



“Cannabis” is a Blue State Word: Marijuana Decriminalization, Keyword Development, and Considering Political Contexts in Search Results

Kevin Seeber and Rachel Stott*

Medical researchers call it “cannabis.” Criminologists call it “marijuana.” Online discovery tools search both terms simultaneously. So how should librarians talk about this, or other politicized topics, when discussing keywords with students? What about when implementing discovery tools with proprietary search algorithms? This paper presents a systematic review of scholarly literature using the terms “cannabis” and “marijuana” to determine the context that has guided what words are used to describe this drug over time. This review’s implications go far beyond one example, however, and speak to the larger need for librarians to be critical of how we describe politically-charged topics.

Introduction

In the summer of 2017, the Auraria Library teaching team set out to develop a shared lesson plan that could be used across multiple sections of a first-year writing course at Metropolitan State University of Denver (MSU Denver). Auraria Library is part of the Auraria Campus, a single location in downtown Denver, Colorado that serves three separate institutions of higher education: the Community College of Denver (CCD), MSU Denver, and the University of Colorado Denver (CU Denver). Building on previous success our department had realized with the first-year writing program at CU Denver,¹ we met with colleagues in the English Department at MSU Denver to determine what learning outcomes would be most appropriate for us to address in a new shared lesson plan for their courses. Based on those conversations, our department concluded that the main focus of the sessions would be on identifying keywords and synonyms, as well as demonstrating an overview of search mechanics in multiple library databases.

This main emphasis of identifying keywords from a research topic or question is very much in-line with established library instruction practice. The ACRL *Information Literacy Competency Standards* state that an information literate student “[i]dentifies keywords, synonyms and related terms for the information needed.”² Likewise, the ACRL *Framework for Information Literacy for Higher Education* calls for learners to “use different types of searching language (e.g., controlled vocabulary, keywords, natural language) appropriately.”³ From the perspective of our teaching team, a first-year instruction session focused on “keywords” seemed consistent with the current professional discourse and our institutional priorities, and we set about creating a lesson plan that spoke to this knowledge practice.

In the fall semester that followed these planning meetings, our department taught more than 40 instruction sections using the shared lesson plan. To address the outcome related to identifying keywords, one of the

* Kevin Seeber is Education and Outreach Services Department Head at the University of Colorado Denver’s Auraria Library, kevin.seeber@ucdenver.edu; Rachel Stott is a Teaching and Learning Librarian at the University of Colorado Denver’s Auraria Library, rachel.stott@ucdenver.edu

examples we presented to students was a research question that has become familiar to our department: “How has legalizing the sale of recreational marijuana impacted drug crime in Colorado?” Very quickly our assessments revealed that students had little trouble identifying the main concepts in this research question. Instead, we found that much more productive discussions came from identifying synonyms for these concepts, in particular “recreational marijuana.” Students were quick to name informal terms such as “pot” or “weed,” but often a conversation was needed before someone would offer the term “cannabis.” Through these discussions we were able to ask students what kinds of information they would expect to find associated with these different terms for the same substance, which was an exercise that yielded interesting results.

In general, most students (and our department) predicted that “cannabis” would be associated more with medical research, and that “marijuana” might be associated more with criminology or legal research. This assumption reinforced our belief that identifying the contexts around the vocabulary we use in research is an important part of critically thinking about information and determining its relevance to our research need. At the same time, we found ourselves asking several questions: What happens when perceptions of a substance change over time? Does the vocabulary used in scholarly research evolve with those perceptions? And even when we discuss context around search terms, what happens as more and more search tools are automatically applying related terms when we run a search, potentially removing some of the contextual choices being made by the searcher? We set out in this paper to answer these questions, as well as connect these concerns to the broader discourse surrounding search algorithms, recommended language, and how concepts change over time.

To start, we looked at the concept of “query expansion,” which is the process by which an information retrieval system expands the scope of a search by automatically applying related search terms to the original query.⁴ In practice, this means a user could enter a search term like “cars” and the system would automatically search with a related term, such as “automobiles.” But in Colorado, a state that had legalized recreational cannabis before the rest of the United States, we found ourselves asking about specific words and how related terms might carry different political connotations. Would searching for “cannabis” yield different results than a search for “marijuana?” And how would those results convey an underlying sentiment to researchers using our library search tools?

