

*Farhad Manjoo*

---

## Throwing Google at the Book

*Google's new search engine puts a world of knowledge at our fingertips. Publishers say the Internet giant is robbing them of their rightful fees. Maybe it's time to call copyright laws history.*

Just the announcement last December elicited a thrill. Google, the young oracle that brought order and sense to the World Wide Web, now planned to take on the printed word, reaching into major university libraries to scan and digitize all the knowledge contained in books. The company promised to make every printed book as accessible as a Web site, allowing anyone with Internet access to search through every page of every book for any particular word or phrase. Google had signed deals allowing it to scan millions of books at Stanford, the University of Michigan, Harvard, Oxford, and the New York Public Library. So ambitious was the effort that its only real analogues were the stuff of legend and fiction: the lost library at Alexandria and Jorge Luis Borges's fantastical Library of Babel. On hearing of Google's effort, one librarian told the *New York Times*, "Our world is about to change in a big, big way."

A year later, Google's grand plan to digitize the world's books still seems as fantastical as it did when it was first pro-

posed. Earlier this year, the company started scanning books at libraries and on November 3 launched an elegant beta version of its book search engine—but the project faces an uncertain future.

At issue is copyright law: Does Google have the legal right to copy library books and make them searchable online? Trade groups for authors and publishers say no. In September, the Authors Guild, a professional society of more than 8,000 writers, filed suit against Google to stop the scanning project; in October, the Association of American Publishers, which represents large publishing houses, also sued. Both groups charge that Google, which does not plan to ask authors and publishers for permission before it scans their books, would engage in massive copyright infringement—and also cost the book industry a great deal of potential revenue—if it goes ahead with its effort.

Google insists that its project is legal, as it would only offer snippets—one or two sentences—of copyrighted works that publishers had not given the company permission to scan. It also argues that its plan would boost, not reduce, book sales and would be a boon to the book industry. But its quest to bring books to the Web now looks certain to spark a major courtroom battle, and it's a battle that Google, however deep its pockets and well-remunerated its lawyers, is not guaranteed to win.

“One of the great things about this conflict is it points out the absurdity of American copyright law,” says Siva Vaidhyanathan, a media scholar and copyright expert at New York University. Vaidhyanathan believes that what Google wants to do may well be illegal under today's copyright regime. At the same time, he notes that Google can't really create a system that relies on publishers' granting permission to digitize their books, precisely because nobody really knows who owns the rights to all the books in the

library. So, Vaidhyanathan says, Google is stuck; scanning books without asking permission may be illegal, and scanning books after asking permission is impractical to the point of impossibility.

But if copyright law stands in the way of Google's grand aim, isn't it time we thought about changing the law? That's the most salient question raised in the fight over Google's effort to build a digital library. The company—and the host of other firms that will surely follow in its path—is poised to create a tool that could truly change the way we understand, and learn about, the world around us. A loss for Google would echo throughout the tech industry, dictating not just how we use technology to improve books but also how other media—movies, music, TV shows, and even Web pages—are indexed online. Can we really afford to let content owners stand in the way of Google's revolutionary idea?

In response to the legal uncertainty, Google put its scanning project on hold for several months in the summer of 2005, but it has now resumed the project. Last week, the company put its first stash of scanned library books online. Diving into this trove is a trip. You could easily lose days in Google's digital labyrinth, not unlike the way you might walk into the stacks at Stanford or Harvard on a Friday and emerge punch drunk on a Sunday, amazed by the breadth of the work you've seen. The difference is that Google's library is searchable; you can find what you want not just by looking up an author or a Dewey decimal subject but also by typing a particular phrase or quote—"The play's the thing," say, or "Let my people go!"—that you're looking for in a book. Google will look for your search term in every page of every volume in the library and instantly show you images of the pages in each book where the phrase appears.

At the moment, Google's library mostly contains a trove of work published before 1923; copyrights on these books

have expired, and the books' contents, therefore, are in the public domain, free for anyone to use in any way. Amid these titles you'll find all manner of books in Google's stash: Among many other things, there's an illustrated first edition of Henry James's *Daisy Miller*, a 1702 history of France with an exceedingly long title; *Debates of the House of Commons, 1667 to 1694*, which records a certain Mr. Finch arguing against the naturalization of aliens; and a 1785 gardening book, which advises farmers to plant hedges of holly around their corn, since "Holly does not fuck the land" and therefore rob the corn of nutrients. (It's possible, though, that this last one is actually "suck the land" and that Google's text-recognition program made a mistake with the old script.)

