SYRACUSE SCIENCE AND TECHNOLOGY LAW REPORTER

The Copyright Implications of Web Archiving and Caching

David M. Ray

Spring 2006

I. Introduction

Despite the best efforts of the drafters of the Copyright Act, new technologies and methods raise new legal questions in the battle between the protection of authors' rights and the availability of information to the public. These debates often begin online in various blogs and web sites, until they eventually find themselves before a court of law. One such issue is the caching and archiving web sites, a practice that has been ongoing for several years but is only now finding its way into copyright suits.

Since 1996, the Internet Archive, through a feature called the "Wayback Machine," has been gathering information published on the Internet by individuals and organizations, both public and private, and storing it in a publicly available repository.[1] Recognizing the transitory nature of the Internet, the Internet Archive's founders foresaw the need to preserve snapshots of web content as a resource for future generations.[2] A user can easily retrieve a cached copy of a web site captured months or years earlier by visiting the Internet Archive web site and entering the appropriate URL.[3]

---

[1] About IA, Internet Archive, http://www.archive.org/about/about.php (visited Oct. 25, 2006).

[2] *Id.*

[3] *See* Wayback Machine, Internet Archive, http://www.archive.org/web/web.php (last visited Oct. 25, 2006).

Similarly, the search engines operated by Google, Yahoo, and MSN retain copies of the web sites they index and make them available to users as part of their search results.[4] A user has the option of viewing the search engine's cached copy of a web page, or visiting the web site directly.[5] These search engines retain a cached copy of each web site they index until it is crawled again, which can be a matter of hours, days, or weeks.[6] For purposes of this note, the analysis will focus on Google, as the pioneer of this technology and a representative example of search engine practices.

Until recently, these apparent violations of U.S. copyright law have gone largely unchallenged by the organizations and individuals who publish content on the web. However, this may be changing in light of recent copyright suits filed against the Internet Archive and Google. In *Field v. Google*, a case recently decided by the District of Nevada, a claim of copyright infringement was brought against Google based upon its cache feature.[7] Similarly, in *Healthcare Advocates v. Harding, Earley, Follmer & Failey*, a case currently pending before the Eastern District of Pennsylvania, The Internet Archive was sued for, inter alia, copyright infringement and violation of the Digital Millennium Copyright Act.[8] Finally, a preliminary

[4] Google Home Page, http://www.google.com (last visited Oct. 25, 2006); Yahoo! Home Page, http://www.yahoo.com (last visited Oct. 25, 2006); MSN Search Home Page, http://search.msn.com (last visited Oct. 25, 2006).

[5] Google Help Center: Google Web Search Features, http://www.google.com/support/bin/static.py?page=searchguides.html&ctx=results (last visited Oct. 25, 2006).

[6] Field v. Google, Inc., 412 F. Supp. 2d 1106, 1124 (D. Nev. 2006) (Google caches information for approximately 14 to 20 days (citation omitted)).

[7] *See id.*

[8] *See* Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Failey, No. 05-03524 (E.D. Pa.. filed Jul. 8, 2005).

ruling in *Perfect 10 v. Google* challenges what was thought to be a solid precedent of fair use for thumbnail images in search engines, as established by the landmark case *Kelly v. Arriba*.[9]

In *Field v. Google*, a Nevada attorney and author filed a copyright infringement suit against Google, alleging that Google's cache feature made copies of his creative works in violation of his exclusive rights under the Copyright Act.[10] Field, the plaintiff, made a number of his works publicly available on his web site, blakeswritings.com, where they were cached by Google.[11] He subsequently removed the works from his web site, and claimed infringement when they continued to be publicly available through the cached copies retained by Google.[12] Finding against Field, Judge Jones found Google's cache feature is fair use, and further upheld Google's other defenses of implied contract and safe harbor for caching under the Digital Millennium Copyright Act (DMCA).[13] It should be noted that Field's actions were interpreted by the court as opportunistic and acting in bad faith, as he admitted he was fully aware of Google's practices at the time he allowed his site to be cached and failed to utilize available code to prevent it.[14] Once made aware of the suit, Google immediately removed the content in question and discontinued caching of blakeswritings.com.[15]

---

[9] Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828 (C.D. Cal. 2006); Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003); *See also* Nate Anderson, *Judge to Google: Some thumbnails are illegal, available at* http://arstechnica.com/news.ars/post/20060222-6234.html (last visited Oct. 25, 2006).

[10] *Field*, 412 F. Supp. 2d at 1109.

[11] *Id.* at 1114.

[12] *Id.*

[13] *Id.* at 1123-25.

[14] *Id.* at 1116.

[15] *Id.* at 1123.

The dispute in *Healthcare Advocates* began with a prior copyright and trademark infringement matter against a competitor, Health Advocate, Inc.[16] The Internet Archive is commonly used as a source of evidence in copyright and trademark infringement matters, as it provides a record as to the appearance and content of a web site on an earlier date, which is especially useful in matters where the first party to use a mark or certain content is a disputed issue.[17] Healthcare Advocates exercised its ability to opt of out inclusion in the archive, relying on the Internet Archive's assertion that through the use of a particular instruction file, the archived versions of their web site would be blocked.[18] During the discovery phase of the prior suit, Harding Early, attorneys for the Health Advocate, were able to access the blocked archives by bombarding the Internet Archive with requests, exploiting a vulnerability in the system.[19] Healthcare Advocates then filed a suit against Harding Early, also naming the Internet Archive, with claims of copyright infringement, breach of contract, and violation of the DMCA's provisions addressing electronic copy protection.[20]

Perfect 10 v. Google has reopened the issue of fair use for thumbnail images utilized in image search functionality offered by Google and other search engines.[21] Perfect 10, whose primary business is nude photographs of women, won a preliminary injunction against Google,

---

[16] Healthcare Advocates, No. 2:05-cv-03524-RK.; Tom Zeller, *Keeper of Expired Web Pages Is Sued Because Archive Was Used in Another Suit*, N.Y. TIMES, July 13, 2005, at C9.

[17] Shannon Duffy, *Law Firm Accused of Internet Hacking*, THE LEGAL INTELLIGENCER, July 13, 2005 *available at* http://www.law.com/jsp/article.jsp?id=1121245511855.

