebook.com hogan**
 - **site:gov “john hogan” 1850..1870** searches for only sites whose extension is .gov
 - **“john hogan” 1850..1870 -site:gov** searches sites except those whose extension is .gov
 - **“john hogan” OR “john * hogan” OR “hogan john” site:raims.com** [Ontario County (NY) archives site]
 - Useful for GenWeb-type sites!!

Google Search Commands & Syntax

➤ **related:** command

➤ Gives the same results as clicking on *Similar Pages* in the results

➤ **related:www.findagrave.com** [no space!]

Google Search Commands & Syntax

➤ **intitle:** command

➤ Searches for words in the title of web pages

➤ **intitle:hogan ~genealogy** [no space!]

➤ Will return results such as The Hogan Family Genealogy, Hogan Family History, etc.

Google Search Commands & Syntax

➤ **cache:** command

- We've all clicked on a link in the Google results and gotten "File Not Found" or something similar. Return back to the google results page and click on the "Cached" link. This will return the page's contents as they were when Google last indexed it.
- Also can do this using **cache:www.abc.com**
- Another alternative is www.archive.org (Wayback Machine)

Google Search Commands & Syntax

➤ **filetype: or ext:**

➤ searches for only files of the type indicated

➤ **~genealogy filetype:xls**

➤ **~genealogy -filetype:pdf**

➤ **hogan filetype:ged**

Google Search Commands & Syntax

- **phonebook:** command
 - Has been removed by google
 - For reverse phone number searches: enter number in search box: **123 456-7890**

Google Search Commands & Syntax

➤ Calculator

- Enter **1930 – 73** (for approximate birth year for someone who died in 1930 at the age of 73).

Operators available: + - * / ()

- Birth date calculators plus more are available at <http://www.searchforancestors.com/utility/>

Google Search Commands & Syntax

➤ **Converter**

➤ To convert one unit of measure to another

➤ **10 rods in miles**

➤ **5 euros in dollars**

➤ **time in Dublin**

Google Search Commands & Syntax

- **Define/spell** command
 - **define:grog**
 - **define grog**
 - **spell nikel** gives “Showing results for [spell *nickel*](#).
Search instead for [spell *nikel*](#)”

Google Search Commands & Syntax

➤ **info:**

➤ Info about a web site:

➤ **Info:**<http://nyrgs.org/> yields

[RGS HomePage - index.htm](http://nyrgs.org/)

nyrgs.org/

This is the homepage of the Rochester (NY) Genealogical Society for those tracing their roots in the upper Genesee River valley.

Google can show you the following information for this URL:

- Show [Google's cache](#) of nyrgs.org/
- Find web pages that are [similar to](#) nyrgs.org/
- Find web pages that [link to](#) nyrgs.org/
- Find web pages [from the site](#) nyrgs.org/
- Find web pages that [contain the term](#) "nyrgs.org/"

Googling for Genealogy

Part II

Other Google tools relevant for genealogy

Google Tools & Apps

- Google search commands and syntax discussed above can be used in all Google tools and applications – Sort of
 - That is a goal which is a work in progress.
 - Some tools are more compliant than others. For example, most any Google web search string can be used in Google Books.

Google Language Tools

- http://www.google.com/language_tools
 - **Translated search** – enter query in native language, search sites of another language and translates back to your native language
 - **Translate text** – type or cut/paste text in box, select to/from languages
 - Babel Fish (<http://babelfish.yahoo.com/>) is also excellent
 - <http://translate.google.com> can detect from language and say the result
 - **Translate a web page** - enter a url, select to/from languages
 - **Use the Google Interface in Your Language** – select preferred language
 - **Visit Google's Site in Your Local Domain**
 - **Local Bias** - Use the google site of the country you're researching
 - Ex: google.com for US, google.ie for Ireland, google.co.uk for UK, etc.
 - See **Visit Google's Site in Your Local Domain** for google sites for individual countries.

Google Books

- <http://books.google.com/>
- Google has digitized millions of books in cooperation with libraries and publishers worldwide. This collection is mainly public domain, out-of-copyright books.
- Results of your search will list books with an indication of their availability: Full View, Limited Preview, Snippet View, No Preview.
- Many books (Full View) can be downloaded free in PDF format and are fully searchable.
- Options are in the column on the left. Note that dates refer to when item was published.
- GooBooGeni (<http://gooboogeni.com/>) is cataloging Google Books content relevant to genealogy.
- Genealogy Book Links (<http://www.genealogybooklinks.com/>) is cataloging ebooks related to US genealogy.