Google's collection also includes a vast number of books published after 1923 that publishers have already given Google permission to include—but because these books are under copyright, Google limits their functionality in order to reduce the chance that the service will negatively affect book sales. For instance, searching for the name Calliope in Jeffrey Eugenides's *Middlesex* will yield several page numbers but not the content of all those pages. That way you can't read the entire book through the search interface.

The main fight between Google and publishers involves a third category of books, those that are still under copyright but that publishers have not given Google permission to include in its library. When Google, in the course of scanning books at a library, comes upon a book published after 1923, publishers insist that the company should set it aside and get permission first; Google says that it has the right to scan these books and make them available online. The company insists that it will soon include such books in its library.

At the moment, though, what this means for you is a truncated library. Right now, no text search in Google will return any phrases contained in many popular titles. For

instance, you can't find such titles as *Lolita*; *The Great Gatsby*; *The Best and the Brightest*; *The Da Vinci Code*; or much of anything by John Updike, Philip Roth, Richard Feynman, John McPhee, Shelby Foote, Terry McMillan, Sharon Olds, Julia Child, or Woody Allen.

This is most problematic for obscure books, books you don't know you're looking for. Take this hypothetical scenario: Let's say that somewhere in the stacks at the University of Michigan there is an essay by a writer you've never heard of, on a subject you didn't know about, in a volume no longer in print, by a publishing house no longer in business; let's say, moreover, that even though you don't really know it, this essay is exactly what you're looking for, the answer to all your searching needs, in much the same way you find Web pages every day by people you don't know that turn out to be just the thing. Ideally, as Google envisions it, you could one day go to its search engine; type in a certain bon mot; and find this book, your book. Because it's still under copyright, Google would only show you a few sentences around your search term as it appeared in the text, not the whole volume, but you'd know it was there in the library, and if you wanted it, you'd be free to check it out or find some way to buy it. Without Google's system, you'll never hear of this book.

In such a scenario, proponents of Google's plan see nothing but good—good for the company, for Internet users, and especially for authors. In most copyright disputes between content companies and tech firms, there is often a legitimate question over which party might benefit more from a new technology, notes Fred von Lohmann, an attorney at the Electronic Freedom Foundation (EFF), which sides with Google in this battle. "Take the Napster case," von Lohmann says. In that situation, Napster claimed that its file-swapping tool could increase CD sales by letting

people preview music before they purchased it; the CD industry, meanwhile, said the system had caused a significant drop in sales. Both sides cited numbers to support their arguments, and each theory sounded at least plausible.

“But with the Google Print situation, it’s a completely one-sided debate,” von Lohmann says. “Google is right, and the publishers have no argument. What’s their argument that this harms the value of their books? They don’t have one. Google helps you find books, and if you want to read it, you have to buy the book. How can that hurt them?”

Obscure books—books that are out of print or otherwise hard to get hold of—would stand to gain the most from such a system, and it turns out that there are plenty of such books in the libraries Google plans to scan. Not long ago, the Online Computer Library Center, a nonprofit library research group, set out to count and catalog the books Google would capture in its project. The OCLC determined that at the five research libraries with which Google had formed deals, about 80 percent of the books in the stacks were published after 1923 and still under copyright. But only a small number of these books are currently in print.

Tim O’Reilly, a computer book publisher and sponsor of influential tech conferences, points out that in 2004 only 1.2 million different book titles were sold in the United States, according to Nielsen Bookscan. This means that while a significant number of library books are protected by copyright, they are also out of print—70 percent or more, O’Reilly estimates. These books, he says, represent the “twilight zone” of the publishing world; someone owns them, but since they’re perceived to have no commercial value (because they’re no longer sold in stores), publishers don’t have any incentive to promote and market them, let alone to go through the expense of scanning them and making them searchable online.