[18] *Id.*

[19] *Id.*

[20] *Id.*

[21] *Perfect 10*, 416 F. Supp. 2d 828, 836 (C.D. Cal. 2006).

who used thumbnails of photographs available on Perfect 10's site in its image search feature.[22] Critical to this finding was the fact that Perfect 10 had signed a licensing agreement with a U.K. company to provide thumbnail versions of Perfect 10's photographs for use on mobile phones, while Google had made its thumbnails available through a mobile phone accessible search feature.[23] The court found that Google's thumbnails could act as a market replacement for those being licensed by Perfect 10, which swayed the court to a finding against fair use.[24] Without such direct competition in the mobile phone arena, it is likely the court would have denied the injunction, which is consistent with *Kelly v. Arriba*.[25]

This note will analyze both web site archiving, as performed by the Internet Archive, and long term web caching, as practiced by Google, in the light of current copyright law. This note supports current practices insofar as they are performed respectfully of the copyright owners, and in a way that advances the public good.

II. Web Caching and Archiving Practices

An important distinction in this discussion is the difference between archiving and caching as it relates to web sites. An archive not only systematically collects information, but *preserves* it for historical use.[26] Information stored in an archive is typically maintained for a

[22] *Id.* at 831.

[23] *Id.* at 832, 849.

[24] *Id.* at 851.

[25] *Id.*; *Kelly*, 336 F.3d 811; Anderson, *supra* note 9.

[26] BLACK'S LAW DICTIONARY 712 (8th ed. 2004). "**archive,** *n*. (*usu. pl.*) **1.** A place where public, historical, or institutional records are systematically preserved. **2.** Collected and preserved public, historical, or institutional papers and records. **3.** Any systematic compilation of materials, esp. writings, in physical or electronic form. -- **archive,** *vb*."

long period of time, or indefinitely.[27] A cache, on the other hand, is temporary by definition.[28] Copies of a web site are made and stored either on another server, or locally on the user's computer, in order to decrease the traffic to the original web site, and to increase the speed at which web pages are transferred.[29] These copies are made when a *user requests* a particular web page.[30] When another user requests the same page within a short period of time, instead of retrieving the page again, the cached copy is served to the second user.[31]

The caching performed by Google is slightly different from this traditional method of caching. Google and other search engines employ a "crawler" is to find and index web pages.[32] This is a fully automated process whereby a program requests web pages, rather than a user. Copies of the pages are then stored in an index, which Google users search when they enter search terms.[33] As a convenience for its users, Google makes these cached copies available to its users through a link labeled "Cached" which appears next to each search result.[34] When the "Cached" link is clicked, Google's computers respond automatically by transmitting the cached

---

[27] *Id.*

[28] Wikipedia, Web Cache, http://en.wikipedia.org/wiki/Web_caching (last visited Feb. 9, 2006). "Web caching is the caching of web documents (e.g., HTML pages, images) in order to reduce bandwidth usage, server load, and perceived "lag." A web cache stores copies of documents passing through it; subsequent requests may be satisfied from the cache if certain conditions are met."

[29] Digital Millennium Copyright Act of 1998, Pub. L. No. 105-304, 112 Stat. 2860. (29 -31 DO NOT COME FROM THE TEXT OF THE DMCA. WE NEED TO ASK AUTHOR FOR CORRECT SITE)

[30] *Id.*

[31] *Id.*

[32] Google.com, Google Information for Webmasters, http://www.google.com/webmasters/1.html (last visited Feb. 18, 2006).

[33] *Id.*

[34] *Id.*(CANNOT VERIFY)

copy of the web site, which is effectively a "snapshot" of the page when it was last crawled by Google.[35] The cached copy includes a prominent notice that the page is cached and may be out of date, and has the relevant search terms highlighted.[36] Cached copies are generally 14 to 20 days old, representing the amount of time between successive crawls by Google.[37]

In implementing this automated process, Google assumes that copyright owners consent to cached copies of their web pages being made available through the cache feature.[38] However, web site owners can prevent Google and other search engines from enabling the cache feature for their pages by adding a tag to the beginning of each page, as follows:

<META NAME="ROBOTS" CONTENT="NOARCHIVE">[39]

The inclusion of this tag will still allow the page to be indexed by Google and included in search results, but will ensure that the "Cached" link is not displayed. In this way, Google has created a caching procedure in which copyright owners must actively opt out of. The reason for this is practical; with billions of pages in its index, obtaining affirmative consent from each copyright holder would be impracticable, if not impossible.[40]

The Internet Archive employs a similar technology to traverse the web, gathering web sites for inclusion in its collection. A "crawler" is employed, which searches the web through an

---

[35] *Id.* (CANNOT VERIFY)

[36] Google.com, Google Help: Search Features, http://www.google.com/features.html#cached (last visited Feb. 18, 2006).

[37] Field, *supra* note 6.

[38] *Field*, 412 F. Supp. 2d at 1116.

[39] Google.com, Google Information for Webmasters, http://www.google.com/webmasters/remove.html#uncache (last visited Feb. 18, 2006).

[40] *Field*, 412 F. Supp. 2d at 1112.

automated process, creating copies of web sites at periodic intervals.[41] Incremental copies of the sites are maintained and made available to users of the Wayback Machine.[42] In contrast to Google, which maintains only the most recent copy of a crawled site, replacing it with each successive crawl, the Internet Archive maintains these copies indefinitely.[43] This is consistent with its stated mission of "prevent[ing] the Internet … from disappearing into the past."[44]

Authors who do not wish to have their content included in the Internet Archive can also opt out through the use of a "robots.txt" file.[45] Proper use of this file, containing an instruction to the Internet Archive, will result in the removal of existing copies of a site from the archive, and will prevent future inclusion.[46] The Internet Archive also accepts requests for removal by email, which is useful for web sites that are no longer available on the Internet. [47]

In summary, both Google and the Internet Archive engage in practices by which copies of web pages are created via an automated process, and those copies are made available to users through their respective services. Both systems rely on an opt-out model, in that permission is not sought from the site authors prior to the copying and distribution of their works, but requests

---

[41] Internet Archive, Internet Archive Frequency Asked Questions, http://www.archive.org/about/faqs.php (last visited Feb. 18, 2006) (follow "How can I get my site included in the Archive?" hyperlink).

[42] Internet Archive, Internet Archive Frequency Asked Questions, http://www.archive.org/about/faqs.php (last visited Feb. 18, 2006) (follow "What is the Internet Archive Wayback Machine?" hyperlink).

[43] *About the Internet Archive*, Internet Archive, http://www.archive.org/about/about.php (last visited Oct. 17, 2006); Stefanie Olsen, *Google cache raises copyright concerns*, CNET News.com, http://news.com.com/2100-1032-1024234.html (last modified Jul. 10, 2003).

[44] Internet Archive, *supra* note 43.

