

COPYRIGHT ACCORDING TO GOOGLE

For Presentation at the Maine Bar Association's
Winter Meeting, January 31, 2008

By: Rufus E. Brown, Esq., Brown & Burke, Program Chair

Panelists: Robert Mittel, Esq., MittelAsen, Daphne Keller, Esq., Google Legal
Department, Elizabeth Horn, HMG Search Marketing

I. INTRODUCTION

- A. Copyright law is strongly shaped by policy considerations and technology.
- B. Nowhere is this more evident than the emerging law of copyright as applied to Internet search engines.
- C. Not surprisingly, Google has been at the forefront of the law on this issue, usually as a defendant.
- D. Now it is pushing the envelope with its Book Project, where it is in the processing of copying works from 5 major libraries to digital format on a massive scale never before undertaken or even imagined.
- E. This program will:
 - 1. Examine the emerging law of copyright law as applied to search engines;
 - 2. Suggest how this law will apply to Google's Book Project; and
 - 3. Finally, examine how owners of copyrighted material can minimize/optimize exposure on the Internet.

II. TECHNICAL BACKGROUND TO THE ISSUES BEING DISCUSSED

- A. What does Google want and how do they get it?
- B. Explain/illustrate:
 - 1. Architecture of search engines.
 - 2. Google's AdSense.
 - 3. WebPages & HTML
 - 4. Crawlers/bots
 - 5. Cache files

6. thumbnails
7. Snippets
8. Metatags
9. Robot .txt
10. In line linking

III. EMERGING LAW OF COPYRIGHT APPLIED TO SEARCH ENGINES.

A. Fair Use Doctrine:

1. Codified in 17 U.S.C. §107:
 - a. Balances the rights of copyright owner with the purpose of copyright law “to promote the Progress of Science and the Useful Arts.” Courts have flexibility to avoid rigid application of copyright when to do so would stifle the very creativity which the law is designed to promote.
 - b. Ultimately, the doctrine is entirely an “equitable rule of reason to be applied in light of the purposes of the Copyright Act that is so flexible as to defy definition.
 - c. Nevertheless, the preamble purposes cite criticism, comment, news reporting, teaching, scholarship, or research as examples.
 - c. Also, the statute sets out 4 (non-exclusive tests) factors to consider.
2. First factor: *purpose and character of use*:
 - a. Including whether use is commercial or for nonprofit educational.
 - b. Commercial uses not an automatic bar. It depends on how “exploitive” the secondary use is.
 - c. And whether and the extent to which the use is “transformative” or “supersedes” or “consumes” original work. See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

3. Second factor: *nature of the copyrighted work*
 - a. Creative works get more protection than factual works.
 - b. Unpublished works more protection than already published.
 4. Third factor: *amount and substantiality of the portion used*
 - a. in relation to what is necessary for the intended use.
 - b. even if wholesale copying necessary as an intermediate step.
 5. Fourth factor: *effect of the use upon the potential market of the copyrighted use.*
 6. Doctrinal issues:
 - a. Section 103(a) of the Copyright Act defines a “derivative work” as “a work based on one or more pre-existing works, such as ... any other form in which a work may be recast, *transformed*, or adapted.”
 - b. The market based factor is circular.
 - c. Ad hoc nature of the inquiry.
- B. Kelly/Perfect 10
1. *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003).
 - a. Facts:
 - i. Photographer posting photos on his website sues Arriba for direct infringement, copying.
 - ii. Arriba operates a search engine that crawls the Internet and finds and downloads the full size the photos on Kelly’s website into Arriba’s database.
 - iii. From the download, a thumbnail is created for display in response to search inquires.
 - iv. Then the full size image is deleted from the database in the server.

- v. When a user finds and double clicks on the thumbnail, in-line linking takes the user to Kelly's full sized photos on his website framed with the search engine banner and ads.
- vi. And links to websites with infringing photos.
- vii. Arriba conceded copying, but argued "fair use" 17 U.S.C. §107.

b. Holdings:

- i. First factor: Not highly exploitive because there is no direct advertising and Kelly's photos are just a few of thousands of images in Arriba's database. Use here is transformative. The original pictures are artistic expression; the search engine functions as a tool to help index and improve access to images on the Internet.
- ii. Second factor and third factor not helpful here.
- iii. Fourth factor: no harm to Kelly's market because the thumbnail use guides user to Kelly's full size pictures on his website rather than away.

