Title: Boolean Redux: Performance of Advanced versus Simple Boolean Searches and Implications for Upper-Level Instruction

Authors:
M. Sara Lowe, IUPUI University Library, 755 W. Michigan St., Indianapolis, IN 46202, mlowe@iupui.edu, 317-274-0349

Sean M. Stone, IU School of Dentistry, 1121 W Michigan St., Indianapolis, IN 46202, smstone@iu.edu, 317-274-5207

Bronwen K. Maxson, University of Oregon Libraries, 1299 University of Oregon, Eugene, OR 97403, bmaxson@uoregon.edu, 541-346-3069

Eric Snajdr, IUPUI University Library, 755 W. Michigan St., Indianapolis, IN 46202, esnajdr@iupui.edu, 317-278-2778

Willie Miller, IUPUI University Library, 755 W. Michigan St., Indianapolis, IN 46202, wmmiller@iupui.edu, 317-274-7365

Acknowledgement:
Special thanks to Kathleen Hanna who contributed to the data gathering and analysis of this manuscript before her retirement. She did not want to be listed as a co-author stating “My CV stops getting added to when I retire.” We miss you Kathleen.

Title

Boolean Redux: Performance of Advanced versus Simple Boolean Searches and Implications for Upper-Level Instruction

Abstract

Building off an earlier study, which examined whether or not it was beneficial to teach Boolean logic to introductory students, the present study examines the efficacy of Boolean OR searching in more advanced search strategies, for example, capstone projects and graduate theses and dissertations. Results show that both simple and advanced Boolean searches yield relevant results. Based on relevance, there is no compelling evidence that either search is superior. To capture all the literature on a topic, however, it is important that upper-level students know the relevant databases for their discipline and perform multiple searches. Results can help inform whether teaching Boolean search skills to upper-division students in disciplinary contexts is time well spent.

Keywords

Boolean Logic, Information Literacy, Instruction

Introduction

The first study on the topic of Boolean had a suggestive title: “The Boolean is Dead, Long Live the Boolean!” (Lowe et al., 2018). This is based on the idea that our profession has given value and prominence to teaching Boolean, but that favor may no longer be as relevant in the first-year context. At the same time, the authors wanted to explore when and how teaching Boolean is still relevant. Drabinski’s (2014) article, “Toward a Kairos of Library Instruction,” asks librarians to consider the sociohistorical context of certain “truths” in our profession and sometimes to resist them. Boolean could be considered part of our “Procrustean bed.”

Additionally, proprietary database algorithms can make search results opaque both to students and librarians. In this light, we should ask: Does, and how can, teaching Boolean logic help students to engage with and intervene in the world? What is the context in which teaching Boolean fits? Is there a disciplinary component or correlation to the usefulness of employing Boolean search operators?

The results of the first study found that natural language, phrase searching outperformed simple Boolean searches for finding relevant articles. The study, though, was limited to the types of research projects first-year students in introductory courses might encounter. The authors were curious if Boolean was necessary for upper-level, advanced students. For example, in the context of more complex research done for a capstone project or graduate thesis or dissertation. While OR is usually employed if only a limited number of results are retrieved or when results are irrelevant, there are larger implications for upper-level, advanced students. The logic would be: an advanced undergraduate or graduate student must learn important searching skills for surveying the literature such as Boolean logic. This will allow them to comprehensively determine what has been done before so they can ask new questions in order to make new discoveries, learn to locate and implement the best methods into their practice, and gather the best (e.g., relevant and high quality) information in order to make an informed decision. But is that logic valid? At its heart, our research question was simple: Should we teach Boolean at all? In practice, our research question was: Is using simple Boolean AND sufficient or will the advanced, upper-level student miss relevant literature that would otherwise turn up if they had searched using more advanced Boolean OR search logic?

Literature Review

Several areas of the literature are relevant to this research and build on the literature review from the original study (Lowe et al., 2018). Areas important to highlight focus on upper-level, or

more advanced, undergraduate and graduate students around: the complexity of Boolean logic for students; and, their research habits, including information and research needs. As with the original study, important to the present study is that Boolean logic has been repeatedly proven to be challenging for students to understand and use correctly or effectively (Hancock-Beaulieu, 1992; Hildreth, 1983; Hildreth, 1997; Jansen et al., 2000; Rousseau et al., 1998). Additionally, multiple studies found only a small percentage of students actually use Boolean in their searches, usually limited to the AND operator (Bloom & Deyrup, 2015; Dinet et al., 2004; Georgas, 2014; Lau & Goh, 2006; Malliari & Kyriaki-Manessi, 2007; Novotny & Cahoy, 2006). This pattern holds for graduate students as well. In a survey of graduate students in Education, Catalano (2010) found they favored the usage of AND over the more advanced Boolean operators NOT and OR. In her survey, 25% didn't know what the Boolean operator NOT was and 27% *rarely* or *never* used it. The operator OR had a slightly lower percentage of students not knowing what it was (24%) and *rarely* or *never* using it (17%) (Catalano, 2010). Korobili et al.'s (2011) survey of graduate Philosophy and Engineering students in Greece found 47% reported *never* using Boolean operators and 28% used them *seldom*. Perrett's (2004) audit of Australian graduate students found less than half (49%) had a firm understanding of how to search databases using Boolean operators. Mehrad and Rahimi (2012) found for a range of search skills, students at both the MA and PhD level, performed lowest on Boolean and synonyms and suitable keywords. However, familiarity and instruction can mediate these deficiencies. Interacting with graduate students over the course of a year, Chu and Law (2007) found that students' use of Boolean operators rose with each meeting as their familiarity and comfort with Boolean logic grew.

