

of 22 projects asks volunteers to do everything from identifying galaxy types in astronomical images to transcribing historical weather records. Robert Simpson, a developer and head of communications for the Zooniverse team, says that the five-year-old scheme has 930,000 registered participants and that there is fairly

consistent interest in new projects.

Quantifying the effects of different motivational tools is difficult, says Grey, whose cyberscience centre has received funding to explore the possible benefits of common rules and credit schemes across different platforms such as BOINC and the Zooniverse. “Because

of its grass-roots nature, everyone’s doing their own thing; there’s no common metric,” he says.

One thing is certain: there is still plenty of spare brainpower to access. “US citizens alone spend 200 billion hours watching television a year,” says Simpson. “We only need to tap a tiny fraction of that.” ■

PUBLISHING

Elsevier opens its papers to text-mining

Researchers welcome easier access for harvesting content, but some spurn tight controls.

BY RICHARD VAN NOORDEN

Academics: prepare your computers for text-mining. Publishing giant Elsevier says that it has now made it easy for scientists to extract facts and data computationally from its more than 11 million online research papers. Other publishers are likely to follow suit this year, lowering barriers to the computer-based research technique. But some scientists object that even as publishers roll out improved technical infrastructure and allow greater access, they are exerting tight legal controls over the way text-mining is done.

A few years ago, scientists complained that publishers were stymieing ambitious plans to use computer software to pull out information from published papers. Some researchers who ran software to harvest data from online articles found their programs blocked, and those who asked for permission found themselves trapped in tortuous case-by-case negotiations — even though they had already paid subscription fees for access. Max Haeussler, a computational biologist at the University of California, Santa Cruz, for instance, spent more than three years arguing with publishers for permission to extract DNA data from 3 million articles to annotate an online map of the human genome (see *Nature* 483, 134–135; 2012).

“It was a legitimate criticism, that people sent text-mining requests in to publishers and they bounced around for a time without any response,” admits Chris Shillum, vice-president of product management for platform and content at Elsevier. The publisher previously considered requests “case by case”, he says — but it now wants to make text-mining permissions quicker and easier to obtain. “What we’ve tried to do is take the practical barriers away.”

Under the arrangements, announced on 26 January at the American Library Association

conference in Philadelphia, Pennsylvania, researchers at academic institutions can use Elsevier’s online interface (API) to batch-download documents in computer-readable XML format. Elsevier has chosen to provisionally limit researchers to 10,000 articles per week. These can be freely mined — so long as the researchers, or their institutions, sign a legal agreement. The deal includes conditions: for instance, that researchers may publish the products of their text-mining work only under a licence that restricts use to non-commercial purposes, can include only snippets (of up to 200 characters) of the original text, and must include links to original content.

“Finally, someone is showing that there is no need to be afraid of text-mining analysis any more,” says Haeussler.

Researchers working on the Human Brain Project — a European consortium that plans to use a supercomputer to recreate everything known about the human brain — have already used Elsevier’s interface to do text-mining, says the project’s spokesman Richard Walker, who is based at the Swiss Federal Institute of Technology in Lausanne. “We are very pleased with it. It resolves genuine technical issues,” he says.

And neuroscientist Shreejoy Tripathy at the University of British Columbia in Vancouver, Canada, worked with Elsevier last year to pull out information on neuron physiology from thousands of articles (see *neuroelectro.org*). Text-mining is not yet well known, he says, but he hopes that the easier access will kick off its greater adoption among scientists. “As more papers get published that use text-mining, other researchers like myself — who

“Finally, someone is showing that there is no need to be afraid of text-mining analysis.”

are neuroscientists and not programmers — will see the need for the technique,” he says.

Shillum says that Elsevier is ahead of the curve — but that other publishers are likely to follow soon. CrossRef, a non-profit collaboration of thousands of scholarly publishers, will in the next few months launch a service that lets researchers agree to standard text-mining terms and conditions by clicking a button on a publisher’s website, a ‘one-click’ solution similar to Elsevier’s set-up.

And, in the past year, large institutions and pharmaceutical companies have started to ask for text- and data-mining rights when renegotiating site licences, says Jessica Rutt, rights and licensing manager at Nature Publishing Group (NPG), the publisher of this journal. Anyone with those rights may mine NPG content. Many publishers are also experimenting with delivering text-minable content to pharmaceutical companies for an extra fee, she adds.

But some researchers feel that a dangerous precedent is being set. They argue that publishers wrongly characterize text-mining as an activity that requires extra rights to be granted by licence from a copyright holder, and they feel that computational reading should require no more permission than human reading. “The right to read is the right to mine,” says Ross Mounce of the University of Bath, UK, who is using content-mining to construct maps of species’ evolutionary relationships.

National governments are also weighing in on the issue. The UK government aims this April to make text-mining for non-commercial purposes exempt from copyright, allowing academics to mine any content they have paid for. And the European Commission, worried that barriers to computational research could hinder scientific innovation, is also examining the issue. It has convened a group chaired by Ian Hargreaves, an intellectual-property specialist at Cardiff University, UK, who recommended the changes to UK law, to examine the economic impact of text- and data-mining for scientific research and barriers to its use. The panel will reach conclusions by the end of February.

“Our plan is just to wait for the copyright exemption to come into law in the United Kingdom so we can do our own content-mining our own way, on our own platform, with our own tools,” says Mounce. “Our project plans to mine Elsevier’s content, but we neither want nor need the restricted service they are announcing here.” ■