Indeed, in many cases the publishers and rights holders of these books are unknown. There is no national registry of copyright holders in the United States, as there is a national registry of patents. Any book published is automatically granted a copyright, and if a book publisher goes out of business, or an author dies, the copyright to the work may well be buried in contracts that long ago turned to dust. “We precluded any possibility of creating a copyright database,” says Vaidhyanathan, and “it’s impossible for a company like Google, or a historian, or a documentary filmmaker, or anyone to find out who owns what. Even publishers don’t know what they own. It’s just impossible.”

O’Reilly is one of few publishers who support Google’s plan, and he likes it precisely because he thinks it will shed light on these little-known titles whose rights holders are hard to track down. “One of the biggest arguments for Google’s approach is that it is the only solution that solves a hard problem,” O’Reilly says. He points out that only 2 percent of books sold in 2004 had more than 5,000 copies purchased; the rest languished in obscurity. And that, he wrote in a recent *New York Times* Op-Ed, “is a far greater threat to authors than copyright infringement, or even outright piracy.” Google, O’Reilly went on to write, “promises an alternative to the obscurity imposed on most books. It makes that great corpus of less-than-bestsellers accessible to all. By pointing to a huge body of print works online, Google will offer a way to promote books that publishers have thrown away, creating an opportunity for readers to track them down and buy them . . . In one bold stroke, Google will give new value to millions of orphaned works.”

But if it’s true that Google’s new system would be good for old books, it’s also true that the system would be a good one for Google, helping to cement its position as the world’s dominant search engine. Nobody knows—and Google isn’t

saying—how much money the company stands to make directly from the library venture. In a recent Op-Ed in the *Wall Street Journal*, Eric Schmidt, Google's CEO, sought to play down the company's profit motive. He pointed out that the company will not place advertisements on search pages for books it scans from libraries. Though Google will place ads on pages for books that publishers have given the company permission to include, it will send publishers the "majority" of revenue for such ads, Schmidt wrote. Google will include a referral link to let people purchase books they find in the library—a "Buy this book" link to several major online bookstores—but the company won't "make a penny on referrals," he wrote.

Rather than making money from the individual book searches, Google's library will pad Google's bottom line by increasing the value of its main search engine. Although Google remains by far the world's most popular search engine, it faces stiff competition from other firms—Yahoo, Microsoft, Amazon, and others—who want a share of its vast audience and are also planning ventures to digitize and offer search systems for print books and other media. By offering something—millions of books—that others are not yet offering, Google will be creating another reason for users to stick with its interface for searching the Web.

But Google's competitors are not far behind. Amazon, which already offers a feature to search inside many books in its store, has just announced a plan to let users buy specific pages of books. Microsoft and Yahoo, meanwhile, have joined the Open Content Alliance, a nonprofit group that includes contributions from the Internet Archive and the University of California at Berkeley and that plans to digitize books only after asking for publishers' permission.

It's Google's profit motive that raises the suspicion of authors and publishers. As they see it, digital technology

provides authors and publishers a new way to make a great deal of money on their back catalogs of books—a huge source of revenue that is currently being untapped. Google is creating a system that exploits that back catalog, so why shouldn't Google pay content owners for the right use of that catalog?

“The author is creating the value here,” says Paul Aiken, executive director of the Authors Guild, “and the author should get some of the money. If there's a new value for books created on the Internet, the authors should be given new incentives to create works for it.”

Aiken compares Google's plan to use books with the way Hollywood uses novels as plots for its movies. When film producers first started making movies from books, “They could have said, ‘Hey, how does it hurt the author if I make a movie from his book?’” Aiken points out. “You could argue, after all, that more people would buy the book because of the movie.” But that's not the way the world works, Aiken says. Hollywood pays publishers for the rights to novels they want to use, and in the same way, Google should pay publishers—who would then distribute money to authors—for the right to add books to its database. Aiken declined to offer a detailed, specific plan by which Google could pay authors for their contributions to its search engine. But he suggested that one idea might look very much like the system that radio stations use to pay musicians. Google could pay an annual licensing fee to publishers, and the money would be distributed to publishers and authors according to how often those books were viewed in the search engine.

