New York Law Journal

Technology Today

VOLUME 235—NO. 1 TUESDAY, JANUARY 3, 2006 ALM

Google's Book Project

Search Engine Seeks to Digitize Industries' Content Cache

INTERNET ISSUES

BY PETER J. PIZZI

Google's Book Library Project (Book Search) proposes to scan into its searchable database the entire collections of four prominent university libraries as well as that of the New York Public Library. Once completed, users will be able to search the text of hundreds of thousands of books through the search engine.[1]

Google will allow users full-text searching of the works, but will only display excerpts of a few sentences each—called "snippets"—for works still subject to copyright protection. A publisher can "opt out" of the program by notifying Google which works to exclude.[2]

Groups of publishers and authors have separately sued Google in the Southern District of New York asserting that the book search project is copyright infringement. Google has responded by asserting that the project is "fair use."[3]

The ability of computers to create perfect copies and, through the Internet, to disperse copies instantly, has lead to continuing conflicts between "content" industries and the forces of Internet-based information technology.

> *In an attempt to conform to the first fair use factor, Google attempts to minimize the commercial aspects of its Book Search by displaying the "snippet" of the copyrighted work without any "sponsored links" such as those displayed with general Google searches.*

Earlier this year, these two forces clashed in *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*,[4] where the plaintiffs challenged the defendants' "peer-to-peer" file sharing software used in exchanging digital copies of recorded music. The court tipped the balance in favor of the media conglomerates, requiring the technology firms to face claims for contributory infringement even though the subject products were capable of non-infringing uses. The court ruled that evidence of evil motives on the part of the software firms in developing their business plans was sufficient to defeat their motions to dismiss the case.

Google's Book Search pits similar interests against each other, but in this instance the technology firm is not the renegade sort as in Grokster. Instead, information technology is represented by the search monolith that has become one of the world's most famous business franchises. Book Search is aimed at an über warehouse of searchable content, namely, all (or nearly all) published books.

As in *Grokster*, the case will involve legal principles of intellectual property law, but the outcome will likely be strongly influenced by judicial perceptions of public policy. Should the courts "let Google be Google" and continue its march toward creating an efficient search of everything known to humankind? Or should courts keep Google away from published books for fear that to do otherwise would threaten the underlying business model

Peter J. Pizzi *is chair of Connell Foley's Internet and information law practice. He also chairs the Internet and Litigation Committee of the New York State Bar Association's Commercial and Federal Litigation Section. (A member of his family is a software engineer employed by Google.)* **Ryan McGonigle,** *an associate at the firm, assisted in the preparation of this article.*

of publishing, as the industry now claims? As in *Grokster*, the pressures each side will bring to bear on the outcome are enormous.

Fair Use Factors

An owner of a copyright has the exclusive right to reproduce, distribute, and publicly display copies of the work. See 17 U.S.C. §106. Fair use is a statutory exception to copyright. See 17 U.S.C. § § 106, 107. The statute sets out four factors to consider in determining whether the use in a particular case is a fair use.

Courts are instructed to balance these factors in light of the objectives of copyright law, rather than view them as definitive or determinative tests. The four factors are: (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit 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. See 17 U.S.C. § 107.

The leading case applying fair use to search engines is the U.S. Court of Appeals for the Ninth Circuit's decision in *Kelly v. Arriba Soft Corporation*, 336 F.3d 811, 820 (2003) ("[search engines] benefit the public by enhancing information-gathering techniques on the Internet.").[5]

In *Arriba Soft*, a professional photographer alleged copyright infringement against an image search engine that yielded thumbnails of plaintiff's images. Considering the first fair use factor, the court acknowledged that the search engine in question was a commercial use but saw this as no obstacle by itself.

Previously, the U.S. Supreme Court in *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994), had rejected the proposition that a commercial use of the copyrighted material ends the inquiry under this factor.

