Google Hopes to Open a Trove of Little-Seen Books  
By MOTOKO RICH  
Published: January 4, 2009

MOUNTAIN VIEW, Calif. — Ben Zimmer, executive producer of a Web site and software package called the Visual Thesaurus, was seeking the earliest use of the phrase "you're not the boss of me." Using a newspaper database, he had found a reference from 1953.

But while using Google’s book search recently, he found the phrase in a short story contained in "The Church," a periodical published in 1883 and scanned from the Bodleian Library at Oxford.

Ever since Google began scanning printed books four years ago, scholars and others with specialized interests have been able to tap a trove of information that had been locked away on the dusty shelves of libraries and in antiquarian bookstores.

According to Dan Clancy, the engineering director for Google book search, every month users view at least 10 pages of more than half of the one million out-of-copyright books that Google has scanned into its servers. Google’s book search “allows you to look for things that would be very difficult to search for otherwise,” said Mr. Zimmer, whose site is visualthesaurus.com.

A settlement in October with authors and publishers who had brought two copyright lawsuits against Google will make it possible for users to read a far greater collection of books, including many still under copyright protection.

The agreement, pending approval by a judge this year, also paved the way for both sides to make profits from digital versions of books. Just what kind of commercial opportunity the settlement represents is unknown, but few expect it to generate significant profits for any individual author. Even Google does not necessarily expect the book program to contribute
significantly to its bottom line.

“We did not think necessarily we could make money,” said Sergey Brin, a Google founder and its president of technology, in a brief interview at the company’s headquarters. “We just feel this is part of our core mission. There is fantastic information in books. Often when I do a search, what is in a book is miles ahead of what I find on a Web site.”

Revenue will be generated through advertising sales on pages where previews of scanned books appear, through subscriptions by libraries and others to a database of all the scanned books in Google’s collection, and through sales to consumers of digital access to copyrighted books. Google will take 37 percent of this revenue, leaving 63 percent for publishers and authors.

The settlement may give new life to copyrighted out-of-print books in a digital form and allow writers to make money from titles that had been out of commercial circulation for years. Of the seven million books Google has scanned so far, about five million are in this category.

Even if Google had gone to trial and won the suits, said Alexander Macgillivray, associate general counsel for products and intellectual property at the company, it would have won the right to show only previews of these books’ contents. “What people want to do is read the book,” Mr. Macgillivray said.

Users are already taking advantage of out-of-print books that have been scanned and are available for free download. Mr. Clancy was monitoring search queries recently when one for “concrete fountain molds” caught his attention. The search turned up a digital version of an obscure 1910 book, and the user had spent four hours perusing 350 pages of it.

For scholars and others researching topics not satisfied by a Wikipedia entry, the settlement will provide access to millions of books at the click of a mouse. “More students in small towns around America are going to have a lot more stuff at their fingertips,” said Michael A. Keller, the university librarian at Stanford. “That is really important.”

When the agreement was announced in October, all sides hailed it as a landmark settlement that permitted Google to proceed with its scanning project while protecting the rights and financial interests of authors and publishers. Both sides agreed to disagree on whether the book scanning itself violated authors’ and publishers’ copyrights.

In the months since, all parties to the lawsuits — as well as those, like librarians, who will be affected by it — have had the opportunity to examine the 303-page settlement document and try to digest its likely effects.

Some librarians privately expressed fears that Google might charge high prices for subscriptions to the book database as it grows. Although nonprofit groups like the Open Content Alliance are building their own digital collections, no other significant private-sector competitors are in the business. In May, Microsoft ended its book scanning project, effectively leaving Google as a monopoly corporate player.

David Drummond, Google’s chief legal officer, said the company wanted to push the book database to as many libraries as possible. “If the price gets too high,” he said, “we are simply not going to have libraries that can afford to purchase it.”

For readers who might want to buy digital access to an individual scanned book, Mr.
Clancy said, Google was likely to sell at least half of the books for $5.99 or less. Students and faculty at universities who subscribe to the database will be able to get the full contents of all the books free.

For the average author, “this is not a game changer” in an economic sense, said Richard Sarnoff, chairman of the Association of American Publishers and president of the digital media investments group at Bertelsmann, the parent company of Random House, the world’s largest publisher of consumer books.

“They will get paid for the use of their book, but whether they will get paid so much that they can start living large — I think that’s just a fantasy,” Mr. Sarnoff said. “I think there will be a few authors who do see significant dollars out of this, but there will be a vast number of authors who see insignificant dollars out of this.”

But, he added, “a few hundred dollars for an individual author can equate to a considerable sum for a publisher with rights to 10,000 books.”

So far, publishers that have permitted Google to offer searchable digital versions of their new in-print books have seen a small payoff. Macmillan, the company that owns publishing houses including Farrar, Straus & Giroux and St. Martin’s Press and represents authors including Jonathan Franzen and Janet Evanovich, offers 11,000 titles for search on Google. In 2007, Macmillan estimated that Google helped sell about 16,400 copies.

Authors view the possibility of readers finding their out-of-print books as a cultural victory more than a financial one.

“Our culture is not just Stephen King’s latest novel or the new Harry Potter book,” said James Gleick, a member of the board of the Authors Guild. “It is also 1,000 completely obscure books that appeal not to the one million people who bought the Harry Potter book but to 100 people at a time.”

Some scholars worry that Google users are more likely to search for narrow information than to read at length. “I have to say that I think pedagogically and in terms of the advancement of scholarship, I have a concern that people will be encouraged to use books in this very fragmentary way,” said Alice Prochaska, university librarian at Yale.

Others said they thought readers would continue to appreciate long texts and that Google’s book search would simply help readers find them.

“There is no short way to appreciate Jane Austen, and I hope I’m right about that,” said Paul Courant, university librarian at the University of Michigan. “But a lot of reading is going to happen on screens. One of the important things about this settlement is that it brings the literature of the 20th century back into a form that the students of the 21st century will be able to find it.”

Google’s book search has already entered the popular culture, in the film version of “Twilight,” based on the novel by Stephenie Meyer about a teenage girl who falls in love with a vampire. Bella, one of the main characters, uses Google to find information about a local American Indian tribe. When the search leads her to a book, what does she do?

She goes to a bookstore and buys it.
Google Hopes to Open a Trove of Little-Seen Books

FROM THE NEW YORK TIMES

Google Signs a Deal to E-Publish Out-of-Print Books
(November 10, 2008)

Google Strikes Deal To Allow Book Scans
(October 29, 2008)

THE MEDIA EQUATION; Mourning Old Media’s Decline
(October 29, 2008)

Google to Digitize Newspaper Archives
(September 9, 2008)

Related Articles

INSIDE NYTIMES.COM