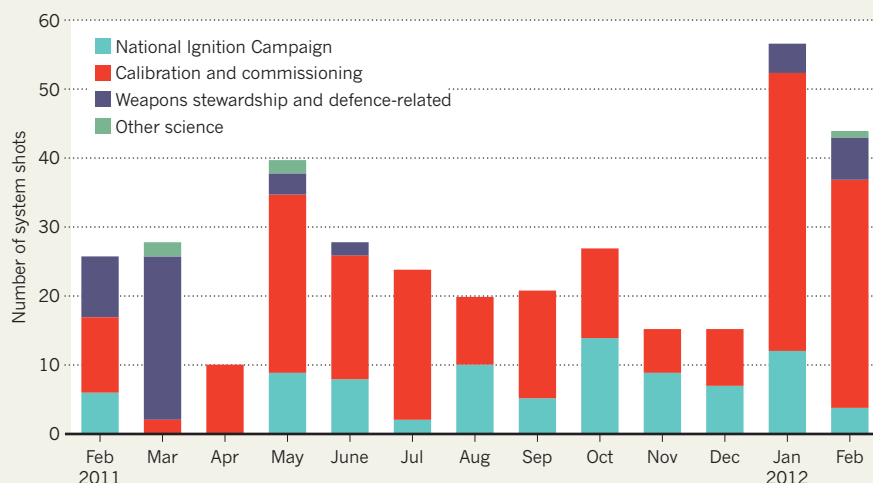


POWER PLAY

The National Ignition Facility's 192-beam laser achieved a record number of shots in January. Most were not directly part of the effort to ignite fusion.



► approach. For example, plasma physicists at the University of Rochester in New York want to adapt NIF's lasers so that they can implode a hydrogen-isotope pellet directly and dispense with the hohlraum.

The NIF scientists aren't waiting for

alternative approaches to catch up, however. Even before achieving ignition, they are racing to plan their next project, a demonstration power plant that they call LIFE, for Laser Inertial Fusion Energy. To be economic, the plant would have to produce more than 50 times

more energy from each shot than it puts in, and would have to boost repetition rates from a few shots a day to 15 per second — no mean feat.

In the quiet, cavernous NIF facility is a mock-up of one of the modular beam lines that would make up LIFE, small enough to fit in the back of a truck. Whereas NIF's set-up uses thousands of bulky flashbulbs to pump energy into the glass lasers, LIFE would use small, transistor-powered light-emitting diodes. Moses dismisses the notion that it's too early to commit to lasers as the drivers of a future power plant. Because of investment in lasers and transistors for consumer electronics, the world has already chosen, he says. Historians will look back, and "they'll see the transistor and the laser as the turning point".

LIFE director Mike Dunne says that the capital costs for the pilot plant would be about \$4 billion, and it could be putting hundreds of megawatts into the grid by the early 2020s — at least a decade earlier than the magnetic-fusion community hopes to deliver a practical power plant. Recalling the first time he presented the LIFE concept to magnetic-fusion researchers at a conference a few years ago, Moses says "The response to it was almost violent: 'This cannot be.' They were shocked at the ambition of it for sure. And they still are." ■

PUBLISHING

Trouble at the text mine

Computers can rapidly scan through thousands of research papers to make useful connections, but work is being slowed by publishers' unease.

BY RICHARD VAN NOORDEN

When he was a keen young biology graduate student in 2006, Max Haeussler wrote a computer program that would scan, or 'crawl', plain text and pull out any DNA sequences. To test his invention, the naive text-miner downloaded around 20,000 research papers that his institution had paid to access — and promptly found his IP address blocked by the papers' publisher.

It was not until 2009 that Haeussler, then at the University of Manchester, UK, and now at the University of California, Santa Cruz, returned to the project in earnest. He had come to realize that standard site licences do not permit systematic downloads, because publishers fear wholesale theft of their content. So Haeussler began asking for licensing terms to crawl and text-mine articles. His goal was to serve science: his program is a key part of the text2genome project, which aims to use DNA sequences in research papers to link the

publications to an online record of the human genome. This could produce an annotated genome map linked to millions of research articles, so that biologists browsing a genomic region could immediately click through to any relevant papers.

