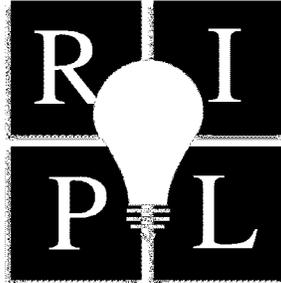# THE JOHN MARSHALL
REVIEW OF INTELLECTUAL PROPERTY LAW

THE CACHE COW: CAN CACHING AND COPYRIGHT CO-EXIST?

NICOLE BASHOR

ABSTRACT

Copyright protection in a digital environment is difficult. Protection is made more difficult because of the scope of the Internet, uncertainty of a governing Internet agency, and copyright enforcement problems. Copyright protection is further complicated because search engines utilize a process involving copying web sites and displaying cached links to web sites without the web site owner's express permission. By analyzing the intersection of copyright law and the search engine caching process, it is clear that search engines are on dangerous ground and possibly violate the copyrights of web site owners. Accordingly, a solution is proposed that compensates web site owners and ensures that search engines operate within the bounds of copyright.

THE CACHE COW: CAN CACHING AND COPYRIGHT CO-EXIST?

NICOLE BASHOR[*]

"If a business cannot be operated within the bounds of the Copyright Act, then perhaps the question of its legitimate existence needs to be addressed."[1]

INTRODUCTION

The millions of people who use Internet search engines[2] want fast searches, relevant results, and results on demand.[3] At the same time, authors, musicians, and artists desire the identical things, but not at the expense of copyright protection for their works.[4] Traditionally, copyright law has struggled to protect its purpose while adapting to technological advances.[5]

In an effort to satisfy the demands of Internet users and to increase access to and speed of web searches, search engines perform a complicated process involving copying, indexing, and displaying cached links to web sites.[6] The web site caching

---

[1] Playboy Enters. Inc. v. Webbworld, Inc., 968 F. Supp. 1171, 1175 (N.D. Tex. 1997).

[2] BARRON'S LAW DICTIONARY 462 (5th ed. 2003) ("SEARCH ENGINE: a computerized system that enables the user to type in various search terms and receive information that fits the designated query."); *see also* Ian Wylie, *The Search for the Fastest Engine: Can an Upstart From Norway Outpace Google and Finish First on the Web?*, Oct. 2002, http://www.fastcompany.com/magazine/63/smartcompany.html (last visited Apr. 27, 2006) (stating Google "pulls in 58 million users a month.").

[3] *See* Danny Sullivan, 2001 Search Engine Watch Awards, Search Engine Watch, Feb. 6, 2002, http://searchenginewatch.com/awards/article.php/2155891 (last visited Mar. 16, 2006) (listing characteristics of the winner for the Outstanding Search Service award as "accurate, fast, comprehensive, easy to use, ads were clearly identified and not intrusive").

[4] BARRON'S LAW DICTIONARY 111 (5th ed. 2003) ("COPYRIGHT: the protection of the works of artists and authors giving them the exclusive right to publish their works or determine who may so publish."); *see also, e.g.*, Copyright.com, Copyright Education and Services for Authors, http://www.copyright.com/ccc/do/viewPage?pageCode=au1 (last visited Oct. 30, 2006) (giving an example of an organization that provides licensing services for copyright holders).

[5] S. REP. NO. 105-190, at 1 (1998) [hereinafter DMCA History].

> Copyright laws have struggled through the years to keep pace with emerging technology from the struggle over music played on a piano roll in the 1900's to the introduction of the VCR in the 1980's. With this constant evolution in technology, the law must adapt in order to make digital networks safe places to disseminate and exploit copyrighted materials.

Id.

[6] *See* Wylie, *supra* note 2 (discussing how search engines compete to have the fastest and most reliable results); *see also* Google Guide: How Google Works, http://www.googleguide.com/

and indexing process involves a search engine's spider[7] copying the original web site and the search engine translating the copied web site text into an index.[8] A cached hyperlink[9] ("link") is normally displayed in search engine search results as well, which is a link to a copy of the website as displayed the last time the search engine spider visited the web site. The search engine caching process[10] has been a source of controversy recently because it allows for potential abuse of copyright holders' interests.