At the outset, we recognized that library workers are not the only people thinking about this particular topic and considering what words to use. Looking to journalism, we found that as recently as 2015 the *Associated Press Stylebook*, which is used by most print media outlets, did not provide guidance on differentiating between drugs, marijuana, cannabis, or recreational versus medicinal uses. Instead, this edition included only a generic entry for “drugs.”⁵ The 2018 edition of the *AP Stylebook*, however, has been updated to include a specific entry for “marijuana, medical marijuana.”⁶ This entry includes descriptions of different parts of the cannabis plant, as well as suggestions for discussing the chemical compounds in marijuana, along with the geographic implications of using marijuana (more popular in the U.S.) and cannabis (more popular internationally). As it relates to this paper, this update to the *AP Stylebook* reflects the changing laws, attitudes, and public awareness of cannabis, and indicates that it is no longer enough to refer to “marijuana” as a “drug.” Context dictates the way that this substance is written about in the U.S. media, and these implications extend to the scholarly sources often cited by college students.

Regarding context, the decriminalization of cannabis has in many ways come to be a stand-in for political polarization in the U.S, with more liberal-leaning states voting to allow for recreational use of the drug. Beginning with the passage of Amendment 64 and Initiative 502 in 2012, cannabis was decriminalized for recreational use in the states of Colorado and Washington, respectively.⁷ Subsequently, some seven states and the District of Columbia have decriminalized cannabis for recreational use, with the 2016 election serving as a breaking point in the debate surrounding legalization. All of these recreational developments have occurred in addition to the 37 states which have legalized some form of medicinal marijuana, going back to the state of California in 1996, when Proposition 215 passed.⁸

The purpose of this study is to determine how the words “cannabis” and “marijuana” have been used in the scholarly literature over a two-decade period, with an eye toward determining how updates to search tools could be changing the ability of database users to locate information relevant to their needs and contexts.

Methods

To better understand how the vocabulary around this topic has shifted over time, we conducted a systematic review of scholarly literature describing either “marijuana” or “cannabis” that was published between 1996 and 2016. These dates correspond with the initial legalization of marijuana for medical purposes in California in 1996 and extend through the 2016 U.S. election. This review was done using EBSCO’s Academic Search Premier database. We selected this database for three reasons: First, Academic Search Premier is a multidisciplinary database, allowing us to review articles from a wide range of scholarly disciplines. Second, Academic Search Premier’s “Advanced Search” functionality allows for users to turn off the “apply related terms” function (EBSCO’s version of query expansion), which subsequently allowed for us to search with the exact terms we entered. Third, the database has an option to export article-level metadata in bulk, allowing us to collect metadata from several thousand articles relatively quickly.

When running the searches to complete the systematic review, the keyword “cannabis” was entered as the search term and the field was limited to “Abstract.” This was done to eliminate articles that may have mentioned cannabis in the full text, but were not focused specifically on the substance. The search results were also limited to “Scholarly (Peer Reviewed) Journals.” Finally, only one year was searched at a time (1996, 1997, 1998, etc.) to gather article-level metadata by year. At the completion of each search, the metadata was downloaded and imported into an Excel spreadsheet, organized by year. The same search protocols were repeated for all 21 years using “marijuana.”

Next, we cleared the dataset of any entries that were not actual research articles, including conference proceedings, letters to the editor, book/product reviews, and other types of articles that had not undergone the peer-review process. Additionally, we removed any results that had a page count of two or fewer, as we assumed these would be “research briefs” or other miscellanea. We also removed articles that were written in a language other than English. The result was a set of 10,479 articles (5,643 in the “cannabis” search and 4,836 in the “marijuana” search). With this full set of articles, we tracked the frequency with which marijuana and cannabis appear in the search results throughout the 21-year time period to determine vocabulary trends in the scholarly conversation.