What are the research habits and information needs of upper-level undergraduate and graduate students? Pickard and Logan's (2013) examination of the research process of seniors versus first-year students found seniors viewed research as an iterative process and were more adept

than first-year students at refining search strategies. Miller's (2013) mixed methods study of the research process of capstone students found "finding useful information" was identified as the most challenging part. Dubicki (2015), in an analysis of advanced undergraduate and graduate student writings about their research process, found even experienced undergraduate researchers reported anxiety when beginning their research. Graduate students were more confident, and used more advanced search strategies such as defining search terms and using discipline-specific databases (Dubicki, 2015). Researching graduate students at Georgetown University, Gibbs et al. (2012) found one of students' most frequently cited frustrations with research was difficulty in searching for journal articles including retrieving sources from keyword searches. Conversely, Patterson (2009), through graduate students' self-assessment at University College Dublin found only 15% lacked confidence in using keywords to search, however, 48% expressed low confidence or no familiarity at all using controlled vocabulary to search. In a small study with Biology graduate students in Italy, Vezzosi (2009) found they did not want to invest time and effort learning complex tools for their research. However, upper-level graduate students did report a preference for advanced searching techniques, such as Boolean, more so than newer graduate students (Vezzosi, 2009).

A mention of the role of algorithms in academic databases is also both necessary and challenging. Necessary because the design of the algorithms affects search results. Challenging because the databases searched in this study are proprietary, owned by corporations who do not openly share information about how they are designed beyond search tips. Broussard (2018) defines an algorithm as "a series of steps or procedures that the computer is instructed to follow," or "a computational procedure for deriving a result, much like a recipe is a procedure for making a particular dish." Algorithms are mysterious things and their effect on searches is difficult, and in some cases, impossible to fully understand. Pasquale (2015) discusses this in-depth. Several studies have been done on recall versus precision (Buckland & Gey, 1994;

Walters, 2009; Walters, 2011), but these, and the present study, are limited by what we can know about the design of the algorithms from the users' perspective. We know that algorithms are designed by humans, humans are subject to bias and can commit errors (for example, Ingersoll, 2019; Lee et al., 2019; Obermeyer et al., 2019), and that algorithms reflect the instruction or design of their creators. What does that mean in terms of library databases? There may be incentive for for-profit parent companies of research databases to refer users to research they publish (for example, Sage Research Methods, 2019) and discovery systems may not index content from competing vendors (Breeding, 2015).

Methods

Eleven databases were included in this study: Academic Search Premier (EBSCO); Dialnet;[1] Google Scholar;[2] Hispanic American Periodicals Index (HAPI); JSTOR; Modern Language Association (MLA) International Bibliography (ProQuest); Nexis Uni; ProQuest Central; PubMed; Scopus; and Web of Science. Databases were selected in an effort to represent a wide array of disciplines as well as a likelihood of being used by upper-level undergraduate students during the completion of research projects.

Although discovery layers are ubiquitous in academic libraries (for example, in a 2018 survey of ARL libraries only one respondent library did not have one (Allison & Mering, 2018)) individual databases were searched instead of discovery layers for several reasons. Discovery layers allow Boolean logic as well as databases and the authors are at a single institution which uses the EBSCO Discovery Layer; an EBSCO database was already being searched and products from the same vendor have similar functionality. As this was not a review of discovery layers, searching them was outside the scope of this project and, regardless of which databases were, or were not searched, the purpose of the study was to determine whether advanced Boolean OR should be used, the databases were a means to an end, not the focal point. Even in a web discovery world, Boolean is still being taught, and some (for example, Lowe et al., 2018;

Seeber, 2015) argue librarians should spend more time on other skills like critical evaluation. Finally, upper-level students are expected to have a basic familiarity with large, general databases as well as at least one or two that would be contextually specific to their research topic based on experience or from suggestions by faculty, librarians, or other sources such as research guides as part of their assignment (for example, a cursory examination of most library guides as well as studies such as Kwadzo, 2015).

Each database was searched in two ways: a Simple Boolean search, and an Advanced Boolean Search (see Table 1). Results were limited using the “scholarly/peer reviewed articles” option or closest equivalent if available. JSTOR was limited to “Articles;” Nexis Uni to “Law Reviews;” ProQuest Central to “Peer reviewed;” Scopus and Web of Science to “Article.” Google Scholar was included because of its ubiquity with an understanding that it would not be directly comparable to the other databases. Limiters used in the other databases were not available in Google Scholar and truncation does not function as it does in most other databases. A student with a basic knowledge of truncation may not be aware that it does not function in Google Scholar and logically would be likely to try such searches in the course of their research. No limiters were used with PubMed and Dialnet. Limiters were not required for HAPI which only contains scholarly journals.

Search topics were generated as part of a previous research project and are based on real scenarios encountered by the authors while working with introductory-level undergraduate student researchers. Three scenarios were chosen to represent searches that would encompass a wide range of subjects and databases and basic, standardized queries were generated to try to closely represent typical, undergraduate research questions. Details of both Boolean searches were added and developed by the interrater pairs evaluating them based on

the original scenarios. The overall goal was not to create perfectly crafted searches but to try to reflect search behaviors that would be typical for undergraduate researchers.

The essence of this research is to explore the value in training students in the use of the relatively advanced Boolean OR. It is understood that under normal circumstances, an advanced researcher would apply OR if initial, simpler searches returned fewer than desired results. However, this experiment intentionally ignores this in order to be able to make as direct a comparison as possible between similar searches derived from a single research question.

