

Record how you search, not just what you find: Thoughtfully constructed search terms greatly enhance the reliability of digital research

blogs.lse.ac.uk/impactofsocialsciences/2013/06/10/record-how-you-search-not-just-what-you-find/

6/10/2013

The way in which digital search results are determined and displayed are continually changing and a lack of a defined approach can have significant repercussions on research. [M. H. Beals](#) recommends employing the Boolean search method because of the flexibility it provides in adjusting and recording search parameters. By creating a permanent record of how you obtained your search results, you can ensure that your methodology is consistent.



The advent of the world-wide-web has changed academic research forever. This is a simple, unavoidable truth. A [2012 study](#) by JISC and the British Library, for example, found that in most disciplines over 70% of doctoral students concluded their most recent query with an electronic journal article, rather than a hard-copy resource or primary data. It also found that the majority had used an electronic database, often [Google Scholar](#), to discover that source.

On the one hand, this is a great victory for the free (*libre*) dissemination of human knowledge. The current generation of scholars can access more information with the click of a mouse than their predecessors could have obtained in a lifetime of archival research. On the other, the inconsistent storage, dispersal and retrieval mechanisms for this information may largely negate its advantages.

There are two reasons for this. The first, largely out of our control, is the [shifting nature of search algorithms](#). In order to provide 'better' results for its users, Google and other search providers continually modify the way in which results are determined and displayed—their relevance to your search parameters. Because of this, even the most meticulously constructed search queries [will not provide replicable results](#). This has significant repercussions for projects that require a systematic review of the existing literature, and less costly but equally frustrating repercussions for any other form of literature review.

The second reason, wholly within our control, is a lack of meticulously constructed search queries. Traditional data collection demands rigorous adherence to an established, documented methodology. Whether recording readings within a laboratory, coding transcribed speech patterns or selecting archival records for an edited collection, researchers develop guidelines to regulate their decisions and maintain consistency within their data set.

As we move into the digital realm, and especially in our retrieval of published material, these methodologies are all too quickly forgotten. This unnecessarily biases our understanding of the existing literature, which can lead to awkward moments during a *viva* or conference, and ultimately prevents us from making effective use of the ever more diverse world of digital content. By thoughtfully constructing sets of search terms, and keeping detailed records of when and where they were used, we can greatly enhance the effectiveness, reliability and comprehensiveness of our digital research.

What we must remember, of course, is that a *thoughtful* set of search terms is not simply a list of words that relate to (or are a rephrasing of) our research question. The vocabulary must be carefully considered, but so too must the form.

First, some basic vocabulary. For the web savvy, mathematicians and engineers, the term Boolean (BOO-le-an) is a familiar one. For those less familiar with ins and outs of electronic database searching, it may be less so. If you are a complete novice, the University of Auckland had produced a straightforward video, available on [YouTube](#). The folks at CommonCraft have also produced a video on [Web Search Strategies](#) that you may find useful.

For those who cannot watch (or listen) to these at the moment, a Boolean search can be most simply described as one that uses the terms **AND**, **OR**, and **NOT** in order to narrow or limit a search. **AND** and **NOT** are also sometimes indicated by the use of a **+** or **-** symbol immediately before a word. Quotation (speech) marks can also be used in order to search for a specific phrase. For example:

“M. H. Beals” AND newspapers NOT migration

or

“M. H. Beals” +newspapers -migration

Boolean searching has a number of consequences for electronic research, but the most important, perhaps, is the flexibility it provides in adjusting and recording search parameters. Take, for example, the *Scotman*'s electronic archive.

