EDUTORIAL

Focused Search of the Literature

Mitch Kapor, one of 1980s most renowned software gurus, said once that getting information off the Internet is like taking a drink from a fire hydrant. Digital development and a widespread culture of communication have pushed us to manage seemingly endless amounts of information. However, the skills needed to distinguish signal from noise and filter it down to a manageable level are often underappreciated, even by well published researchers.[1] This short article recalls some tips and tricks to make sense of the massive amounts of scientific information ("the water") pouring out of online sources ("the fire hydrant"), without compromising both precision and exhaustiveness.

SQUEEZING THE JUICE TO MEDLINE

From 1879 to 2004, the National Library of Medicine (NLM) of the United States published a monthly compilation of over 5,000 selected medical journals, the so called *Index Medicus*. This huge database, together with others covering nursing, dentistry and pharmacy, is available online from 1971 under the name of MEDLINE. The database contains around 30 million records indexed from 1950, and adds about a million new records yearly.

Searches can be conducted in MEDLINE from a variety of web interfaces. However, PubMed (its own NLM sponsored search tool) is nowadays the most popular interface[2] and offers a higher sensitivity in the search for systematic reviews.[3] More than 20 other interfaces exist and offer interesting features;[4,5] some are focused on using natural language (semantic search), answer structured clinical questions, cluster results, identify similar publications, or look for expert networking. For example, GoPubMed (http://gopubmed.org) and Medsum (http://webtools.mf.uni-lj.si/public/medsum.html) allow searching with the same architecture and commands as PubMed, but offering extra information on bibliometrics.

The interaction with PubMed (http://www.pubmed.gob) can take place using natural language, but with poor results. Indexers use a particular language with precise and selected terms (Medical Subject Headings, MeSH) to categorise and classify each new article that enters MEDLINE. Looking for the exact correspondence of our natural language idea in the MeSH database always enhances precision[6] and avoids conflicting with term variations, plurals, or synonyms[7] (Table 1). However, it requires a proper use from authors and journals when submitting papers. To look for a MeSH term, just select "MeSH" in the dropdown menu left of the search box; mark your term choice in the results and a new window will pop up, allowing to move the MeSH and selected subheading to the search box. Besides, it is also useful to look at how MeSH terms were assigned by the indexers to a relevant paper, and it can be done by clicking on the "MesH terms" tab right below the abstract.

Combining MeSH terms and other search strings can be accomplished by means of Boolean operators (AND, OR, and NOT) and nesting. This last one involves the use of parentheses, which will always be read by PubMed from left to right. Also, some codes between square brackets can help to target one specific field in the MEDLINE record (Table 2). Using the search tag "[ti]" is especially handy and can greatly improve the search focus by looking just at the title. Also the tag "[pt]" allows one to retrieve only a certain type of publication, such as case reports, reviews, practice guidelines, or clinical trials. This feature can also be set by using the filters on PubMed's left sidebar, but the small effort made learning to use tags is rewarded with quicker results and avoids the risk of forgetting to deactivate filters after being set.

PubMed provides a free NCBI account that allows saving searches;[8] as of January 2019, the login can also be done using Google Gmail credentials. PubMed will not send any email unsolicited by the user. Saving a search is useful for building a bundle of major journals within a knowledge field (e.g. eur j vasc endovasc surg[ta] OR j vasc surg [ta] OR j endovasc ther[ta] or vasc endovasc surg[ta] OR ann vasc surg[ta]). This search can be run and thereupon combined with others (e.g. #1 AND #2), allowing one to dive only into the pre specified group of journals. Furthermore, searches can be scheduled to be re-conducted every so often, receiving the new records as an email. This feature is convenient for keeping up to date during lengthy projects (e.g. PhD thesis) or to be automatically informed of new publications within a particular topic.

Other fire hydrants to drink from

Aside from MEDLINE, some other databases deserve a look when trying to be exhaustive.[9] Bramer et al.[10] and Falagas et al.[11] found that a combination of searches in MEDLINE, Embase, Web of Science Core Collection, and Google Scholar offers the best performance for a systematic review approach.

