Title
Navigating Machine-Driven Research Landscapes: A Comparative Approach

Authors
Tranfield, Wynn
Caldwell, Christy

Publication Date
2024-11-11

DOI
10.1145/3678884.3681842

Navigating Machine-Driven Research Landscapes:

A Comparative Approach

Wynn Tranfield[†]
University Library
University of California, Santa Cruz
Santa Cruz, CA, USA
mtranfie@ucsc.edu

Christy Caldwell
University Library
University of California, Santa Cruz
Santa Cruz, CA, USA
caldwell@ucsc.edu

ABSTRACT

The growth of scientific literature poses a significant challenge for researchers and librarians. It is part of many librarians' core responsibilities to be able to identify and utilize appropriate research tools and databases to assist and advise scholars engaged in research. The recent proliferation of machine learning supported tools has created a tremendous gap in literature addressing the actual efficacy of these new tools, even as compared to conventional library databases. This work aims to build knowledge of these new tools (Scite, Elicit, SciSpace, Epsilon) by comparing search results within defined parameters, and evaluating results for format, topic relevance and uniqueness. By understanding the strengths and limitations of these tools, researchers will be better positioned to make informed decisions about their literature search.

CCS CONCEPTS

• human centered computing • artificial intelligence • walkthrough evaluations • consumer products

KEYWORDS

citation analysis; bibliographic evaluation; academic research; information seeking; digital libraries; artificial intelligence

ACM Reference format:

Wynn Tranfield and Christy Caldwell. 2024. Navigating Machine-Driven Research Landscapes: A Comparative Approach. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24), November 9–13, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA 6 pages. https://doi.org/10.1145/3678884.3681842

[†]Corresponding Author

CSCW Companion '24, November 9-13, 2024, San Jose, Costa Rica.

ACM ISBN 979-8-4007-1114-5/24/11.
https://doi.org/10.1145/3678884.3681842

1 INTRODUCTION

Those engaged in research have long sought to expedite and optimize the literature review process - an essential, yet often arduous phase of communicating one's research. Scholars at various levels consult with librarians who assist researchers in identifying and using appropriate research tools and databases. Accordingly, there is a great deal of literature discussing the relative effectiveness of various research databases, disciplinary strengths, and user preferences. Of particular interest has been the information seeking behaviors of novice researchers with aims to reduce the cognitive complexity of the task. Four especially complex tasks were identified and examined by Erin Matas [11]: identifying keywords, facing a large number of results, scanning results for relevant articles, and searching the right place.

A new wave of machine-learning supported scholarly search tools are attempting to meet researcher needs by providing alternatives to both Google Scholar and conventional library subject databases. At this point, there are several very similar tools available, each promoting optimized literature search and retrieval. The tools claim to provide researchers with a greater number of highly relevant scholarly articles in a shorter amount of time, with less friction. At present, there is a tremendous gap in literature addressing the actual efficacy of these new tools, even as compared to conventional databases.

For the past twenty years, much of the literature on searching in the right place has centered on the dichotomy between Google Scholar and library subject databases as they "coexist" in our research ecosystem. The simple web-based single search bar provided by Google Scholar (GS) when it launched in 2004 contrasted with many library subject databases' more complex multiple search box interfaces. Initial comparisons between the results of GS searches and library subject database searches revealed that library databases provided higher quality results [4]. A study by Oh and Colón-Aguirre [13] determined that researchers perceive academic library discovery systems as more comprehensive than Google Scholar, however, Google Scholar is easier to use and generally leads to high levels of satisfaction.

There is a consensus that Google is where students begin their searches, and students are unlikely to venture beyond the first page of results [3, 8, 19]. Students are also reluctant to engage in a wide range of information sources, preferring instead to adopt conservative strategies that minimize time spent researching [20]. Students report that GS requires less "mental effort" to use, and are familiar with the interface [13]. Libraries have responded by adopting the single search bar interface in response to user preferences for web-scale discovery tools with a single search box, facets for refining results, and relevance rankings [1, 15].