For additional analysis, we selected a representative sample of articles. To obtain the sample, articles were arranged alphabetically by title and numbered; we then selected every 20th article (starting with #1, then #21, then #41, and so on) for each year, resulting in 540 articles, representing 5.2% of the total dataset. With this smaller subset of article-level metadata, we coded for: term used to describe the substance (“marijuana” or “cannabis”); application of the substance; geographic location of research; research/academic discipline; and sentiment toward the substance. Each researcher coded 540 articles on their own, then met to find consensus and re-examine any coding categories where there was initial disagreement.

TABLE 1
Coding: Sentiment

Positive	The research discussed the substance in a largely positive way, including medicinal and therapeutic benefits.
Negative	The research discussed the substance in a negative way, including references to addiction, abuse, detrimental health effects, social disorders, risks to public health and safety, or other generally bad things.
Mixed	The research discussed the substance in terms of inconclusive findings, balance of benefit and harm, doesn’t adjudicate on the benefits or drawbacks of cannabis.

TABLE 2
Coding: Application of Substance

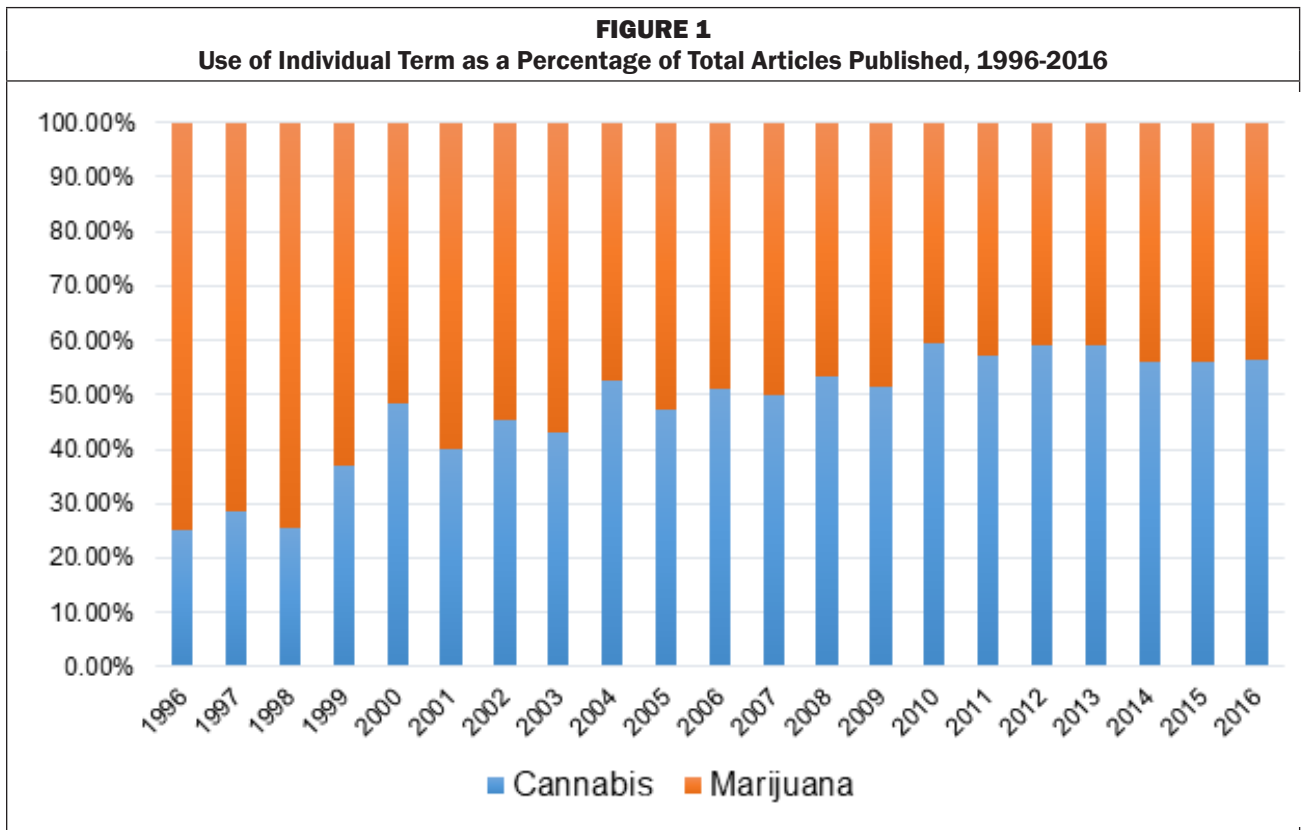
Medical	Specifically mentioned being prescribed or its utility as a therapeutic intervention
Recreational	Other use where medical advice/treatment wasn't mentioned, including "self-medication"
Criminal	Talked about arrests, detection, or seizure
Laboratory	Discussed experiments that dealt only with chemical compounds of substance or testing on non-human subjects
Agriculture	Discussed cultivation of cannabis or related crops, such as hemp