Sample Topic	**Search Queries**	**Databases Searched**
What are the effects of television advertising on children?	Simple Boolean query: Television advertising AND food AND children Advanced Boolean query: (Television OR TV) AND (Advertis* OR Promot*) AND (Food choice* OR Dietary choice*) AND (Pre-teen OR Child* OR Tween*)	Academic Search Premier Google Scholar JSTOR Nexis Uni
Is social capital a determinant of health outcomes among the elderly?	Simple Boolean query: social capital AND health outcomes AND elderly Advanced Boolean query: (social capital OR social network* OR social support) AND (health OR medical) AND (elderly OR older	Google Scholar ProQuest Central

	people OR aged OR gerontolog*) AND outcomes	
What is the effect of stress on women in the workplace?	Simple Boolean query: Stress AND women AND workplace Advanced Boolean query: (stress OR suffering OR anguish OR distress) AND (women OR woman OR girl OR gender OR sex-specific) AND (workplace OR worksite OR employment OR profession)	Google Scholar PubMed Scopus Web of Science
How are women's gender roles represented in Spanish (or Latin American/language) theater?	Simple Boolean query: Women AND role AND spanish theater Advanced Boolean query: (women OR female OR wife) AND ("gender role" OR role OR "sex role") AND (spa* OR latin americ*) AND (theater OR theatre OR drama OR play)	HAPI MLA International Bibliography
How are women's gender roles represented in Spanish theater? (spanish-language)	Simple Boolean query: mujeres AND papel AND teatro español Advanced Boolean query: (mujer OR feminino OR expose) AND papel AND (españ* OR "América latin*" OR laintoamerican* OR "América Central") AND (teatr* OR drama)	Dialnet

Table 1: Search queries and databases searched

The authors divided into pairs so that each sample topic in each database was searched twice and independently by both members of an interrater pair. Searches were completed in two rounds. The first round, June 12-October 19, 2018, searched all databases except Google Scholar which was searched between February 7-22, 2019.

Citations and abstracts were captured for the first 50 results of each search. Only the first 25 results of each search were analyzed for this study. Each search result was scored on a rubric for relevance (see Table 2) with scores ranging from 0 (not relevant) to 3 (very relevant). Relevance scores were based on the title and abstract of each search result.

Not Relevant (0)	**Less Relevant (1)**	**Relevant (2)**	**Very Relevant (3)**
0 of total concepts represented OR false hits, terms are there but used in different ways (e.g.,	Less than half concepts represented OR concepts are there but not relevant to research question	Majority or all of concepts represented either in title or abstract but when looking at abstract,	All concepts represented in title or abstract and abstract is relevant

social work instead of social rejection)		may be tangential to research question	

Table 2: Article relevance evaluation rubric

After performing the searches, interrater pairs met and normed their relevance rankings for the paired sets of search results of each database. Overlap of results was also indicated for the paired sets of search results for each database.

Results

Overall Results

As might be expected, the advanced Boolean search returned a higher number of results than the simple Boolean (see Table 3). The only exception is the TV Advertising search in Google Scholar, where there were more results with the simple than the advanced Boolean, although this may be because of the extensive use of truncations not supported by Google Scholar.

When averaging the mean results from all databases (excluding the Spanish-language searches in HAPI and Dialnet) the simple Boolean search outperforms the advanced search in relevance of the first 25 results, 2.39 compared to 2.14 (see Figure 1). A one-way between subjects ANOVA ($F(1, 1042) = 20.60$) was conducted and the difference is statistically significant with $P < 0.05$. However, Cohen's $d = 0.01$ indicating a small effect size. In other words, although the difference between the two searches is statistically significant, the small effect size indicates that the difference is trivial.

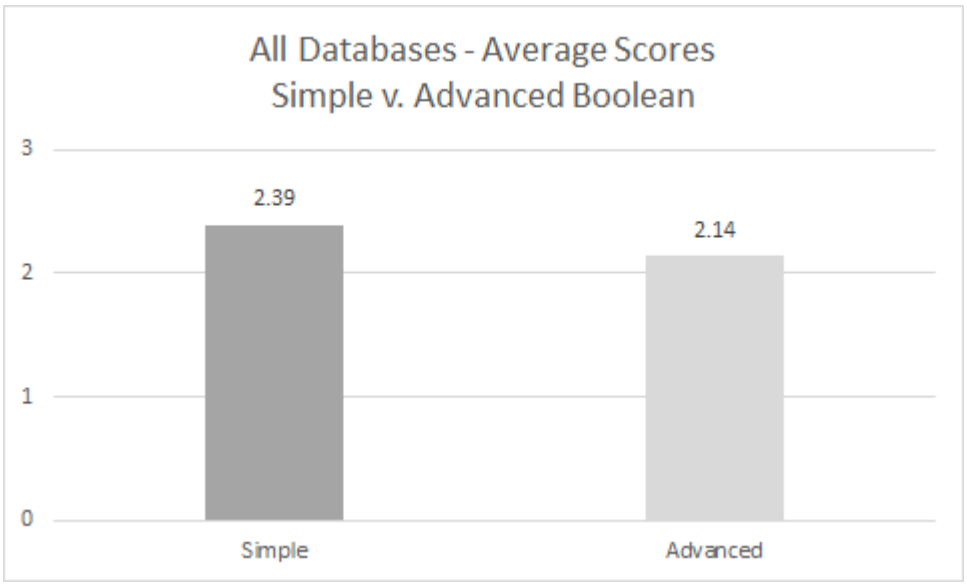

Figure 1: Average database scores - Simple versus advanced Boolean [for all databases except Spanish language searches (HAPI and Dialnet)]

Results by Database

When examined individually, interesting differences between databases emerge. In about one-third of the databases (Academic Search Premier, Dialnet, and two of three Google Scholar searches) the advanced search outperformed the simple search. In the other two-thirds (Google Scholar (TV), JSTOR, MLA International Bibliography, Nexis Uni, ProQuest Central, PubMed, Scopus, and Web of Science) the simple search outperformed the advanced search (see Figure 2 and Table 4). In Dialnet, Google Scholar (Social Capital), JSTOR, MLA International Bibliography, ProQuest Central, and Scopus, the difference is statistically significant $P < 0.05$. ProQuest Central has a Cohen's $d = 0.1$, indicating a large effect size. JSTOR has a Cohen's $d = 0.07$, indicating a medium to large effect size. Dialnet, Google Scholar, MLA International Bibliography, and Scopus all have a Cohen's $d = 0.05$, indicating a medium effect size.