Once a search is executed, scanning and evaluating resources is an enduring challenge for novice and experienced researchers alike. Rowlands et al. found that though digital natives demonstrate ease and familiarity with search engines, they have less assessment skills than previous generations of scholars [17]. Novice researchers engage in satisficing, content with resources that seem good enough [13, 20], and most accessible to them [9]. Convenience as a chief factor in information retrieval was found to matter significantly regardless of users age, gender, or academic role [5].

Students' facet use, search behavior, and quality of articles selected is directly influenced by the type of tool used, and the way that tool was configured [6]. Training novice researchers to optimize tool use is a way to reduce the perceived complexity of a research database, or optimize the performance of familiar tools. However, with limited time and resources, clear instructions from faculty and librarians are important. Having a more robust understanding of the strengths and weaknesses of discrete services will allow instructors to better train novice researchers.

Machine learning supported tools could very well improve the search experience for novice researchers by incorporating Retrieval Augmented Generation (RAG). Machine learning supported tools rely heavily on Open Access (OA) publications, which make up an increasing number of publications but do not cover all journals or research areas evenly [16]. Some disciplines are better supported by OA publications or preprints than others. Machine learning database tools are vague about their retrieval and ingestion of scientific articles. Scite.ai reports using OA repositories such as PubMed Central, and utilizing Unpaywall and Crossref TDM to identify articles. They also report to ingest indexes from "over a dozen" publishers, and process updates "anywhere between daily and monthly" [12]. The tool Elicit limits results to publications indexed in Semantic Scholar [10].

The novelty of many of these database tools is the generated explanation and summarization, such as that provided by SciSpace's Copilot [18]. Easing the search process and providing keyword matching is also highlighted as a feature of these tools. Elicit, for example, allows users to provide a search query in question format. However, one librarian reviewer of the tool pointed out that this feature could be a limitation since researchers would need to make sure their question was well developed [10]. There is potential and enthusiasm for AI evaluation and analysis, however, early studies have demonstrated that the overall accuracy of Scite's assessments were low [2]. There is still much to learn about coverage and gaps that might emerge in searches, but a full assessment of these tools' scope would be challenging [7]. An estimation of Google Scholar alone is an enduring effort [14].

2 STUDY OBJECTIVE

This work aims to build knowledge of these literature search tools by comparing search result compositions for currency, relevancy, and uniqueness in ten discrete scientific disciplines. Understanding the strengths and limitations of these tools will allow librarians and researchers to develop best practices for incorporating these new tools into their research practice.

We aim to: explore a methodology for evaluating the effectiveness of AI/RAG research tools in retrieving relevant scientific literature, compare relevancy of literature search results to conventional library database search methods, and to identify strengths and limitations of each literature search tool. Having a more robust understanding of the strengths and weaknesses of discrete services will allow instructors to better train novice researchers. This preliminary work will encourage more rigorous examinations of this new wave of literature search tools, and encourage an expansion of disciplinary focuses.

3 METHODOLOGY

Building on search designs for systematic review preparations, this study uses search translation and bibliographic comparisons to assess discrete citations. Database tools assessed for comparison were Web of Science, Google Scholar, Scite, SciSpace, Elicit, and Epsilon.

This preliminary study involved examination of ten distinct sub-disciplines within the sciences. A convenience sample of STEM disciplines represented on our R1 research campus were selected. One research question was crafted for each major discipline, based on another convenience sampling of actual thesis topics submitted in those disciplines over the past five years. Scope notes about what a hypothetical researcher with the selected discipline would consider within scope or out of scope based on the topic were added to assist evaluators. Librarians designed naive search strategies for each research question, and translated their search into formats appropriate to each database tool. Questions were crafted with the novice researcher in mind.

As noted in Figure 1, Search returns from each database tool were documented, and the first fifty citations displayed in a relevancy ranking were downloaded. All literature searches were conducted from May 10-17, 2024. Caches for each subject and database were assessed for static variables, including range of citation age and resource format. Caches for respective searches were combined and deduplicated using SR Accelerator. The proportion of unique results for each database tool was then examined.

