

Keywords, discoverability, and impact

Tanja Bekhuis, PhD, MS, MLIS, AHIP

DOI: <http://dx.doi.org/10.3163/1536-5050.103.3.002>

Keywords: Keywords, Discoverability, and Impact

Editors' Note: Keywords will improve the article impact and are now necessary for Journal of the Medical Library Association articles. Here is a brief editorial with background information.

When the editor of the *Journal of the Medical Library Association (JMLA)* asked me to write a piece about author keywords in MEDLINE structured abstracts [1], the first thing I did was search for “keywords” in Google and Google Scholar. This exercise was reminiscent of the lithograph *Drawing Hands* by M. C. Escher [2]. Think about it. I used Google, an über-web search engine, to write keywords to find keywords. Google and Google Scholar returned a deluge of information (721 million and 4.35 million hits, respectively). Not surprisingly, at the top of the first page in Google were hits for Google AdWords and their Keyword Planner Tool [3].

I then did more focused searches in the ACL Anthology, a digital archive of papers in computational linguistics, and in *Scientometrics* or journals with similar coverage to confirm that keyword analysis is thriving in the text-mining and bibliometrics communities; for example, see Ventura and Silva [4] or Yao et al. [5].

Despite the circularity of my initial searches and too narrow follow-up attempts, I learned that the meaning of the concept varies depending on the domain. For example, in the search engine optimization (SEO) domain, keywords are terms that improve page rank. Shrewd selection and placement of words or phrases visible to the user or buried in hypertext markup language (HTML) can move a hit toward the top of a list returned by a search engine. Regarding these terms, the world of SEO has some curious neologisms, such as spamdexing, which refers to keyword stuffing, search engine spam, or black-hat SEO [6]. In contrast, white-hat SEO is ethical; its practitioners eschew black-hat techniques. In corpus linguistics, keywords discriminate between collections of documents to identify what is unique about, say, general versus scientific prose, or British versus American English [7]. Text miners and other computational scientists extract informative keywords to classify documents or improve retrieval.

This brings us to why you, as an author, should carefully consider the list of keywords that you will assign to your *JMLA* article and its relationship to your title and abstract. Think of optimization principles for discoverability of your article beyond MEDLINE and potential effects on the impact of your work. Overall, enhancing discoverability of *JMLA* articles should improve journal visibility, subsequent

citation counts, and its impact. This is a desirable outcome for you and the profession.

Discoverability could depend on how well the title, abstract, and keyword list form a miniaturized version of your paper. This is why a good structured abstract resembles a paper written in the “Introduction, Methods, Results And Discussion” (IMRAD) format (see Cooper’s editorial in the April 2015 *JMLA*). The title includes the most important concepts in your paper and, ideally, the study design; the abstract summarizes the components of your paper; and the keyword list includes relevant concepts but with more detail than in the title. If keywords are too broad or too narrow, they are useless. All three pieces are important because web search engines and text-mining applications target these sections and sometimes overweight text, depending on location. Additionally, when presented to the reader, the title, abstract, and keyword list must be laden with relevant information to capture attention.

To write the keyword list for the *JMLA*, channel your inner indexer. Select the Medical Subject Headings (MeSH) that best characterize your topic to improve retrieval in MEDLINE [8]. Additionally, find words and phrases not covered by MeSH but known to practitioners and researchers in your field. The MeSH terms you proffer could improve decisions that a National Library of Medicine human indexer makes—after all, you are likely to know more about the topic of your paper than the indexer does. Adding non-MeSH terms could improve discoverability of your article by web search engines and by users who search digital repositories aside from PubMed and PubMed Central.

For example, in a recent paper we wrote for the *JMLA* on building gold standard datasets as a prelude to developing search filters [9], my coauthors and I reported that “oral squamous cell carcinoma” is not covered by MeSH, even though it is the most common cancer of the oral cavity. However, the term is a synonym for “mouth squamous cell carcinoma” in Emtree, the controlled vocabulary for Embase [10]. It also appears in the National Cancer Institute Thesaurus as “oral cavity squamous cell carcinoma” [11]. Any of these terms would have been good keywords for our paper.

In sum, if terms from controlled vocabularies beyond MeSH seem useful, consider adding them to your keyword list. Additionally, consider free-text terms for which users are likely to search. By carefully constructing your title, abstract, and keyword list, you will enhance discoverability of your article and its potential impact.



Tanja Bekhuis, PhD, MS, MLIS, AHIP, tcb24@pitt.edu, Assistant Professor, Department of Biomedical Informatics and Department of Dental Public Health, School of Medicine and School of Dental Medicine, University of Pittsburgh, 5607 Baum Boulevard, Suite 514, Pittsburgh, PA 15206-3701

REFERENCES

1. US National Library of Medicine, National Institutes of Health. Structured abstracts [Internet]. Bethesda, MD: The Library [cited 18 Nov 2014]. <http://www.nlm.nih.gov/bsd/policy/structured_abstracts.html>.
2. Escher MC. Drawing hands [Internet]. Lithograph, 332mm x 282mm. 1948 [cited 25 Feb 2015]. <<http://www.mcescher.com/gallery/most-popular/drawing-hands/>>.
3. Google Ads. Google AdWords keyword planner [Internet]. Google; 2015 [cited 25 Feb 2015]. <<https://adwords.google.com/KeywordPlanner>>.
4. Ventura J, Silva J. Automatic extraction of explicit and implicit keywords to build document descriptors. In:

- Correia L, Reis LP, Cascalho J, eds. Progress in artificial intelligence. Springer; 2013. p. 492–503.
5. Yao Q, Chen K, Yao L, Lyu PH, Yang TA, Luo F, Chen SQ, He LY, Liu ZY. Scientometric trends and knowledge maps of global health systems research. Health Res Policy Syst. 2014 Jun 5;12(1):26.
6. Weideman M. Website visibility: the theory and practice of improving rankings. Chandos Publishing, Elsevier; 2009.
7. McEnery T, Hardie A. Corpus linguistics: method, theory and practice. Cambridge, UK: Cambridge University Press; 2011.
8. US National Library of Medicine, National Institutes of Health. MeSH: Medical Subject Headings [Internet]. Bethesda, MD: The Library [cited 28 Oct 2014]. <<http://www.nlm.nih.gov/mesh/>>.
9. Frazier JJ, Stein CD, Tseytlin E, Bekhuis T. Building a gold standard to construct search filters: a case study with biomarkers for oral cancer. J Med Lib Assoc. 2015 Jan;103(1):346–54. DOI: <http://dx.doi.org/10.3163/1536-5050.103.1.005>.
10. What is Emtree? [Internet]. Elsevier BV [cited 28 Oct 2014]. <<http://www.elsevier.com/online-tools/embase/about/emtree>>.
11. US National Cancer Institute. NCI thesaurus [Internet]. Bethesda, MD: The Institute [cited 28 Oct 2014]. <<http://ncit.nci.nih.gov>>.

Your Library—eImagined for Health Sciences

- 24/7 user access to health science database
- eBook Database for fast content search
- Purchase and Patron Driven Acquisition models available
- Pay only once—for the life of the edition

Visit R2Library.com
 Start your free 30 day trial today!

R2 DIGITAL LIBRARY

Visit R2Library.com to learn more
 Or call us at 800.345.6425 x600

Rittenhouse Book Distributors, Inc.
 511 Fehely Drive | King of Prussia, PA 19406
www.rittenhouse.com