This article explores search engines' copying, indexing, and caching links, and how these processes contradict web site owners' copyright interests in some aspects, but further the policy goals of copyright in other aspects. First, the background section discusses how copyright laws intersect with the search engine caching process. Next, the analysis section examines conflicting views and policy considerations with regards to the search engine caching process and presents a hypothetical scenario illustrating these opposing views. Finally, the proposal section provides a multi-part solution to balance the interests of both copyright holders and search engines.

I. BACKGROUND

In order appreciate the conflict, it is important to understand copyright law and the search engine caching process. Section A discusses the pertinent aspects of copyright law including the goals of copyright law, copyright infringement, and defenses to infringement. Section B explores the search engine caching process along with the advantages and disadvantages of each aspect of the process. Section C discusses the current state of the law and the applicability of the current case law to the issue of caching and copyright.

---

google_works.html (last visited Mar. 10, 2006) (explaining the entire process of how Google sends out its robots to copy the web sites, stores the web sites in the index, and recalls web sites when a search engine user enters the matching search terms) [hereinafter Google Guide].

[7] Ebay, Inc. v. Bidder's Edge, Inc., 100 F. Supp. 2d 1058, 1060 (N.D. Cal. 2000).

> A software robot is a computer program which operates across the Internet to perform searching, copying, and retrieving functions on the web sites of others ... [p]rograms that recursively query other computers over the Internet in order to obtain a significant amount of information are referred to in the pleadings by various names including software robots, robots, spiders, and web crawlers.

Id.

[8] *See generally* Google Guide, *supra* note 6.

[9] Stacey L. Dogan, *Infringement Once Removed: The Perils of Hyperlinking to Infringing Content,* 87 IOWA L. REV. 829, 837 (explaining the technology of hyperlinking and how links connect the Internet).

[10] *See* MICHAEL D. SCOTT, INTERNET AND TECHNOLOGY LAW DESK REFERENCE 86 (Aspen 2004) ("[C]aching: a technique for temporarily storing digital information closer to a requesting computer on the Internet to speed up access to that information.").

A. Copyright Law Generally

The objective of copyright law has remained the same from its foundation in the Constitution:[11] promote creation, dissemination, and access to new works through the incentive of copyright protection.[12] However, in the past few years, the fast growth of technology and the Internet has created diverse and challenging problems for copyright holders because copyrights are difficult to protect in a digital environment.[13]

The cornerstone of copyright protection is embodied in a copyright holder's exclusive rights to reproduce works, distribute copies[14] of works, and prepare derivative works.[15] When a copyright holder believes one of their exclusive rights have been violated, she can attempt to enforce it through a cause of action for infringement.[16] Granting protection to authors' works for a limited amount of time serves as an incentive system with the goal of producing greater creative activity.[17]

---

[11] U.S. CONST. art. I, § 8, cl. 8. "The Congress shall have Power . . . to promote the Progress of Science and the useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." *Id.*

[12] *See* Sony Corp. of Am. v. Universal City Studios, 464 U.S. 417, 431 (1984).

> The limited scope of the copyright holder's statutory monopoly . . . reflects a balance of competing claims upon the public interest: [c]reative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts . . . the ultimate aim is . . . to stimulate artistic creativity for the public good.

Id.

[13] *See generally* Ian C. Ballon, *Copyright, Digital Entertainment and the Impact of New Technologies: A Primer,* 711 PLI/PAT 9, (2002) (giving an overview of the complicated issues that exist throughout the areas of copyright, trademark, patent, and trade secrets as they intersect with the Internet).

[14] *See* Marobie-FL, Inc. v. Nat'l Ass'n of Fire Equip. Distribs., 983 F. Supp. 1167 (N.D. Ill. 1997).

> Under the Copyright Act . . . "copies" are defined as "material objects . . . in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.

Id.

[15] 17 U.S.C. § 106 (2000).

> Subject to sections 107 through 122, the owner of copyright under this title has exclusive rights to do and to authorize any of the following: to reproduce the copyrighted work . . . to prepare derivative works . . . to distribute copies . . . of the copyrighted work to the public by sale or other transfer of ownership.

Id.

[16] *See* Playboy Enters., Inc. v. Webbworld, Inc., 968 F. Supp. 1171, 1175 (N.D. Tex. 1997) (discussing how a copyright holder can assert his rights through an infringement cause of action).

[17] *See* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 546 (1985).

1. Copyright Infringement

Copyright infringement occurs when one of the copyright holder's exclusive rights is violated.[18] Copyright law provides liability for infringement through various causes of action including direct, contributory, and vicarious infringement.[19] This section focuses on the requirements of direct infringement.