Google News Archive

- Content from present to ~200 years old, mainly US based newspapers.
- Google has terminated development of this project but is keeping content available.
- Recently Google has made historical content harder to access.
 - <http://news.google.com/newspapers> (list of papers & dates)
 - Enter a search string and click *Search Archive* (yields recent news)
 - In the column on the left, click *Custom Range...* under *Archives*
 - Enter a range of years, for example 1800 to 1900, click *Search*
- Prices are listed in results, some items are free. Most of the fee-based services will display a subset of an item for free.

Google Blog Search

- <http://blogsearch.google.com/>
- Might also want to use the following (I think these sites provide more significant results).
 - <http://blogfinder.genealogue.com/>
 - http://googleyourfamilytree.com/genealogy_blog_search.php
- Also most RSS readers (see next slide) can search

Google Reader (RSS)

- RSS (~Real Simple Syndication) is an efficient way to keep tabs on blogs & other sites that you check out frequently to see what's new.
- http://www.commoncraft.com/rss_plain_english
- Set up a reader to accumulate feeds from selected sites.
<http://www.google.com/reader/>
- Then when you see the RSS symbol on a site, you can easily subscribe to a feed from that site.
- The originating sites have no contact info for you therefore you cannot be spammed. You don't have to spend time going to the originating sites. Updates are fed to your reader. You can read them or not.

Google Images

- <http://images.google.com/> includes LIFE photo archives
- After a search, click on Images in Google bar to bring up related images
- **rochester ny ~vintage**
- **rochester ny postcards ~vintage**
- Click the camera icon on the right of the Images search box. You can enter one of your images of a person or building and Google will display similar images.
- The column on the left offers options such as *Black and White* (for vintage photos), *Faces*, and *Line Drawings* (for maps).
- Picasa (<http://picasa.google.com>) is a great google app for managing photos on your computer (& uploading).

Google Video

- <http://video.google.com/>
- <http://youtube.com> (owned by google) only finds youtube videos, google video searches all video sites
- Genealogy training, interviews with genealogy professionals, etc.

Google Alerts

- <http://www.google.com/alerts>
- Automates queries to be done 24/7 for you. So once you've tested your "perfect" search string set it up in google alerts
- Its easy to do. Just enter the search string and how often you want to be notified: immediately, once a day, once a week. Also indicate whether you want the results delivered to your email or to a RSS reader.
- Additional functionality is available if you login to your gmail account...particularly useful to keep track of many alerts (unlimited number of alerts are allowed).
- Results do not include sponsor links or spam.
- **Results are not cumulative.** For example, using "once a day" frequency, results will reflect items added to the google index since the previous day.

Google Maps

- <http://maps.google.com/>
- Maps: enter an address in the search box **Stanley, NY** When the map is displayed, click on a lettered balloon or anywhere on the map. Then click on Street View to see an interactive photo of that address (not all addresses have photos). Street View is available in some foreign countries including Ireland.
- Because of local bias, initial map is for the country site you're using.
- **catholic church gorham ny** or **gorham ny** and then click Search Nearby
- **cemeteries canandaigua ny**
- For any view including Street View you can Print, Send (emails a link) and Link (provides link to include in an email or IM & html to imbed on a website).
- Don't know of a feature in google that allows you to save an image of the map. Can use a screen shot utility like The Snipping Tool or MWSnap to accomplish this. Also can use the Print feature and then select "Print to File."
- Entering a county and state yields a map with the county outlined.
- May be able to pinpoint obsolete addresses, for example property taken by eminent domain which now exists in the middle of an expressway.
- Google My Maps lets you create custom private or public maps.

Google Earth

- <http://www.google.com/earth/index.html>
- Google Maps on steroids
- Free download
- Can overlay current images with historic (Rumsey Historical Maps)
- *The Genealogists Google Toolbox* by Lisa Louise Cook
 - (Cook, Lisa Louise *The Genealogist's Google Toolbox*. San Ramon, CA:Genealogy Gems Publications, 2011.)
 - Also Google Earth for Genealogy Volume I & II (DVD) by Lisa Louise Cook

Google Scholar

- <http://scholar.google.com/>
- Searches peer-reviewed journals, theses, dissertations, etc. Can be useful to genealogists when researching history, sociology and anthropology.
- (Also see <http://www.doaj.org>)

gmail

- gmail.google.com, gmail is a free email service with tagging.
 - **Best Practice: Use a permanent email address (gmail or yahoo) for all genealogy contacts.**
- Do NOT use an email address provided by your ISP (Internet Service Provider) like roadrunner or aol.
- Do NOT use an email address provided by an employer or other temporary supplier.