[45] Internet Archive, *Removing Documents from the Wayback Machine*, http://www.archive.org/about/exclude.php.

[46] *Internet Archive Frequently Asked Questions*, *supra* note 42. (follow "How can I remove my site's pages from the Wayback Machine?" hyperlink).

[47]Internet Archive, s*upra* note 45. ("The Internet Archive is not interested in offering access to Web sites or other Internet documents whose authors do not want their materials in the collection.").

to be omitted are honored. The Internet Archive is a non-profit organization, and provides its archival service without compensation, depending on donations and fundraising. Google is advertisement driven, and would stand to profit from increased advertising revenue due to additional traffic to their web site.[48] Google does not, however, place its own advertisements on the cached pages themselves.

In addition to the copyright questions this practice raises, there are several other concerns raised by this copying and redistribution of content. First, it affects the ability of authors and other copyright owners to control the information they distribute by removing it from their web sites. For example, the New York Times and the Wall Street Journal once offered old articles free of charge, but have changed their policies to adopt a paid model.[49] However, the newspapers' decision to change the use of content is thwarted if these same articles are still available through Google's cache or on the Internet Archive.[50] This allows users to circumvent the paid system established by the copyright owner and access them for free.

A second concern is lost advertising revenue, which many web sites depend on for survival. While information is provided without direct charge to web site visitors, it is done so with the expectation that a certain percentage of them will click on the advertisements and generate revenue for the site.[51] If the content is hosted elsewhere, the advertising may no longer

[48] *Perfect 10 v. Google, Inc.*, 416 F. Supp. 2d 828, 834, (D. Cal. 2006).

.

[49] Olsen, *supra* note 43.

[50] *Id.*

[51] Denis T. Rice, *Current Issues in Online Copyright Liability*, 765 PRACTICING L. INST. 335, 362 (2003).

function or be current, diverting revenue away from the web site owners.[52] This results in a loss of revenue for the copyright owner and a potential financial gain for the infringing party.

A third concern is that such repositories may increase liability for web site owners. Old web sites maintained at the Internet Archive have been used as evidence in trademark, copyright and trade secret suits.[53] They can also interfere with a web site's compliance with new legislation or court rulings that require that certain content be removed.[54] In such cases, the offending content may well still reside in web caches and online archives. Thus, content owners no longer have control over the distribution of their copyrighted materials, making them exposed to increased liability, and undermining both business and legal strategies.

III. Copyright Protection on the World Wide Web

Qualifying Works

Works published on the Internet are fully protected under the Copyright Act, subjecting them to the same qualifications and limitations applicable to works fixed in other media.[55] Copyright protection can be claimed only on the expression of the work of authorship and not in

---

[52] *Id.*

[53] Attig v. DRG, Inc., 2005 U.S. Dist. LEXIS 5183 (E.D.Pa. 2005); Flynn v. Health Advocate, Inc., 2004 U.S. Dist. LEXIS 12536 (E.D.Pa. 2004); Malletier v. Burlington Coat Factory Warehouse Corp., 2005 U.S. App. LEXIS 21963 (2d Cir. 2005).

[54] Ellison v. Robertson, 357 F.3d 1072, 1074 (9th Cir. 2004).

[55] In order to qualify for protection under the 1976 Copyright Act, three criteria need to be satisfied. First, the item in question must be a qualifying work of authorship, as specified in section 102. Second, it must satisfy a basic level of originality. Third, it must be fixed in a tangible medium. If these are satisfied, the copyright owner holds exclusive rights over the work, including reproduction, distribution, and public display, as defined under section 106. According to section 102 (a) works of authorship include in particular literary, musical, pictorial, and audiovisual works as well as sound recordings. Web sites generally contain a mixture of text, images, and multimedia elements which correspond to literary, pictorial, and audiovisual works, respectively.

its underlying ideas.[56] In web caching and archiving, the entire page is copied, including all elements of expression ranging from graphics to word placement. On first blush, it appears that most web sites would readily meet the standards necessary to qualify for copyright protection, especially given the level of expression available to authors in this versatile medium.

The area where the Internet differs most from traditional media is in fixation.[57] The communication of information via the Internet depends on storing that information in computer memory devices,[58] including the ISP's servers and the viewer's hard disk cache and computer memory. It appears to be settled law that works of authorship, stored in the hard disk of a computer, are fixed pursuant to Section 101, because a work of authorship, stored in the hard disk of a computer, is a stable copy of the work which, with the aid of a machine, can be perceived for a period of more than a transitory duration. Courts have long recognized that computer memory is a tangible means of expression and that a computer program imprinted in a memory is fixed in a tangible medium.[59]

Infringement

Among the exclusive rights granted to copyright owners, the following apply to caching and archiving: the reproduction of the copyrighted work in copies; the distribution of copies of the copyrighted works to the public by sale or other transfer of ownership, or by rental, lease, or

[56] *See generally* Baker v. Selden, 101 U.S. 99 (1880).

[57] 17 U.S.C. § 101 (1976). ("A work is 'fixed' in a tangible medium of expression when its embodiment in a copy . . . is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than a transitory duration.").

[58] Rice, *supra* note 51, at 349.

[59] MAI Sys. Corp. v. Peak Computer Inc., 991 F.2d 511, 519 (9th Cir. 1993) (A "copy created in the RAM can be 'perceived, reproduced, or otherwise communicated,' we hold that the loading of software into the RAM creates a copy under the Copyright Act.").

lending; and the display of the copyrighted work publicly.[60] Copyright infringement occurs when one or more of these exclusive rights is violated, and both actual copying and improper appropriation can be shown.[61] Actual copying can be proven either through direct or indirect evidence, and improper appropriation requires that sufficient material be copied.[62] As a practical matter, it is the latter that is of concern in regard to the systematic caching or archiving of web pages because access and actual copying are not disputed in the caching and archiving scenarios under analysis.

Web caching and archiving, as practiced by Google and the Internet Archive, respectively, violate three of the exclusive rights granted to the copyright owner. First, copies are made of the individual web pages during the crawl process.[63] Second, the distribution right is violated as both services make the works in question available on their respective sites, in effect offering their own publication of the work.[64] Third, by making these pages available on their respective sites, both of which are open to the public, these services engage in the public display of protected works. Making a work available to be received or viewed by the public over an electronic network constitutes a public performance or display of the work.[65]

---

[60] 17 U.S.C. § 106 (2002).

[61] Arnstein v. Porter, 154 F.2d 464, 468 (2d Cir. 1946).