But Haeussler and his text2genome colleague Casey Bergman, a genomicist at the University of Manchester, have spent more than two years trying to agree terms with publishers — and often being ignored or rebuffed. "We've learned it's a long, hard road with every journal," says Bergman.

Many publishers say that they will allow their subscribers to text-mine, subject to contract and the text-miners' intentions, and point to a number of successful agreements. But like many early advocates of the technology, Haeussler and Bergman complain that publishers are failing to cope with requests, and so are holding up the progress of research. What is

more, they point out, as text-mining expands, it will be impractical for individual academic teams to spend years each working out bilateral agreements with every publisher.

With his frustration boiling over, Haeussler last week started a project to e-mail all the main science publishers for permission to mine their content. He will log their responses online (at <http://text.soe.ucsc.edu>) in the hope of raising awareness of the problem.

Academia is abuzz with excitement over text-mining. Thanks to growing computer power, software can recognize, extract and index scientific information from vast amounts of plain text, allowing computers to read and organize a body of knowledge that is expanding too fast for any human to keep up. 'Semantic software' is starting to record the relationships between scientific 'entities' — for example, between a particular drug and a specific enzyme.

For pharmaceutical firms, text-mining is "a basic necessity" that assists drug development, says Raul Rodriguez-Esteban, a computational

NATURE.COM
To read more about
text-mining, see:
go.nature.com/rubywp

biologist at the drug giant Boehringer Ingelheim in Ridgefield, Connecticut. Companies routinely create custom databases of proteins, drugs, cell types and the interactions between them, all gleaned from text-mining, he explains. The technology still needs human oversight, but most enthusiasts expect text-mining to be the key to a new kind of scientific discovery based on rich, computer-readable representations of knowledge gathered from plain-text research articles.

But, as Haeussler has discovered, there is a major roadblock. Freely available patents and article abstracts are open for text-mining, but material behind paywalls is not — even when institutions have paid for a site licence. “The licence is oriented towards permitting the human to download and read an article, but not to text-mine it,” says John McNaught, deputy director of the National Centre for Text Mining at the University of Manchester. Even freely accessible papers may not come with permissive licences: of the 2.4 million abstracts listed by PubMedCentral, only 400,000 (17%) are licensed for text-mining.

ILLCIT PROSPECTING

Software programmers can circumvent publishers’ detection systems, for example by ensuring that papers aren’t crawled or downloaded in one batch. This breaches the normal site licence terms, but Haeussler says that papers derived from such technically illegal text-mining have been published in leading journals.

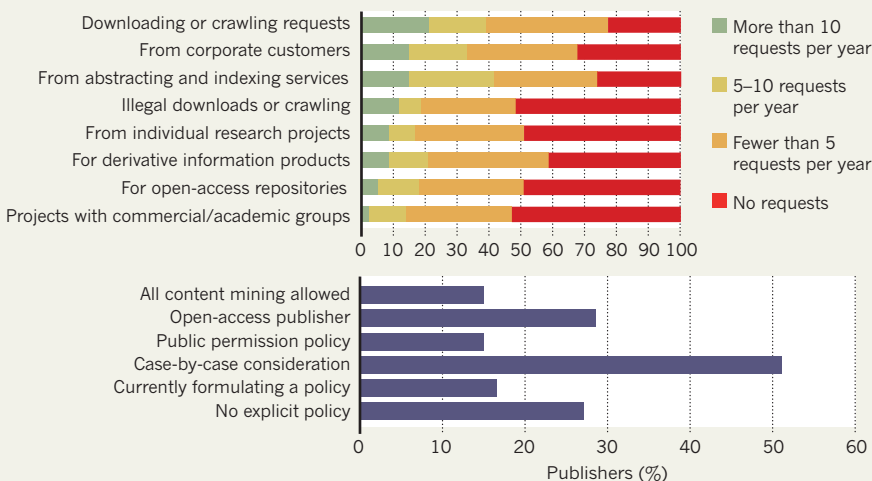
Those wishing to text-mine within the rules must agree contracts with the publishers, and sometimes pay a fee. Haeussler got permission to mine the corpus of Amsterdam-based publisher Elsevier for free. But another academic text-mining project, BioNOT, based at the University of Wisconsin–Milwaukee, was not so fortunate. The collaboration was charged extra for its contract to search Elsevier papers to automatically extract negative results, potentially useful for showing that genes are not related to a disease, for example.

