



The Google Print Library Project: A Copyright Analysis

Journal of Internet Banking and Commerce, December 2005, vol. 10, no.3
(<http://www.arraydev.com/commerce/jibc/>)

By **Jonathan Band**

Email: jband@policybandwidth.com

Mr. Band represents Internet companies and library associations on intellectual property matters. He does not represent any entity with respect to the Google Print project.

On August 11, 2005, Google announced that it would not scan copyrighted books under its Print Library Project until November, so that publishers could decide whether they want to opt their in-copyright books out of the project. Given the confusion in press reports describing the project, publishers should carefully study exactly what Google intends to do and understand the relevant copyright issues. This understanding should significantly diminish any anxiety publishers possess about the project.

The Google Print Project

The Google Print project has two facets: Print Publisher Program and the Print Library Project. Under the Publisher Program, a publisher controlling the rights in a book can authorize Google to scan the full text of the book into Google's search database. In response to a user query, the user receives bibliographic information concerning the book as well as a link to relevant text. By clicking on the link, the user can see the full page containing the search term, as well as a few pages before and after that page. Links would enable the user to purchase the book from booksellers or the publisher directly, or visit the publisher's website. Additionally, the publisher would share in contextual advertising revenue if the publisher has agreed for ads to be shown on their book pages. Publishers can remove their books from the Publisher Program at any time. The Print Publisher Program raises no copyright issues because it is conducted pursuant to an agreement between Google and the copyright holder.

Under the Print Library Project, Google plans to scan into its search database materials from the libraries of Harvard, Stanford, and Oxford Universities, the University of Michigan, and the New York Public Library. In response to search queries, users will be able to browse the full text of public domain materials, but only a few sentences of text around the search term in books still covered by copyright. This is a critical fact that bears repeating: for books still under copyright users will be able to see only a few sentences on either side of the search term. Users will not see a few pages, as under the Publisher Program, nor the full text, as for public domain works. Indeed, a full page of the book is never seen for an in-copyright book scanned as part of the Library Project unless a publisher decides to transfer their book into their Publisher Program account, in which case it would be under the agreement between Google and the copyright holder.

Google's August 11th Announcement

The Association of American Publishers reacted negatively to the Print Library Project. In response to

the AAP's concerns, Google announced on August 11, 2005, that if a publisher provided it with a list of its titles that it did not want Google to scan at libraries, Google would respect that request, even if the book were in the collection of one of the participating libraries. To allow publishers to determine whether they wanted to exclude any of their titles from the Library Project, Google stated that it would not scan any more copyrighted works until November.

Patricia Schroeder, AAP President, stated that 'Google's announcement does nothing to relieve the publishing industry's concerns.' She claimed the Google's optout procedure 'shifts the responsibility for preventing infringement to the copyright owner rather than the user, turning every principle of copyright law on its ear.' The AAP expressed continued 'grave misgivings about ' the Project's unauthorized copying and distribution of copyright-protected works.'

Analysis of the AAP's Copyright Claims

The Print Library Project involves two actions that raise copyright questions. First, Google copies the full text of books into its search database. Second, in response to user queries, Google presents users with a few sentences from the stored text. Because the amount of expression presented to the user is de minimus, this second action probably would not lead to liability. But even if a court did not view the second action as de minimus, both actions fall within the scope of the fair use privilege.

The leading decision that considered the fair use issues relating to search engine operations is *Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003). Arriba Soft operated a search engine for Internet images. Arriba compiled a database of images by copying pictures from websites, without the express authorization of the website operators. Arriba reduced the full size images into thumbnails, which it stored in its database. In response to a user query, the Arriba search engine displayed responsive thumbnails. If a user clicked on one of the thumbnails, she was linked to the full size image on the original website from which the image had been copied. Kelly, a photographer, discovered that some of the photographs from his website were in the Arriba search database, and he sued for copyright infringement. The lower court found that Arriba's reproduction of the photographs was a fair use, and the Ninth Circuit affirmed.

With respect to the first factor, 'the purpose and character of the use, including whether such use is of a commercial nature,' 17 U.S.C. '107(1), the Ninth Circuit acknowledged that Arriba operated its site for commercial purposes. However, Arriba's use of Kelly's images was more incidental and less exploitative in nature than more traditional types of commercial use. Arriba was neither using Kelly's images to directly promote its web site nor trying to profit by selling Kelly's images. Instead, Kelly's images were among thousands of images in Arriba's search engine database. Because the use of Kelly's images was not highly exploitative, the commercial nature of the use weighs only slightly against a finding of fair use.