Aiken's argument is echoed by publishers. Google, notes Pat Schroeder—the former Democratic congresswoman from Colorado who now heads the Association of American Publishers (AAP)—is rich! Both Schroeder and Aiken

fingered Google's latest earnings report, which showed that the company recorded a 700 percent increase in profits in the third quarter of 2005 compared with the same quarter in 2004. In 2004, the company's revenues exceeded \$3 billion and the firm could make double that much by the end of the current fiscal year. In other words, Google is sitting on a gold mine, and authors and publishers, notably, are not. "They try to sound like they have this high moral purpose so they can't be bothered with permission," Schroeder says of Google. "They tell us it's good for the world, and it's good for publishers. The thing they leave out is it's really good for Google."

The Authors Guild hasn't conducted a survey of its members to determine what they think of the Google plan, but Aiken says the e-mails he gets from authors run overwhelmingly in support of the guild's lawsuit against the company. There's no reason not to believe Aiken; it's not hard to find authors who are deeply suspicious of what Google plans to do. Take Peter Salus, a veteran author of computer books who lives in Toronto. Salus has authored, coauthored, or edited about two dozen books, some of which are in print and some of which are not. He says that he understands the benefits of Google Print—but he just wants the company to do one thing in return: ask his permission.

"I think it's absurd that they think the authors should have to come to them to opt out of the database," rather than the other way around, Salus says. Because Google's project directly benefits from his and his fellow authors' work, Salus says, it's incumbent upon the company to make sure that authors are OK with what it's doing. And what if it's too logistically difficult for Google to find every author of every book in the library and ask his or her permission? "That's tough—it really breaks my heart," he says. "But there is no

burden on me or anybody else to make it easier for them to make money.”

Many authors feel differently. One is Julian Dibbell, author of *My Tiny Life*, a memoir of the author’s life in the virtual computer world called LambdaMOO. When told of Aiken’s theory that Google’s database would use authors like him in the same way that Hollywood might use them and that authors should get paid for allowing their books to go to Google, Dibbell said, “My blood is boiling just as you relay this to me.” As Dibbell sees it, “Google is not piggy-backing on my creative effort in the same parasitic way that a movie based on a novel might be doing.” To Dibbell, Google is acting not like the Hollywood producer who steals an author’s ideas but instead like a book reviewer who popularizes an author’s work. After all, Dibbell notes, book reviewers routinely use snippets from books in their reviews, and magazines and newspapers make loads of money from advertisements they run alongside book reviews. Authors don’t feel entitled to any of that money, he says, so why should they get a slice of the money Google will make from its service? “Given what’s at stake here, which is the creation of a resource that nobody is denying is a good thing, their stance seems wrong to me,” Dibbell says of his fellow authors.

Whether Google is acting more like a book reviewer or like a movie producer in its use of other people’s books may turn out to be a key question in the legal battle to come. Google, which did not make a company attorney available to *Salon*, has insisted that, as with book reviewers, copying books falls within the “fair use” exception of American copyright law. Google essentially argues that because it is copying books as a step toward a larger goal—the creation of a search engine of library books—its actions are permitted. After all, the company points out, it does exactly the

same thing with Web pages. To create a searchable index, the Google Web search engine copies entire Web sites—all of *Salon*, for instance, resides in Google’s servers—without their permission. If copying a Web site is OK, why is copying a book not?

In his *Wall Street Journal* editorial, Google CEO Schmidt defended the company’s legal interpretation. “The aim of the Copyright Act is to protect and enhance the value of creative works in order to encourage more of them—in this case, to ensure that authors write and publishers publish,” he wrote. “We find it difficult to believe that authors will stop writing books because Google Print makes them easier to find, or that publishers will stop selling books because Google Print might increase their sales.”

Fred von Lohmann of the EFF agrees with Google’s view of the law, and he says that several federal court rulings uphold what Google is doing. Courts have already ruled, for instance, that it’s OK for companies to make copies of video games as part of their efforts to reverse-engineer those video games; because the copied video games weren’t meant to be sold and were instead used for some other purpose (the creation of a reverse-engineered product), the copies were considered fair use. Then there’s *Kelly v. Arriba Soft*, a 2002 case in which a judge ruled that it was legal for a search engine company to copy photographs from other sites online and display “thumbnails” of those photos as part of its search results. The thumbnails, the court said, were quite different from the original photographs—they were smaller and of lower resolution—and were unlikely to be used for the same purposes as the originals.