With regard to the first factor, the Ninth Circuit held the thumbnails qualified as fair use because the use was transformative of the nature, purpose and character of the purpose the images served on the plaintiff's Web site. ("Kelly's images are artistic works intended to inform and engage the viewer in an aesthetic experience. His images are used to portray scenes from the American West in an aesthetic manner. Arriba's use of Kelly's images in the thumbnails is unrelated to any aesthetic purpose."). In seeking evidence of a transformative use of the copyrighted materials, the court followed the Supreme Court's fair use decision in *Campbell*.

Google may argue that its display of mere excerpts of a copyrighted work in a non-commercial, not-for-profit setting is a fair use under the Copyright Act.

The second factor suggests that the scope of fair use is narrower for creative works as opposed to non-fiction or factual works. *Campbell*, 510 U.S. at 586.

The third factor focuses on the extent of the copying of the copyrighted work, and inquires whether the use is reasonable in light of the purpose of the copying. In addressing the final factor—the impact on the market for the original work—the court in *Arriba Soft* understood its charge to assess "not only the extent of market harm caused by the particular actions of the alleged infringer, but also whether unrestricted and widespread conduct of the sort engaged in by the defendant would result in a substantially adverse impact on the potential market for the original."

The court also recognized that a "transformative" work is less likely to have an adverse impact on the market for the original than a work that supersedes the copyrighted work. This factor favored the search engine because it served to increase attention to the copyrighted works and to attract business to the plaintiff's site.

In an attempt to conform to the first fair use factor, Google attempts to minimize the commercial aspects of its Book Search by displaying the "snippet" of the copyrighted work without any "sponsored links" such as those displayed with general Google searches.

The publishing industry is unimpressed by Google's snippets-only display, arguing that the entire purpose of the Book Search is commercial in nature in that it seeks to draw Internet users to the Google site and thereby create more advertising revenue for Google.

This proposition could be argued to distort the first fair use factor. That an alleged infringer is a for-profit entity that gains revenue based on the number of customers using its services is not the relevant inquiry.[6] The first factor looks to whether the use of the copyrighted material is itself commercial in nature. Google certainly can argue that it gains nothing commercially by producing snippets of one work or another in response to search queries and that its Book Search educates the user regarding published works relevant to her search. Analysis of this factor appears to favor Google.

Analysis of the second factor—the nature of the copyrighted work—will vary depending on the text in question, making it difficult to reach an overall conclusion on this factor in the context of Book Search. Google will also use only small snippets of the copyrighted works, making the third factor one that should be resolved in Google's favor.

The Supreme Court has held that the most important fair use factor is the fourth—the impact of the use on the market for the original work. *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 566 (1985). Google argues this factor must be resolved in its favor because its Book Search results will only serve to attract members of the public to the copyrighted work and should lead to increased sales. Indeed, Book Search includes links to major booksellers where copies of works retrieved by search results may be purchased without any revenue to Google.

Recognizing that they have a difficult row to hoe in proving that Book Search causes any damage to the market for books, publishers may attempt to shift the focus away from the market for the sale of books entirely.

In a recent conference on Book Search,[7] industry lawyers argued that technology companies such as Yahoo! and Microsoft were prepared to pay publishers to license books from publishers for online use but that, if Google were to continue digitizing books without consent of publishers, such companies will be forced to follow suit, amassing their own digital book library also without paying anything.

Industry lawyers argue that Google's decision to bypass licensing deals will destroy all other licensing opportunities for publishers, requiring the court to resolve the fourth fair use argument in favor of the publishers.

There are several problems with this argument. First, that other companies were prepared to spend large sums to license massive publishing industry libraries is a dubious proposition. If this market were viable, why have not such transactions occurred during the 10 years since the Internet revolution began? When Yahoo! and Microsoft announced an agreement with publishers involving scanning public domain works and other copyrighted materials for which consent from the publisher had been obtained, there was no indication either was paying money to publishers.[8]

Further, even if there were companies prepared to pay publishers to license their content for online use, they would likely do so only in exchange for complete works. Book Search does not present complete works but only snippets. It is not clear how the market for complete digital works would be relevant to a search engine that only produces snippets of copyrighted material.