Direct infringement requires the plaintiff (1) prove ownership of the copyrighted material and (2) prove the defendant copied the material.[20] Additionally, the plaintiff must prove the defendant's conduct was volitional.[21] Volitional acts are conscious and purposeful actions committed by the infringer demonstrating their objective.[22]

At this time, there is limited case law discussing copying on the Internet involving volitional conduct. One example, in *Costar Group v. Loopnet*, Loopnet, an Internet Service Provider ("ISP"), was held not to be a direct copyright infringer because it passively stored copyrighted material posted by others.[23] In *Costar*, Loopnet subscribers posted Costar's copyrighted photographs on the Loopnet website.[24] Because the Loopnet system uploaded the copyrighted pictures at the request of the users, the *Costar* court held volitional conduct does not include the automatic response of an ISP's equipment to a request from a user's input.[25] As a result, the court relieved Loopnet from liability.[26] However, because *Costar* was directed towards passive ISPs, its holding is not applicable to the search engine's because search engine's are not passive in the search engine caching process.[27]

2. Defenses to Copyright Infringement

In response to an allegation of copyright infringement, the accused infringer can assert several defenses. Some commonly used defenses to an infringement claim

---

[18] *See* 17 U.S.C. § 501(b) (2000) ("The legal or beneficial owner of an exclusive right under a copyright is entitled ... to institute an action for any infringement of that particular right committed while he or she is the owner of it.").

[19] *See* Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc., 907 F. Supp. 1361, 1367–78 (N.D. Cal. 1995) (explaining the three different types of copyright infringement liability and the statutory requirements of each).

[20] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991) ("To establish infringement, two elements must be proven: (1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original.").

[21] *See Religious Tech.*, 907 F. Supp. at 1370 (explaining in addition to the statutory elements for a claim of copyright infringement, volitional conduct must be proven).

[22] Costar Group, Inc. v. Loopnet, Inc., 373 F.3d 544, 550 (4th Cir. 2004) (describing volitional acts as "actual infringing conduct with a nexus sufficiently close and causal to the illegal copying that one could conclude that the machine owner himself trespassed on the exclusive domain of the copyright owner.").

[23] *Id.* at 555.

[24] *Id.* at 546.

[25] *Id.* at 555.

[26] *Id.*

[27] *See id.* at 554 (stating "liability is ruled out for passive, automatic acts engaged in though a technological process initiated by another [person].").

include: non-infringement,[28] implied license,[29] fair use,[30] and the Digital Millennium Copyright Act ("DMCA") safe harbor provisions.[31]

a. Fair Use

Fair use is a copyright infringement defense accorded to potential infringers giving them a limited right to use copyrighted material in a reasonable manner without the owner's consent.[32] A court applies the fair use defense after weighing different factors relevant to the nature of the interests in the case.[33] When applying the fair use defense, courts examine (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.[34] Courts weigh the factors differently, but the purpose and character of the use and the effect of the use on the potential market or value of the work factors are generally given more weight.[35] Finally, the fair use defense is negated if a copyright holder can show the challenged use will "adversely affect the *potential* market for the copyrighted work" if the use becomes widespread.[36]

b. DMCA Safe-Harbor Defense

While most of the defenses would be available to a search engine in a claim against it for infringement, it is unclear as to whether the DMCA safe harbor provisions would be applicable. Congress adopted the DMCA, in part, to update copyright law to embrace the technological advances of the Internet.[37] The DMCA

---

[28] *See* 17 U.S.C. § 501(a) (2000).

[29] *See* Effects Assocs. Inc. v. Cohen, 908 F.2d 555, 558 (9th Cir. 1990) (explaining that an implied license is a defense for copyright infringement and that an implied license can be granted orally, or implied by conduct).

[30] Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 549 (1985) ("Fair use was traditionally defined as 'a privilege in others than the owner to use the copyrighted material in a reasonable manner without his consent.'").

[31] 17 U.S.C. §§ 512(a)–(d), 512(j) (2000) (listing the requirements for an Internet Service Provider to qualify for DMCA safe harbor protection).

[32] Field v. Google Inc., 412 F. Supp. 2d 1106, 1117 (D. Nev. Jan. 19, 2006).

[33] *See* 17 U.S.C. § 107 (2000); *see also Harper & Row, Publishers, Inc.*, 471 U.S. at 553 ("The nature of the interest at stake is highly relevant to whether a given use is fair.").