Even powerful drug firms find the negotiations a burden. “When we have licensed and paid for the full text, we feel that we should also have the right to mine it,” says Henning Nielsen, head of the Library and Information Centre at the Danish pharmaceutical firm Novo Nordisk in Bagsværd, Denmark, and president of the Pharma Documentation Ring (PDR), an association of information managers covering 21 of the world’s largest drug firms.

Publishers deal with text-mining requests in various ways. Last year, the Publishing Research Consortium (PRC), a trade body that supports research on scholarly communication,

MINE ALL MINE

About 75% of the 190 publishers polled by the Publishing Research Commission have received text-mining requests (top), but they have a range of policies (some have more than one) on granting access (bottom).



commissioned a survey about content-mining, for which it polled 190 journal publishers (E. Smit and M. Van der Graaf *Learn. Publ.* 25, 35–46; 2012). Of these, 48% said that they had detected illegal crawling and downloads of their content, and 51% had received requests from individual research projects — although most had received fewer than five requests per year (see ‘Mine all mine’). More than half of publishers said that they decide on a case-by-case basis whether to allow access. Of these, one-third said that they would charge for it if the request was for commercial purposes. For example, some publishers seem concerned that if someone text-mines their content to produce a marketable product, it could compete with or supplant their own content. Nature Publishing Group in London, which publishes this journal, says that it does not charge existing subscribers to mine content to which they already have access, subject to contract.

There are signs that policies may soon be clarified. Nielsen says that the PDR hopes to hammer out a solution with major publishers this year, to allow drug firms to text-mine the literature more easily. And last August, the UK government accepted the recommendations of an intellectual-property review that said scientists should be allowed to mine text and data from journal articles without having to ask permission from a copyright owner — although this has not become law, and does not trump current licence agreements, which tend to bar systematic downloading of papers.

On 8 March, the Copyright Clearance Center — an organization based in Danvers, Massachusetts, that works with publishers on rights licensing — is holding a forum in Amsterdam to discuss what publishers should do about text-mining. And the International Association of Scientific Technical and Medical Publishers, a trade body based in Oxford, UK, says that it is working to agree a shared

position on text and data mining, which it expects to resolve by the summer.

Increasingly, publishers are starting to recognize the opportunities of text-mining, and to mine their own content. The PRC survey found that just under half of publishers said that they already do so, with almost one-third of the rest planning to start this year. The work — often contracted out to the same third-party text-mining firms that are employed by the pharmaceutical industry — typically involves computer programs picking out all the chemicals, genes or proteins from a research paper, and in some cases uploading them to online databases.

LIMITED ACCESS

Elsevier is now actively inviting text-miners, including BioNOT, to write programs (or ‘apps’) that crawl through the full text of its research articles to pick out information. Subscribers to Elsevier’s website can access more than 100 of these apps — including Haeussler’s program. But the apps run only within the website, and contracts usually stipulate that the mined content cannot be used elsewhere. This, says Bergman, is of limited use, because the publisher covers only a small amount of the research literature. He and others shudder at the prospect of individual publishers making text-miners adopt different standards, or stipulating that a particular text-mining program can be used only on their papers — effectively destroying the technology’s potential to crawl across the entire research literature.

Publishers are still working out how to take advantage of text-mining, but none wants to miss out on the potential commercial value. “The technology is progressing so quickly that publishers haven’t had time to think it through,” says David Haussler of the University of California, Santa Cruz, who leads the text2genome project. “As soon as they do, they will realize this is a wonderful opportunity.” ■ [SEE EDITORIAL P.124](#)