Moreover, Google does itself pay to license complete works pursuant to its Print for Publishers program, and will apparently continue to do so even though it is moving ahead with its digitization project. Having access to complete works through this program enables Google to produce more expansive search results to users. In other words, the market for licensing complete works will remain viable despite the Book Search project because the latter only makes snippets of copyrighted works available.

Moreover, as suggested by Jonathan Band, Google may argue that its "opt out" mechanism—whereby any publisher can elect to remove titles from Book Search—is analogous to the implied consent crucial to function of Internet search engines.[9] According to Band, a search engine copies vast amounts of material without the express consent of individual sites, under the rationale that a site that makes material available on the Web is impliedly licensing the search engine to capture that content.[10]

When a search engine such as Google "crawls" the Internet for content and copies material on published sites, it is able to do so in part because the site/copyright owner has not placed the digital equivalent of a "Do Not Enter" tag on the site.

Band aptly compares Google's decision to allow copyright owners to "opt out" of Book Search to the implied consent upon which the crucial search engine industry depends.

This last point leads to another issue courts will confront: What is the copying at issue? Publishers argue that Book Search copies the entire text of each work and that this is the step at which the fair use factors should be analyzed. Google may argue that the digitization stage of Book Search is an intermediary stage incidental to a fair use—the display of snippets in response to a search query.

Alternatively, Google may argue that the fact that an entire work was copied does not preclude a finding of fair use, see *Sony Corp. of Am. v. Universal City Studios*, 464 U.S. 417, 449-50 (1984), and that its display of mere excerpts of a copyrighted work in a non-commercial, not-for-profit setting is a fair use under the Copyright Act.

••••••••••••••●●●••••••••••••••

1. See http://books.google.com/googlebooks/library.html. (last visited Dec. 27, 2005).
2. See http://books.google.com/googlebooks/publisher_library.html #options3. (last visited Dec. 27, 2005).
3. The *McGraw-Hill Companies v. Google Inc.*, Civil Action No. 05Civ8881 (SDNY Oct. 19, 2005); *The Author's Guild et al. v. Google, Inc.*, Civil Action No. 05CV8136 (SDNY, Sept. 20, 2005).
4. 545 U. S. ____, 125 S. Ct. 2764 (2005).
5. This case was discussed in more detail in P. Pizzi, "Perfect 10 v. Google," NYLJ, Jan. 4, 2005.
6. *Campbell*, 510 U.S. at 584 (if commerciality carried presumptive force against a finding of fairness, none of the examples of fair use in § 107 would qualify as fair use); *Belmore v. City Pages, Inc.*, 880 F. Supp. 673 (D. Minn. 1995).
7. At a Dec. 14, 2005 program at the Association of the Bar of the City of New York, Allan Adler, Vice president, legal and governmental affairs, Association of American Publishers, and Paul Aiken, Executive Director, Authors' Guild, presented views of publishers and authors, respectively, concerning Google Book Search.
8. See http://news.com.com/Yahoo+to+digitize+public+domain+books/2100-1038_3-5887374.html and See http://www.libraryjournal.com/article/CA6279264.html.
9. See Jonathan Band, "The Google Print Library Project: A Copyright Analysis" at 5, http://www.policybandwidth.com/doc/googleprint.pdf (last visited Dec. 20, 2005).
10. Band at 5 ("billions of dollars of market capital represented by the search engine companies are based primarily on the fair use doctrine").

This article is reprinted with permission from the January 3, 2006 edition of the NEW YORK LAW JOURNAL. © 2006 ALM Properties, Inc. All rights reserved. Further duplication without permission is prohibited. For information, contact ALM Reprint Department at 800-888-8300 x6111. #070-01-06-0002