Himmelfarb Health Sciences Library, The George Washington University

Health Sciences Research Commons

[Himmelfarb Library Faculty Publications](#)   [Himmelfarb Health Sciences Library](#)

2021

Scite: Contextualizing the Citation Count

Stacy Brody

Scite: Contextualizing the Citation Count
Stacy Brody

scite (https://scite.ai/)
scite Inc.
hi@scite.ai
334 Leonard St
Brooklyn, NY 11211

Pricing:
Tiered pricing model with free, basic($7.99/month or $50/year), premium ($14.99/month or $100/year), and enterprise plans.

scite (https://scite.ai/), founded by Josh Nicholson and Yuri Lazebnik and previously funded by the National Science Foundation (NSF) and National Institute on Drug Abuse (NIDA) [**1, 2, 3**], is a database of over 800 million citation statements [**4**] tagged by a machine learning algorithm as supporting, mentioning, or contrasting the findings of cited articles [**5**] and by where the statements appear in the citing articles (introduction, results, methods, discussion, or other). scite also provides a count of editorial notices for each article. Users can search the website and install plug-ins for browsers Chrome and Firefox and reference management tools such as Zotero. Additional tools include reports and dashboards, badges, and automated reference checks.scite can be used by researchers to locate evidence and evaluate references; librarians to enhance research impact projects; publishers and editors to check reference lists of submissions [**6, 7**]; and journals, publishers, and databases to create context and showcase impact [**4, 8, 9**].

scite aims to contextualize the citation count
Citation counts are commonly used to determine the impact of articles or bodies of work. Academic search engines might use citation counts in relevancy ranking algorithms [**10**]. In reference consultations and library instruction sessions, I have learned that students use citation counts to select articles. Faculty and researchers are evaluated on their scholarly works, and citation counts are often included in tenure packets.

Citation count is a unidimensional metric. Each citation statement, no matter where it appears in the citing article or whether it provides supporting or contrasting evidence, counts equally as 1.

A highly-cited article may be a seminal paper all related articles cite in the introduction to be taken seriously by reviewers. Perhaps the highly-cited article clearly describes an established method or best practices (i.e. [11]) and so is commonly mentioned in methods sections. Alternatively, an article may be cited often because the results are repeatedly being overturned [12]. The use and citation of disputed articles can impact research trends, clinical practice, and patient health and well-being [13, 7]. It is important to note here that, because "(a) scite classifies citation statements at the level of the claim, not the full paper, and (b) the citing article making the contradicting claim itself could be without merit", "a contradicting citation statement does not necessarily mean the cited paper is wrong" [14].

Ultimately, citation count alone does not show how or why an article is being referenced. Yet, this count is used to determine the impact of scholarship (i.e. [15]). The plethora of LibGuides on research impact and citation counts, among other metrics, suggests librarians are highly engaged in this space, that faculty seek new ways to demonstrate their impact, and that administrators and funders are keenly interested in tools to evaluate faculty and funded researchers. We and our researchers could benefit from using tools like scite to contextualize these counts.

scite uses text mining and artificial intelligence
scite uses machine learning to classify citation statements as supporting, mentioning, or contrasting. "While the scite classification model is proprietary, it is based on an open-source deep learning classifier... The corpus on which the model was trained included 43,665 citation statements classified by trained annotators with experience in a variety of scientific fields" [14].

scite continues to build partnerships with publishers to gain access to article full-text for text mining. They signed an agreement with Sage in early 2021 [16], adding to pre-existing indexing agreements with publishers from BMJ to Wiley [4].

scite tools for researchers

Researchers use scite to see how others cite their publications and how their results fit into the larger landscape. Are their works being referenced in methods sections? Are their results being supported or contrasted in new works? Researchers can display this information by customizing and adding the scite badge to their websites.

With a free account, researchers can create a very limited number of reports and visualizations and set up author alerts. A paid account allows researchers to access the reference check feature, which alerts authors of the uploaded manuscript to references that may have been disputed, retracted, or otherwise received an editorial notice. Most publishers do not check reference lists for retracted or disputed papers; authors can ensure they are citing articles appropriately by using scite.ai's tool to alert them to references that may be in dispute [7].

Cite responsibly

Using the scite browser and Zotero plug-ins, librarians, researchers, and students can see more nuanced citation counts, which can help them better evaluate articles. For instance, a student who sees a high citation count might opt to reference an article. However, if the student notes a high number of contrasting citation statements, the student might be encouraged to select a different piece, the findings of which have not been in such dispute.