[62] *Id.*

[63] MAI Sys.,.991 F.2d at 518 (holding that electronic copies of computer materials infringe on the exclusive rights of the copyright owner).

[64] *In re* Napster, Inc. Copyright Litig., 377 F. Supp. 2d 796, 802-804 (N.D. Cal. 2005) (finding that Napster had not "distributed" songs, because they did not reside on the Napster system; here, the copyright protected works reside on Google's servers).

[65] *See* Kelly v. Arriba Soft Corp., 336 F.3d 811, 817 (9th Cir. 2003); *see also* Playboy Enter., Inc. v. Frena, 839 F. Supp. 1552, 1556-7 (M.D. Fla. 1993).

In regard to the amount copied, there are some small variations between the practices of Google and The Internet Archive. Google makes an independent copy of each web page indexed; the links within that page remain unchanged and continue to point to the live site, not the cached versions.[66] A number of pages from a single web site may be cached by Google, but the user can only access these cached pages one at a time from the search results.[67] In contrast, The Internet Archive creates a copy of the entire web site and further modifies links to internal pages to point to their archived version.[68] This Internet Archive thus allows users to browse an entire web site, as it would have appeared on the date it was captured.[69]

The main difference between the two practices depends on whether the protected work is considered to be an individual web page or the entire web site. When registering a claim of copyright in a web site, the entire site is submitted as the work, as it exists at the time of submission to the Copyright Office, which considers the entire site, as submitted, as a work.[70] Thus, the service may or may not copy the whole work, which consists of the entire web site. The Internet Archive generally copies more of a web site, quantitatively, as it endeavors to

---

[66] Olsen, *supra* note 43.

[67] Id.

[68] Internet Archive, Internet Archive Frequently Asked Questions, http://www.archive.org/about/faqs.php (last visited Oct. 6, 2006).

[69] The difference is analogous to a print magazine that is no longer in publication with a particular article that you would like to read. You call your local library with your request. They do not have the magazine, but offer to obtain a photocopy from another library. In the Google scenario, the library provides you with a copy of the single most relevant page from that article. In the Internet Archive model, the library obtains a photocopy of the entire magazine for your perusal.

[70] *Copyright Registration for Online Works*, U.S. COPYRIGHT OFFICE CIRCULAR 66 (UNITED STATES COPYRIGHT OFFICE), Jul. 2006, *available at* http://www.copyright.gov/circs/circ66.html ("For all online works other than computer programs and databases, the registration will extend only to the copyrightable content of the work *as received in the Copyright Office and identified as the subject of the claim*.").

preserve as much of the work as possible.[71] Google will capture multiple pages from a web site and present them not as a complete work but individual cached pages from the broader site.[72]

In either case, a large amount of the work is copied. Even in the caching technique employed by Google, many pages from an individual site will be made available through its cache. From a qualitative standpoint, if the page is returned in response to the user's query, it is likely to be the most important part of the work to that user, or the "heart of the work", which increases the likelihood of a finding of improper appropriation. Thus, both practices copy large amounts of the protected works, qualitatively and quantitatively. Under a strict interpretation of copyright law, it is likely that both practices would meet the threshold of infringement, making the services liable in the absence of a defense.

IV. Defenses

Implied License

There is a common mistaken belief that information published on the Internet has been released into the public domain and the rights to control that content have been abandoned. Of course, this is not the case. However, it does the raise the question as to whether authors grant certain rights to online readers by making their works freely available online. Among the strongest arguments defending the use of copyrighted works made available online is the concept of implied license.[73] It is well established that "'a nonexclusive license may be granted orally, or

[71] Internet Archive, *supra* note 68 ("How did I end up on the live version of a site? or I clicked on X date, but now I am on Y date, how is that possible?").

[72] *See generally* Olsen, *supra* note 43.

[73] Effects Assocs. v. Cohen, 908 F.2d 555, 557 (9th Cir. 1990).

may even be implied from conduct.'"[74] Whether an implied license exists is determined on a case-by-case basis because it relies on the intent of the author as evidenced by conduct and circumstances.[75]

Arguably, an author would not post information on the Internet if he or she did not intend for others to read it. This necessitates making a temporary copy of the web site on the user's computer, which would be within the implied license granted by the author.[76] The extent of such a license is largely implied by the author's conduct. For example, web sites such as the New York Times include links next to the articles they publish entitled "E-Mail This" and "Printer-Friendly".[77] Through these tools provided by the content owner, a user can e-mail the article to others, or print a copy for her own use. The inclusion of these tools implies permission to copy and distribute the work. In the absence of these links, it is not clear whether a license exists to copy and distribute the page through printing or e-mail, and doing so may be a violation of copyright law. While the personal copying of a web site is unlikely to become the subject of litigation, corporate copying may well attract a suit. In the help section of the New York Times web site, the following limitations to the implied license granted to the users of the New York Time web site are discussed:

> **CUSTOMER SERVICE**
>
> **May I post New York Times articles in a newsgroup environment?**

[74] *Id.* at 558 (quoting Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 10.03A, at 10-36 (1989)).

[75] *See* Herbert v. United States, 32 Fed. Cl. 293, 298 (1994) ("The existence and scope of an implied license . . . necessarily depends of the facts of each individual case." (citation omitted)).

[76] Rice, *supra* note 51.

[77] NY Times Home Page, http://www.NYTimes.com (select an article; then the available tools will appear on the right) (last visited Oct. 5, 2006).

> Posting a New York Times article in a non-commercial newsgroup environment for the purpose of discussion is permitted only if it is not possible to link to the article on www.nytimes.com. It is not permitted, even in a newsgroup environment, to create an archive of New York Times articles. New York Times articles posted in compliance with this policy must be removed within 30 days, and include the following copyright notice:
>
> **© 2004 The New York Times Company. Reprinted by Permission.**
> If the discussion group is e-mail based, only the URL may be distributed with a link back to the article on www.nytimes.com. Under no circumstances is e-mail distribution of New York Times articles to discussion groups allowed without permission.
>
> The "E-mail this Article" feature located on the pages of most articles allows the e-mail distribution of an article to up to 20 recipients. The use of this feature is limited to personal, non-commercial purposes only. [78]

It is important to note that this information is not presented to the user when utilizing the "E-Mail This" feature, and is only found through a search of the web site's help area. If it were, this would transform it into an express license. However, if the user is not made aware of these terms, and acts only on the implication that she may read, copy, and distribute the content based on the tools provided on the web site, implied license applies.