Undergraduate student screeners were then provided with systematic review screening software to quickly scan provided abstracts to judge the relevancy of an article to a prescribed research question and topic. Screeners were blinded to which database tool retrieved a given resource.

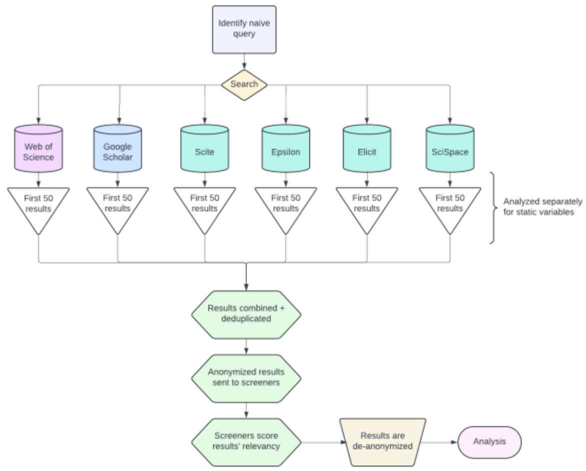

Figure 1: Flowchart Citation analysis process.

4 PRELIMINARY RESULTS

The results for this study are preliminary, but intriguing. At this point in the analysis, patterns have emerged in the results. The overwhelming majority of items returned from all searches were journal articles. Web of Science yielded the highest number of unique results for each search (See Figure 2). SciSpace searches yielded the fewest unique results, with the highest proportion of citations duplicated in other databases (See Figure 3).

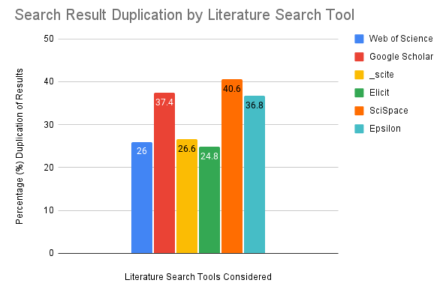

Figure 3: Search Result Duplication by Literature Search Tool (Percentage Duplicated Results)

As expected, there were disciplinary differences in result formats (See Table 1). Physics and Computer Science and Engineering searches yielded the highest number of conference proceedings across search databases. The preprint server, Biorxiv, was very well represented in Molecular, Cell and Developmental Biology.

Journal articles returned in Elicit, Epsilon, SciSpace and Scite were overwhelmingly open access, indexed in institutional repositories or Semantic Scholar. Elicit stood out as an outlier when examining overall publication years of returned citations, returning fewer overall recently published materials than other literature search tools.

