

New York Law Journal



Web address: <http://www.nylj.com>

VOLUME 234—NO. 90

TUESDAY, NOVEMBER 8, 2005

ALM

COMPUTER LAW

BY RICHARD RAYSMAN AND PETER BROWN

Search Engines, Digital Cataloging and Copyright Law

Earlier this year, Google Inc. began a large-scale book-scanning project, commonly known as the “Google Library Project.” Essentially, the project entails the digitization of substantial parts of the collections of selected major libraries to build a comprehensive, searchable database of millions of books.

Users would gain complete access to older books already in the public domain. Moreover, by entering desired search terms, users also would be able to search copyrighted works and view results that contain only the snippets of text that surround the search term, along with bibliographical and online sales information.

In response to the project, the Author’s Guild, a professional organization for published writers, filed suit against Google in the U.S. District Court for the Southern District of New York in September, seeking declaratory and injunctive relief and claiming that Google “has reproduced a digital copy of the Works without the copyright holders’ permission and in violation of the authors’ rights under copyright laws.”¹ A month later, the Association of American Publishers (AAP)² filed a similar lawsuit, also in the Southern District of New York.

Richard Raysman and Peter Brown are partners at Brown Raysman Millstein Felder & Steiner. They are co-authors of “Computer Law: Drafting and Negotiating Forms and Agreements” (Law Journal Press).



Richard Raysman

Peter Brown

In both cases, the copyright holders are wary that digital copies of books, even excerpts of books, being available online might lead to piracy similar to that faced by the music industry. The Author’s Guild is also concerned that Google has not offered to pay royalties from any potential ad revenue generated from its Library Project,³ while the AAP has expressed concern that the project adversely affects its own digital library initiatives.⁴ This litigation raises an important question, namely whether a search engine can be liable for copyright infringement when its search results show only a small piece of text of copyrighted work, without the permission of the content owner.

Google’s Library and Similar Projects

Google has plans to scan the libraries of Harvard, Oxford and Stanford universities and of the University of Michigan, as well as of the New York Public Library in creating complete digital copies of the works that will reside on Google’s servers and be searchable online.

The results of the searches will only be accessible via Google.com. Google has not asked permission from the copyright holders to digitize the works, but has instituted an opt-out program whereby any publisher who does not want its works included can notify Google. In contrast, Yahoo!, along with the Internet Archive, has announced a similar project to scan books, called The Open Content Alliance, which will begin scanning public domain works and thereafter other copyrighted materials on an opt-in basis, only after receiving the copyright owner’s permission. Microsoft Corp. recently said that it would participate in the project by agreeing to pay about \$5 million toward the digitization of 150,000 books.

The ‘Arriba Soft’ Case

The seminal case to have addressed the application of copyright law to a “smaller” piece of a copyrighted work available on the Internet via a search engine is the U.S. Court of Appeals for the Ninth Circuit’s decision in *Kelly v. Arriba Soft Corp.*, 336 F3d 811 (9th Cir. 2003).⁵ In *Arriba Soft*, the plaintiff, a professional photographer, sued the defendant, a search-engine company that displayed its results in the form of small pictures (i.e., “thumbnails”).⁶ The defendant obtained its database of pictures by copying images from other Web sites. By clicking on one of the thumbnails, a user could view a larger version of the same picture within the context of Arriba Soft’s Web page.

To operate its site and gather thumbnail images from the Internet, Arriba Soft used a Web crawler program that the company developed. Essentially, this crawler downloaded full-size copies of millions of images from the World Wide Web onto Arriba's server. The program then used these copies to generate smaller, lower-resolution thumbnails. Once the thumbnails were created, the program deleted the full-size originals from the server. Although a user could copy these thumbnails to his computer, he could not increase the resolution of the thumbnail; any enlargement of the image would result in a loss of clarity and would lack commercial value.

Ruling in favor of the defendant, the Ninth Circuit held that the creation and use of the thumbnails in the search engine was a fair use, but remanded the case to the district court to determine whether the display of the large images was a violation of the plaintiff's exclusive right to publicly display his works. The court reasoned that the defendant's use of the thumbnail images was "transformative" and added a further purpose or different character to the original, thereby creating something of value and not stifling artistic creativity. The court further noted that the use of the images served a different function than that of the plaintiff's, namely improving access to information on the Internet.