Embase (Excerpta Medica Database) is an European database with over 2,800 journals not included in MEDLINE and indexed from 1970. It can be accessed standalone or integrated in Scopus, a platform created by Elsevier (Elsevier, Amsterdam, The Netherlands) in 2004. Both are commercial service providers and require a subscription,

1078-5884/
https://doi.org/10.1016/j.ejvs.2019.03.038

Table 1. Example of search trimming. Looking for first class evidence on hand sanitizing with alcoholic solutions in PubMed

Search term	No. of results	Explanation
hand washing	8239	First approach in natural language
"hand disinfection"	5429	Search for a specific word block
"hand disinfection"[ti]	228	Search for a specific word block, only in the title
"Hand Disinfection"[Mesh]	5246	MeSH term
"Hand Disinfection"[Mesh] AND alcohol	973	Focus on alcohol based solutions
"Hand Disinfection"[Mesh] AND alcohol[ti]	309	Focus on specific papers with "alcohol" in the title
"Hand Disinfection"[Mesh] AND alcohol*[ti]	337	28 extra results accounting for "alcoholic" in the title
"Hand Disinfection"[Mesh] AND "Ethanol"[Mesh]	296	Additional trimming using the MeSH term for alcohol
("Hand Disinfection"[Mesh] AND "Ethanol"[Mesh]) AND clinical trial[pt]	57	Focus on clinical trials (RCTs)

MeSH = Medical Subject Heading; RCT = randomised controlled trial.

but they can often be accessed through institutional repositories. Compared with PubMed, Scopus almost doubles the number of entries (over 54 million as of 2018), including books, book chapters, and conference abstracts; around a 25% of these entries cannot be found elsewhere. Scopus excels when looking for literature published outside the United States and in non-English languages. It also provides useful bibliometric graphics under the area "Analyze search results", concerning type of publication, country, affiliation, author, and temporal trend.

The search interaction with Scopus can be done in a similar way to PubMed, by means of Boolean operators, nesting, quotation marks, and field tags. Some singularities are described in Table 2. For example, an equivalent search to that in Table 1 regarding hand disinfection would be TITLE-ABS-KEY("Hand disinfection") AND INDEX-TERMS(Alcohol), and it provides 321 results (vs. 296 in PubMed). A very useful trick is adding the sequence "AND NOT INDEX(medline)", that excludes results from PubMed. Incorporating this into the search string above offers 44 documents, including five notes, four reviews, three conference papers, and three letters. Of these all, 29 are written in English, 15 in German, and three more in French, Bulgarian, and Spanish.

The Web of Science (WoS**)** is an online database created by Thomson Reuters (Thomson Reuters Corp., Toronto, Canada) back in 2004, integrated in the ISI (Institute for Scientific Information) Web of Knowledge (WoK). ISI WoK is well known thanks to the Journal Citations Report, the tool that allows calculation of the Journal Impact Factor, highly criticised but still a standard in bibliometrics.[12] WoS includes the SciELO database, focused on the Latin American and Brazilian environment. However, the search syntax is slightly limited and it does not include pre-print versions of

Table 2. Boolean operators and other advanced search tags to interact with databases

Definition	PubMed operator	Web of Science operator	Scopus operator	Google Scholar operator
Each citation contains all the search terms (both A and B), and not only one of them	AND			
Each citation contains at least one of the search terms (search A + search B)	OR			
Eliminates from the results all the citations with the second search term (A − B)	NOT		AND NOT	-(minus sign)
Shows only exact matches with the content of the quotations marks (e. g. "Gore Excluder")	"…"			
Wildcard symbol: allows multiple spellings of a word (e. g. sul?ur for sulphur or sulfur)	?			n/a
Truncation symbol: uses just the root of the word (e. g. aneury* for aneurysm, aneurysmal, aneurysmatic…)	*			n/a
Searches in a particular tag of the article's record:				
- Author (e. g. debakey me[au])	[au]	AU=	AUTH	author:"…"
- Searches only in the title of the paper (e. g. gore excluder[ti])	[ti]	TI=	TITLE	n/a
- Searches only in a particular journal (e. g. eur j vasc endovasc surg[ta] AND phlebography[ti])	[ta]	SO=	SRCTITLE	n/a
- Searches for publications in a given year (2010[dp]) or range (e.g. 2000:2005[dp])	[dp] *1995:2000	PY = * 1995−2000	PUBYEAR * No range, just > or <	Menu
- Offers results of a given article type (e. g. review, practice guideline, case report, clinical trial)	[pt]	n/a	n/a	n/a
Nesting: processes the terms inside the parentheses as a unit (block)	(…)			n/a