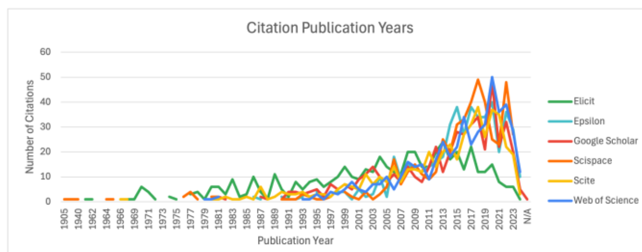

Figure 4: Citation Publication Years

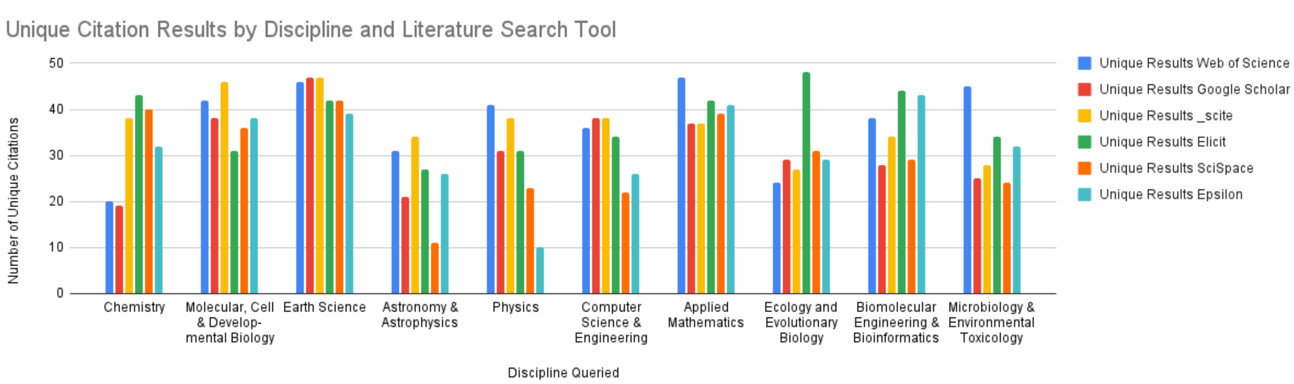

Figure 2: Unique Citation results by database tool

Subject	Database	Book or Book Chapter	Conference Paper	Journal Article	Patent	Preprint	Report	Supplemental Materials	Thesis or Dissertation	Unknown	Grand Total
Applied Math	Elicit		4	42			1		2	1	50
	Epsilon		10	36		4					50
	Google Scholar			50							50
	Scispace		6	36		5			3		50
	Scite		2	47		1					50
	Web of Science		4	46							50
Applied Math Total			**26**	**257**		**10**	**1**		**5**	**1**	**300**
Astro	Elicit		1	49							50
	Epsilon		6	35		7			2		50
	Google Scholar		1	46		3					50
	Scispace	1		31		18					50
	Scite		15	29		6					50
	Web of Science		3	47							50
Astro Total		**1**	**26**	**237**		**34**			**2**		**300**
BME	Elicit	3		47							50
	Epsilon			50							50
	Google Scholar	1		48		1					50
	Scispace	1		46		1			2		50
	Scite	1		47		2					50
	Web of Science	1	2	47							50
BME Total		**7**	**2**	**285**		**4**			**2**		**300**
Chemistry	Elicit	4	1	42	2				1		50
	Epsilon		4	43	2				1		50
	Google Scholar		1	48			1				50
	Scispace			50							50
	Scite			50							50
	Web of Science			50							50
Chemistry Total		**4**	**6**	**283**	**4**		**1**		**2**		**300**
CSE	Elicit	3	24	23							50
	Epsilon		16	23		10			1		50
	Google Scholar		2	48							50
	Scispace	3	19	21	2	4				1	50
	Scite		6	44							50
	Web of Science		12	38							50
CSE Total		**6**	**79**	**197**	**2**	**14**			**1**	**1**	**300**
Earth	Elicit			50							50
	Epsilon		8	36		2	1		3		50
	Google Scholar	3		46					1		50
	Scispace			48		1			1		50
	Scite			50							50
	Web of Science			50							50
Earth Total		**3**	**8**	**280**		**3**	**1**		**5**		**300**
EEB	Elicit	1		47					2		50
	Epsilon		6	40			2	1	1		50
	Google Scholar			45					5		50
	Scispace	1	2	42			1	1	3		50
	Scite	1		47		1			1		50
	Web of Science		1	49							50
EEB Total		**3**	**9**	**270**		**1**	**3**	**2**	**12**		**300**
MET	Elicit	1	2	46			1				50
	Epsilon	1	5	39		2	1		2		50
	Google Scholar	1		49							50
	Scispace	3	3	39		3			1	1	50
	Scite		1	48		1					50
	Web of Science			49		1					50
MET		**6**	**11**	**270**		**7**	**2**		**3**	**1**	**300**
MCDB	Elicit	3	1	41		5					50
	Epsilon	1	3	33		11			2		50
	Google Scholar	2	3	39		3			2	1	50
	Scispace	6	2	29		8			4	1	50
	Scite	1		33		15			1		50
	Web of Science		3	47							50
MCDB Total		**13**	**12**	**222**		**42**			**9**	**2**	**300**
Physics	Elicit	2	21	22		1			4		50
	Epsilon		22	15					13		50
	Google Scholar	2	3	24		3	2		16		50
	Scispace		7	18					25		50
	Scite		7	42		1					50
	Web of Science		22	27				1			50
Physics Total		**4**	**82**	**148**		**5**	**2**	**1**	**58**		**300**
Grand Total		**47**	**261**	**2449**	**6**	**120**	**10**	**3**	**99**	**5**	**3000**

