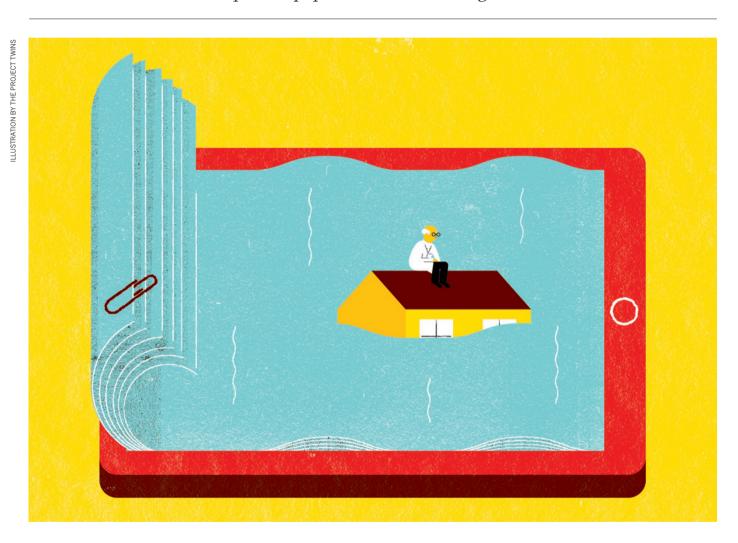# HOW TO TAME THE FLOOD OF LITERATURE

Recommendation services claim to help researchers keep up with the most important papers without becoming overwhelmed.

BY ELIZABETH GIBNEY

Casey Bergman's daily research routine used to include checking all his e-mails and web alerts to pick out fresh papers in his field. But he grew dissatisfied with table-of-contents alerts from journals, RSS (Rich Site Summary) feeds and automated e-mails from the PubMed database. The flow of content was manageable, but if he left it for more than a day, "it became a burden", he says.

So last year Bergman, a computational geneticist studying fruit flies (*Drosophila*) at the University of Manchester, UK, turned to a fresh approach: an automated Twitter account (or 'twitterbot') that he named FlyPapers. The bot trawls PubMed and the arXiv preprint server to find papers containing the word *Drosophila*, and spits them out into its followers' feeds. Bergman finds it much easier to catch up with FlyPapers popping up in his Twitter feed — and his idea has spawned

around 55 twitterbots in other disciplines.

It is no surprise that academics are coming up with their own ways to keep on top of the flood of literature. "It's a common struggle," says Bergman. A staggering 6,000 papers are published every day — and although no one wants to be overloaded with recommendations, missing key papers is "mortifying", says Sally Burn, a developmental geneticist at Columbia University in New York City. She uses a service called Scizzle, which regularly sends her the results of saved

RECOMMENDATION ENGINES
A guide to reading

Google Scholar (scholar.google.com)
Sends alerts about recommended papers on the basis of a user's publication history.

ReadCube (readcube.com)
Mendeley (mendeley.com)
Reference managers with recommendation engines (Readcube is backed by Digital Science, a sister firm to Nature Publishing Group).

PubChase (pubchase.com)
Recommends papers based on the libraries of users with similar interests.

Sparrho (sparrho.com)
Asks the user to train its recommendations engine by approving or rejecting suggestions.

Faculty of 1000 Prime (f1000.com/prime)
Sends alerts about biomedical articles, using the ratings of 5,000 senior scientists.

Twitter (twitter.com)
Automated twitterbots can track keywords (see go.nature.com/bppzew for instructions), or users can follow colleagues, as they can on other social networks such as ResearchGate.

Nowomics (nowomics.com)
Users 'follow' biological keywords such as specific genes, proteins or processes.

Scizzle (myscizzle.com)
Automates the process of making multiple PubMed searches with keywords and filters, and allows users to bookmark relevant papers.

PubMed searches. "Unless you have all day, and ten people working for you trawling the literature, I think it's the best situation you're going to get," she says.

But a stream of papers based on keywords only scratches the surface of what is technologically possible. Emerging literature-recommendation engines promise not only to filter the flood of papers to a trickle, but also to learn from their users' interests to add personalized suggestions (see 'A guide to reading'). "In spirit, it's similar to what Netflix or Amazon do," says Matthew Davis, a computational biologist at the University of Texas at Austin who wrote the algorithm for one such service, PubChase — now owned by ZappyLab, a firm in Berkeley, California, that makes web- and phone-based tools for scientists.