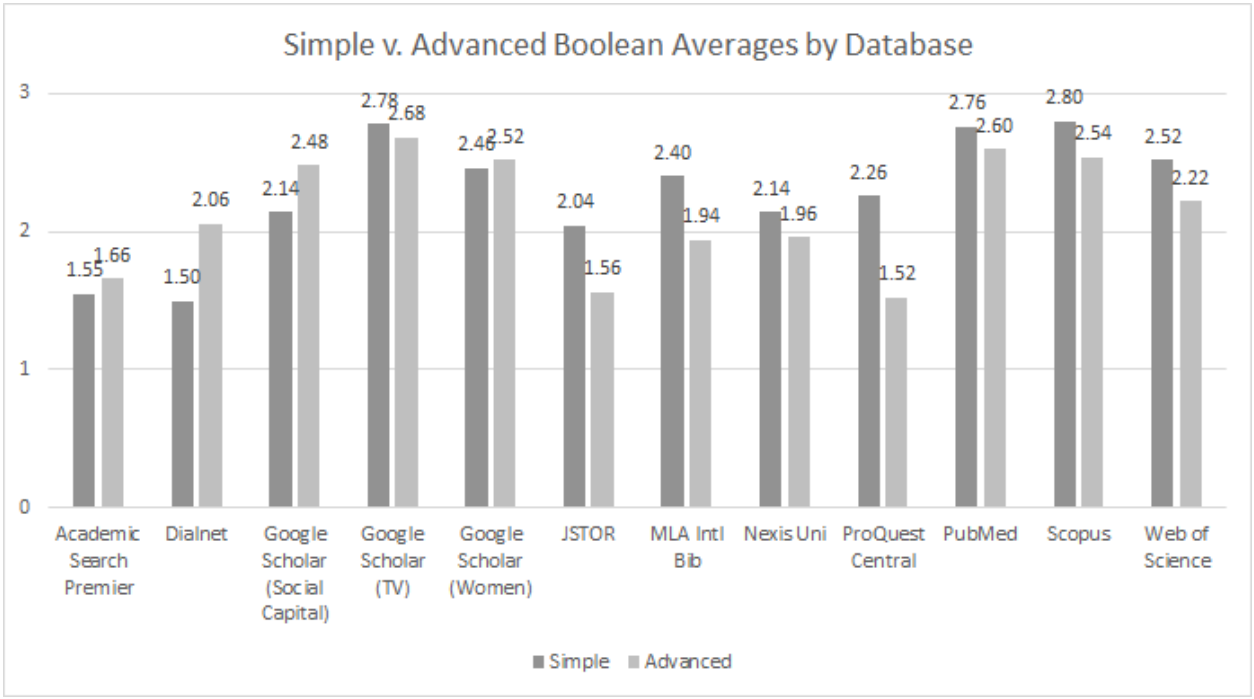

Figure 2: Simple versus advanced Boolean average of rater 1 and rater 2 scores by database

	Rater 1		Rater 2	
	Simple Boolean	Advanced Boolean	Simple Boolean	Advanced Boolean
Academic Search Premier	11	1,407	11	1,502
Dialnet	26	46	26	59
Google Scholar (Social Capital	257,000	1,020,000	255,000	975,000

Search)				
Google Scholar (TV Advertising Search)	300,000	26,600	304,000	26,300
Google Scholar (Women Search)	1,190,000	2,810,000	1,230,000	2,810,000
HAPI	1	8	1	0
JSTOR	5,998	8,627	8,641	12,920
MLA International Bibliography	17	164	18	187
Nexis Uni	556	310	559	310
ProQuest Central	16,786	317,712	16,583	307,646
PubMed	804	6,372	755	4,473
Scopus	980	4,507	985	4,540
Web of Science	790	4,012	786	3,981

Table 3: Number of total results per search by database for each rater

		Simple Boolean		Advanced Boolean	
		Mean	Standard Deviation	Mean	Standard Deviation
Academic Search Premier	Rater 1	1.45	1.128	1.48	1.046
	Rater 2	1.64	1.027	1.84	0.943
Dialnet	Rater 1	1.60	1.231	2.20	0.951
	Rater 2	1.40	1.225	1.92	1.077
Google Scholar (Social Capital Search)	Rater 1	2.40	0.577	2.44	0.651
	Rater 2	1.88	0.881	2.52	0.714
Google Scholar (TV Advertising Search)	Rater 1	2.64	0.638	2.52	0.770
	Rater 2	2.92	0.400	2.84	0.624
Google Scholar (Women Search)	Rater 1	2.32	0.627	2.28	0.792
	Rater 2	2.60	0.645	2.76	0.436
HAPI	Rater 1	1.00	n/a	2.50	0.756
	Rater 2	1.00	n/a	n/a	n/a
JSTOR	Rater 1	2.00	0.816	1.48	1.046
	Rater 2	2.08	0.812	1.60	0.866

MLA International Bibliography	Rater 1	2.29	0.920	1.92	1.077
	Rater 2	2.50	0.985	1.96	1.060
Nexis Uni	Rater 1	2.00	0.816	1.76	0.779
	Rater 2	2.28	0.614	2.16	0.688
ProQuest Central	Rater 1	2.20	0.707	0.88	1.013
	Rater 2	2.32	0.945	2.16	0.898
PubMed	Rater 1	2.72	0.542	2.40	0.764
	Rater 2	2.80	0.408	2.80	0.408
Scopus	Rater 1	2.83	0.482	2.33	0.856
	Rater 2	2.77	0.429	2.75	0.444
Web of Science	Rater 1	2.56	0.712	2.36	0.810
	Rater 2	2.48	0.714	2.08	0.954

Table 4: Mean and standard deviation by database by rater for each search

Next we examined the overlap percentage by database. This is the number of results that were the same between the simple and advanced Boolean search for each rater. The total overlap average was 25%, meaning only one in four articles were similar between simple and advanced. The overlap percentage ranged by database from 0% to 60% and was largely similar between raters (see Figure 3).