ProQuest | ProQuest Historical Newspapers: The Scotsman (1817-1950)
Basic Search | Advanced | Obituaries | Publications

Advanced Search

Look Up Citation | Command Line | Find Similar

Field codes | Search tips

australia	OR	("new south wales")	OR	("van diemen's land")	in	All fields + text
	OR	("botany bay")	OR	new holland	in	All fields + text
	OR	("swan river")	OR	"new zealand"	in	All fields + text
	OR	("van diemen's land")	OR		in	All fields + text
	NOT	("naval intelligence")	OR		in	All fields + text

⊞ Add a row | ⊞ Remove a row

Search Clear form

Search options

Date range: Specific date range...
Search for a range of years, years and months, or specific dates

Start
1 January 1817 (yyyy)

End
31 December 1844 (yyyy)

Search subject areas
Use search forms customized for each subject.
Genealogy
Historical Newspapers

This ProQuest-based service offers users a powerful, but accessible, interface for limiting searches. However, the complexity of my search—which also included 'article types' limitations not shown here—meant that a typographical error or thoughtless omission required a tedious re-entering of these details. Yet, the open nature of the ProQuest search function offered a solution.

Once my search parameters had been entered, the URL in the browser's address bar displayed the following:

[http://search.proquest.com/hnpscotsman/results/139A19D38223DCAD31D/1/\\$bqueryType\\$3dadvanced:hnpscotsman\\$3b+sortType\\$3drelevance\\$3b+searchTerms\\$3d5b\\$3cAND\\$7ccit:dRANGE:1817-0-1,1844-11-31,+ftblock\\$3d55199+670835+670834+7+660829+199+55007+55000+670831+670828+660845+670829+660843+660840,+removeDuplicates\\$3dtrue\\$7d\\$3b+metaData\\$3d\\$7bUsageSe](http://search.proquest.com/hnpscotsman/results/139A19D38223DCAD31D/1/$bqueryType$3dadvanced:hnpscotsman$3b+sortType$3drelevance$3b+searchTerms$3d5b$3cAND$7ccit:dRANGE:1817-0-1,1844-11-31,+ftblock$3d55199+670835+670834+7+660829+199+55007+55000+670831+670828+660845+670829+660843+660840,+removeDuplicates$3dtrue$7d$3b+metaData$3d$7bUsageSe)

\$3dRecordType\$7d\$5d

I have highlighted the search terms above to demonstrate the syntax used. Although obviously cluttered by a great deal of additional information, the key terms are easily identifiable. More importantly, they are easily editable. By discovering where your terms are within the URL, you can quickly edit your search (correcting errors or refining details) without having to return to the search screen.

Understanding the URL also allows you to continue a single systematic search over multiple visits. ProQuest allows users to save searches, but many other databases do not. Moreover, saving the URL itself means you will have quick access to your search, rather than navigating through layers of option menus. Finally, it allows you to plan your search queries in a methodical fashion.

Other databases are less open with their search protocols. [British Library Newspapers](#), for example, takes the following advanced search

and returns the URL

<http://find.galegroup.com.lcproxy.shu.ac.uk/bncn/advancedSearch.do>

Unfortunately, with no Boolean syntax, the URL cannot be saved or refined. Nonetheless, there are some short-cuts to be found. By crafting an **AND—OR—NOT** statement such as

australia OR "new south wales" OR "van diemen's land" OR "botany bay" OR "new holland" OR "swan river" OR "new zealand" NOT "naval intelligence"

and placing it within the basic search box, a single line can still return carefully limited results.

As with the Boolean-laden URLs, these phrases can be stored and crafted into a comprehensive search methodology. This is particularly useful if you are searching for a resource in a variety of different repositories. By creating a spread-sheet or other permanent record of *how* you obtained your results, you can ensure that your methodology is consistent and (theoretically) replicable, both within a single database and as you move from one archive to another.

A similar version of this post appeared on M.H. Beals' personal blog and can be [found here](#).

Note: This article gives the views of the author, and not the position of the Impact of Social Science blog, nor of the London School of Economics.

About the Author

M. H. Beals is a Senior Lecturer in History at Sheffield Hallam University. She has published on Scottish Diaspora, and the continuing role of sending communities in the migration process, including her recent monograph *Coin, Kirk, Class and Kin: Emigration, Social Change and Identity in Southern Scotland*. She is currently researching the role of pre-Victorian newspapers networks in the development of imperial identity, as well as the role of digital methodologies in media history.

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.