Infringement and Fair Use

Under the 1976 Copyright Act, certain uses of a copyright protected work, which might otherwise be considered infringement, may be deemed fair use and not result in liability. The use of a copyrighted work for purposes such as criticism, comment, news reporting or educational research may not be a copyright infringement in certain cases as set forth in 17 USC §107. Section 107 delineates four nonexclusive factors to

consider whether a particular use is a "fair use." They are: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for certain educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for, or value, of the copyrighted work. A

Th[e] litigation raises an important question, namely whether a search engine can be liable for copyright infringement when its search results show only a small piece of text of copyrighted work, without the permission of the content owner.

reviewing court must balance these factors on a case-by-case basis in light of the objectives of the Copyright Act, which is intended to promote creativity, thereby benefiting the artist and the public alike.

Relying on the Supreme Court's decision in *Campbell v. Acuff-Rose Music Inc.*, 510 US 569, 579, 114 SCt 1164 (1994), the *Arriba Soft* court made clear that even though the defendant's use of the copyrighted material was of a commercial nature, the fair use inquiry does not end. Instead, the central purpose of the investigation is to determine whether the new work merely supersedes the object of the original creation, or adds something new, with a further purpose or different character, altering the first work with a new expression, meaning or message. In short, the question is whether and to what extent the new work is "transformative." Under the second factor, some works are recognized as

closer to the core of intended copyright protection than others. *Campbell*, 510 US at 586. For example, fictional, creative works come closer to this core of protection than do primarily factual works. *Harper & Row Publishers Inc. v. Nation Enters.*, 471 US 539, 563, 105 SCt 2218 (1985).

With respect to the third factor, if the secondary user copies only as much as is necessary for his intended use, then this factor will not weigh against him. *Arriba Soft* at 820-21. The fourth factor requires the court to consider "not only the extent of market harm caused by the particular actions of the alleged infringer, but also "whether unrestricted and wide-spread conduct...would result in a substantially adverse impact on the potential market for the original." *Campbell* at 590.

Library Project and Fair Use

Google claims that its Library Project is consistent with the fair use doctrine under copyright law.⁷ The Author's Guild and the AAP, however, contend that Google is copying works for commercial purposes without permission, which amounts to copyright infringement.

Both *Arriba-Soft* and the current Google litigation involve commercial search engines; however, whether the Library Project itself is deemed commercial is a matter for the court. Google's project involves snippets of copyrighted text, while *Arriba Soft* concerned lower-resolution photographs. In *Arriba-Soft*, the entire image, despite its poor resolution, was still capable of being displayed, whereas Google will only display a portion of the copyrighted work to the end user. The effect that the Library Project will have on the literary market (e.g., increased exposure and sales versus harm to the market and encouragement of piracy) may be considered under the fair use analysis,

along with the project's alleged public good (e.g., the benefit of having a digital accessible library versus the harm suffered by the copyright holders).

Yet another factual distinction is that the Library Project's copyrighted works were generally unavailable, in whole or in part, on the Internet, whereas in *Arriba Soft*, the copyright owner had posted his photographs on his Web site. The plaintiffs contend that literary copyright holders, unlike Web-site owners who openly post Internet content, have not expressly or implicitly allowed their works to be searchable by Google. According to the AAP complaint, however, "Google analogizes the Google Library Project's scanning of entire books to its reproduction of the content of Web sites for search purposes."⁸ In the end, this factual distinction might implicate one or more of the fair use factors.

Which "copies" will be subject to the court's inquiry? In conducting its fair use analysis, the court may arguably subject not only the Library Project's final search results (displaying small portions of text around the search term) to fair use scrutiny, but also other levels of copying.⁹

To understand this point requires some technological explanation. To allow searches that produce snippets of text around inputted terms requires "levels" of copying. For Google to successfully complete its project, it will need to scan entire books and then use Optical Character Recognition (OCR) technology to recognize the text and create searchable digital copies. Along the way, other temporary or cached copies incidental to computer and Internet functioning are likely to be created. Moreover, when the snippets are displayed in the search results, still another copy is made on the user's computer screen.

