Google Scholar

From Wikipedia, the free encyclopedia

Google Scholar is a freely accessible web search engine that indexes the full text or metadata of scholarly literature across an array of publishing formats and disciplines. Released in beta in November 2004, the Google Scholar index includes most peer-reviewed online academic journals and books, conference papers, theses and dissertations, preprints, abstracts, technical reports, and other scholarly literature, including court opinions and patents.[1] While Google does not publish the size of Google Scholar's database, third-party researchers estimated it to contain roughly 160 million documents as of May 2014[2] and an earlier statistical estimate published in PLOS ONE using a Mark and recapture method estimated approximately 80-90% coverage of all articles published in English.[3]

Google Scholar is similar in function to the freely available CiteSeerX and getCITED. It also resembles the subscription-based tools, Elsevier's Scopus and Thomson Reuters' Web of Science.

	Google Scholar
Web address	scholar.google.com (http://scholar.google.com)
Type of site	Bibliographic database
Registration	Optional
Owner	Google
Launched	November 20, 2004
Current status	Active

Contents

- 1 History
- 2 Features and specifications
- 3 Ranking algorithm
- 4 Limitations and criticism
- 5 See also
- 6 References
- 7 External links

History

Google Scholar arose out of a discussion between Alex Verstak and Anurag Acharya,[4] both of whom were then working on building Google's main web index.[5][6] Their goal was to "make the world's problem solvers 10% more efficient" [7] by allowing easier and more accurate access to scientific knowledge. This goal is reflected in the Google Scholar's advertising slogan – "Stand on the shoulders of giants" – taken from a quote by Isaac Newton and is a nod to the scholars who have contributed to their fields over the centuries, providing the foundation for new intellectual achievements.

Scholar has gained a range of features over time. In 2006, a citation importing feature was implemented supporting bibliography managers (such as RefWorks, RefMan, EndNote, and BibTeX). In 2007, Acharya announced that Google Scholar had started a program to digitize and host journal articles in agreement with their publishers, an effort separate from Google Books, whose scans of older journals do not include the metadata required for identifying specific articles in specific issues.[8] In 2011, Google removed Scholar from the toolbars on its search pages,[9] making it both less easily accessible and less discoverable for users not already aware of its existence. Around this period, sites with similar features such as CiteSeer, Scirus, and Microsoft Windows Live Academic search were developed. All three of these are now defunct.

A major enhancement was rolled out in 2012, with the possibility for individual scholars to create personal "Scholar Citations profiles", public author profiles that are editable by authors themselves.[10] Individuals, logging on through a Google account with a bona fide address usually linked to an academic institution, can now create their own page giving their fields of interest and citations. Google Scholar automatically calculates and displays the individual's total citation count, h-index, and i10-index. According to Google, "three quarters of Scholar search results pages [...] show links to the authors' public profiles" as of August 2014.[10]

A feature introduced in November 2013 allows logged-in users to save search results into the "Google Scholar library", a personal collection which the user can search separately and organize by tags.[11] A metrics feature now supports viewing the impact of academic journals,[12] and whole fields of science, via the "metrics" button. This reveals the top journals in a field of interest, and the articles generating these journal's impact can also be accessed.

Features and specifications

Google Scholar allows users to search for digital or physical copies of articles, whether online or in libraries.[13] It indexes "full-text journal articles, technical reports, preprints, theses, books, and other documents, including selected Web pages that are deemed to be 'scholarly.'"[14] Because many of Google Scholar's search results link to commercial journal articles, most people will be able to

access only an abstract and the citation details of an article, and have to pay a fee to access the entire article.[14] The most relevant results for the searched keywords will be listed first, in order of the author's ranking, the number of references that are linked to it and their relevance to other scholarly literature, and the ranking of the publication that the journal appears in.[15]

Using its "group of" feature, it shows the available links to journal articles. In the 2005 version, this feature provided a link to both subscription-access versions of an article and to free full-text versions of articles; for most of 2006, it provided links to only the publishers' versions. Since December 2006, it has provided links to both published versions and major open access repositories, but still does not cover those posted on individual faculty web pages; access to such self-archived non-subscription versions is now provided by a link to Google, where one can find such open access articles.