Title	Creator	Year	Publi...	Suppo...	Contra...	Mention...	Total Smart Citations	Total Distinct Citing Public
Anthocyanin profile of young red wines...	González...	2016	Intern...	⊘2	⊘0	⊘17	19	▥26
Information needs of poultry farmers in...	Msoffe a...	2016	Infor...	⊘2	⊘0	⊘8	10	▥4
The Economics of Agricultural Informati...	Diekman...	2009	Revie...	⊘0	⊘0	⊘10	10	▥18
Communicating Agricultural Informatio...	Mugwisi	2015	Libri: I...	⊘0	⊘0	⊘2	2	▥4
Information use in rural development	Meyer	2003	New ...	⊘0	⊘0	⊘0	1	▥24
Use and Contribution of Information So...	Kapoor a...	2015	Journ...	⊘0	⊘0	⊘1	1	▥2

Figure 1: In the desktop version of Zotero, the scite plug-in displays columns with counts of supporting, contrasting, and mentioning citation statements.

scite smart citations are beginning to appear on databases, including EuropePMC, and journal websites [4, 9]. Smart citations add one more way to evaluate articles in the broader context of research and scholarly communication. Librarians should keep up with these developments to prepare for questions from their research communities.

Publications posted on non-publisher websites or stored in personal reference libraries may lack updated information and context, i.e. links to errata or retraction notices [17]. The scite browser

plug-in provides insights across databases and platforms, so long as the underlying metadata is present, contextualizing, to some extent, works appearing on non-publisher sites. The scite Zotero plug-in offers access to the scite data, embedded in a reference library and regularly updated with information on citation statements to saved articles. This can help users quickly check reference libraries for papers, the findings of which have been contrasted.

Support scholarly impact work
A scite dashboard provides citation counts, with the nuance of supporting, mentioning, and contrasting citations, as well as counts of editorial notices related to a set of papers. When using dashboards, there are options to set notifications for events such as new citations to papers within the collection or search results set. Additionally, the dashboard calculates the scite index for 2-year, 5-year, and lifetime citations. The scite index is calculated by dividing the number of supporting citation statements by the sum of supporting and contrasting citation statements [**18**]. As in all impact metrics work, papers must be old enough to have accumulated cites of a significant amount. scite also offers a searchable table of scite index values for journals in their system [**19**].

scite can also be used to conduct research
The scite corpus has been used to check the quality of references cited in Wikipedia articles [**14**]. scite data could be used in reviews looking at how datasets, software, and research tools are used, by restricting a citation search to publications citing tools in their Methods sections [**20**]. While some databases, for instance PubMed Central, offer the ability to restrict searches to the Materials and Methods sections of papers, this feature is not widely available. The use of scite to locate studies employing various tools and techniques as part of their methodology could be a unique and valuable use of the database.

Another potential avenue of research could involve the use of the reference check feature, for instance, examining evidence syntheses, i.e. systematic reviews and clinical guidelines, for the inclusion of references with significant numbers of contrasting statements from other publications or editorial concerns.

Use scite with a grain of salt
The algorithm which classifies citation statements as supporting, mentioning, or contrasting is still evolving. To reflect this, scite displays confidence levels for classifications and offers

opportunities to flag misclassified cites [**21**]. Currently, ""Most citations [in scite] are classified as 'mentions', because the classifier is trained to be cautious..." says [Giovanni] Colavizza [an AI scientist at the University of Amsterdam and visiting researcher at the Alan Turing Institute in London], who is a user of the platform and whose team has analysed data from the start-up [scite] in the past" [**22**]. A high count of mentioning citation statements may provide little more nuance or detail than a total citation count.

As with all citation indexes, methods and data sources vary. While scite continues to partner with publishers to expand the database, the corpus to which they have access is incomplete. In scite, the count is the number of individually extracted citation statements, not the number of citing articles. Due to this variation, as well as the corpus from which the statements are extracted, different numbers are found across platforms like Scopus, Web of Science, PubMed, Google Scholar, and Dimensions, among others. By way of example, the article, "PRESS peer review of electronic search strategies: 2015 guideline statement" [**23**], has been cited by 493 articles in PubMed, 835 publications with 590 extracted citation statements in scite, 740 publications in Scopus, 671 publications in Web of Science, and 1,078, according to Google Scholar (as of June 4, 2021).