In applying the theory of implied license to web caching and archiving, the author's knowledge of an infringing practice and acquiescence in that use can constitute an implied license.[79] For creation of implied license under this theory, it must be demonstrated that the copyright owner was aware of the particular practice (archiving or caching), and took no steps to prevent it.[80] In *Field v. Google*, Field, a web site owner, conceded knowledge of Google's caching practice and the ability to opt, but failed to attempt to prevent the caching, which acquiescence was found to license Google to make a cached copy of his site. As this case

[78] Customer Service, *May I post New York Times articles in a newsgroup environment?* N.Y. TIMES, http://www.nytimes.com/ref/membercenter/faq/rightsperqa5.html (last visited Feb. 14, 2006).

[79] Field, 412 F. Supp. 2d at 1116 (citing Keane Dealer Servs., Inc. v. Harts, 968 F. Supp. 944, 947 (S.D.N.Y. 1997)).

[80] *Id.*

demonstrates, findings of implied license need to be made on a case-by-case basis, depending largely on evidence as to the knowledge and conduct of the copyright owner. A question not answered by the court in this case is a broader one – at which point are such practices considered generally known among those who publish on the Internet, thereby creating a blanket implied license?

Once Google was made aware of Field's lawsuit through the service of his complaint, it removed the "Cached" links from his search engine listings, in compliance with its policy not to include web pages against the wishes of the copyright holder.[81] This is consistent with the law governing implied licenses, since Field's suit terminated any implied license that existed.

DMCA Safe Harbor

The Digital Millennium Copyright Act ("DMCA") includes a provision designed to address the issues that arise in traditional system caching scenarios.[82] However, this provision would not apply to the form of caching and archiving that is the subject of this note. In a traditional caching scenario, the network provider acts as a silent intermediary between a web site and its visitors.[83] Copies of web pages requested by users are stored by the network provider, and served to subsequent users instead of obtaining a fresh copy, reducing network

[81] Field, 412 F. Supp. 2d at 1123.

[82] 17 U.S.C. § 512(b) (2006).

[83] H.R. REP. NO. 105-551 pt.2, at 51-52 (1998).

traffic and improving response time. This entire process occurs automatically at a system level, and is invisible to users.[84]

In particular, section 512(b) provides a safe harbor for system caching, allowing for the "intermediate and temporary storage" of content.[85] While the indefinite archiving of web sites would clearly fall outside this definition, it has been argued that the caching of web sites as practiced by Google may be exempted under this section of the DMCA.[86] In order for the section 512(b) safe harbor to apply, the content must be:

- intermediate and temporary;
- made available online by a person other than the service provider;
- transmitted to a third person at his or her direction; and
- stored through an automatic technical process so as to make the material available to users of the system who request access to it from the person who initially made the material available.[87]

In addition, the content must be transmitted without modification per section 512(c)(1), and where material is made available online without the authorization of the copyright owner of the

[84] Perfect 10, Inc. v. CCBill, LLC, 340 F. Supp. 2d 1077, 1086 (June 22, 2004) ("The DMCA was enacted both to preserve copyright enforcement on Internet and to provide immunity to service providers from copyright infringement liability for 'passive,' 'automatic' actions in which service provider's system engages through technological process initiated by another without knowledge of service provider."); H.R. REP. NO. 105-551, at 50-51 (1998).

[85] 17 U.S.C.A. § 512(b).

[86] *Field*, 412 F. Supp. 2d at 1123.

[87] Rice, *supra* note 51, at 9.

material, the service provider must respond "expeditiously" to remove or disable access to the allegedly infringing material upon notification in a manner prescribed by section 512(c)(3).[88]

The first question raised is what is meant by "intermediate and temporary" in section 512(b). In *Field*, the court looked to *Ellison v. Robertson* for an analogous situation.[89] In *Ellison*, the court analyzed whether the safe harbor provisions of Section 512(a) applied to messages posted to online bulletin boards (known as Usenet), which were hosted by America Online ("AOL") for a period of approximately fourteen days.[90] The court found the posting to be "intermediate and *transient*" under the provisions of Section 512(a).[91] The court in *Field* then applied this finding to the interpretation of "intermediate and temporary", and determined that web caching for a similar period of time (fourteen to twenty days in the case of Google) was "intermediate and temporary" within the meaning of Section 512(b).[92]

This interpretation of "intermediate and temporary" in Section 512(b) is questionable, as Sections 512(a) and 512(b) refer to different activities, activities that Congress did not consider to be analogous. Section 512(a) was written to address the transmission of material over networks, such as e-mail or other directed messages. By the use of "intermediate and transient" in Section 512(a), Congress was referring to "a copy made and/or stored in the course of a transmission, not a copy made or stored at the points where the transmission is initiated or

[88] 17 U.S.C.A. §512(c)(3).

[89] *Field*, 412 F. Supp. 2d at 1124 (citing *Ellison v. Robertson*, 357 F.3d 1072, 1081 (9th Cir. 2004)).

[90] *Id.*

[91] *Field*, 412 F. Supp. 2d at 1124.

[92] *Id.*

received."[93] Usenet, the service at issue in *Ellison*, operates by transmitting e-mail messages which are received by a targeted "group" and made available to users who subscribe to that group. [94] The messages are held for a short period of time on the network provider's servers so that the subscribers have the opportunity to view these messages.[95]

Section 512(b) was designed to deal with a somewhat different scenario, in which information requested by a user is stored to improve performance or accessibility.[96] Thus, the terms "intermediate and temporary" take on a different meaning. An e-mail message may be held on a network for months before a user retrieves it, making it intermediate and transitory under the meaning of 512(a). However, under 512(b), the safe harbor allowance for the caching of material, where the network in question acts as more than a simple "pass through" conduit.[97] As such, different standards may well apply, and the court's analysis in *Field* is questionable.

Sections 512(b)(1)(B) and (C) were similarly analyzed in *Field*, finding that these requirements of the safe harbor provision are satisfied by the caching as practiced by Google. Section 512(b)(1)(B) requires that three parties be involved in the caching process: 1) the "originator" of the content, 2) the network that transmits and caches the information, and 3) the "user" who requests the information.[98] In order for the caching to qualify for the safe harbor, the "user" must initiate the process by requesting the information from the "originator." The

---

[93] Digital Millenium Copyright Act of 1998, H.R. REP. NO. 105-551, at 50-51 (1998).

[94] Wikipedia, *Usenet*, http://en.wikipedia.org/wiki/Usenet.

[95] *Id.*

[96] *See* H.R. REP. NO. 105-551, at 51-52.

[97] *Id.* at 51.