[34] 17 U.S.C. § 107 (2000).

[35] *See Field*, 412 F. Supp. 2d at 1118 (explaining the varying importance of each factor).

[36] *Harper & Row, Publishers, Inc.*, 471 U.S. at 568.

[37] *See In re* Aimster Copyright Litig., 252 F. Supp. 2d 634, 656 (N.D. Ill. 2002). The DMCA was "enacted both to preserve copyright enforcement in the Internet and to provide immunity to service providers from copyright infringement liability for 'passive,' 'automatic' actions in which a service provider's system engages through a technological process initiated by another *without the knowledge of the service provider.*" *Id.* (emphasis added).

affects the rights of all of the key participants in the online world including copyright holders and ISPs.[38]

Some of the main provisions of the DMCA apply to ISPs, including the safe harbor provisions, which serve to limit the ISPs' liability based on actions of third parties.[39] To qualify for these safe harbors, the ISPs must meet certain threshold requirements that revolve around "account holders" and "subscribers."[40] This is problematic for search engines because they do not have account holders or subscribers. Another requirement for DMCA protection is that ISPs cannot interfere with standard technical measures copyright owners have implemented to protect copyrighted material.[41] This is also a challenging requirement for search engines, because cached links interfere with standard technical measures. A web site owner may try to protect her work using a registration system that cached links help usurp, thereby defeating the technical measure of the registration system.[42] Even though some courts have liberally applied the DMCA safe harbor provisions to search engines, this defense is still challenging for search engines because they fail to meet the threshold requirements.[43]

---

[38] *See* DMCA History, *supra* note 5, at 6.

> Legislation implementing the treaties provides this protection and creates the legal platform for launching the global digital on-line marketplace for copyrighted works . . . by limiting the liability of service providers, the DMCA ensures that the efficiency of the Internet will continue to improve and that the variety and quality of services on the Internet will continue to expand.

Id.

[39] *See* 17 U.S.C. §§ 512(a)–(d), (j) (2000); *see also* Perfect 10, Inc. v. Cybernet Ventures, Inc., 213 F. Supp. 2d 1146, 1174 (C.D. Cal. 2002). "The DMCA created a series of four 'safe harbors' to protect 'providers of online services' from liability, primarily monetary, based on claims of copyright infringement attributable to the actions of users." *Id.*

[40] *See Perfect 10, Inc.,* 213 F. Supp. 2d at 1174.

> In order to qualify for these safe harbors, a provider of online services must: 1) adopt a policy that provides for the termination in appropriate circumstances of subscribers and account holders of the service provider's system or network who are repeat infringers; 2) reasonably implement the policy; and 3) inform subscribers and account holders of the service provider's system or network about the policy.

Id.

[41] *Id.* at 1174 ("Standard technical measures includes a requirement that they be 'developed pursuant to a broad consensus of copyright owners and service providers in an open, fair, voluntary, multi-industry standards process.'").

[42] *See* Stephanie Olsen, *Google Cache Raises Copyright Concerns,* NEWS.COM (July 9, 2003) http://news.com.com/2100-1038_3-1024234.html (last visited May 14, 2006). This article discusses the problems with cached links in terms of web sites requiring registration and fees. *Id.* Cached links allow search engines users to enter these sites through the cached link without having to pay the registration fee. *Id. See also* Wiki How, http://wiki.ehow.com/Bypass-Registration-on-Web sites-using-the-Google-Cache (last visited Mar. 12, 2006) (giving detailed instructions on bypassing web site registrations).

[43] *See* Craig W. Walker, *Application of the DMCA Safe Harbor Provisions to Search Engines,* 9 VA. J.L. & TECH. 2 (2004). "Unfortunately, applying the DMCA safe harbor provisions to search engines is problematic. Key portions of the statute refer to 'subscribers' and 'account holders,' making their application to search engines unclear because search engines typically do not have subscribers or account holders." *But see In re* Aimster Copyright Litig., 252 F. Supp. 2d 634, 658 (N.D. Ill. 2002) ("A plain reading of both definitions reveals that 'service provider' is defined so

B. Website Copying, Indexing, and Caching Links

One of the purposes of the search engine caching process is to store web sites for use at a later time.[44] As the Internet expands, the demand for high-speed web site retrieval increases.[45] Search engines are competing to have the fastest and the most accurate results because these qualities attract more users,[46] more advertising,[47] and thus, more money for the search engines.[48]