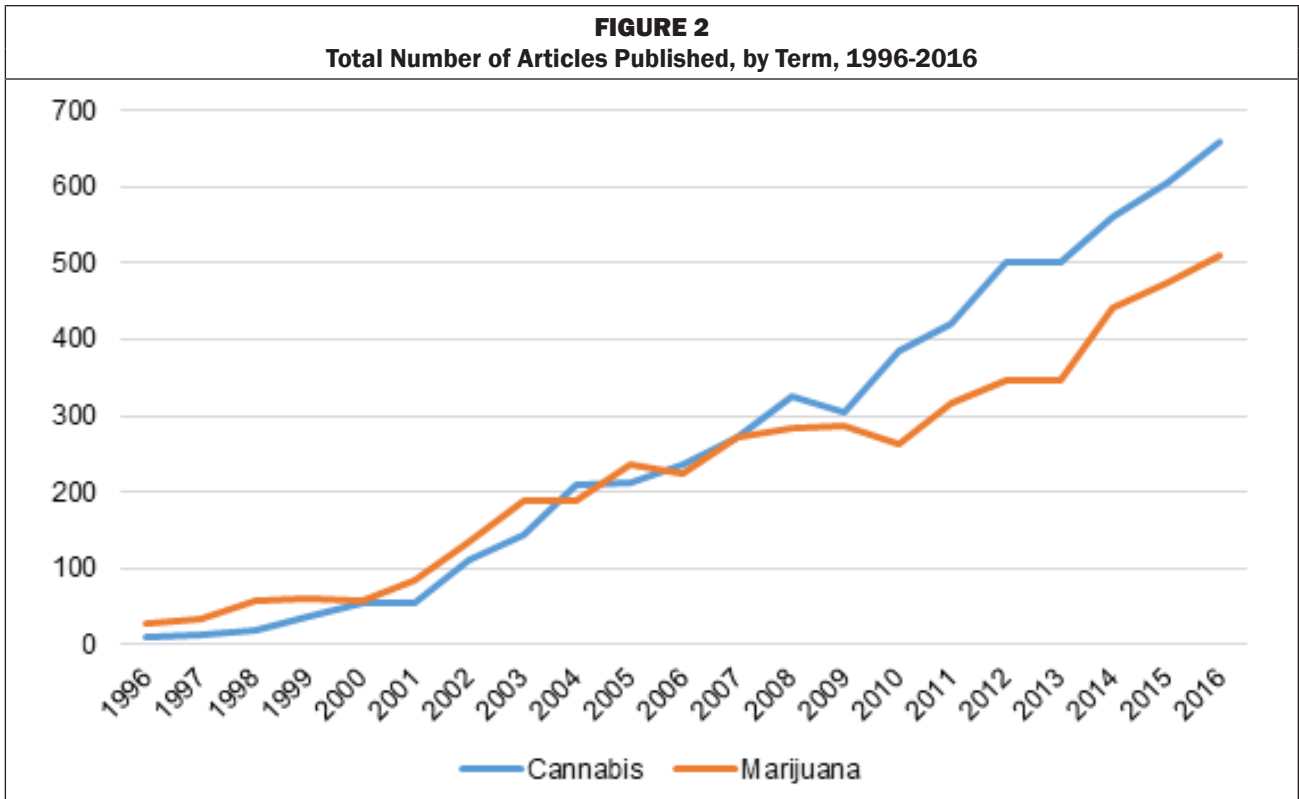
TABLE 3
Coding: Site of Research*

United States	Specifically mentioned in the article title, subject headings, or abstract the research being completed in the U.S.
International	Specifically mentioned in the article title, subject headings, or abstract the research being completed in a country outside the U.S., or the U.S. and an additional country(ies)