[98] 17 U.S.C. § 512(b)(1)(B) (2006).

network's purpose in retaining a copy must be to serve future users through an automated technical process, when the information is requested by another "user" from the same "originator," as required by section 512(b)(1)(C). In *Field*, the court reasoned that Google could act as both the "user" and the network, as Google requests and stores the pages for later use.[99] The court similarly found that Google satisfied Section 512(b)(1)(C), as the pages were cached by an automated process, and made available to users if they should find that their attempts to reach the originating site were unsuccessful for any reason.[100] Both interpretations contravene the legislative intent laid out in the House Report 105-551(II):

> New Section 512(b) applies to a different form of intermediate and temporary storage than is addressed in subsection (a). In terminology describing current technology, this storage is a form of ''caching,'' which is used on some networks to increase network performance and to reduce network congestion generally, as well as to reduce congestion and delays to popular sites. **This storage is intermediate in the sense that the service provider serves as an intermediary between the originating site and the ultimate user.** The material in question is stored on the service provider's system or network for some period of time to facilitate access by users subsequent to the one who previously sought access to it. For subsection (b) to apply, **the material must be made available on an originating site, transmitted at the direction of another person through the system** or network operated by or for the service provider **to a different person**, and stored through an automatic technical process so that users of the system or network who subsequently request access to the material from the originating site may obtain access to the material from the system or network.[101]

The critical missing element in Google's caching is the requesting "user." For the safe harbor to apply, a "user" must request a particular page, which is stored by the network. Further, in order to distribute the cached page to a future user, it must be transmitted at the future user's request,

[99] Field, 412 F. Supp. 2d at 1124.

[100] *Id.*

[101] H.R. REP. NO. 105-551, pt. 2, at 51-52 (1998) (Emphasis added).

from the "originator." Neither condition is met in the case of Google's cache feature, and therefore the safe harbor provision of 512(b) was not intended to protect this kind of activity.

As stated in the notes of section 512, the "Digital Millennium Copyright Act (DMCA) was enacted both to preserve copyright enforcement on Internet and to provide immunity to service providers from copyright infringement liability for passive, automatic actions in which service provider's system engages through technological process initiated by another without knowledge of service provider."[102] The activities undertaken by Google are neither passive nor without knowledge, but rather are active and intentional, with the goal of providing an additional feature to search engine users. Therefore, the safe harbor provided under section 512(b) does not apply to Google's cache feature.

Reproduction by Libraries and Archives

Section 108(a) provides that under certain conditions,

> [I]t is not an infringement of copyright for a library or archives, or any of its employees acting within the scope of their employment, to reproduce [or distribute] no more than one copy or phonorecord of a work . . . [provided] (1) the reproduction or distribution is made without any purpose of direct or indirect commercial advantage; (2) the collections of the library or archives are (i) open to the public, or (ii) available not only to researchers affiliated with the library or archives . . . but also to other persons doing research in a specialized field; and (3) the reproduction or distribution of the work includes a notice of copyright . . .[103]

Section 108(a), however, does not apply to materials the library or archive does not own.[104]

---

[102] 17 U.S.C. § 512, Note 1 (1999).

[103] 17 U.S.C. § 108(a) (2005).

[104] Peter B. Hirtle, *Digital Preservation and Copyright,* Stanford Copyright & Fair Use, http://fairuse.stanford.edu/commentary_and_analysis/2003_11_hirtle.html (last visited Oct. 4, 2006). ("[Section 108(a)] does not help them, however, preserve materials that they do not own, such as networked resources or Web sites.").

Slightly different rules apply under section 108 to works preserved in digital form. Libraries and archives may make up to three copies of an unpublished work currently held in their collection for the purposes of preservation and deposit at another library.[105] Similarly, they may make up to three copies of a published work to replace a work in their collections that is damaged, deteriorating, or lost, or whose format has become obsolete, if the library determines that an unused replacement cannot be obtained at a fair price.[106] In either case, copies in digital format, like those of unpublished works, may not be made available to the public outside the library premises.[107] The argument could be made that due to the rapid rate of change on the Internet, combined with the lack of artifacts due to its electronic nature, that (confused as to what "it" is -> replacing it would make the sentence flow better) it is deteriorating, and cannot be replaced at a fair price. However, by making the copies available to the public in the comfort of their own homes, the safe harbor rules under sections 108(b) and 108(c) do not apply to Google and the Internet Archive.

However, section 108 expressly states that nothing therein shall in any way affect "the right of fair use as provided by section 107 . . . "[108] Thus, other acts of library and archival reproduction and distribution of works may constitute fair use, particularly if they are for purposes of scholarship, are non-commercial in nature, and do not usurp the copyright owner's market for the original.[109]

---

[105] 17 U.S.C. § 108(b) (2005).

[106] *Id.*

[107] *Id.*

[108] 17 U.S.C. § 108(f)(4) (2006).

[109] 17 U.S.C. § 107 (2006).

Fair Use

To apply the affirmative defense of fair use, courts consider the four statutory factors when weighing the rights of copyright owners against the public good:

> (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.[110]

Of the available defenses, fair use is by far the strongest and most applicable. [<-should this sentence be cited?] Both Google and The Internet Archive provide services that are open to the public and increase the availability of content published on the Internet, thus serving the policies of access and social benefit the fair use doctrine was designed to promote.

Section 107 favors certain categories of use which provide public value, such as scholarship, research, and teaching.[111] These categories of use are not determinative in finding fair use, but instead provide general guidelines. Non-profit use is also favored, but this is not dispositive, and commercial uses may also be found to be fair use.[112]

---

[110] *Id.*

[111] *See Id*. ("Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright.").

[112] June M. Besek, *Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment*, (Council on Library and Information Resources 2003) ("Certain uses are favored in the statute; they include criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, and research. A nonprofit digital archive for scholarly or research use, for example, would be favored by the law. However, favored uses are not automatically deemed fair, and other uses are not automatically deemed unfair. The four factors discussed earlier must be evaluated in each case.").

The first factor, the purpose and character of the use, turns largely on three elements. The first is the purpose of the work, which includes whether the work is for commercial or non-profit purposes. The second is whether the use fits in one of the categories laid out in the statute such as criticism, comment, news reporting, teaching…scholarship, or research, or some other socially beneficial purpose.[113] The third element is whether a use is transformative, or whether it puts the original work to a new and substantially different use.[114]

The Internet Archive is a non-profit organization with goals of supporting research and scholarship.[115] As such, it ranks highly on the first and second elements of factor one in that its use is non-profit and beneficial to society.