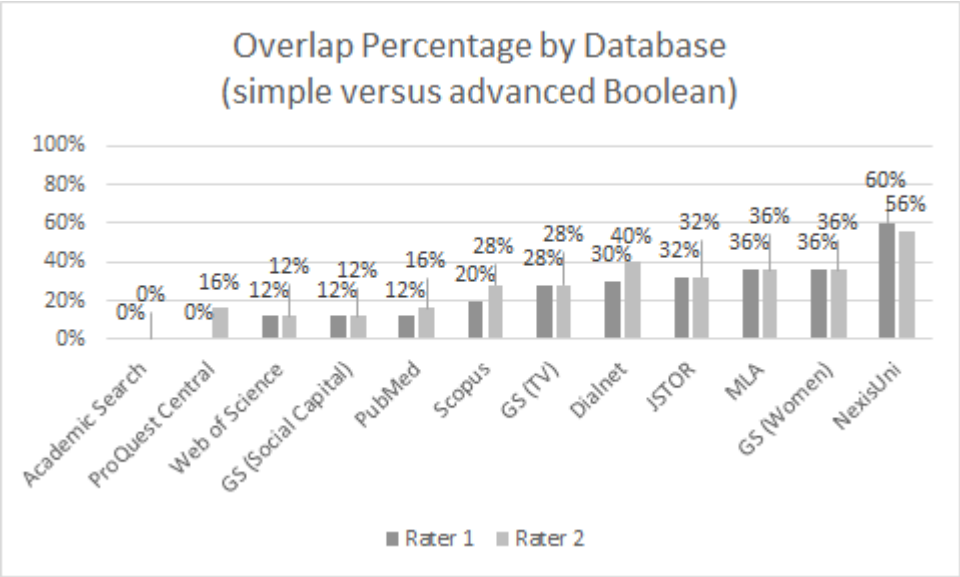

Figure 3: Percentage of results that were the same between simple and advanced search (rater 1 and rater 2)

Overlap percentage is not the full story, however. If results are unique, but not relevant, then it is not necessarily imperative to do the search. This is why we also explored the percentage of results that scored a 3-*very relevant* and that only appeared in one of the searches. Here the total *very relevant* and unique results averaged 36% for the simple search and 33% for the advanced search. When examined by database, again, percentages vary widely, ranging from eight to 67 percent (see Figure 4). The simple search had a higher percentage of relevant unique results in Nexis Uni, MLA, JSTOR, Academic Search, ProQuest Central, Google Scholar (TV), Web of Science, PubMed, and Scopus. The advanced Boolean search had a higher percentage of unique, relevant results in Dialnet and Google Scholar (Social Capital and Women). The higher the percentage, the more likely that *very relevant* unique results will be found in both simple and advanced searches. This means a user would want to do both searches to be assured of retrieving all relevant materials in databases such as Web of Science, PubMed, and Scopus. A lower unique, relevant percentage means there was more

overlap between the simple and advanced searches, and users would not necessarily find a lot of additional, relevant results by doing both styles of searches in databases such as Nexis Uni, MLA, JSTOR, Academic Search, and ProQuest Central.

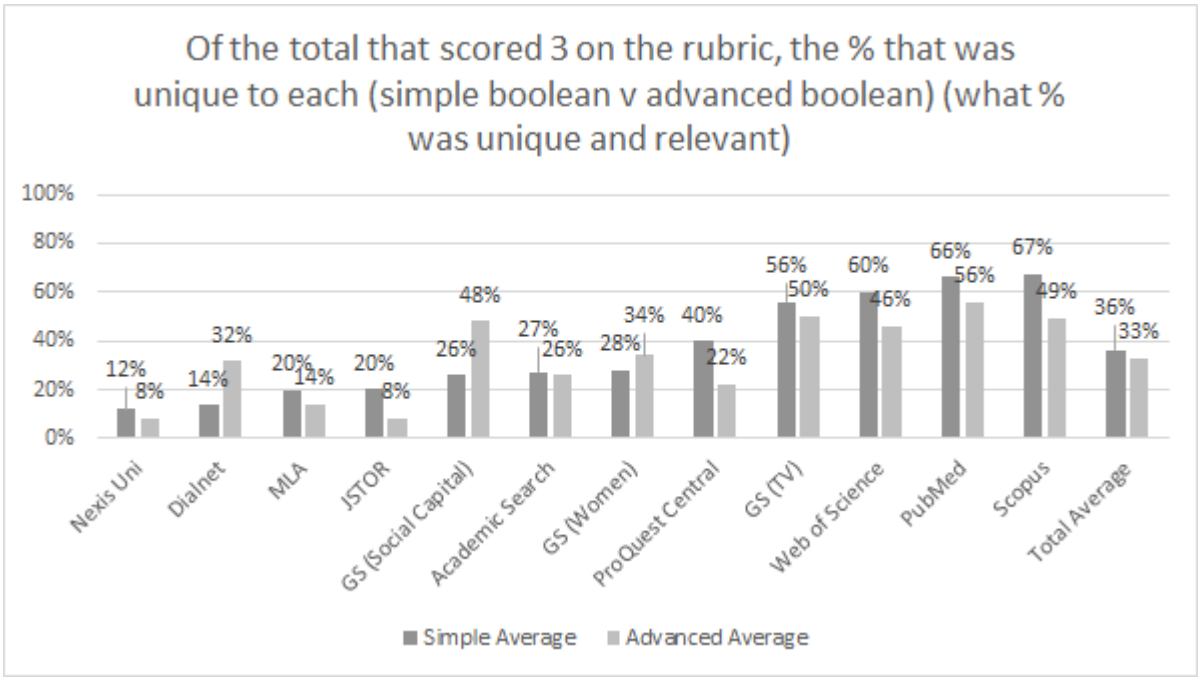

Figure 4: Of the results that scored a 3-*very relevant* on the rubric, the percentage that was unique to each search (average of rater 1 and rater 2)

