

Computing giants launch free science metrics

New Google and Microsoft services promise to democratize citation data.

BY DECLAN BUTLER

Mapping the landscape of science is about to get easier than ever before. Google and Microsoft are rolling out free tools that will enable researchers to analyse citation statistics, visualize research networks and track the hottest research fields.

The systems could be attractive for scientists and institutions that are unable — or unwilling — to pay for existing metrics platforms, such as Thomson Reuters' Web of Knowledge and Elsevier's Scopus database.

Launched in 2004 as a search engine for academic publications, Google Scholar last month added Google Scholar Citations (GSC), which lets a researcher create a personal profile showing all their articles in the Google Scholar database (go.nature.com/7wkpea). The profile also shows plots of the number of citations these papers have received over time, and other citation metrics including the popular *h*-index, which attempts to measure both the productivity of a scientist and the overall impact of their publications. The service is currently in invitation-only beta testing, but Google intends eventually to roll it out to all researchers.

Meanwhile, Microsoft Academic Search (MAS), which launched in 2009 and has a tool similar to GSC, has over the past few months added a suite of nifty new tools based on its citation metrics (go.nature.com/u1ouut). These include visualizations of citation networks (see 'Mapping the structure of science'); publication trends; and rankings of the leading researchers in a field.

But although Microsoft's platform has many more features, Google Scholar has an enormous size advantage at present that makes its metrics far more accurate and reliable, say researchers. Google Scholar has indexed much more of the literature than has Microsoft, or indeed Web of Knowledge or Scopus. By contrast, MAS often turns up only a fraction of an author's true publications, which can result in its citation metrics having "absurdly low" values, says Péter Jacsó, an information scientist at the University of Hawaii in Honolulu.

"Microsoft Academic Search is still a nascent offering to the community," explains Lee Dirks, director of education and scholarly

communication at Microsoft Research Connections, the academic-collaboration arm of Microsoft Research. MAS's content surged from 15.7 million to 27.1 million publications between March and June, and that pace will continue, says Dirks. Anne-Wil Harzing at the University of Melbourne, Australia, who develops tools to extract citation metrics from Google Scholar, says that MAS has "great potential".

Some researchers question whether purely computational approaches can ever generate reliable bibliographic databases and citation

Scholar is so low that they do not greatly affect the accuracy of more robust metrics calculations such as the *h*-index.

Google Scholar also has an advantage over commercial providers in its extensive coverage of books — a significant research output in the social sciences and humanities — as well as conference proceedings, which are important outputs in the computing and engineering fields. Covering these is "crucial" to producing accurate metrics in these fields, says Ton van Raan, a bibliometrics expert at the Centre for Science and Technology Studies at Leiden

University, the Netherlands. Joel Hammond, director of product development at Thomson Reuters, points out that the Web of Knowledge already indexes conference proceedings, and that it plans to launch a book-citation index this autumn. Scopus has similar plans.

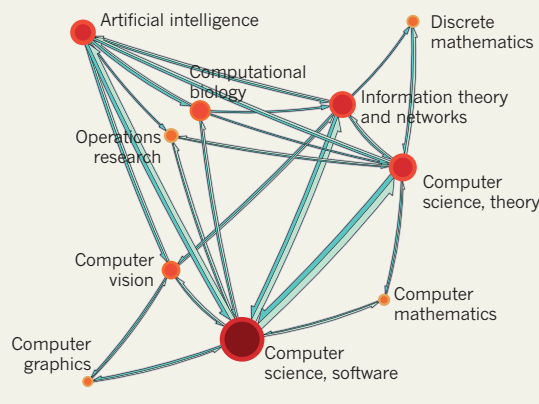
Neither MAS nor GSC see themselves as direct competitors of Web of Knowledge or Scopus, however. "This is not about competition, this is about providing an open platform for academic research," says Dirks. Acharya, who was born in India, says that he is driven by a humanitarian goal: making available to everybody services that were previously accessible only to those at richer institutions. He says he finds it "satisfying" that Google Scholar's server logs reveal widespread use by researchers in poorer countries, where commercial services are often unavailable.

Hammond says that Thomson Reuters controls which publications it indexes more strictly than do the free services, and argues that this makes its metrics calculations more reliable. Scopus takes a similar line. But others say GSC and MAS might eventually become good enough for many users. "They have the major advantage of being freely available to anyone, and with continued development I think they have the potential to become serious competitors to the commercial products," says Carl Bergstrom, a biologist at the University of Washington, Seattle, who collaborates with both Microsoft Research and Thomson Reuters to analyse citation data.

Van Raan agrees. "It is clear that the commercial citation index producers will be more and more in competition with these free-access facilities," he says. ■

MAPPING THE STRUCTURE OF SCIENCE

Analysis of citations from 1.7 million computer-science publications in Microsoft Academic Search data reveals the relative importance of research fields, and the flow of citations between them.



metrics without some human intervention to clean up and check the data. Jacsó points out that the text-mining software used by MAS and GSC can sometimes extract erroneous bibliographic information from publications, for example by misidentifying author names or affiliations (P. Jacsó *Online Inform. Rev.* **34**, 175–191; 2010).

Anurag Acharya, the Google engineer behind Google Scholar and its new metrics system, counters that it has long since dealt with such issues, and that a stack of recent improvements means that his system is working "better and better". Harzing adds that

NATURE.COM
For more on
metrics see:
www.nature.com/metrics

critics often focus too much on such extreme bibliographic errors. She estimates that the overall level of errors in Google