Collaboration Tools

- plus.google.com Google+
- <http://googleplusforgenealogists.com/> Daniel Lynch's blog about using google+ for genealogy
- [Cyndi's List Uses of Facebook for Genealogy](#)

Googling for Genealogy

Part III

Other search engines relevant for genealogy

Daniel Lynch's Power Searches

- http://googleyourfamilytree.com/genealogy_powersearch.php
 - Genealogy PowerSearch #1 (ancestor search)
 - Genealogy PowerSearch for Google Images
 - Genealogy BlogSearch
 - GEDCOM Search and Other File Types
 - Genealogy PowerSearch using Google Maps
 - Genealogy Database Search on Leading Sites
 - Directed Search on Government Archive Web Sites
- (Also see <http://www.searchforancestors.com/google/searcher.html>)

Other Search Engines

➤ 15 Most Popular Search Engines – March 2012
(<http://www.ebizmba.com/articles/search-engines>)

1. google.com
2. <http://www.bing.com/>
3. <http://www.search.yahoo.com/>
4. <http://www.ask.com/>
5. <http://search.aol.com>
6. <http://home.mywebsearch.com/>
7. <http://www.lycos.com/>
8. <http://www.dogpile.com/>

Other Search Engines

➤ Best Search Engines 2012

(<http://www.listofsearchengines.info/>)

1. google.com
2. <http://www.search.yahoo.com/>
3. <http://www.bing.com/>
4. <http://www.ask.com/>
5. <http://www.scrubtheweb.com/>
6. <http://blekko.com/>
7. <http://duckduckgo.com/>
8. <http://entireweb.com/>

Other Search Engines

- Metasearch Tools – searches using multiple search engines. At <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html>, UC Berkeley is generally critical of metasearches, but is “tolerant” of the following 4 metasearch tools:
 - www.dogpile.com ... google, yahoo, bing, and ask
 - <http://search.yippy.com/> ...(formerly clusty) direct link/integration with Wikipedia, clusters results in “logical” groups, searches Ask, Open Directory, Gigablast and others
 - www.surfwax.com ...for “SERIOUS Deep Digging”
 - www.copernic.com ...for “SERIOUS Deep Digging”, can search web from home page but a free download of Copernic Agent is also available.

Other Search Engines

➤ Best Meta-Search Engines 2012 (<http://www.listofsearchengines.info/>)

1. <http://infospace.com/>
2. <http://www.info.com/>
3. <http://www.dogpile.com/>
4. <http://excite.com/>
5. <http://www.search.com/>
6. <http://www.webcrawler.com/>
7. <https://ixquick.com/>
8. <http://www.mamma.com/>

Genealogy Specific Search Engines

- Mocavo, <http://www.mocavo.com>
 - “All related results from industry sources such as genealogy message boards, family trees, state and local historical societies, the Library of Congress, National Archives, Ellis Island, Find A Grave, the Internet Archive, various U.S. state archives, and many tens of thousands of genealogy sites built by individuals will be displayed.”
- UK/Ireland-specific Mocavo, <http://www.mocavo.co.uk>
- **“john hogan”**
 - Mocavo automatically searches for **“hogan john”** and “jumps over” initials and middle names
- **NM OR “New Mexico” OR “N. M.”**
 - Do NOT have to include all forms of state names and abbreviations.
- Can upload a GEDCOM tree and Mocavo will report matches

Genealogy Specific Search Engines

- MyHeritage, <http://www.myheritage.com/research>
 - Search portion only works with Internet Explorer (IE) browser
 - Searches 1500+ sites. Use the Advanced Search to focus the search on relevant sites (not all 1500+).
 - For each search you can pick up to 5 spelling variations for the surname. Also have a choice of Exact, Soundex, Megadex.
 - Can save search results & can schedule recurring searches

Genealogy Specific Search Engines

- Archives.com, <http://www.archives.com/>
 - US GenWeb Archives, Interment.net, findagrave, cousinconnect, immigrants.net (\$)
- Ancestor Hunt, <http://www.ancestorhunt.com/>
 - A portal to many sites
- Super Search, <http://surnamesupersearch.com/>
 - A portal to many searches

Possible Best Practice for Searching

➤ “Automatic”

- Google with your “perfect” search string(s) in a google alert
- Mocavo
- MyHeritage

➤ “Manual”

- FamilySearch
- Mocavo
- Google
- Metasearch tools – dogpile, yippy, copernic

Happy Hunting