**If location where research was completed was not apparent from article title, subject headings, or abstract, the entry was not coded for site of research*

Disciplines were assigned based on each article's title, abstract, and subject headings. The final list of disciplines includes: Agriculture, Biology, Business, Chemistry, Counseling, Criminology, Education, Environmental Science, Forensics, Gender/Sex, Law, Medicine, Philosophy, Policy, Psychology, Public Health/Safety, Religion, Sociology, Sports, and Substance Use.





Results

Frequency of Term Use

We found that in the years 1996 through 2003, “marijuana” was more commonly used in academic literature than “cannabis.” In the mid-2000s, use was more or less equal (both coming up 271 times in 2007). But from 2008 forward, cannabis returned more search results than marijuana. We also saw a substantial increase in publications for both terms beginning in 2012, corresponding with the legalization of recreational marijuana in Colorado and Washington state. Specifically, the number of articles mentioning “cannabis” in their abstract increased from 501 to 660 (32% increase), while the “marijuana” articles increased from 346 to 509 (47% increase).

Application of Substance

For both search terms, the vast majority of scholarly articles (76% with “marijuana” and 70% with “cannabis”) focused on recreational use of the substance. Articles describing medicinal applications account for 12% of the marijuana articles, whereas they were only 10% of the articles in the cannabis dataset. Studies detailing criminal use accounted for 6% of the marijuana articles and 3% of the cannabis articles. We also coded for articles that discussed the substance in contexts where it was not consumed, such as agricultural and laboratory settings. The full results are included in Table 4.

TABLE 4
Application of Use

Use	Marijuana (n 250)		Cannabis (n 290)	
	Count	Percentage	Count	Percentage
Recreation	190	76.00%	202	69.66%
Medical	31	12.40%	28	9.66%
Criminal	14	5.60%	10	3.45%
Laboratory	13	5.20%	39	13.45%
Agriculture	2	0.80%	11	3.79%

Geographic Location of Research

Not every article included a geographical reference in their title, subject headings, or abstract. For “marijuana” articles, 217 mentioned a location; of these, 168 (77%) were focused on the U.S. exclusively, with the remaining 49 (23%) addressing non-U.S. settings. For the “cannabis” dataset, 147 articles included a geographical reference, with only 38 (26%) focusing on the U.S., while 109 (74%) discussed settings outside of the U.S.

Research/Academic Discipline

The three most common disciplines associated with both cannabis and marijuana were Substance Use, Medicine, and Psychology. Substance Use appeared in nearly equal proportions across both terms (34.00% of marijuana articles and 33.79% of cannabis articles were coded as Substance Use). Medicine was the next most common discipline for both search terms, though it showed up more often in cannabis results (31% for cannabis compared to 22% for marijuana). Finally, 14% of marijuana articles were coded for Psychology, while 12% of cannabis articles were also coded in this discipline. All Agriculture articles fell into the cannabis dataset (2% of cannabis articles), and the majority of Criminology articles were in the marijuana set (5% compared to 2%). Public Health/Safety articles were also more likely to use marijuana (6% compared to 3%), while the Forensics discipline was more likely to use cannabis (4% compared to 1%). The full results are included in Table 5.