Finally, we examined precision. This is the number of results that scored 2-*relevant* or 3-*very relevant* on the rubric divided by the total number of results. As might be expected, because there are fewer synonyms to potentially bring up irrelevant results, the simple search outperformed the advanced search, but not by much (83% versus 73%, respectively). This would tend to indicate that relevancy ranking in database algorithms works well, regardless of the complexity of the search. Here too, differences between databases were interesting (see Figure 5). While some databases saw very little difference in precision between the searches

(for example, Academic Search, Google Scholar, Nexis Uni, PubMed, Web of Science, and Scopus) with others the simple search was vastly more precise (JSTOR, MLA International Bibliography (ProQuest), and ProQuest Central). As both MLA International Bibliography and ProQuest Central are from the same vendor that could be a likely factor for their similarity.

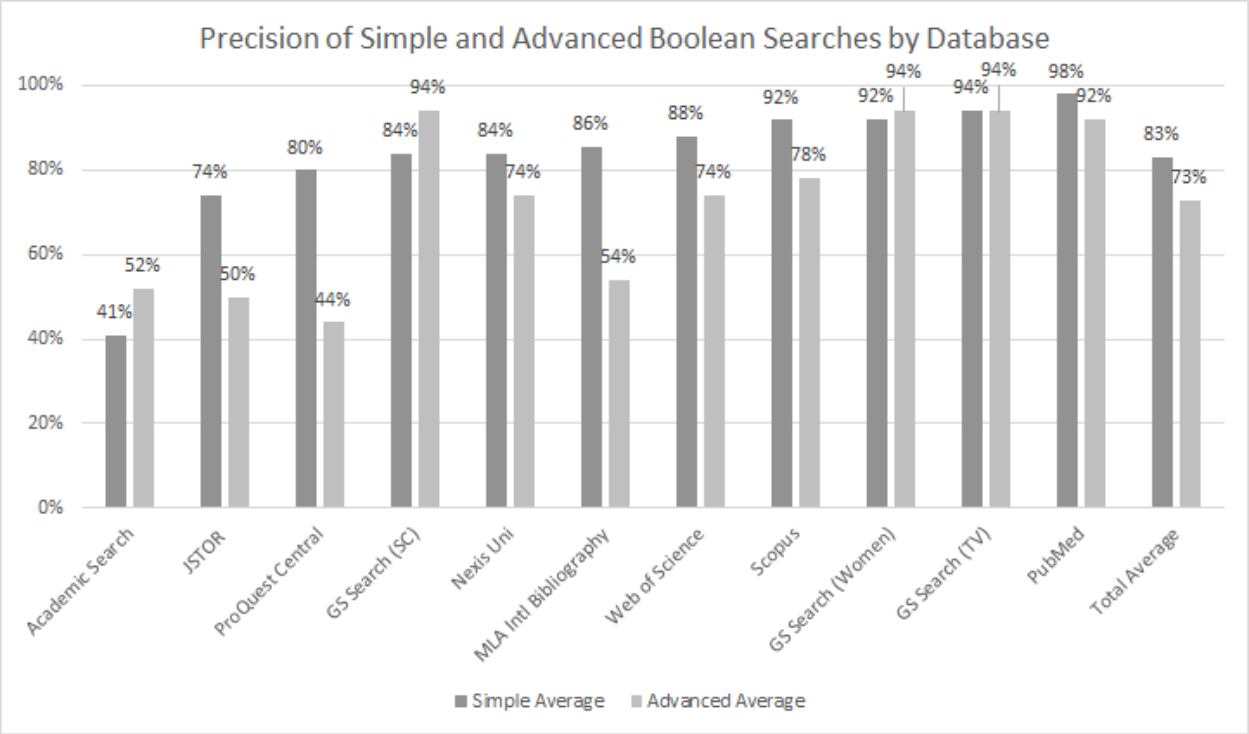

Figure 5: Precision of simple and advanced search results (results scoring 2 or 3 divided by total number of results) by database

Discussion & Limitations

To return to our research question: Is using simple Boolean, AND, sufficient or will the advanced, upper-level student miss relevant literature that would otherwise turn up if they had used more advanced Boolean (e.g., using OR for synonyms) search logic? In other words, which performs better, simple or advanced Boolean searches? As expected, the overall number

of results using the advanced searches were higher across all databases (with the exception of Academic Search Premier). However, this did not necessarily translate into increases in relevance of articles when compared to the simple search. Overall relevance of simple searches was significantly higher than advanced searches but with little effect. Results vary for relevance for searches in individual databases. Four of the databases show more relevant results with the advanced search. Each of these four results comes from a single database instance from each of the four test questions so it is challenging to generalize these results as being indicative of database or search subject matter. Overlap of results is generally low, but comparable between simple and advanced searches, again with differences between individual databases. Precision was largely similar between advanced and simple searches.

The results show that while there may be database specific differences, overall, simple searches yield relevant results but so do advanced searches. Based on relevance, there is no compelling evidence that either search is superior. Given the low level of overlap in several databases, an upper-level student would be well served to do both types of searches, regardless of number of results, if they want to perform a comprehensive literature search.

How does this translate into teaching Boolean logic to advanced student researchers? It is clear from previous studies mentioned in the literature review that even graduate level researchers are at best unfamiliar with, and, at worst, completely ignorant of some if not all of the aspects of Boolean searching beyond using the operator AND. Given this, the simple search is likely to be much more realistic in anticipating a student approach to searching for resources given the test scenarios. The advanced searches in this study are by no means perfect or comprehensive, but even in their current state, model a behavior that, as noted in the literature review, would be rare, if not absent, in the majority of advanced student researchers. Instructors would have to

weigh the pros and cons of including this kind of instruction as part of the information literacy competencies they deliver to students.