Under copyright law, any copy without express permission of the copyright

holder is unauthorized, but not all unauthorized copies are prohibited, given the availability of a fair use defense. Thus, the Southern District might look beyond the final copy in conducting its infringement analysis and may scrutinize all copies to determine whether the fair use defense is available in each instance. To aid its analysis, the court may look to the *Arriba Soft* decision and similar cases for guidance, along with existing Supreme Court precedent, in applying the fair use factors. Ultimately, it remains to be seen, should the matter proceed to motion practice or trial, how the Southern District will balance the equities to reach its conclusion in these two cases, given the novelty of the issue and the lack of case law directly on point.

Conclusion

The outcome of this litigation and the ultimate resolution of the issue will likely have an impact on the day-to-day functions of search engines as they crawl and index the World Wide Web. Currently, the primary function of Google and other search engines is to lead the user to information that is already available on the World Wide Web. As search engines branch out from their core businesses, so too will the ways in which they gather information. For example, Google and other search engines, among other things, provide e-mail services and satellite image maps. As for the digitized library, this attempted expansion will begin to test the boundaries of fair use on the Internet, as this medium presents new challenges to existing copyright law, portions of which were written prior to the electronic age.



1. See ¶31, Author's Guild Complaint, *The Author's Guild v. Google, Inc.*, Doc. No. 05CV8136 (SDNY, Sept. 20, 2005) available at: news.findlaw.com/hdocs/docs/google/agggoog92005cmp.pdf.

2. The named plaintiffs include The McGraw-Hill Companies, Inc., Pearson Education, Inc., Penguin Group (USA) Inc., Simon & Schuster, Inc., and John Wiley & Sons, Inc.

3. At the time of this writing, Google has stated that it will not place ads on Google pages for books from its Library Project. However, the AAP plaintiffs contend that "when Google makes...other digital copies available to the public...it does so in order to increase user traffic to its site, which then enables it to increase the price it charges its advertisers." See ¶6, AAP Complaint, *The McGraw-Hill Companies, Inc. v. Google, Inc.* Doc. No. 05Civ8881 (SDNY Oct. 19, 2005) available at news.findlaw.com/wsj/docs/google/mcggog101905cmp.pdf.

4. See ¶5, AAP Complaint.

5. Similar to *Arriba Soft*, there is a case currently pending in the U.S. District Court for the Central District of California, between Perfect 10 and Amazon.com. Perfect 10, an adult magazine publisher, sued Amazon.com (and Google, Inc. in a related suit) over the display of thumbnails of Perfect 10 images on Amazon.com's own A9 search engine. See "Adult Magazine Seeks Ban Against Google, Amazon.com," Techweb, Aug. 25, 2005 (available at: <http://www.techweb.com/wire/ebiz/170100404>). Another relevant precedent is *Video Pipeline, Inc. v. Buena Vista Home Entertainment*, 342 F3d 191 (3rd Cir. 2003), in which the court found copyright infringement in *Video Pipeline's* use of homemade video movie clips for commercial motives.

6. In its original opinion in 2002 (which was withdrawn), the court also held that the linking and framing of full-size versions of the same images violated the copyright owner's right of public display. See *Kelly v. Arriba Soft Corp.*, 280 F3d 934 (9th Cir. 2002). Those portions of the opinion concerning the right of public display were removed from the revised opinion, and the case was remanded for resolution of that issue.

7. Official Google Blog: "Google Print and the Authors Guild," Sept. 20, 2005 available at: <http://googleblog.blogspot.com/2005/09/google-print-and-authors-guild.html>.

8. See ¶29, AAP Complaint.

9. Under federal copyright law, copies are "material objects...in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device." 17 USC §101. As one court noted, "[T]he digitization or input of any copyrighted material, whether it be computer code or visual imagery, may support a finding of infringement notwithstanding only the briefest of existence in a computer's RAM." *Tiffany Design, Inc. v. Reno-Tahoe Specialty, Inc.*, 55 FSupp2d 1113 (D. Nev. 1999).

This article is reprinted with permission from the November 8, 2005 edition of the NEW YORK LAW JOURNAL. © 2005 ALM Properties, Inc. All rights reserved. Further duplication without permission is prohibited. For information, contact ALM, Reprint Department at 800-888-8300 x6111. #070-11-05-0008