Further, the archiving of web pages is transformative, as it puts the content it captures to a new use, which is the archival and display of web sites for the purposes of historic preservation. Transformative use was an important element in finding fair use in *Kelly v. Arriba Soft Corp.*, where a search engine sued over the inclusion of thumbnail images from a site it indexed. In Kelly, the court found that by improving access to information on the Internet, the use of thumbnails in search engine results was transformative fair use.[116] Similarly, using the "Wayback Machine" on The Internet Archive, visitors can see how the web, and specific web

[113] 17 U.S.C. § 107 (2006).

[114] Campbell v. Acuff-Rose Music, Inc. 510 U.S. 569, 579 (1994). ("Whether the new work merely "supersede[s] the objects" of the original creation or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message … Although such transformative use is not absolutely necessary for a finding of fair use, the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works").

[115] Internet Archive, *About IA*, http://www.archive.org/about/about.php#why (last visited Oct. 1, 2006).

[116] *Kelly*, 336 F.3d at 819.

sites, has evolved over time, as the Internet has matured.[117] The Internet Archive enables efficient access to older web sites, which, as previously discussed, have been used as evidence in trademark and copyright infringement suits, serving as an official record of a web site as it existed on a particular day.

Google, on the other hand, is a public, for-profit company, which, while not determinative, weakens a finding of fair use.[118] Google benefits from its cache feature in the form of increased traffic, which results in increased advertising revenue.[119] However, it should also be noted that Google has a unique role among search engines, as it is the dominant player in the search engine market.[120] Millions of users, both individual and organizational, depend on Google to help navigate an ever increasing number of web sites.[121] It is used for education and research alongside traditional research tools.[122] Therefore, Google arguably provides a public service, even though it is a commercial organization.[123]

Google also makes transformative use of the web sites it caches. First, it makes content available when it cannot otherwise be reached for technical or other reasons.[124] Second, the

---

[117] Internet Archive, *About IA*, http://www.archive.org/about/about.php#future (last visited Oct. 1, 2006).

[118] *Kelly*, 336 F.3d at 818.

[119] Perfect 10 v. Google, 416 F. Supp. 2d 828, 846-47 (2006).

[120] Google.com, Google Corporate Information: Our Philosophy; http://www.google.com/corporate/tenthings.html (last visited Oct. 12, 2006).

[121] *Perfect* 10, 416 F. Supp. 2d at 849.

[122] *Id.* at 845.

[123] *Id.* at 848-849.

[124] Field v. Google, 412 F. Supp. 2d 1106, 1119 (D.Nev. 2006).

"Cached" links allow users to see what recent changes have been made to a particular site.[125] Third, by highlighting the search terms in the body of the cached page, it helps users understand why a particular page was responsive to their query.[126] In addition, by requiring that users affirmatively choose to see the cached copy, which includes a notice that it is not current, it makes clear that it is not intended as a replacement for visiting the original site.[127] Like the use of thumbnails in *Kelly v. Arriba*, Google's web caching improves the ability for users to quickly access information on the Internet, putting the content to an entirely new use. As such, a finding of transformative use is likely to be found, as it was in *Kelly*.

The second factor is the nature of the copyrighted work.[128] This, of course, depends on the nature of the work copied. It traditionally favors greater protection of creative works, and lesser protection of fact based works. This factor is the least significant of the four, and is not particularly determinative, as

> even within the field of fact works, there are gradations as to the relative proportion of fact and fancy. One may move from sparsely embellished maps and directories to elegantly written biography. The extent to which one must permit expressive language to be copied, in order to assure dissemination of the underlying facts, will thus vary from case to case.[129]

The third factor is amount and substantiality of the portion used as compared to the work as a whole.[130] The Internet Archive stores entire web sites; Google stores entire web pages. As

---

[125] Field v. Google, 412 F. Supp. 2d 1106, 1119 (D.Nev. 2006).

[126] *Id.*

[127] *Id.*

[128] 17 U.S.C. § 107 (2006).

[129] Robert A. Gorman, *Fact or Fancy? The Implications for Copyright*, 29 J. COPYRIGHT SOC'Y 560, 563 (1982).

[130] *Id.*

previously discussed, Google will cache many pages from the same site, but does not store and display them as a unified site in the way that the Internet Archive does. However, even the copying of an entire work does not necessarily result in a finding against fair use on this factor.[131] It can be further argued that in order to serve the respective transformative purposes of web site archival and web page caching, each needs to copy as much of the original as possible.[132] This is analogous to parody, where much if not all of a work is necessary for the transformative use to be successful.[133] As noted in Kelly, "[i]f 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."[134]

The final and often most important factor is the effect on the potential market for the original work.[135] Both the Internet Archive and Google's Cache feature have drawn criticism in the past that users might bypass paying for content from commercial sites by seeking older content through the use of these services.[136] Similarly, if users visit the cached or archived versions of a web site, the affected web sites may lose advertising revenue.[137] Many web sites depend on advertising revenue to support the availability of "free" content on their sites.[138]

---

[131] Sony Corp. v. Universal City Studios, Inc., 464 U.S. 417, 449-50 (1984).

[132] Kelly v. Arriba Soft Corp., 336 F.3d 811, 821 (9th Cir. 2003).

[133] Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994).

[134] *Kelly*, 336 F.3d at 820-21.

[135] Harper & Row, Publishers, Inc. v. Nation Enter., 471 U.S. 539, 568, (1985).

[136] Olsen, *supra* note 43.

[137] Rice, *supra* note 51.

[138] *Id.*

The question is whether people will use these services as a replacement for the original sites.[139] For example, due to Google's highlighting of the responsive search terms, some users may well use the cached copy instead of the original in order to facilitate rapid searching.[140] The cached link is displayed directly next to the live site in the search results, increasing the likelihood that a user will select it when seeking a particular web site.[141] However, due to the transformative nature of both archiving and caching, it is unlikely that users will opt to use these tools as replacements for the live web site.[142] In the case of Google, the cache feature is intended to be used when the original site is unavailable, and the user is aptly notified that the page he or she is viewing is not the original.[143] The Internet Archive is intended to be used when a past version, and not the current version of a site is sought, and the archived copies can only be accessed through the Wayback Machine. In fact, content less than six months old is not included in the archive, strengthening the argument that it does not serve as a commercial substitute for the sites in its archive.[144] In either case, in order to prevail on this point, a plaintiff would need to demonstrate a potential loss of traffic and/or revenue.[145]

Interestingly, in *Field*, the court added a fifth factor, good faith.[146] Here they found that by offering the ability to opt out of inclusion in the cache, Google makes a sincere effort not to

---

[139] Olsen, *supra* note 43.
[140] *Id.*

[141] *Id.*

[142] *Id.*

[143] *Id.*

[144] Internet Archive Frequency Asked Questions, Internet Archive, http://www.archive.org/about/faqs.php (last visited Oct. 11, 2006).