As scite's founders note, "Our results should be considered with caution given the limitations of the model precision, the current limited coverage of articles analyzed by scite, and the fact that articles that could not be linked to a DOI in the data set were excluded" [**14**].

scite Alternatives

While scite is imperfect, the algorithm, corpus, and accompanying tools continue to evolve to provide nuanced insights into how and where works are being used. scite fits into a larger landscape of tools to contextualize research articles and impact. While scite restricts its focus to publication types with DOIs, mainly research articles and scholarly publications, altmetrics, for instance, expands the scope to social media and blogs, reference management tools like Mendeley, policy documents, and other publication types.

As of April 2021, Edifix, a reference editing tool, "identifies retracted or corrected articles and articles published in journals with fraudulent or unprofessional practices", as determined by Cabells Predatory Reports [**24**]. In contrast to scite, the Cabells check provided by Edifix relies on journal-level, rather than article-level, data.

CrossRef's Cited By service works with publishers to create and maintain metadata-enabled connections among citing and cited articles [**25**]. This program enables linking and updates to citation counts but does not provide the same nuance as scite does in terms of where the citation statement is located in the citing paper and whether the citation statement is supporting, contrasting, or mentioning the cited article.

Depending on the features of interest, free and low-cost alternatives to scite are available. If the most important feature is the identification of retracted papers, the Retraction Watch plug-in for Zotero flags papers that have been retracted and identified as such by Retraction Watch [**26**].

Pricing Options
scite offers four pricing tiers, including a free option.

With a free account, users can search the smart citation database, generate one scite report and one visualization per month, create one custom dashboard, and set author and paper alerts. Search results up to 1,000 can be exported as CSVs, which include bibliographic metadata and unique identifiers; supporting, contrasting, mentioning, and total cites; and links for the full scite reports. As described above, a dashboard provides data on smart citations for a set of articles. A scite report provides detailed results, including each extracted citation statement and its classification as supporting, mentioning, or contrasting, for an individual article. The free account does not include the ability to export scite reports. Due to the one per month report limit and the restriction on report downloads, real and regular use is severely restrained under a free account, though alerts can be useful for current awareness and monitoring.

A basic plan, costing $7.99 per month or $50 per year, offers unlimited reports and visualizations, plus the option to export reports as CSV files, a valuable feature for librarians supporting research impact and bibliometrics projects.

The premium plan, at $14.99 per month, or $100 per year, adds unlimited custom dashboards with up to 1000 DOIs, unlimited reference checks, unlimited alerts, and unlimited saved search alerts with up to 1000 DOIs. With higher limits, librarians or principal investigators can monitor and regularly report the impact of article sets, such as those generated by larger departments.

Researchers can use the reference check tool to ensure they are not citing articles that have been highly disputed in their reference lists.

Enterprise plans provide higher article limits, access to data and functionalities, such as APIs, and additional support.

Summary

As the technology, tools, and articles available for text mining continue to evolve, the full value of scite remains to be seen. Use of the free browser and reference manager plug-ins can provide some interesting insights to reference librarians, students, and researchers. Any purposeful uses require paid subscriptions.

Acknowledgments

References

1. Nicholson J. A deep learning platform to evaluate the reliability of scientific claims by citation analysis [Internet]. NIH RePORTER; 2019 [cited 10 June 2021]. <https://reporter.nih.gov/project-details/9885663>.
2. Nicholson J. A deep learning platform to evaluate the reliability of scientific claims by citation analysis [Internet]. NIH RePORTER; 2020 [cited 10 June 2021]. <https://reporter.nih.gov/project-details/10136941>.
3. Lazebnik Y. Award #1913619 - SBIR Phase I: scite: a deep learning platform to identify confirming and refuting citations [Internet]. National Science Foundation; 2020 [cited 10 June 2021].
4. Herther NK. Scite.ai update part 1: creating new opportunities [Internet]. Charleston Hub; 2021 [cited 10 June 2021]. <https://www.charleston-hub.com/2021/02/scite-ai-update-part-1-creating-new-opportunities-an-atg-original/>.

5. scite help desk. How are citations classified? [Internet]. scite; 2019 [cited 10 June 2021]. <https://help.scite.ai/en-us/article/how-are-citations-classified-1a9j78t/>.

6. STM Publishing News. Aries systems and scite partner to evaluate research validity of manuscripts in EM [Internet]. STM Publishing News; 2021 [cited 10 June 2021]. <https://www.stm-publishing.com/aries-systems-and-scite-partner-to-evaluate-research-validity-of-manuscripts-in-em-2/>.