Student needs for searching instruction is not only driven by the demands of the individual project but by subject area and related databases which may vary significantly in how they perform. On one hand, there is clearly a significant gap, even in advanced student understanding of the subject matter, and it would be a significant investment of time to adequately address the topic in class. This would require more time than a traditional one-shot allows. On the other hand, if the students had a significant need for a truly exhaustive literature search, to support a dissertation or major grant proposal for example, it might be worth the investment of time. It may be possible to deliver this kind of instruction in a primarily asynchronous or flipped classroom format with students expected to come to a classroom session with at least a basic familiarity of the concepts which can be expanded on by the librarian. Advanced Boolean training might also be better suited to one-on-one sessions as needed, addressing the needs of only the most intensive students. In short, knowing how to use the advanced Boolean OR adds value, but this value must be weighed against the cost involved in teaching it.

With respect to understanding how algorithms affect searches and how we might teach to students, the *Framework for Information Literacy for Higher Education* (Association of College & Research Libraries, 2016) gives us a template. Though we may not have viable or efficient alternatives, librarians can at least work with students to understand that intellectual property is a legal and social construct that varies by culture, and how and why some individuals or groups of individuals may be underrepresented or systematically marginalized within the systems that produce and disseminate information. One noteworthy example is Safiya Umoja Noble's (2018) book, *Algorithms of Oppression: How Search Engines Reinforce Racism*. In addition, library

subscription databases can be used to teach students to understand how the commodification of their personal information and online interactions affects the information they receive and the information they produce or disseminate online; and, how to make informed choices regarding their online actions in full awareness of issues related to privacy (Singley, 2020) and the commodification of personal information. Broussard (2018) describes algorithmic accountability as a type of computational journalism within the frame of social justice work. “Sometimes,” she argues, “algorithmic accountability reporting means writing code to investigate the algorithms that are being used increasingly to make decisions on our behalf. Other times, it means looking at badly designed technology or falsely interpreted data and raising a red flag.”

There is an increasing need to address the ramifications of algorithms as part of information literacy instruction. Students are often presented with an array of databases and may have trouble deciding which one to start with. Students would be more sophisticated searchers and consumers of information if they had a better understanding of how algorithms can affect search outcomes. Additionally, this would help them better understand the appropriateness of searching multiple databases for a particular information need. Students can benefit from not only understanding the importance of scope in databases but also the differences that can occur in mechanisms of searching. Students have no logical reason to assume that different databases will behave in different ways, however, it is crucial for them to understand that searches in different databases, even those containing similar resources, may not give the same results. This is particularly difficult for the instructor as they may not have an explanation as to why this happens in certain databases making communicating this information challenging. One example of a tool that addresses this is the Social Media Collective’s (SMC) (2016) Critical Algorithm Studies Website and Reading List. The SMC is a network of social science and humanistic researchers who have created a resource for information consumers

and educators dedicated to researching algorithms as social concerns and ongoing collection of related resources.

As with all research, there are limitations to the present study. Some of the limitations of the first study apply here as well. While questions were based on actual student research questions, actual student researchers were not included so there is the possibility of benevolent bias. Expanding these studies and testing them based on real student searching behavior is a future possibility and could provide additional confirmation or questioning of the results. Another limitation was that the operator OR was included as part of “advanced” searching but NOT was not. As pointed out in the literature review, however, there is an almost total lack of knowledge about and use of NOT in advanced student research practice. Given this, it is unlikely that students would construct searches including NOT but it could be included in future analyses of the databases to give a more comprehensive evaluation.

Conclusion

Unlike the first study, results of simple versus advanced Boolean searches were more complex and nuanced. Overall, while the simple Boolean outperformed the advanced Boolean search, the effect size was not large and varied widely by database. For complex research topics a disciplinary approach to librarianship is helpful. When working with students in or out of the classroom, a balance must be struck between the needs of the individual project and the complications related to the subject area and related databases. Librarians need to be familiar with disciplinary research challenges and, by extension, details of searching in certain databases to best advise and teach upper-level students. For some disciplines, results highlight that keyword development is a less useful way to spend limited instruction time, but with others that time may be absolutely necessary to capture all relevant results. Complicating this are the proprietary algorithms which make it challenging for librarians to know exactly what the most

effective search might be for individual databases. In addition, algorithms change and we cannot see behind the curtain. One additional observation is that multiple databases from the same vendors behaved in similar ways that diverge from the average. It may be that the factors that make the performance of certain databases differ from others may be as much an artifact of the company that produces them as anything having to do with content.

It is important for library and information science professionals to continually review search efficacy across databases, because our field relies on databases and their search algorithms to provide information to our users. The efficacy of search queries has a direct impact on teaching and research. Further study in this area might focus more deeply into a discipline-specific approach of investigating the merits of simple and advanced Boolean searches. Such studies would enable subject librarians to make more data-informed decisions when weighing the benefits of using instructional time to teach Boolean versus other pressing topics.

References

Allison, D. A., & Mering, M. (2018). Use of discovery tools in ARL libraries. *Faculty Publications, UNL Libraries,* 366. https://digitalcommons.unl.edu/libraryscience/366

Association of College and Research Libraries. (2016). *Framework for information literacy for higher education*. http://www.ala.org/acrl/standards/ilframework

Bloom, B., & Mestrovic Deyrup, M. (2015). The SHU research logs: Student online search behaviors trans-scripted. *Journal of Academic Librarianship, 41*(5), 593–601. https://doi.org/10.1016/j.acalib.2015.07.002

Breeding, M. (2015 February). The Future of Library Resource Discovery: A White Paper Commissioned by the NISO [National Information Standards Organization] Discovery to Delivery (D2D) Topic Committee. https://groups.niso.org/apps/group_public/download.php/14487/future_library_resource_discovery.pdf

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Buckland, M., & Gey, F. (1994). The relationship between recall and precision. *Journal of the American society for information science*, *45*(1), 12-19.