n/a = not available.

articles. The most remarkable feature is that it indexes papers form 1900, being especially convenient for the retrieval of results before 1945, that cannot be found elsewhere (not PubMed nor Scopus). An interesting example using this feature is identifying articles published by Michael Ellis DeBakey before 1950.

PubMed debakey me [au] AND 1900:1950 [dp] 26 results, earliest 1945

WoS AU=(debakey me) AND PY=(1900−1950) 38 results, earliest 1934

Finally, Google Scholar was also launched in 2004 and offers additional pre-prints, theses, news, and press releases, diving deep into the so called gray literature.[13] Unlike the other search tools described here, it is a machine curated database and it is unclear how many items it includes per search as the directory behind is the whole Internet. It makes no use of tags, index terms, or controlled language (Table 2). In return, it allows sorting the results by relevance (which is not exactly equivalent to the number of citations used in Scopus or WoS), a significantly greater access to full text resources, and twice as many relevant articles, with similar precision.[14] Despite this broadness, most experts do not recommend using it as a standalone search tool.[15]

REFERENCES

1 Corrao S, Natoli G. The case of homeopathy, "how to search PubMed" may be a first step. *Eur J Intern Med* 2017;**41**:e8−9.
2 De Leo G, LeRouge C, Ceriani C, Niederman F. Websites most frequently used by physician for gathering medical information. *AMIA Annu Symp Proc* 2006;**2006**:902.
3 Boeker M, Vach W, Motschall E. Semantically equivalent PubMed and Ovid-MEDLINE queries: different retrieval results because of database subset inclusion. *J Clin Epidemiol* 2012;**65**:915−6.
4 Keepanasseril A. PubMed alternatives to search MEDLINE: an environmental scan. *Indian J Dent Res* 2014;**25**:527−34.
5 Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)* 2011;**2011**:baq036−6.
6 Kim S, Yeganova L, Wilbur WJ. Meshable: searching PubMed abstracts by utilizing MeSH and MeSH-derived topical terms. *Bioinformatics* 2016;**32**:3044−6.
7 Chapman D. Advanced search features of PubMed. *J Can Acad Child Adolesc Psychiatr* 2009;**18**:58−9.
8 Fatehi F, Gray LC, Wootton R. How to improve your PubMed/MEDLINE searches: 3. advanced searching, MeSH and My NCBI. *J Telemed Telecare* 2014;**20**:102−12.
9 Goossen K, Tenckhoff S, Probst P, Grummich K, Mihaljevic AL, Büchler MW, et al. Optimal literature search for systematic reviews in surgery. *Langenbecks Arch Surg* 2018;**403**:119−29.
10 Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Syst Rev* 2017;**6**:245.
11 Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *FASEB J* 2008;**22**:338−42.
12 Kiesslich T, Weineck SB, Koelblinger D. Reasons for journal Impact factor changes: influence of changing source items. *PLoS ONE* 2016;**11**:e0154199.
13 Paez A. Gray literature: an important resource in systematic reviews. *J Evid Based Med* 2017;**10**:233−40.
14 Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Haynes RB, Weyr MA, et al. Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *J Med Internet Res* 2013;**15**:e164.
15 Boeker M, Vach W, Motschall E. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. *BMC Med Res Methodol* 2013;**13**:131.

Francisco Álvarez Marcos
Department of Vascular Surgery, Asturias University Central Hospital (HUCA), Oviedo, Spain
Email-address: alvarezmarcos@seacv.es