A similar argument can be made on behalf of Google’s book search engine, von Lohmann points out. Google is not giving readers the exact copy of the books it scans from the library; rather, it’s just giving them a snippet in the same

way that a graphical search engine may show users a thumbnail image of a picture.

But if there are cases that support Google's view of copyright law, there are also federal court cases that line up against it. The main case involves MP3.com, a boom-time Internet company that copied tens of thousands of CDs to its internal database without getting record labels' permission to do so. (MP3.com planned to make digital tracks of songs available to anyone who could prove they'd purchased a physical CD of the music.) In 2000, a federal judge in New York ruled that MP3.com's copying of CDs without permission infringed upon the record labels' copyrights. A reasonable judge may look at Google's actions as being essentially no different; just as MP3.com copied CDs without asking, so too is Google copying books.

Both authors and publishers sued Google in the Southern District of New York, where the MP3 case is still an important legal precedent; the choice of locale, says Vaidhyanathan, was not an accident. "They picked the court that is likely to rule along the lines of the MP3.com case and less likely to think that *Kelly v. Arriba Soft* was a good decision," he says.

Aiken of the Authors Guild, for one, is sanguine about his lawsuit's prospects. "I don't think if you were to survey copyright lawyers you'd find their view prevailing," he says of Google. "Our case is very, very strong. It would shock the copyright bar if it was decided against us." Still, both Aiken and Schroeder say they are open to settlement discussions with Google. "There's a lawsuit," says Aiken, "but if they wanted to negotiate a license we could work something out."

Google has not yet filed a legal response in either case, and it's unclear whether the company is open to a settlement. But several observers say Google may also be inclined

to think about sitting down with the authors and publishers and talking about the case, including discussing a possible permission-based system for getting books into its library. For one thing, Google wouldn't want to risk losing in this case, says Vaidhyanathan, as a badly worded ruling in this case could put its other operations in legal jeopardy.

O'Reilly says that he was recently at an event sitting between Larry Page, one of Google's cofounders, and John Sargent, who sits on the board of the AAP. "I'm not going to tell you what they said," O'Reilly says, "but I think it's fair to characterize what they're doing now as negotiating by lawsuit and press release." Each side is hoping for some early legal decisions to go its way, O'Reilly says, and then the real settlement talks will begin.

But whatever happens with Google's venture, a more lasting outcome from this case may be a change in the way we think about how much control an author or publisher, musician or record label, filmmaker or studio is allowed to exert over works they create—a question that has been cast into stark relief in the digital age.

Lawrence Lessig, a Stanford law professor and copyright scholar, likes to tell the story of Thomas Lee and Tinie Causby, two North Carolina farmers who in 1945 cast themselves at the center of a case that would redefine how society thought of physical property rights. The immediate cause of the Causbys' discomfort was the airplane; military aircraft would fly low over their land, terrifying their chickens, who flew to their deaths into the walls of the barn. As the Causbys saw it, the military aircraft were trespassing on their land. They claimed that American law held that property rights reached "an indefinite extent, upwards"; that is, they owned the land from the ground to the heavens. If the government wanted to fly planes over the Causbys' land, it needed the Causbys' permission, they insisted.

The case, in time, came to the Supreme Court, where Justice William O. Douglas, writing for the Court, was not kind to the Causbys' ancient interpretation of the law. Their doctrine, he said, "has no place in the modern world. The air is a public highway, as Congress has declared. Were that not true, every transcontinental flight would subject the operator to countless trespass suits. Common sense revolts at the idea. To recognize such private claims to the airspace would clog these highways, seriously interfere with their control and development in the public interest, and transfer into private ownership that to which only the public has a just claim."

Google supporters say the publishers' objection resembles that of the Causbys'. Just as the airplane rendered the Causbys' rights to the skies incompatible with the modern world, the Web has rendered publishers' right to the digital universe out of tune with modern technology and society. The public benefit of making millions of books, or excerpts of books, readily available to people worldwide "could be the most important contribution to the spread of knowledge since Jefferson dreamed of national libraries," Lessig wrote recently on his blog. "It is an astonishing opportunity to revive our cultural past, and make it accessible. Sure, Google will profit from it. Good for them. But if the law requires Google (or anyone else) to ask permission before they make knowledge available like this, then Google Print can't exist." And if Google Print can't exist, maybe it's time to reexamine the law.