Catalano, A. J. (2010). Using ACRL standards to assess the information literacy of graduate students in an education program. *Evidence Based Library and Information Practice*, *5*(4), 21-38.

Chu, S. K. W., & Law, N. (2007). Development of information search expertise: Postgraduates' knowledge of searching skills. *portal: Libraries and the Academy*, *7*(3), 295-316.

Dinet, J., Favart, M., & Passerault, J-M. (2004). Searching for information in an online public access catalogue (OPAC): The impacts of information search expertise on the use of Boolean operators. *Journal of Computer Assisted Learning, 20*, 338–46.

Drabinski, E. (2014). Toward a Kairos of library instruction. *Journal of Academic Librarianship, 40*(5), 480–485. https://doi.org/10.1016/j.acalib.2014.06.002

Dubicki, E. (2015). Writing a research paper: students explain their process. *Reference Services Review*, *43*(4), 673-688.

Georgas, H. (2014). Google vs. the library (Part II): Student search patterns and behaviors when using Google and a federated search tool. *portal: Libraries and the Academy, 14*(4), 503-532.

Gibbs, D., Boettcher, J., Hollingsworth, J., & Slania, H. (2012). Assessing the research needs of graduate students at Georgetown University. *The Journal of Academic Librarianship*, *38*(5), 268-276.

Hancock-Beaulieu, M. (1992). User friendliness and human-computer interaction in online library catalogues. *Program, 26*(1), 29–37.

Hildreth, C. R. (1983). To Boolean or not to Boolean. *Information Technology and Libraries, 2*(3), 235–37.

Hildreth, C. R. (1997). The use and understanding of keyword searching in a university online catalog. *Information Technology and Libraries, 16*(2), 52–62.

Ingersoll, A. (2019, September 9). How the algorithms running your life are biased. *Washington Post.* https://www.washingtonpost.com/business/how-the-algorithms-running-your-life-are-biased/2019/09/08/58e515a2-d206-11e9-a620-0a91656d7db6_story.html

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management, 36*(2), 207–27.

Korobili, S., Malliari, A., & Zapounidou, S. (2011). Factors that influence information-seeking behavior: The case of Greek graduate students. *The Journal of Academic Librarianship*, *37*(2), 155-165.

Kwadzo, G. (2015). Awareness and usage of electronic databases by geography and resource development information studies graduate students in the University of Ghana. *Library Philosophy and Practice (e-journal)*, 1210. http://digitalcommons.unl.edu/libphilprac/1210

Lau, E. P., & Goh, D. H. L. (2006). In search of query patterns: A case study of a university OPAC. *Information Processing & Management, 42*(5), 1316-1329.

Lee, N. T., Resnick, P., & Barton, G. (2019, May 22). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings Institute*. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

Lowe, M. S., Maxson, B. K., Stone, S. M., Miller, W., Snajdr, E., & Hannah, K. (2018). The Boolean is dead, long live the Boolean! Natural language versus Boolean searching in introductory undergraduate instruction. *College & Research Libraries, 79*(4), 517-534. https://doi.org/10.5860/crl.79.4.517

Malliari, A., & Kyriaki-Manessi, D. (2007). Users' behaviour patterns in academic libraries' OPACs: a multivariate statistical analysis. *New Library World, 108*(3/4, 107–22.

Mehrad, J., & Rahimi, F. (2012). Online search skills of Shiraz University post graduate students: A survey. *International Journal of Information Science and Management (IJISM)*, *7*(1), 1-13.

Miller, R. E. (2013, April). The almost experts: Capstone students and the research process. In *Imagine, innovate, inspire: Proceedings of the Association of College & Research Libraries (ACRL) 2013 conference* (pp. 16-22).

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Novotny, E., & Cahoy, E. S. (2006). If we teach, do they learn? The impact of instruction on online catalog search strategies. *portal: Libraries and the Academy*, *6*(2), 155-167.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Patterson, A. (2009). A needs analysis for information literacy provision for research: a case study in University College Dublin. *Journal of Information Literacy*, *3*(1), 5-18.

Perrett, V. (2004). Graduate information literacy skills: the 2003 ANU skills audit. *The Australian Library Journal*, *53*(2), 161-171.

Pickard, E., & Logan, F. (2013). The research process and the library: First-generation college seniors vs. freshmen. *College & Research Libraries, 74*(4), 399-415.

Rousseau, G. K., Jamieson, B. A., Rogers, W. A., Mead, S. E., & Sit, R. A. (1998). Assessing the usability of on-line library systems. *Behaviour & Information Technology, 17*(5), 274–81.

Sage Research Methods. Retrieved May 27, 2020. https://us.sagepub.com/en-us/nam/sage-research-methods-books-and-reference

Seeber, K. P. (2015). Teaching “format as a process” in an era of Web-scale discovery. *Reference Services Review, 43*(1): 19-30.

Singley, E. (2020). A holistic approach to user privacy in academic libraries. *The Journal of Academic Librarianship, 46*(3), 102151.

Social Media Collective. (2016 December 15). *Critical algorithm studies: A reading list*. https://socialmediacollective.org/reading-lists/critical-algorithm-studies/

Vezzosi, M. (2009). Doctoral students' information behaviour: an exploratory study at the University of Parma (Italy). *New Library World, 110*(1/2), 65-80.

Walters, W. H. (2009). Google Scholar search performance: Comparative recall and precision. *portal: Libraries and the Academy*, *9*(1), 5-24.

Walters, W. H. (2011). Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases. *portal: Libraries and the Academy*, *11*(4), 971-1006.

[1] Dialnet is an open access bibliographic database created and maintained by the University of La Rioja in Spain. Dialnet, https://dialnet.unirioja.es/, accessed October 10, 2019.

[2] While Google Scholar is a search engine, for the purposes of this study we refer to it as a database.