Through its "cited by" feature, Google Scholar provides access to abstracts of articles that have cited the article being viewed.[16] It is this feature in particular that provides the citation indexing previously only found in CiteSeer, Scopus and Web of Science. Through its "Related articles" feature, Google Scholar presents a list of closely related articles, ranked primarily by how similar these articles are to the original result, but also taking into account the relevance of each paper.[17]

As of July 2013, Google Scholar is not yet available to the Google AJAX API.

Google Scholar's legal database of US cases is extensive. Users can search and read published opinions of US state appellate and supreme court cases since 1950, US federal district, appellate, tax and bankruptcy courts since 1923 and US Supreme Court cases since 1791.[16] Google Scholar embeds clickable citation links within the case and the How Cited tab allows lawyers to research prior case law and the subsequent citations to the court decision.[18] The Google Scholar Legal Content Star Paginator extension inserts Westlaw and LexisNexis style page numbers in line with the text of the case.[19]

Ranking algorithm

While most academic databases and search engines allow users to select one factor (e.g. relevance, citation counts, or publication date) to rank results, Google Scholar ranks results with a combined ranking algorithm in a "way researchers do, weighing the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature".[15] Research has shown that Google Scholar puts high weight especially on citation counts[20] and words included in a document's title.[21] As a consequence, the first search results are often highly cited articles.

Limitations and criticism

Quality — Some searchers consider Google Scholar of comparable quality and utility to commercial databases.[22][23] The reviews recognize that its "cited by" feature in particular poses serious competition to Scopus and Web of Science. An early study, from 2007, limited to the biomedical field, found citation information in Google Scholar to be "sometimes inadequate, and less often updated".[24] The coverage of Google Scholar may vary by discipline compared to other general databases.[25]

Coverage — Especially early on, some publishers did not allow Scholar to crawl their journals. Elsevier journals have been included since mid-2007, when Elsevier began to make most of its ScienceDirect content available to Google Scholar and Google's web search.[26] As of February 2008 the absentees still included the most recent years of the American Chemical Society journals. Google Scholar does not publish a list of scientific journals crawled, and the frequency of its updates is unknown. It is therefore impossible to know how current or exhaustive searches are in Google Scholar, although a recent study[3] estimates that Google Scholar can find almost 90% (approximately 100 million) of all scholarly documents on the Web written in English. Nonetheless, it allows easy access to published articles without the difficulties encountered in some of the most expensive commercial databases.[27]

Matthew effect — Google Scholar puts high weight on citation counts in its ranking algorithm and therefore is being criticised for strengthening the Matthew effect;[20] as highly cited papers appear in top positions they gain more citations while new papers hardly appear in top positions and therefore get less attention by the users of Google Scholar and hence fewer citations.

Google Scholar effect – It is a phenomenon when some researchers pick and cite works appearing in the top results on Google Scholar regardless of their contribution to the citing publication because they automatically assume these works' credibility and believe that editors, reviewers, and readers expect to see these citations.[28]

Incorrect field detection — Google Scholar has problems identifying publications on the arXiv preprint server correctly. Interpunctuation characters in titles produce wrong search results, and authors are assigned to wrong papers, which leads to erroneous additional search results. Some search results are even given without any comprehensible reason.[29][30]

Vulnerability to spam — Google Scholar is vulnerable to spam.[31][32] Researchers from the University of California, Berkeley and Otto-von-Guericke University Magdeburg demonstrated that citation counts on Google Scholar can be manipulated and complete non-sense articles created with SCIgen were

indexed from Google Scholar.[33] They concluded that citation counts from Google Scholar should only be used with care especially when used to calculate performance metrics such as the h-index or impact factor. Google Scholar started computing an h-index in 2012 with the advent of individual Scholar pages. Several downstream packages like *Harzing's Publish or Perish* also use its data.[34] The practicality of manipulating h-index calculators by spoofing Google Scholar was demonstrated in 2010 by Cyril Labbe from Joseph Fourier University, who managed to rank "Ike Antkare" ahead of Albert Einstein by means of a large set of SCIgen-produced documents citing each other (effectively an academic link farm).[35]

