Google, Amazon, and Libraries

Workshop #1

European Library Automation Group (ELAG) 2007
May 9-11, 2007
Barcelona, Spain

Tamar Sadeh, Ex Libris

Introduction

For centuries, libraries were the guardians of human knowledge. Libraries collected, described, stored, and offered their patrons every manifestation of human-created intellectual works—manuscripts, books, musical scores, and journals, to name just a few. Furthermore, libraries provided the infrastructure for people to access these materials and use them, an infrastructure extending from carefully designed catalogs to professional assistance to reading rooms and circulation services. With the advent of the electronic era, automated systems brought an improvement in the management of library collections and services and, later, enabled libraries to provide materials in electronic form. However, it was only with the Internet revolution that the way in which library users perceive and draw on library collections and services underwent a drastic change.

Looking at the changes that have occurred in the last fifteen years, we can discern several interesting trends that are threatening the hegemony of libraries as knowledge authorities and as the informed providers of quality data.

First and foremost, the Internet has opened direct channels for users to access information and has eliminated the need for them to go to the library, or even through the library, in their quest for materials. Not only is a huge amount of information available online, but users can obtain even physical items through various Internet services.

Second, the process of searching has changed: no longer do users need to adapt to the way in which libraries describe and present the world of knowledge; Internet services such as Web search engines and Internet bookstores have created new searching models and practices, offering a search process that is simpler and more intuitive. Furthermore, with these new searching models, users can overcome a lack of searching expertise and still gain immediate gratification, even though the results may be less accurate or of lesser quality than those obtained through the library services.

Third, the use of the Internet has created alternative means of human interaction. Instead of consulting an objective, knowledgeable reference librarian when looking for specific materials, users often base their selections on the opinions and selections of others. Users may do this knowingly; for example, they might check the number of times that an article has been cited, they might follow recommendations ("people who bought this book also bought…"), or they might use tags that other users have coined and applied. However, these patterns of selection have a great influence on the way in which people find information, because Internet tools and services base their presentation of search results on factors related to their business strategies or on analyses of how other users have employed the materials.

A fourth trend that is threatening the authority of libraries is the recent shift in the Internet's role to that of a platform for collaboration, as expressed in various Web 2.0-inspired tools. This platform enables users not only to benefit from other people's advice but also to contribute their own. Users are willing to share their materials, opinions, and expertise, a development that has made the research environment interactive and much more engaging. Librarians' hegemony in describing materials and assessing them is now being replaced by a collective effort that is not limited to describing existing materials but also encourages users to share new, self-published materials and take part in initiatives such as community-created and maintained encyclopedias.

In the last 15 years, many players have helped transform the well-established library arena. Google and Amazon are often mentioned as representatives of Web search engines and Internet bookstores. Although Google was not the first search

engine, it has become the most popular and influential and has expanded its services in a way that is unmatched by other rivals. Amazon.com, the first Internet bookstore, has presented a new, revolutionary model for acquiring books. Furthermore, Amazon has retained its position as the leading Internet bookstore and has expanded its services in terms of content and business models.

During the workshop, we will deal with various services now offered by Google and Amazon and the way in which these companies interact with libraries and library users. We will look at the role of the library in this changing world and examine how libraries are responding to the changes. We will also explore hypothetical directions that libraries may pursue.

Some facts and figures

According to Google's Web site [1], the company opened its first facility in September 1998, at which time it hired its first employee--Craig Silverstein, now Google's director of technology. In December of that year, *PC Magazine* named Google one of its Top 100 Web Sites and Search Engines for 1998. By the early 2000s, the number of Google searches per day reached 700 times the number of searches per *month* in the University of California library catalog [2].

In just a few years, Google has become the most popular Web search engine. SearchEngineWatch.com [3], citing data from comScore Networks' qSearch, states that of the 6.9 billion searches that U.S. Internet users conducted in February 2007, 3.3 billion were conducted with Google, 2 billion with Yahoo, 730 million with MSN, 348 million with Ask.com, and 338 million with Time Warner sites, including AOL.

Viewed from the business perspective, Google is also extremely impressive: the company reported revenues of $3.21 billion for the quarter ended December 31, 2006, an increase of 67% compared to the fourth quarter of 2005 and an increase of 19% compared to the third quarter of 2006 [4].

The Google Labs page [5] lists various services that it now provides and projects that are currently under development. The services that are most applicable to our discussion are Google Scholar [6] and Google Book Search [7]—the former enabling users to search for scholarly materials, and the latter, enabling users to search in digitized books. These two services are Google's first explicit attempts to step into the library's domain.

The Google Scholar site states that "Google's mission is to organize the world's information and make it universally accessible and useful. Facilitating library access to scholarly texts brings us one step closer to this goal. We're thankful to the libraries and librarians who make it possible" [8]. Realizing that good discovery tools are mandatory but not sufficient, Google Scholar partners with libraries when it comes to delivering materials to users; only a few months after its launch in November 2004, the Google Scholar team added OpenURL capabilities to Google Scholar, thus offering libraries the option to direct their users to the libraries' delivery services even when those users are looking for materials through Google Scholar. Other library-oriented features added to Google Scholar include the connection to libraries via OCLC WorldCat®, the inclusion of national library catalogs, document delivery via the British Library Direct service, and links to citation management tools such as RefWorks.