2. *Perfect 10 v. Google*, 508 F.3d 1146 (9th Cir, 2007).

a. Facts:

- i. Almost identical facts as *Kelly*.
- ii. Perfect 10 markets and sells subscriptions for on-line subscriptions for photos of models.
- iii. It claimed that Google's thumbnails in its search engine database infringed on its display and distribution rights because users are directed to websites with infringing copies.

b. Holdings:

- i. A person makes a "copies" when it downloads images from another's hard drive for storage in its own hard drive database and

“displays” the copy when the stored images are sent electronically to a user.

- ii. So here the thumbnails prima facie violate display rights.
- iii. But no violation of display or distribution rights re full size pictures because Google does not store a copy of the full image, but simply uses HTML to direct user to original website where full size images are stored.
- iv. Like *Arriba* in the *Kelly* case, Google argues “fair use” concerning thumbnails.
- v. Court agrees with *Kelly* “fair use” analysis, with emphasis.
- vi. First factor: It describes the use of the thumbnails by Google as “*highly transformative*” providing access to information on the Internet (as an “electronic reference tool”) as opposed the entertainment value of the full size photos as artistic expression. It is more transformative than parody because the search engine is “an *entirely new use* for the original work.” Advertising is used but the “significantly transformative nature of Google’s search engine, particularly in light of the public benefit, outweighs Google’s superseding and commercial uses.”
- ii. Second factor: Pictures creative, but previously published.
- iii. Third factor: Google does not display more than necessary.
- iv. Fourth factor: Thumbnails do not hurt Perfect 10’s full size pictures. Perfect 10’s cell phone market “hypothetical” and outweighed by the highly transformative use.
- v. Overrules District Court that found no “fair use” because Google’s Ad Sense Program more commercially exploitive than *Arriba*’s in *Kelly* and

because of potential harm to Perfect 10's cell phone market for thumbnails.

C. *Field v. Google*, 412 F.Supp. 2d 1106 (D.Nev. 2006).

1. Facts:

- a. Field is an author and an attorney. He published over 50 of his copyrighted articles on his personal website and then sued Google for copyright infringement based on Google's alleged copying and distribution of his works.
- b. Google's uses an automated "Googlebot" to crawl the Internet and indexes and stores HTML code from websites in a temporary "cache" file.
- c. The results of a search give the title of the webpage followed by a "snippet" from the webpage. Following the snippet, Google provides the URL to link to Field's articles stored on his hard drive and then supplies another link the cached archival copy of the webpage stored in Google's data base.
- d. It is not possible to contact each webpage owner to see if the owner wants the site listed in its search results or stored through cache links.
- e. So the industry has developed protocols to communicate with Google's crawler with metatag instructions in HTML code.
 - i. E.g., instructions can allow a given page to be indexed but not cached by use of a "no archive metatag."
 - ii. Or use "robot.txt" to prevent crawling of webpage all together.
- f. Field was familiar with search engine process and decided to manufacture a claim to make money from Google's standard practice.
- g. He used settings on his website to allow Google to crawl and index and did not use "no archive" metatags.

2. Holdings of the court:

- a. No direct infringement because defendant did not engage in volitional conduct. The automated non-volitional conduct by Google in response to a user's request does not constitute direct infringement. Google is passive in the process. *Accord, Parker v. Google*, 2007 U.S.App. LEXIS 16370 (3d Cir.2007).
- b. The website owner is deemed to have granted an *implied license* by posting copyrighted material on his website without instructions to prevent crawling and archiving.
- c. Field estopped from claiming infringement under the circumstances.
- d. Also, Google's crawling, indexing and cache archiving constitutes "fair use."
 - i. First factor: Google's caching is "transformative" because it serves different and socially important purposes (e.g., access to information that may be inaccessible; allowing users to understand why a page was responsive to inquiry) and does not "supersede" the original creations.
 - ii. Second factor: Not important and in any event favors fair use because work is posted for free use by Field.
 - iii. Third factor: Google copies no more than necessary.
 - iv. Fourth factor: no market for cached links.
 - v. *Added factor*: Google's good faith.
- e. Digital Millennium Copyright Act:
 - i. Section 512(b) safe harbor for system caching by a service provider if conditions met.
 - ii. Here conditions met: caching "intermediate" and "transitory", material transmitted from person who makes material available on line to another person at the direction of another person, through "an automated technical process."

