

How three MIT students fooled the world of scientific journals

April 14 2015



Left to right: Dan Aguayo, Max Krohn, and Jeremy Stribling in 2005. Credit: Frank Dabek

In recent years, the field of academic publishing has ballooned to an

estimated [30,000 peer-reviewed journals](#) churning out some 2 million articles per year. While this growth has led to more scientific scholarship, critics argue that it has also spurred increasing numbers of low-quality "predatory publishers" who spam researchers with weekly "calls for papers" and charge steep fees for articles that they often don't even read before accepting.

Ten years ago, a few students at MIT's Computer Science and Artificial Intelligence Lab (CSAIL) had noticed such unscrupulous practices, and set out to have some mischievous fun with it. Jeremy Stribling MS '05 PhD '09, Dan Aguayo '01 MEng '02 and Max Krohn PhD '08 spent a week or two between class projects to develop "[SCIgen](#)," a program that randomly generates nonsensical computer-science papers, complete with realistic-looking graphs, figures, and citations.

SCIgen emerged out of Krohn's previous work as co-founder of the online study guide SparkNotes, which included a generator of high-school essays that was based on "context-free grammar." SCIgen works like an academic "Mad Libs" of sorts, arbitrarily slotting in computer-science buzzwords like "distributed hash tables" and "Byzantine fault tolerance."

The program was crude, but it did the trick: In April of 2005 the team's submission, "[Router: A Methodology for the Typical Unification of Access Points and Redundancy](#)," was accepted as a non-reviewed paper to the World Multiconference on Systemics, Cybernetics and Informatics (WMSCI), a conference that Krohn says is known for "being spammy and having loose standards."

When the researchers revealed their hoax, calls started coming in from the likes of The Boston Globe, CNN, and the BBC. Stribling's phone was ringing off the hook thanks to his name being listed first on the paper. ("Randomly listed first," he adds proudly.)

In the wake of the international media attention, WMSCI withdrew the team's invitation to attend. Not to be deterred, the students raised \$2,500 to travel to Orlando, Florida, where they rented out a room inside the conference space to hold their own "session" of randomly-generated talks, outfitted with fake names, fake business cards, and fake moustaches.

At the time the stunt may have seemed like nothing more than a silly "gotcha" moment in the tradition of the "Sokal affair," in which an NYU physicist wrote a nonsense paper that was accepted by a journal of postmodern cultural studies. But SCIGen has actually had a surprisingly substantial impact, with many researchers using it to expose conferences with low submission standards. The team's antics spurred the the world's largest organization of technical professionals, the Institute of Electrical and Electronics Engineers (IEEE), to pull its sponsorship of WMSCI; in 2013 IEEE and Springer Publishing removed more than 120 papers from their sites after a French researcher's analysis determined that they were generated via SCIGen. (Just a few weeks ago Springer announced the release of "[SciDetect](#)," an open-source tool that can automatically detect SCIGen papers.)

The trio of CSAIL alumni have since moved on to other things: Aguayo is a technical lead at Meraki; Krohn, who co-founded both SparkNotes and the dating site OKCupid, now runs Keybase, a startup aimed at making cryptography more accessible; and Stribling had stints at IBM, Google, and Nicira before joining Krohn's team at Keybase this month.

But even a decade later, the team's creation improbably lives on. Stribling says the generator still gets 600,000 annual pageviews that manage to crash their CSAIL research site every few months. The creators continue to get regular emails from computer science students proudly linking to papers they've snuck into conferences, as well as notes from researchers urging them to make versions for other disciplines.

"Our initial intention was simply to get back at these people who were spamming us and to maybe make people more cognizant of these practices," says Stribling, before deadpanning: "We accomplished our goal way better than we expected to."

SCIpher

For the 10-year anniversary, the team reconvened for a project that's once again aimed at predatory publishers.

"[SCIpher](#)" lets you hide secret messages inside randomly-generated calls for papers (CFPs) that appear to be coming from (fictional) conferences with names like "the LYGNY Symposium on relational, software-defined technology."

Entering a secret message into SCIpher create text for a ready-to-send CFP that the CFP's recipient can throw back into the generator to recover the original message.

Stribling says he views SCIpher as a cheeky way to trade secrets—not to mention, to poke fun at conferences' ridiculous, jargon-filled names.

"We combined almost-pronounceable acronyms with random buzzwords cribbed from the SCIgen grammar to evoke the kind of niche specialization that results from thousands of concurrent conferences clamoring for authors," says Stribling. "Plus, while an encrypted email would be a big red flag for some investigators, in our experience when you send out a call for papers, it's very unlikely that anyone will read it."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: How three MIT students fooled the world of scientific journals (2015, April 14)
retrieved 19 April 2024 from

<https://phys.org/news/2015-04-mit-students-world-scientific-journals.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--