TABLE 5
Marijuana and Cannabis by Discipline

Discipline	Marijuana (n 250)		Cannabis (n 290)	
	Count	Percentage	Count	Percentage
Agriculture	0	0.00%	7	2.41%
Biology	0	0.00%	6	2.07%
Business	2	0.80%	0	0.00%
Counseling	1	0.40%	0	0.00%
Chemistry	0	0.00%	1	0.34%
Criminology	12	4.80%	2	2.41%
Education	7	2.80%	1	0.34%
Environmental Science	0	0.00%	1	0.34%
Forensics	3	1.20%	13	4.48%
Gender/Sex	9	3.60%	3	1.03%
Law	7	2.80%	1	0.34%
Medicine	57	22.80%	89	30.69%
Policy	9	3.60%	16	5.52%
Psychology	34	13.60%	34	11.72%
Public Health/Safety	16	6.40%	8	2.76%
Religion	1	0.40%	1	0.34%
Sociology	5	2.00%	4	1.38%
Sports	1	0.40%	0	0.00%
Substance Use	85	34.00%	98	33.79%

Sentiment toward the Substance

Overall, articles about marijuana had more negative sentiment than articles about cannabis. Of the 250 articles coded for marijuana, 150 (60%) discussed marijuana and/or its effects negatively; 82 (33%) had mixed sentiment; and 18 (7%) articles were positive. We coded 290 articles in the cannabis set: 142 (49%) had negative sentiment; 115 (40%) were mixed; and 33 (11%) were positive.

Discussion and Implications

The results of this study challenged some of our assumptions and also provide us with evidence for making changes to our work as librarians. As it relates to these two words, there were clear differences in the geographic setting of the research, with the term “cannabis” returning nearly three times as many international articles as it does articles focused only on the U.S. The inverse was true with “marijuana” as well, with 78% of the results found with that term focusing on the U.S. In our teaching practice, we need to be deliberate in our discussion of geographical context and how different search terms could drastically alter what kinds of locations one will see represented in their search results.

Regarding the use of these two terms in different academic disciplines, we were surprised to see the lack of uniformity in the published literature. Specifically, publications in the fields of Substance Use and Psychology use the words “cannabis” and “marijuana” somewhat interchangeably. In the case of Substance Use literature, it represented approximately one-third of all published literature we studied for both terms. Although it is not addressed in this analysis, future research could focus on this discipline and others to determine what other factors might be guiding the use of one term over the other. We were also surprised that the differences in word choice were not more pronounced in the areas of Criminology and Medicine. Although medical articles were more likely to use the word “cannabis,” they still represented nearly a quarter of the articles in the “marijuana” dataset.

Perhaps the most notable result we observed was the marked increase in scholarly publication around this topic between 1996 and 2016. Although confounds such as limited metadata for older articles might explain the smaller number of results from the 1990s and early 2000s, there was a tremendous increase just between 2012 and 2016, with the total number of articles for both terms growing from 847 to 1169, an increase of 38% in just five years. This growth in research can likely be attributed to regulatory changes in the U.S., but it was still surprising to us and we are curious to see if the trend will continue as legalization spreads. And while we did not do an in-depth analysis on change in sentiment over time, increases in research on therapeutic applications and public policy implications could result in more positive sentiments associated with both marijuana and cannabis literature in the future.

Lastly, these results also demonstrate that, frankly, words matter. Instruction librarians spend a lot of time talking about “keywords” during information literacy sessions, and the present study reinforces our belief that doing so is worthwhile. That said, we are even more convinced that we must explore broader political context around specific words and acknowledge to students that even when words might seem to be interchangeable, there are systems of power that guide how different terms are applied in different settings. We conducted this study looking at a (sometimes) illicit substance and saw that there are indeed differences in how it is described. This does not end with cannabis, however. We live in a culture where politics and identity collide, leading to under-representation and marginalization of groups of people. “Cannabis” and “marijuana” might be used interchangeably at times, but our analysis shows variations in the sentiment attached to these words. This same concept very likely applies when considering the words used to describe people who are marginalized, especially as it relates to their race, ethnicity, gender, sexuality, class, ability-status, and other identities.