Inability to shepardize case law — As of 2010, Google Scholar was not able to shepardize case law, as Westlaw and Lexis can.[36]

Lack of screening for quality — Google Scholar strives to include as many journals as possible, including predatory journals, which "have polluted the global scientific record with pseudo-science, a record that Google Scholar dutifully and perhaps blindly includes in its central index."[37]

See also

- List of academic databases and search engines
- Citation index
- CiteSeer
- getCITED
- Institute for Scientific Information's Web of Science
- Live Search Academic
- Microsoft Academic Search
- Paperpile
- Scirus
- Scopus
- Google Scholar and academic libraries

References

1. "Search Tips: Content Coverage". *Google Scholar*. Google. Retrieved 27 April 2016.
2. Orduña-Malea, E.; Ayllón, J.M.; Martín-Martín, A.; Delgado López-Cózar, E. (2014). *About the size of Google Scholar: playing the numbers.* Granada: EC3 Working Papers, 18: 23 July 2014. ArXiV eprint (http://arxiv.org/abs/1407.6239)
3. *Trend Watch* (2014) Nature **509**(7501), 405 – discussing Madian Khabsa and C Lee Giles (2014) *The Number of Scholarly Documents on the Public Web* (http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0093949), PLOS ONE **9**, e93949.

4. Giles, J. (2005). "Science in the web age: Start your engines". *Nature* **438** (7068): 554–555. Bibcode:2005Natur.438..554G. doi:10.1038/438554a. PMID 16319857.
5. Hughes, Tracey (December 2006). "An interview with Anurag Acharya, Google Scholar lead engineer". *Google Librarian Central*.
6. Assisi, Francis C. (3 January 2005). "Anurag Acharya Helped Google's Scholarly Leap". INDOlink.
7. Steven Levy (2015) The gentleman who made Scholar (https://medium.com /backchannel/the-gentleman-who-made-scholar-d71289d9a82d). "Back channel" on Medium.
8. Quint, Barbara (August 27, 2007). "Changes at Google Scholar: A Conversation With Anurag Acharya". Information Today.
9. Madrigal, Alexis C. (3 April 2012). "20 Services Google Thinks Are More Important Than Google Scholar". Atlantic.
10. Alex Verstak: "Fresh Look of Scholar Profiles (http://googlescholar.blogspot.ch /2014/08/fresh-look-of-scholar-profiles.html)". Google Scholar Blog, August 21, 2014
11. James Connor: "Google Scholar Library (http://googlescholar.blogspot.de/2013/11 /google-scholar-library.html)". Google Scholar Blog, November 19, 2013
12. "International Journal of Internet Science - Google Scholar Citations". Scholar.google.com. Retrieved 2014-08-22.
13. Google Scholar Library Links (https://scholar.google.com/intl/en/scholar /librarylinks.html)
14. Vine, Rita (January 2006). "Google Scholar". *Journal of the Medical Library Association* **94** (1): 97–9. PMC 1324783.
15. "About Google Scholar". Scholar.google.com. Retrieved 2010-07-29.
16. "Google Scholar Help". Google.
17. Official Google Blog: Exploring the scholarly neighborhood (http://googleblog.blogspot.com/2006/08/exploring-scholarly-neighborhood.html)
18. Dreiling, Geri (May 11, 2011). "How to Use Google Scholar for Legal Research". Lawyer Tech Review.
19. "Google Scholar Legal Content Star Paginator". Chrome.google.com. Retrieved 2011-06-06.
20. Jöran Beel and Bela Gipp. Google Scholar's Ranking Algorithm: An Introductory Overview. In Birger Larsen and Jacqueline Leta, editors, Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09), volume 1, pages 230–241, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935. (http://www.sciplore.org /publications/2009-Google_Scholar%27s_Ranking_Algorithm_-- _An_Introductory_Overview_--_preprint.pdf)
21. Beel, J.; Gipp, B. (2009). "Google Scholar's ranking algorithm: The impact of citation counts (An empirical study)". *2009 Third International Conference on Research Challenges in Information Science*: 439–446. doi:10.1109/RCIS.2009.5089308.
22. Bauer, Kathleen; Bakkalbasi, Nisa (September 2005). "An Examination of Citation Counts in a New Scholarly Communication Environment". *D-Lib Magazine* **11** (9). doi:10.1045/september2005-bauer.