Table 1: Table demonstrates item type counts for a portion of experimental searches

5 LIMITATIONS AND FUTURE WORK

The initial results indicate clear differences in the composition of search results from each database, despite identical search strategies. The role of the search algorithm has a tremendous impact on research surfaced, which may have downstream implications on the visibility of research. On one hand, these algorithms could potentially increase the visibility of research from smaller publishers not indexed in library subject databases. On the other, they could favor heavily cited resources, reinforcing existing silos.

For novice researchers, query development still presents a challenge. Built in suggestions such as the "Related Questions" feature provided by SciSpace could be helpful for students developing their research topic. Though beyond the scope of our current research, examining how user experience attributes of each tool could assist or confuse novice researchers would be useful for future instruction planning.

Machine learning research tools exhibit a clear reliance on open access journals, free servers, and preprints. This means fields like physics that rely heavily on preprints are well represented, and fields like chemistry are not. This finding is deeply relevant to researchers seeking a broad perspective of published research. The value-add of machine learning databases lies in their contextual and summary data; however, their indexing capabilities may be limiting. Our work highlights the importance of understanding the type of results likely from a given search tool, so researchers can make informed decisions about their literature review process.

This work is preliminary, but novel and important for librarians and information professionals. Empirical comparisons of commercial tools impacting research are essential for maintaining expertise in literature search practices. We are conducting a relevancy analysis to learn more about how well search results from each database match the previously outlined scope of each research query. Our analysis has been limited to ten disciplines within the sciences, however, we encourage future analysis to consider disciplines within the wider academy as certain fields' discrete research and publication conventions may impact the suitability of certain literature search tools.

6 CONCLUSIONS

There is a tremendous amount of work yet to be done in this space, however, this work begins describing strengths and limitations of using Retrieval Augmented Generation (RAG) tools for literature reviews. Disciplinary conventions must be considered, as must the skill and information need of the researcher. Given the low citation duplication between tools, users are advised against relying on one tool for comprehensive literature searches.

ACKNOWLEDGMENTS

We would like to acknowledge the support of the Librarians Association of the University of California (LAUC).

REFERENCES

[1] Asher, A.D., Duke, L.M. and Wilson, S. 2017. Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources | Asher | College & Research Libraries. (Apr. 2017). DOI:https://doi.org/10.5860/crl-374.

[2] Bakker, C., Theis-Mahon, N. and Brown, S.J. 2023. Evaluating the Accuracy of scite, a Smart Citation Index. *Hypothesis: Research Journal for Health Information Professionals*. 35, 2 (Sep. 2023). DOI:https://doi.org/10.18060/26528.

[3] Bloom, B. and Deyrup, M.M. 2015. The SHU Research Logs: Student Online Search Behaviors Trans-scripted. *The Journal of Academic Librarianship*. 41, 5 (Sep. 2015), 593–601. DOI:https://doi.org/10.1016/j.acalib.2015.07.002.

[4] Brophy, J. and Bawden, D. 2005. Is Google enough? Comparison of an internet search engine with academic library resources. *Aslib Proceedings*. 57, 6 (Dec. 2005), 498–512. DOI:https://doi.org/10.1108/00012530510634235.

[5] Connaway, L.S., Dickey, T.J. and Radford, M.L. 2011. "If it is too inconvenient I'm not going after it:" Convenience as a critical factor in information-seeking behaviors. *Library & Information Science Research*. 33, 3 (Jul. 2011), 179–190. DOI:https://doi.org/10.1016/j.lisr.2010.12.002.