IF YOU LIKE THAT, YOU'LL LIKE THIS
One of the first, and still best-known, services comes from Google Scholar. Its Updates tool suggests articles by applying a statistical model to a record of a researcher's authored papers and citations. "The recommendations are almost scarily good," says Roger Schonfeld, programme director at Ithaka S+R, a non-profit consultancy based in New York City that advises academia on digital technology. But graduate students may not have a sufficient body of work for the site to help, notes Patrick Mineault, a computational neuroscientist at the University of California, Los Angeles.

PubChase suggests articles from PubMed on the basis of a user's publishing record, but it also learns from the articles that the user has read and stored in his or her online library. And it adds another machine-learning technique: comparing this library with other people's collections, with the logic that people with common research interests might benefit from each others' preferences. "I've been really impressed: nearly every article it has recommended has been relevant to my research," says Kelsey Wood, a geneticist at the University of California, Davis, who uses the service along with reference-manager tool Mendeley, owned by Amsterdam-based publisher Elsevier.

Ross Mounce, an evolutionary biologist at the University of Bath, UK, says that PubChase is not useful for those whose interests fall outside the boundaries of PubMed. He prefers Sparrho, a fledgling London-based venture that generates recommendations with a keyword-based feed, and asks users to train the tool by flagging suggestions as relevant or irrelevant. It

includes articles, grants, patents, posters and conference proceedings from all the sciences. "The breadth is a real strength," says Mounce. As with PubChase, recommendations are based on connections between similar users. "We're allowing intelligent curators, humans, to join the scattered dots," says chief executive Vivian Chan, who co-founded Sparrho after she struggled to keep up with the literature while studying for a biochemistry PhD at the University of Cambridge, UK.

As start-ups seeking investment, PubChase and Sparrho are guarded about how many users they have. It is clear that numbers are small. (A *Nature* survey of more than 3,000 scientists found that only 8% had heard of PubChase, and fewer than 1% visited it regularly; see *Nature* **512,** 126–129 (2014).) But both say that their user base is growing quickly.

BACK TO BASICS
Bergman is wary of algorithm-based searches. A machine that learns and tailors recommendations can become like "blinders on your intellectual scope", he says. And he has found that the interdisciplinary nature of his work, which melds genomics and text-mining, confused Google Scholar — the tool threw up irrelevant papers and missed important ones. But Davis says that this narrowing is counteracted by the new doors opened by recommendations based on the profiles of people with similar interests.

Many researchers eschew algorithms altogether, and simply follow colleagues on social networks to find out what is worth reading. "Twitter is the unsung hero of the paper-recommendation world," says Cassie Ettinger, a geneticist in the same research group as Wood. Other scientists check which papers rise to the top in online communities or among users of reference-management services such as Faculty of 1000 Prime and Mendeley.

But the desire to share recommendations or upload libraries to find new papers is hardly universal. Derek Lowe, a chemist at Vertex Pharmaceuticals in Boston, Massachusetts, who writes the blog In the Pipeline, remains a fan of RSS feeds from journal websites. And Burn says that she does not have the time to train a recommendation engine. Mineault acknowledges that automated learning devices will never find all the papers a scientist wants, but he thinks that they will improve. Techniques for gleaning meaning from content will become more sophisticated, he says, and will eventually have a significant role in guiding scientists' reading choices.

For Bergman, a lot of this is a matter of taste. His twitterbot has convened an online fruit-fly community; its suggestions have been retweeted by researchers in other disciplines, and even by non-scientists. Bergman has not ruled out trying further technologies, but he is sticking to FlyPapers for now. "I haven't felt the need to try any others. It's working for me, and that's all that matters," he says. ■ **SEE EDITORIAL P.6**

→ MORE ONLINE

Q&A

Software carpentry: the boot camp for scientific software.
go.nature.com/eijvxi

SHUTTERSTOCK/ISAK55

ARTICLE COLLECTIONS
● Culture of code: are scientists' practices up to the task?
● Online collaboration: what makes a winning formula?
● Open data: how to share research online.

For more on scientific software, apps and online tools, visit nature.com/toolbox