Amazon Plan Would Allow Searching Texts of Many Books

By DAVID D. KIRKPATRICK

Executives at Amazon.com are negotiating with several of the largest book publishers about an ambitious and expensive plan to assemble a searchable online archive with the texts of tens of thousands of books of nonfiction, according to several publishing executives involved.

Amazon plans to limit how much of any given book a user can read, and it is telling publishers that the plan will help sell more books while better serving its own online customers.

Together with little-publicized additions to Amazon's Web site, like listings of restaurants and movie showings, the plan appears to be part of a strategy to compete with online search services like Google and Yahoo for consumers' time and attention. Providing a searchable online database of the contents of books could make Amazon a more authoritative source of information, drawing additional traffic to its online retail store.

A spokeswoman for Amazon declined to comment and would not confirm any details of the plan. The publishing executives said Amazon had asked them to keep the plan confidential until the start of the service, which was scheduled for the fall.

Amazon is calling its program Look Inside the Book II, the publishers said. It would expand on a current program that lets shoppers read a table of contents, a first chapter or a few selected pages provided by the publishers of certain books. But Look Inside the Book II would let online browsers search by terms like "Caravaggio," "sans-culottes," or "Osama bin Laden," and then see a list of books mentioning the term along with the sentence that contains it. Browsers could then choose to see several pages around that citation.

But to see those pages Amazon would require users to register, and it plans to limit the amount of any single book a browser can view.
The publishers said they have been guardedly cooperative. Some said they were willing to let Amazon experiment with works of narrative nonfiction, but not reference books, cookbooks or poetry where shoppers might be satisfied with the few pages produced by a search. Others are holding out for further assurances on preventing piracy and guarantees that they will be able to pull their books from the service.

But some publishing executives also noted that Amazon, by far the largest online bookseller, stands to benefit far more than they do. Now, in addition to books, they said, Amazon can sell music, electronics, clothing and other goods to users drawn to the site by the chance to search its digital archive, and the publishers would not receive a cut of that revenue.

How authors will react is another question. Most book contracts allow publishers to give away excerpts for promotional purposes, but authors may contend that Amazon's search service more closely resembles some kind of research system. "This sounds like an anthology right, and that has to be specifically approved by the author, and if a publisher is going to license the electronic rights to the whole work there has be to reasonable compensation for that," said Paul Aiken, the executive director of the Authors Guild.

Amazon appears to be betting heavily on the idea. At a time when Amazon is squeezing hard to lower its costs everywhere else, the company is paying to enter thousands of texts into its searchable database, the publishers said. Although there are many works already in digital format, others would have to be scanned at a cost of more than $200 a book, executives in the industry said. It is unclear how many books Amazon is paying to scan.

Amazon's searchable archive also appears to be part of the company's response to the four-year-old Google, which is changing the way people use the Internet, gather information, advertise and shop. Microsoft is developing its own alternative search engine as part of its online service and just last week Yahoo agreed to acquire Overture, Google's principal rival in the search business, for $1.6 billion.

The challenge for Web sites like Amazon is that Google's search service is increasingly cutting in front of them as the first place consumers go online, diverting shoppers who might have gone straight to Amazon, the most popular online retailer. Google's Web site accounted for about 32 percent of the four billion Internet searches conducted in May, compared to about 25 percent from Yahoo, 19 percent through American Online and 15 percent with Microsoft's MSN service, according to the research company ComScore Networks.

"Everybody is afraid of Google," said Jordan Rohan, an analyst at SoundView Technology Group.

Search-result pages are also experiencing the fastest growth in online advertising. Advertisers can pay to place short text links next to search results like "New York wedding" or "Hawaiian vacation," and through Overture they can pay to appear among the results themselves.

Amazon recently reached an agreement with Google to use its search-based advertising system for searches made on Amazon. (Google sells the advertising and splits the fees with Amazon.)
Likewise, Google also appears to be taking aim at online retailers like Amazon by testing a spinoff called Froogle.com, which makes it easy to browse and comparison shop at scores of online stores at once. Many consumers already begin searching for products or services at Google, and Froogle looks like an attempt to capitalize on Google's technology and reputation to become a shopping destination to rival Amazon, Yahoo or eBay. If shoppers start at Froogle, "What value does Amazon add?" Mr. Rohan asked.

But Google searches only Web sites, not the contents of books, and Amazon's proposed archive might offer a more authoritative alternative source of information. And Google's list of most popular searches on the site suggests that Amazon's planned combination of a digital library with a retail store and film and restaurant listings might be a good place to look for answers to many of the questions.

Earlier this month, a list of "top gaining queries" on Google included "Pirates of the Caribbean," "Angelina Jolie," "Bastille Day" and "Stonehenge." Amazon already offers local movie times and customer reviews for "Pirates" and lists the opening day for Angelina Jolie's coming film, a sequel to "Lara Croft: Tomb Raider." It sells DVD's, video games and other products related to both films. For after the show, Amazon offers searchable database of restaurant names by location with menus and some customer reviews in New York, San Francisco, Boston and other cities.

But with the addition of its planned archive, Amazon could also offer detailed information about Bastille Day and Stonehenge, as well as related books for sale.

A spokesman for Google, David Krane, said the company had a strong relationship with Amazon, which is an important advertiser on Google's search results as well as a showcase for Google's search-related advertisements. And he noted that the two companies remain in different and complementary businesses, one primarily a retailer and the other a research tool.

Amazon is hardly the first to imagine building a searchable archive of the contents of thousands of books. The start-ups NetLibrary, Questia and Ebrary have all pursued similar goals.

NetLibrary, which initially set out to help libraries provide online access to their patrons, filed for bankruptcy protection and was taken over by a consortium of libraries. Questia, which paid to digitize about 45,000 works of nonfiction, charges fees to use its database, and is directed mostly toward students. It advertises on Overture and Google, where plugs for its service turn up next to searches for student paper topics like "postmodernism," for example.

Ebrary, which is partly owned by the Random House division of Bertelsmann, the publisher Pearson, and textbook giant McGraw-Hill, started with a plan similar to Amazon's, an online library that also sold books. Ebrary now offers the contents of about 30,000 documents, including a combination of books, sheet music and other publications, but it has shifted away from its original business. "Generating revenue in that fashion would have taken a much longer period of time," Christopher Warnock, its chief executive, said. Now, Ebrary mainly helps libraries make books available to patrons over the Internet.

Thomas D. Turvey Jr., a vice president of business development for Ebrary, said Amazon's plan was risky because of the potential copyright issues, the threat of
online piracy and the cost of scanning so many books into digital form.

But Mr. Warnock said he spoke with Amazon about collaborating. "It would have been great if we could have figured out a way that we could have worked together on it," he said. Mr. Warnock said that often one Internet company comes up with a good idea and then another, with a lot of traffic, puts it to use. "It seems like the way the world works," he said.