IV. THE GOOGLE BOOK PROJECT

- A. Two projects.
- B. Partner Program:
 - 1. Copying with permission of publishers who have rights.
 - 2. Full text scanned into Google's database.
 - 3. Search produces bibliographic information and a link to relevant text (full page).
 - 4. Also gives a few pages before and after.
 - 5. Links to enable user to purchase book.
 - 6. Publishers share in Google advertising revenue.
 - 7. No copyright issues.
- C. Library Project:
 - 1. Participating libraries: Harvard, Stanford, University of Michigan, New York Public Library, and Oxford University.
 - 2. Full text scanned into Google's database.
 - 3. Nonconsensual
 - 3. All public domain books copies and full text available in response to search.
 - 4. Copyrighted materials copied, but only a few words surrounding search term ("snippets") displayed in response to a search. Only 3 snippets available per book and not any snippets for reference works.
 - 5. Search identifies library where book can be found and where the book can be purchased.
 - 6. Library gets a free digital copy.
 - 7. In response to objections by publishers and the Author's Guild, Google suspends copying from August to November 1, 2005 to allow copyright owners to opt out.

8. Publishers (McGraw Hill, Pearson Education, Penguin, Simmon & Schuster and John Wiley & Sons) and Author’s Guild bring suit in S.D.N.Y. for copyright infringement in September and October 2005.
- D. Unlike cases cited, here there is literally copying on a massive scale and permanent storage in Google’s hard drive.
 - E. So the issue is whether Google Library Project is protected by “fair use”?
 1. Arguments in favor of “fair use”:
 - a. The Library Project’s purpose is to create a digital indexing for published books, which is both highly “transformative” as ruled by *Kelly*, *Perfect 10*, and *Field*, but extraordinarily important to the public, meeting the underlying purpose of the “fair use” doctrine.
 - b. The commercial exploitation through advertising revenues is outweighed by the high degree of transformative value.
 - c. The books copied have previously been published and are available to the public in libraries. Many of the books copied are factual in nature.
 - d. Entire books are copied, but this is necessary to accomplish transformative functions.
 - e. The user only sees “snippets”.
 - f. The market of publishers of the books are not damaged; their interest are enhanced by greater book sales.
 - g. The “opt- out” program demonstrates Google’s good faith even if does not establish an implied license ala *Field v. Google*.
 2. Arguments against “fair use”:
 - a. The *Kelly/Perfect 10/Field* trilogy analysis of “fair use” takes place in a different environment – a digital on-line environment where Google is indexing materials posted on-line and “copied” in transient cache files, which in the case of *Kelly* and *Perfect 10* are compressed in the form of low

resolution images, in contrast to the Book Project which takes full text images.

- i. In the digital environment, the expectations of copyright ownership is diminished in comparison to the analog world.
 - ii. This point is emphasized in *Field v Google*, implied license and estoppel discussions.
- b. The uses of copying are not “transformative” in the terms of preamble uses, comment, criticism, etc., where the doctrine has been applied in the analog world.
 - i. Google is essentially using unauthorized copies of books to be retransmitted, in part, in another medium.
 - ii. This purpose was firmly rejected as fair use in *UMG Recordings, Inc. v. Sony Music*, 92 F.Supp. 2d 349 (S.D.N.Y. 2000).
- c. The “opt out” program is meaningless in an analog world; it stands copyright on its ear.
- d. Google will benefit commercially from its nonconsensual copying at the expense of copyright owners who get no direct economic benefit. This is inequitable and unfair.
- e. Copyright owners have a market that is being taken from them by others who might be willing to pay for indexing rights. Others are already in the market.
- f. The Library Project is only a step, a giant step, to the ultimate goal of a fully accessible digital library where Google, as a result of its copying, will have enormous, unfair market power (a monopoly) in relation to other providers and copyright owners.
- g. New legislation is needed to fairly sort out the rights and obligations of the parties. The “fair use” doctrine is not equipped to sort out and allocate the complex equities.

V. EXPOSURE OF COPYRIGHTED MATERIALS POSTED ON WEBSITES

A. Protecting against crawlers.

B. Optimizing exposure.