In December 2004, Google initiated its Google Print Library Project, a book digitization effort on an unparalleled scale, through a partnership with five large libraries: four university libraries—those of Harvard University, the University of Michigan, Oxford University, and Stanford University—and the New York Public Library. The combined collections of these institutions are estimated to exceed 15 million volumes [9]. According to Google, "the Library Project's aim is simple: make it easier for people to find relevant books—specifically books they wouldn't find any other way such as those that are out of print—while carefully respecting authors' and publishers' copyrights. Our ultimate goal is to work with publishers and libraries to create a comprehensive, searchable, virtual card catalog of all books in all languages that helps users discover new books and publishers discover new readers." Later, Google changed the name of Google Print to Google Book Search to reflect the way in which people use this digitized collection. Other libraries have joined the project, today forming a group of 13 library partners,

including the Bavarian State Library, the National Library of Catalonia, and the library of the Complutense University of Madrid.

Librarians are watching Google with a mixture of admiration and anxiety. In various forums, some librarians explain how libraries can benefit from Google projects, while other librarians say they are worried by issues such as Google's control over the provision of materials; the way in which it handles copyright matters; and its business motives, which are not entirely clear. Regardless of such reservations, many libraries are cooperating with Google on various levels, such as providing their electronic holdings and even entire catalogs for Google to harvest.

In recent years, we have seen more and more alliances between information providers and Google, although some information providers, such as Elsevier and the American Chemical Society (ACS), are still reluctant to make their data accessible to Google Scholar. Many other providers, including JSTOR, MetaPress, and the Institute of Physics (IOP), have embraced Google and regard Google and Google Scholar as additional entry points—sometimes even key entry points—to their collections and as a means to reach a wider population. According to Gary Coker, the director of research and development at MetaPress, "statistics from early 2006 show that close to 30 per cent of all linking traffic to MetaPress publisher sites originated from Google and Google Scholar" [10]. Sources at JSTOR and IOP estimate unofficially that the traffic coming from Google and Google Scholar constitutes about 75% of their total traffic.

Amazon.com, Inc., was founded in 1994 by Jeff Bezos, who is still the CEO, president, and chairman of the company, and began offering its service in July 1995. The company's revenues in 2006 were $10.71 billion, with a net income of $190 million. It now has about 14,000 employees and expects its net sales in 2007 to be close to $14 billion. [11].

According to a company overview, "Amazon.com strives to be Earth's most customer-centric company where people can find and discover virtually anything they want to buy online. By giving customers more of what they want—low

prices, vast selection, and convenience—Amazon.com continues to grow and evolve as a world-class e-commerce platform" [12].

Jeff Bezos was the first person to completely remove the walls around physical collections. Bezos introduced a catalog that is not bound to actual holdings and that takes advantage of the Internet as an infrastructure. Not only can customers browse through a collection of millions of book titles from their home but they can also complete the entire transaction for obtaining the materials without getting up from their chair. Moreover, the Amazon services guarantee that materials will be sent to the customer's doorstep regardless of the physical distance between the customer and the required materials.

Time Magazine named Bezos its 1999 Man of the Year in recognition of the company's success in popularizing online shopping [13]. Amazon, indeed, has grown in terms of content, offering a large variety of consumer merchandise and services. The Amazon.com site continuously changes and provides diverse user-centric services: long before the library world grasped Web 2.0 concepts, Amazon had already leveraged the data that it gathers about its users' behavior to provide additional services such as recommendations. Furthermore, by extending its offering to used books and enabling customers to sell books as well as to buy them, Amazon has managed to present a collection that is unrivaled in size and availability.

Libraries have regarded Amazon as less of a threat because the Amazon collection is of a slightly different nature—typically focused on new and more popular materials and thus less competitive when compared to the collections of academic libraries. As long as Amazon was offering only new books, it was less of a threat to public libraries, since the Amazon service was more costly; however, now that Amazon.com enables customers to sell back the books that they have acquired, we may see a greater effect on public libraries.

Suggested topics for discussion

1. What is a library collection? Is it what the library manages? Is it what the library users have access rights to? Is it what the library recommends?
2. In today's world of disintermediation, what is the role of the library? Is it limited to providing services, once the discovery process is complete?
3. Is there a need for specialized libraries, tailored to the needs of a specific user community, or can these be replaced by subsets of large services, such as WorldCat?
4. To what extent do today's users need libraries?
5. Can we rely on commercial players to be guardians of human knowledge?
6. Will library practices, mainly cataloging, remain relevant?
7. How might Google develop its "library" products? Do you think it matters if advertisements accompany such products?

References

[1] http://www.google.com/corporate/history.html

[2] Markey, K. The Online Library Catalog: Paradise Lost and Paradise Regained? *D-Lib Magazine*, January/February 2007, Volume 13 Number 1/2. http://www.dlib.org/dlib/january07/markey/01markey.html

[3] http://searchenginewatch.com/showPage.html?page=3625336

[4] http://investor.google.com/releases/2006Q4.html

[5] http://labs.google.com/

[6] http://scholar.google.com/intl/en/scholar/about.html

[7] http://books.google.com/

[8] http://scholar.google.com/intl/en/scholar/libraries.html

[9] http://books.google.com/googlebooks/newsviews/

[10] http://www.researchinformation.info/features/feature.php?feature_id=44

[11] http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=989711&highlight

[12] http://media.corporate-ir.net/media_files/irol/17/176060/OverviewQ12006.pdf

[13] http://www.time.com/time/magazine/article/0,9171,992927,00.html