

## AI science search engines expand their reach

## Semantic Scholar triples in size and Microsoft Academic's relaunch impresses researchers.

## Nicola Jones

11 November 2016



Stuart Isett/Polaris/eyevine

The AI-based 'Semantic Scholar' is led by Oren Etzioni, chief executive of the Allen Institute for Artificial Intelligence in Seattle, Washington.

A free AI-based scholarly search engine that aims to outdo Google Scholar is expanding its corpus of papers to cover some 10 million research articles in computer science and neuroscience, its creators announced on 11 November. Since its [launch last year](#), it has been joined by several other AI-based academic search engines, most notably a relaunched effort from computing giant Microsoft.

[Semantic Scholar](#), from the non-profit Allen Institute for Artificial Intelligence (AI2) in Seattle, Washington, unveiled its new format at the Society for Neuroscience annual meeting in San Diego. Some scientists who were given an early view of the site are impressed. "This is a game changer," says Andrew Huberman, a neurobiologist at Stanford University, California. "It leads you through what is otherwise a pretty dense jungle of information."

The search engine first launched in November 2015, promising to sort and rank academic papers using a more sophisticated understanding of their content and context. The [popular Google Scholar](#) has access to about 200 million documents and can scan articles that are behind paywalls, but it searches merely by keywords. By contrast, Semantic Scholar can, for example, assess which citations to a paper are most meaningful, and rank papers by how quickly citations are rising — a measure of how 'hot' they are.

When first launched, Semantic Scholar was restricted to 3 million papers in the field of computer science. Thanks in part to a collaboration with AI2's sister organization, the Allen Institute for Brain Science, the site has now added millions more papers and new filters catering specifically for neurology and medicine; these filters enable searches based, for example, on which part of the brain part of the brain or cell type a paper investigates, which model organisms were studied and what methodologies were used. Next year, AI2 aims to index all of PubMed and expand to all the medical sciences, says chief executive Oren Etzioni.

"The one I still use the most is Google Scholar," says Jose Manuel Gómez-Pérez, who works on semantic searching for the software company Expert System in Madrid. "But there is a lot of potential here."

## Microsoft's revival

Semantic Scholar is not the only AI-based search engine around, however. Computing giant Microsoft quietly released its own AI scholarly search tool, [Microsoft Academic](#), to the public this May, replacing its predecessor, Microsoft Academic Search, which the company stopped adding to in 2012.

Microsoft's academic search algorithms and data are available for researchers through an application programming interface (API) and the [Open Academic Society](#), a partnership between Microsoft Research, AI2 and others. "The more people working on this the better," says Kuansan Wang, who is in charge of Microsoft's effort. He says that Semantic Scholar is going deeper into natural-language processing — that is, understanding the meaning of full sentences in papers and queries — but that Microsoft's tool, which is powered by the semantic search capabilities of the firm's web-search engine Bing, covers more ground, with 160 million publications.

Like Semantic Scholar, Microsoft Academic provides useful (if less extensive) filters, including by author, journal or field of study. And it compiles a leaderboard of most-influential scientists in each subdiscipline. These are the people with the most 'important' publications in the field, judged by a recursive algorithm (freely available) that judges papers as important if they are cited by other important papers. The top neuroscientist for the past six months, according to Microsoft Academic, is Clifford Jack of the Mayo Clinic, in Rochester, Minnesota.

Other scholars say that they are impressed by Microsoft's effort. The search engine is getting close to combining the advantages of Google Scholar's massive scope with the more-structured results of subscription bibliometric databases such as Scopus and the Web of Science, says Anne-Wil Harzing, who studies science metrics at Middlesex University, UK, and [has analysed the new product](#). "The Microsoft Academic phoenix is undeniably growing wings," she says. Microsoft Research says it is working on a personalizable version — where users can sign in so that Microsoft can bring applicable new papers to their attention or notify them of citations to their own work — by early next year.

Other companies and academic institutions are also developing AI-driven software to delve more deeply into content found online. The Max Planck Institute for Informatics, based in Saarbrücken, Germany, for example, is developing an engine called DeepLife specifically for the health and life sciences. "These are research prototypes rather than sustainable long-term efforts," says Etzioni.

In the long term, AI2 aims to create a system that will answer science questions, propose new experimental designs or throw up useful hypotheses. "In 20 years' time, AI will be able to read — and more importantly, understand — scientific text," Etzioni says.

*Nature* | doi:10.1038/nature.2016.20964