Which brings us back to “query expansion,” the technology deployed by database vendors to automatically search for related terms. Work by Safiya Noble has already demonstrated the concept of algorithmic bias—the

mechanism through which online search tools reinforce systems of oppression that exist elsewhere in our culture.⁹ If we accept that the different terms used to identify individuals convey different sentiments, what does that mean when library search tools return results containing language that is not consistent with the searcher's identity? Put another way, we could give comprehensive instruction around this topic and tell students to be thoughtful about the words they select, but with our current systems a student who identifies as "Chicano" could see results come back that discuss "Hispanics." This is just one example of how perceived synonyms connected to identity could be problematic, but there are myriad other possibilities of how this technology could reinforce existing societal injustices. This is similar to other discussions in the profession surrounding Library of Congress Subject Headings.¹⁰

Based on this analysis, we would ideally like for online search tools to publish their thesauri and inform users when a search algorithm has applied synonyms and other related search terms. Here is where we run into concerns with the "black box" approach to developing search algorithms, however. The purpose of these thesauri is to improve search quality, essentially making them "trade secrets," it is highly unlikely that vendors will share them with users. With that in mind, as instruction librarians we must continue to discuss algorithms and applied synonyms with our students, building awareness of how these systems work and how they might impact an individual's research. Further, we can continue to work with our colleagues in electronic resources and metadata positions to build solutions and customize search interfaces where possible to make it easier for our users to control search behavior. Being able to pick and choose among related terms is not necessarily a new concept within databases—"explode" functionality in search tools such as PubMed have allowed for this type of searching previously, but it has not made the migration to web-scale discovery systems.

In conclusion, this process of designing a lesson plan and exploring its implications sparked a lot of broader discussions for our teaching team. The keywords we use when conducting research do matter. Synonyms can have different political, socio-economic, and cultural meanings, and continuing to discuss these issues—and how they may have changed over time—with our students is important. Helping students understand how search interfaces employ query expansion and the potential ways it impacts search results should be part of this instruction, as well.

Endnotes

1. Zoe Fisher and Kevin Seeber, "Finding Foundations: A Model for Information Literacy Assessment of First-Year Students," *In the Library with the Lead Pipe*, (2017), <http://www.inthelibrarywiththeleadpipe.org/2017/finding-foundations-a-model-for-information-literacy-assessment-of-first-year-students/>
2. Association of College and Research Libraries, *Information Literacy Competency Standards for Higher Education*, (2000), <https://alair.ala.org/handle/11213/7668>.
3. Association College and Research Libraries, *Framework for Information Literacy for Higher Education*, (2015), <http://www.ala.org/acrl/standards/ilframework>.
4. Claudio Carpineto and Giovanni Romano, "A Survey of Automatic Query Expansion in Information Retrieval," *ACM Computing Surveys (CSUR)* 44, no. 1 (2012): 1-50, <https://doi.org/10.1145/2071389.2071390>.
5. The Associated Press, *The Associated Press Stylebook and Briefing on Media Law 2015*. New York: Basic Books 2015.
6. The Associated Press. *The Associated Press Stylebook and Briefing on Media Law 2018*. New York: Basic Books 2018.
7. Keith Coffman and Nicole Neroulis, "Colorado, Washington first states to legalize recreational pot," *Reuters*, November 6, 2012, <https://www.reuters.com/article/us-usa-marijuana-legalization/colorado-washington-first-states-to-legalize-recreational-pot-idUSBRE8A602D20121107>.
8. John Balzar, "Voters Approve Measure to Use Pot as Medicine," *Los Angeles Times*, November 6, 1996, http://articles.latimes.com/print/1996-11-06/news/mn-62740_1_voters-approve-measure.
9. Safiya Umoja Noble, *Algorithms of Oppressions: How Search Engines Reinforce Racism*. New York: New York University Press 2018.
10. Emily Drabinski, "Queering the Catalog: Queer Theory and the Politics of Correction," *The Library Quarterly* 83, no. 2 (2013): 94-111, <https://doi.org/10.1086/669547>; Library of Congress, "Library of Congress to Cancel the Subject Heading 'Illegal Aliens,'" March 22, 2016, <https://www.loc.gov/catdir/cpsol/illegal-aliens-decision.pdf>