23. Kulkarni, A. V.; Aziz, B.; Shams, I.; Busse, J. W. (2009). "Comparisons of Citations in Web of Science, Scopus, and Google Scholar for Articles Published in General Medical Journals". *JAMA: The Journal of the American Medical Association* **302** (10): 1092–1096. doi:10.1001/jama.2009.1307. PMID 19738094.

24. Falagas, M. E.; Pitsouni, E. I.; Malietzis, G. A.; Pappas, G. (2007). "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses". *The FASEB Journal* **22** (2): 338–342. doi:10.1096/fj.07-9492LSF. PMID 17884971.

25. Kousha, K.; Thelwall, M. (2007). "Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis". *Journal of the American Society for Information Science and Technology* **57** (6): 1055–65. doi:10.1002/asi.20584.

26. Brantley, Peter (3 July 2007). "Science Direct-ly into Google". *O'Reilly Radar*.

27. Jacsó, P. (2008). "Google Scholar revisited". *Online Information Review* **32**: 102–114. doi:10.1108/14684520810866010.

28. Serenko, A.; Dumay, J. (2015). "Citation classics published in knowledge management journals. Part II: Studying research trends and discovering the Google Scholar Effect" (PDF). *Journal of Knowledge Management* **19** (6): 1335–55. doi:10.1108/JKM-02-2015-0086.

29. Jacso, Peter (24 September 2009). "Google Scholar's Ghost Authors, Lost Authors, and Other Problems". *Library Journal*.

30. Jacsó, P.�t. (2010). "Metadata mega mess in Google Scholar". *Online Information Review* **34**: 175–191. doi:10.1108/14684521011024191. replacement character in `|first1=` at position 3 (help)

31. Preprint PDF (http://www.sciplore.org/publications/2010-On_the_Robustness_of_Google_Scholar_against_Spam--preprint.pdf)

32. Scholarly Open Access – Did A Romanian Researcher Successfully Game Google Scholar to Raise his Citation Count? (http://scholarlyoa.com/2015/01/20/did-a-romanian-researcher-successfully-game-google-scholar-to-raise-his-citation-count/)

33. Beel, Joeran; Gipp, Bela (December 2010). "Academic search engine spam and google scholar's resilience against it". *Journal of Electronic Publishing* **13** (3). doi:10.3998/3336451.0013.305.

34. "Publish or Perish - Anne-Wil Harzing". Harzing.com. Retrieved 2013-06-15.

35. Labbe, Cyril (2010). "Ike Antkare one of the great stars in the scientific firmament" (PDF). *Laboratoire d'Informatique de Grenoble RR-LIG-2008 (technical report)*. Joseph Fourier University.

36. Oliver Benn (March 9, 2010). "Is Google Scholar a Worthy Adversary?". Law Technology News.

37. Beall, Jeffrey (November 2014). "Google Scholar is Filled with Junk Science". *Scholarly Open Access*. Retrieved 2014-11-10.

External links

- Google Scholar WebSite (http://scholar.google.com)
- Google Scholar Blog (http://googlescholar.blogspot.com)

	Wikidata has a property, ***P1960***, for Google Scholar ID (see uses)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Google_Scholar&oldid=722965122"