Genealogical Searching with Google

The Google search engine is a wonderful tool for genealogists who are trying to mine the massive amount of data that is rapidly being made available to people the world over on the Internet. Because Google is so simple to use, we have a tendency to just use it without thinking. If we take a few minutes to learn a bit about the syntax of the search engine, we might make our time spent looking for those hidden ancestors of ours even more productive.

The number of hits (the things we find) and the order of the display of those hits is determined by Google's page ranking system. The system calculates a page's importance, or value to searchers, based on the importance of the pages that link to it. The science behind the algorithm is not of interest to me as a genealogical searcher. I do need to realize, though, that there is a method to the madness that pulls those hits to me when I tell Google to "search". The syntax of the search query is all that Google has to go on when it looks for pages out there in the "clouds".

Information about the syntax is all available on Google itself. Just click on the Advanced Search link on the Google front page, and then Advanced Search Tips. The syntax items shown here are my favorites when doing genealogical searching.

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.

"andrew robeson"

Terms you want to exclude (-)
Attaching a minus sign immediately before a word/phrase 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.

"andrew robeson" –"robeson county"

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. Note that the * operator works only on whole words, not parts of words.

"andrew * robeson"   ➔ To allow for middle names. This will not find places where there is not a middle name.

Search exactly as is (+)
Google employs synonyms automatically, so that it finds pages that mention, for example, 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 attaching a + immediately before a word (remember, don't add a space after the +), you are telling Google to match that word precisely as you typed it. Putting double quotes around a single word will do the same thing. But remember, spelling errors will not be caught by Google in this case.

"andrew robeson" –"robeson county" +pennsylvania

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).

"andrew robeson" OR "andy robeson"

Numeric Range (..)
Google will search within a range of numbers. Put two periods between the numbers in question. There must not be any spaces between the numbers. These numbers can be dates! Use this option with other keywords to make the date range meaningful.

"andrew robeson" OR "andy robeson" born 1820..1830
"andrew robeson" OR "andy robeson" married 1840..1849

Specific document type (filetype:)
Find documents of the specified type. This will tell Google to look for files, like PDFs or DOC files, any legal file type will work. Suppose we are looking for gedcoms with my ancestor's name inside. Type something like this:

"andrew robeson" OR "andy robeson" filetype:ged

Syntax Order
Ranking favors pages with your words in phrases, close together and in the order typed. Order is important to the outcome of your search. If one phrasing doesn't work, try re-ordering the words.

"andrew robeson" born 1820..1830
vs.
born 1820..1830 "andrew robeson"
vs.
1820..1830 "andrew robeson" born

Similar words (~)
Using the tilde character ~ before a word will tell Google to search for words that are similar to that word. A good keyword to use for our genealogy searches is ~genealogy. That will tell Google to find pages that include words like: ancestry, family, family history, family tree, genealogical, genealogy, records, roots, surname, tree, vital records. This one may help when the original search returns too many records.

"andrew robeson" –"robeson county" – "paul robeson" pennsylvania ~genealogy

When my searches get long like this, I like to store them in a document I have set aside just for various searches I have tried. This allows me to cut-n-paste the next time I search. And it also helps me remember what I have already tried! And it makes adjusting the phraseology order easier.

From MOCOGENSO Newsletter dated 10/6/2009
By Jim Robeson