[6] Dahlen, S.P.C., Haeger, H., Hanson, K. and Montellano, M. 2020. Almost in the Wild: Student Search Behaviors When Librarians Aren't Looking. *The Journal of Academic Librarianship*. 46, 1 (Jan. 2020), 102096. DOI:https://doi.org/10.1016/j.acalib.2019.102096.

[7] Gusenbauer, M. 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*. 118, 1 (Jan. 2019), 177–214. DOI:https://doi.org/10.1007/s11192-018-2958-5.

[8] Hamlett, A. and Georgas, H. 2019. In the Wake of Discovery: Student Perceptions, Integration, and Instructional Design. *Journal of Web Librarianship*. 13, 3 (Jul. 2019), 230–245. DOI:https://doi.org/10.1080/19322909.2019.1598919.

[9] Kim, K.-S. and Sin, S.-C.J. 2011. Selecting quality sources: Bridging the gap between the perception and use of information sources. *Journal of Information Science*. 37, 2 (Apr. 2011), 178–188. DOI:https://doi.org/10.1177/0165551511400958.

[10] Kung, J.Y. 2023. Elicit. *The Journal of the Canadian Health Libraries Association*. 44, 1 (Apr. 2023), 15–18. DOI:https://doi.org/10.29173/jchla29657.

[11] Matas, E. 2023. Undergraduate Students Experience Cognitive Complexity in Basic Elements of Library Research. (Jun. 2023).

[12] Nicholson, J.M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N.P., Grabitz, P. and Rife, S.C. 2021. scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*. 2, 3 (Nov. 2021), 882–898. DOI:https://doi.org/10.1162/qss_a_00146.

[13] Oh, K.E. and Colón-Aguirre, M. 2019. A Comparative Study of Perceptions and Use of Google Scholar and Academic Library Discovery Systems | Oh | College & Research Libraries. (Sep. 2019). DOI:https://doi.org/10.5860/crl.80.6.876.

[14] Orduna-Malea, E., Ayllón, J.M., Martín-Martín, A. and Delgado López-Cózar, E. 2015. Methods for estimating the size of Google Scholar. *Scientometrics*. 104, 3 (Sep. 2015), 931–949. DOI:https://doi.org/10.1007/s11192-015-1614-6.

[15] Pearce, A. 2019. Discovery and the Disciplines: An Inquiry into the Role of Subject Databases through Citation Analysis. *College & Research Libraries*. 80, 2 (Mar. 2019), 195–214. DOI:https://doi.org/10.5860/crl.80.2.195.

[16] Piwowar, H., Priem, J. and Orr, R. 2019. The Future of OA: A large-scale analysis projecting Open Access publication and readership. bioRxiv.

[17] Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H.R., Dobrowolski, T. and Tenopir, C. 2008. The Google generation: the information behaviour of the researcher of the future. *Aslib Proceedings*. 60, 4 (Jan. 2008), 290–310. DOI:https://doi.org/10.1108/00012530810887953.

[18] Roy, T., Kumar, A., Raghuvanshi, D., Jain, S., Vignesh, G., Shinde, K. and Tondulkar, R. 2024. SciSpace Copilot: Empowering Researchers through Intelligent Reading Assistance. *Proceedings of the AAAI Conference on Artificial Intelligence*. 38, 21 (Mar. 2024), 23826–23828. DOI:https://doi.org/10.1609/aaai.v38i21.30578.

[19] Spievak, E.R. and Hayes-Bohanan, P. 2016. Creating Order: The Role of Heuristics in Website Selection. *Internet Reference Services Quarterly*. 21, 1–2 (Apr. 2016), 23–46. DOI:https://doi.org/10.1080/10875301.2016.1149541.

[20] Warwick, C., Rimmer, J., Blandford, A., Gow, J. and Buchanan, G. 2009. Cognitive economy and satisficing in information seeking: A longitudinal study of undergraduate information behavior. *Journal of the American Society for Information Science and Technology*. 60, 12 (2009), 2402–2415. DOI:https://doi.org/10.1002/asi.21179.