Kelly at 818.

The court then considered the transformative nature of the use - whether Arriba's use merely superseded the object of the originals or instead added a further purpose or different character. The court concluded that 'the thumbnails were much smaller, lower resolution images that served an entirely different function than Kelly's original images.' *Id.* While Kelly's 'images are artistic works intended to inform and engage the viewer in an aesthetic experience,' Arriba's search engine 'functions as a tool to help index and improve access to images on the internet.' *Id.* Further, users were unlikely to enlarge the thumbnails to use them for aesthetic purposes because they were of lower resolution and thus could not be enlarged without significant loss of clarity. In distinguishing other judicial decisions, the Ninth Circuit stressed that '[t]his case involves more than merely a transmission of Kelly's images in a different medium. Arriba's use of the images serves a different function than Kelly's use - improving access to information on the internet versus artistic expression.' *Id.* at 819. The court closed its discussion of the first fair use factor by concluding that Arriba's 'use of Kelly's images promotes the goals of the Copyright Act and the fair use exception' because the thumbnails 'do not supplant the need for the originals' and they 'benefit the public by enhancing information gathering techniques on the internet.' *Id.* at 820.

Everything the Ninth Circuit stated with respect to Arriba applies with equal force to the Print Library Project. Although Google operates the program for commercial purposes, it is not attempting to profit from the sale of a copy of any of the books scanned into its database, and thus its use is not highly exploitative. The Google search index functions as a tool that makes 'the full text of all the world's books searchable by everyone.' Neither the full text copies in the index, nor the few sentences displayed to users in response to queries, will supplant the original books. Rather, they will bring the books to the user's attention.

With respect to the second fair use factor, the nature of the copyrighted work, the Ninth Circuit observed that '[w]orks that are creative in nature are closer to the core of intended copyright protection than are more fact-based works.' Kelly at 820. Moreover, '[p]ublished works are more likely to qualify as fair use because the first appearance of the artist's expression has already occurred.' Id. Kelly's works were creative, but published. Accordingly, the Ninth Circuit concluded that the second factor weighed only slightly in favor of Kelly. The Print Library Project involves only published works. And while some of these works will be creative, the vast majority will be non-fiction.

The third fair use factor is 'the amount and substantiality of the portion used in relation to the copyrighted work as a whole.' 17 U.S.C. '107(3). The Ninth Circuit recognized that 'copying an entire work militates against a finding of fair use.' Kelly at 820. Nonetheless, the court states that 'the extent of permissible copying varies with the purpose and character of the use.' Id. Thus, 'if the secondary user only copies as much as is necessary for his or her intended use, then this factor will not weigh against him or her.' Id. at 820-21. In Kelly, this factor weighed in favor of neither party:

although Arriba did copy each of Kelly's images as a whole, it was reasonable to do so in light of Arriba's use of the images. It was necessary for Arriba to copy the entire image to allow users to recognize the image and decide whether to pursue more information about the image or the originating web site. If Arriba copied only part of the image, it would be more difficult to identify it, thereby reducing the usefulness and effectiveness of the visual search engine.

Kelly at 821.

In the Print Library Project, Google's copying of entire books into its database is reasonable for the purpose of the effective operation of the search engine; searches of partial text necessarily would lead to incomplete results. Moreover, unlike Arriba, Google will not provide users with a copy of the entire work, but only with a few sentences surrounding the search term. And if a particular term appears many times in the book, the search engine will allow the user to view only three instances - thereby preventing the user from accessing too much of the book. Thus, at least with respect to the search results, the third factor weighs in favor of Google.

The Ninth Circuit decided that the fourth factor, 'the effect of the use upon the potential market for or value of the copyrighted work,' 17 U.S.C. '07(4), weighed in favor of Arriba. The court found that the Arriba 'search engine would guide users to Kelly's web site rather than away from it.' Kelly at 821. Additionally, the thumbnail images would not harm Kelly's ability to sell or license full size images because the low resolution of the thumbnails effectively prevented their enlargement.

Without question, the Print Library Project will increase the demand for some books. The project will expose users to books containing desired information, which will lead some users to purchase the books or seek them out in libraries (which in turn may purchase more copies of books in high demand). It is hard to imagine how the Library Project could actually harm the market for certain books, given the limited amount of text a user will be able to view. To be sure, if a user could view (and print out) many pages of a book, it is conceivable that the user would rely upon the search engine rather than purchase the book. Similarly, under those circumstances, libraries might direct users to the search engine rather than purchase expensive reference materials. But when the user can access only a few sentences before and after the search term, any displacement of sales is unlikely.