[145] 17 U.S.C. § 107(4) (2006).

[146] *Field*, 412 F. Supp. 2d at 1122.

cache content against the wishes of the author.[147] The court further noted that Google provided ample notice to users that a cached page was not the original, and facilitated easy access to the live page.[148] Finally, the court contrasted this with Field's conduct, where he knowingly allowed his pages to be cached for the purpose of bringing a lawsuit.[149] The inclusion of this factor is particular to the Ninth Circuit, and it is questionable as to whether other circuits would further supplement the statute in this way.[150] However, the argument that the opt-out requests of copyright owners are honored is applicable to both Google and the Internet Archive, and if applied, increases the likelihood of a finding of fair use.

In regard to the Internet Archive, a finding of fair use is likely based on the four factors under section 107. The first factor, the purpose and character of the use, is particularly persuasive. The Internet Archive's mission to create a publicly available archive of Internet history appears to fall squarely within "nonprofit educational purposes" under section 107(1).[151] As discussed, the use is also transformative, serving a very different purpose than the original web sites. The third factor, "the amount and substantiality of the portion used" is appropriate to the purpose of the Internet Archive, which is to create a complete copy of a web site as it existed at a particular point in time.[152] Finally, as to the effect on the potential market for or value of the

---

[147] *Id.*

[148] *Id*.

[149] *Id. at* 1123.

[150] *Id. at* 1122.

[151] 17 U.S.C. § 107(1) (2006).

[152] 17 U.S.C. § 107(3); *see also* About IA, Internet Archive, http://www.archive.org/about/about.php (last visited Feb. 14, 2006).

original work, the two serve entirely different purposes. The only real conflict occurs when web sites choose to offer their own archival service, and the ability to opt out of inclusion in the Internet Archive addresses this concern.[153]

As to Google, the factors are somewhat less conclusive. Google is a for-profit company, but the cache feature arguably supports research and educational use by making information available to users when the original cannot be accessed.[154] It is also transformative, as discussed above. Under the third factor, less of the copyrighted work is used, with single pages made available through its cache as opposed to entire web sites, as with the Internet Archive. This is still a substantial amount, as the pages are copied whole, but this is necessary to the purpose of the use. Finally, as to the effect on the potential market for the original work, the results are less clear for Google. As discussed above, by presenting the cached link directly next to the live site in its search results, and adding value by highlighting the search terms, some users may well choose to access the cached copy instead of the live site. However, this is rather speculative, and many courts would probably find the issue of market substitution remote. In total, it is likely that a court would find Google's cache feature fair use, as the District Court of Nevada did in *Field*.[155]

V. Conclusion

Both the Internet Archive and Google infringe the exclusive rights of copyright owners under a strict interpretation of copyright law. They each also provide an important public service by making the constantly changing content on the web more available to users through archiving

[153] Olsen, *supra* note 43.

[154] *Id.*

[155] *Field*, 412 F. Supp. 2d at 1123.

and caching. While they fall outside the web caching provisions of the DMCA, both practices are fair use. Further, as both support the removal of copyrighted material by its owner, courts are likely to find the public benefit of such services outweighs the potential harm to authors and content creators. However, decisions limiting fair use, such as the injunction granted in *Perfect 10 v. Google*, demonstrate the unreliability of fair use, even in light of relevant precedent.

Fair use is a defense, and still exposes organizations like the Internet Archive and Google to costly litigation. This is especially problematic for non-profit organizations, like the Internet Archive, the very type of organization fair use is designed to protect. As such, Congress should consider amendments to the Copyright Act to explicitly allow for such beneficial uses. Protection for online archives should be addressed under section 108, and the caching provisions under section 512 should be expanded to include caching as practiced by Google. This would allow for the intent of congress to be known, and would promote useful innovations such as the caching and archiving features discussed in this note.

The recent decisions in *Field* and *Perfect 10* illustrate the subjective nature of fair use, and provide conflicting signposts as to its application to Google's library project.[156] In order to promote innovation for the public good, both the courts and Congress need to ensure that the Copyright Act supports such endeavors, rather than provide an unnecessarily chilling effect through strict interpretation against valuable uses that were not envisioned by its drafters.

[156] Edward Wyatt, *Ruling May Undercut Google in Fight Over Its Book Scans*, N.Y. TIMES, Feb. 25, 2006, at C4.

Table of Authorities

Cases

Statutes & Administrative Codes

Other

Internet Archive Frequency Asked Questions, Internet Archive, at http://www.archive.org/about/faqs.php.....8, 13
June M. Besek, *Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment*, Council on Library and Information Resources, *at* http://www.clir.org/pubs/reports/pub112/contents.html (n.d. January, 2003)..... 24
Nate Anderson, Judge to Google: Some thumbnails are illegal, Arstechnica, at http://http://arstechnica.com/news.ars/post/20060222-6234.html (Feb. 26, 2006)
Peter B. Hirtle, *Digital Preservation and Copyright*, Stanford Copyright & Fair Use, *at* http://fairuse.stanford.edu/commentary_and_analysis/2003_11_hirtle.html (visited October 9, 2005) 22
Robert A. Gorman, Fact or Fancy? The Implications for Copyright, 29 J. Copyright Soc. 560, 563 (1982)……..
Stefanie Olsen, *Google cache raises copyright concerns*, CNET News.com, *at* http://news.com.com/2100-1032-1024234.html (Jul. 9, 2003)..... 8, 9, 13, 14, 28, 30
Shannon Duffy, Law Firm Accused of Internet Hacking, The Legal Intelligencer, available at http://http://www.law.com/jsp/article.jsp?id=1121245511855 (Jul. 14, 2005)……...4
The New York Times, http://www.NYTimes.com.....
Tom Zeller, Keeper of Expired Web Pages Is Sued Because Archive Was Used in Another Suit, THE NEW YORK TIMES, July 13, 2005……..4
US Copyright Office Circular 66: Copyright Registration for Online Works, available at http://www.copyright.gov/circs/circ66.html.....70
Web Cache, Wikipedia, *at* http://en.wikipedia.org/wiki/Web_caching (last modified Feb. 9, 2006) 6