7. Piller C. Disgraced COVID-19 studies are still routinely cited. Science, 2021 Jan 22; 371(6527): 331–332. DOI: <https://doi.org/10.1126/science.371.6527.331>

8. Nicholson J. All citations aren't created equal [Internet]. The Wiley Network; 2020 [cited 10 June 2021]. <https://www.wiley.com/network/latest-content/all-citations-aren-t-created-equal>.

9. Araujo D. Europe PMC integrates smart citations from scite.ai [Internet]. EuropePMC; 2020 [Cited 26 Feb 2021]. <http://blog.europepmc.org/2020/01/europe-pmc-integrates-smart-citations.html>.

10. Rovira C, Codina L, Guerrero-Solé F, & Lopezosa C. Ranking by relevance and citation counts, a comparative study: Google Scholar, Microsoft Academic, WOS and Scopus. Future Internet, 2019; 11(9): 202. DOI: <https://doi.org/10.3390/fi11090202>,

11. Moher D, Liberati A, Tetzlaff J, Altman DG, & PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine. 2009; 6(7): e1000097.

12. Suelzer EM, Deal J, Hanus KL, Ruggeri B, Sieracki R, & Witkowski E. Assessment of citations of the retracted article by Wakefield et al with fraudulent claims of an association between vaccination and autism. JAMA Network Open. 2019; 2(11): e1915552. DOI: <https://doi.org/10.1001/jamanetworkopen.2019.15552>.

13. Steen RG. Retractions in the medical literature: How many patients are put at risk by flawed research? Journal of Medical Ethics. 2011; 37(11): 688–692.

14. Nicholson JM, Uppala A, Sieber M, Grabitz P, Mordaunt M, & Rife SC. Measuring the quality of scientific references in Wikipedia: an analysis of more than 115M citations to over 800 000 scientific articles. FEBS J. 2020. DOI: https://doi.org/10.1111/febs.15608.

15. Muller JZ. The tyranny of metrics [Internet]. The Chronicle of Higher Education; 2018 [cited 10 June 2021]. <https://www.chronicle.com/article/the-tyranny-of-metrics/>.

16. SAGE Publishing. Scite and Sage Publishing sign indexing agreement to increase discoverability and reproducibility through smart citations [Internet]. SAGE Publishing; 2021 [cited 10 June 2021].

    <https://group.sagepub.com/press-releases/scite-and-sage-publishing-sign-indexing-agreement-to-increase-discoverability-and-reproducibility-through-smart-citations>.

17. Davis PM. The persistence of error: a study of retracted articles on the Internet and in personal libraries. Journal of the Medical Library Association: JMLA. 2012; 100(3): 184-9. DOI: https://doi.org/10.3163/1536-5050.100.3.008.

18. Rosati D. How are journals cited? characterizing journal citations by type of citation. ArXiv. 2021. http://arxiv.org/abs/2102.11043.

19. scite Inc. Scite: smart citations for better research [Internet]. scite.Ai; 2021 [cited 10 June 2021]. <https://scite.ai>.

20. Brody S. The Unified Medical Language System: a scoping review of its use in research [Internet]. National Library of Medicine; 2020 [cited 26 Apr 2021] <https://www.nlm.nih.gov/about/training/associate/associate_projects/BrodyScopingReview2019.pdf>.

21. scite help desk. How do I fix a mis-classified citation? [Internet]. scite; 2019 [cited 10 June 2021]. <https://help.scite.ai/en-us/article/how-do-i-fix-a-mis-classified-citation-7no39b/>.

22. Khamsi R. Coronavirus in context: scite.ai tracks positive and negative citations for COVID-19 literature [Internet]. Nature; 2020 [cited 10 June 2021]. <https://doi.org/10.1038/d41586-020-01324-6>.

23. McGowan J, et al. PRESS peer review of electronic search strategies: 2015 guideline statement. Journal of Clinical Epidemiology. 2016 July; 75: 40–46. DOI: https://doi.org/10.1016/j.jclinepi.2016.01.021.

24. Edifix [Internet]. Edifix; 2021 [cited 10 June 2021]. <https://edifix.com/features>.

25. Cited-By [Internet]. Crossref; 2021 [cited 10 June 2021]. <https://www.crossref.org/services/cited-by/>.

26. Stillman D. Retracted item notifications with Retraction Watch integration [Internet]. Zotero; 2019 [cited 10 June 2021]. <https://www.zotero.org/blog/retracted-item-notifications/>.