Publishers might argue that the Library Project restricts their ability to license their works to search engine providers. The existence of the Print Publisher Program, however, undermines this argument. By participating in Print Publisher Program, publishers receive revenue streams not available to them under the Library Project. And Google presumably prefers for publishers to participate in the Publisher

Program; Google saves the cost of digitizing the content if publishers provide Google with the books in digital format. In sum, under the Ninth Circuit's analysis in *Kelly*, Google's Print Library Project satisfies the requirements of the fair use doctrine.

The Big Picture

Stepping back from the technicalities of the four fair use factors, it becomes clear that the Print Library Project is similar to the everyday activities of Internet search engines. A search engine firm sends out software 'spiders' that crawl publicly accessible websites and copy vast quantities of data into the search engine's database. As a practical matter, each of the major search engine companies copies a large (and increasing) percentage of the entire World Wide Web every few weeks to keep the database current and comprehensive. When a user issues a query, the search engine searches the websites stored in its database for relevant information. The response provided to the user typically contains links both to the original site as well as to the 'cache' copy of the website stored in the search engine's database.

Significantly, the search engines conduct this vast amount of copying without the express permission of the website authors. Rather, the search engine firms believe that the fair use doctrine permits their activities. In other words, the billions of dollars of market capital represented by the search engine companies are based primarily on the fair use doctrine.

In addition to fair use, search engine firms rely on the concept of implied license. Search engine firms assume that if information is posted on a website, the website operator wanted the information to be found by users, and search engines are the most efficient means for users to find the information. Thus, search engine firms assume that most website operators want their sites copied into the search engine database so that users will be able to find the site. If an operator does not want his site crawled and copied, he can use an exclusion header, a software 'Do Not Enter' sign, which most search engine firms respect. But if a website operator does not use an exclusion header, a search engine will assume that the operator wants the site included in the search database.

This implied license theory has not yet been tested in court, and could actually constitute an element of a fair use defense. Courts have described fair use as an 'equitable rule of reason,' *Stewart v. Abend*, 495 U.S. 207, 237 (1990), and industry practice is considered relevant in assessing the reasonableness of a defendant's conduct. Accordingly, a court is likely to excuse as fair use a search engine's copying of a website that did not use an exclusion header, provided that the search engine could show that it typically respected exclusion headers when website operators did employ them.

In the Print Library Project, Google is relying on fair use just as it and its search engine competitors rely on fair use when they copy millions of websites every week. Moreover, by giving publishers the opportunity to opt-out of the Print Library Project, Google is replicating the exclusion header feature of the Internet. Most authors want their books to be found and read. Moreover, authors are aware that an ever increasing percentage of students and businesses conduct research primarily, if not exclusively, online. Thus, if books cannot be searched online, many users will never locate them. The Print Library Project is predicted upon the assumption the authors generally want their books to be included in the search database so that readers can find them. But if a copyright owner does not want Google to scan her book, Google will honor her request.

Contrary to the AAP's assertion, this opt-out feature does not turn 'every principle of copyright law on its ear.' Rather, it is a reasonable implementation of a program based on fair use.

International Dimensions

Fair use under the U.S. Copyright Act is generally broader and more flexible than the copyright exceptions in other countries, including fair dealing in the U.K. Thus, the scanning of a library of books might not be permitted under the copyright laws of most other countries. However, copyright law is territorial; that is, one infringes the copyright laws of a particular country only with respect to acts of infringement that occurred in that country. Since Google presumably will be scanning the books in the United States, the only relevant law with respect to the scanning is U.S. copyright law.

Nonetheless, the search results will be viewable in other countries. This means that Google's distribution of a few sentences from a book to a user in another country must be analyzed under that country's copyright laws. (Google arguably is causing a copy of the sentences to be made in the random access memory of the user's computer.) While the copyright laws of most countries might not be so generous as to allow the reproduction of an entire book, almost all copyright laws do permit short quotations. These exceptions for quotations should be sufficient to protect Google's transmission of Library Project search results to users.

Conclusion

The Google Print Library Project will make it easier than ever before for users to locate the wealth of information buried in books. By limiting the search results to a few sentences before and after the search term, the program will not conflict with the normal exploitation of works nor unreasonably prejudice the legitimate interests of rightsholders. To the contrary, it often will often increase demand for copyrighted works.

15+ million members. 118+ million publications. 700k+ research projects. Join for free. Recommended publications. Discover more publications, questions and projects in Google. Article. Google Flu Trends.