1. The Details of the Caching Process

At the heart of caching lies a complex process involving software robots, indexing, queries, and servers.[49] Initially, Internet search engines send out their software robots to make copies of web sites and to process the information stored on these web sites.[50] The information processing includes making a high-tech index.[51] When a user enters a search term into the search engine, the search engine checks the index and returns the most relevant results.[52] As part of this indexing process, the search engine document servers[53] retain a copy of the web site as it looked when the robot last visited the web site.[54] Finally, when a search engine user makes a query, the search engine goes to the document server and retrieves the "stored documents" that provide the closest match to the search terms.[55]

Without express permission from web site owners, copyright infringement issues against search engines arise at two stages of the overall search engine process: (a)

---

broadly that we have trouble imagining the existence of an online service that would not fall under the definitions, particularly the second.").

[44] Brian D. Davison, *Web Caching Overview,* http://www.web-caching.com/welcome.html (last visited Apr. 27, 2006) ("Web caching is the temporary storage of web objects (such as HTML documents) for later retrieval.").

[45] *See* Wylie, *supra* note 2 (discussing the importance of a search engine index and speed and how the search engine companies are competing to compile the largest online index of web sites in an effort to make more money).

[46] *See* Sullivan, *supra* note 3 (stating Internet users prefer search engines that are fast and return relevant results).

[47] Adi Ignatius, *In Search of the Real Google,* TIME, Feb. 20, 2006, at 36. "The holy grail turned out to be advertising, and it's not an exaggeration to say that Google is now essentially an advertising company, given that that's the source of nearly all its revenue." *Id.*

[48] *Id. See also* Sullivan, *supra* note 3; *see also* Google Investor Relations, http://investor.google.com/fin_data.html (last visited Mar. 10, 2006) (showing that Google makes most of its money from advertising).

[49] *See* Google Guide, *supra* note 6.

[50] *See id.*

[51] *See id.*

[52] *See* Avi Rappoport, *Anatomy of a Search Engine: Inside Google,* SEARCH ENGINE WATCH, Oct. 30, 2002, http://searchenginewatch.com/searchday/article.php/2161091 (last visited Apr. 27, 2006) (explaining the construction of the search engine index and how the search engine retrieves documents from the index based on specific criteria).

[53] Ignatius, *supra* note 47, at 49 ("Google maintains tens of thousands of servers to store all those cached Web sites it searches.").

[54] Google Guide, *supra* note 6.

[55] Google Guide, *supra* note 6.

the initial copying and (b) the displaying of cached links.[56] The first area of potential liability is when the search engine's robot makes the initial copy of the web site during its indexing efforts.[57] This initial copy is made unless the web site owner programs HyperText Markup Language ("HTML") code into the web site and revokes permission from the search engine.[58] The second area of potential liability is when search engines post a link to the copy of the web site as it looked the last time the robot copied the site.[59] These links appear in a user's search results normally under the heading "cached."[60] The reason these areas cause potential liability for the search engines is that the steps are performed without the express permission of the website owner, which is required under the exclusive rights granted to copyright holders.[61]

2. The Robot Exclusion Standard

There are some industry procedures that apply to search engine robots and that define steps webmasters must perform to have their web sites excluded from the search engine copying, indexing, and caching process.[62] The Robot Exclusion Standard begins with the premise that the search engines have free reign to copy web sites and to post links to the cached web sites.[63] To have a web site excluded, a webmaster must go through a process that involves adding lines of HTML code to the web site so the robot knows to ignore the site.[64] Additionally, the webmaster can add HTML code that sends a message to the search engine indicating not to archive the web site via a cached link.[65] This "opt-out" process was developed because of the

---

[56] See Parker v. Google, Inc., 422 F. Supp. 2d 492, 496 (explaining Google's potential liability at the initial copying and indexing stage of the search engine caching process); see also Field v. Google, Inc., 412 F. Supp. 2d 1106, 1115 (D. Nev. 2006) (explaining Field claims that Google has liability based on a cached link).

[57] See Forget Google Print Copyright Infringement; Search Engines Already Infringe, http://blog.searchenginewatch.com/blog/050525-093716 (last visited Mar. 16, 2006) (explaining that the Google search engine infringes upon copyright holders when the Googlebot makes the initial copy of the web site).

[58] See Robots Exclusion, http://www.robotstxt.org/wc/exclusion.html (last visited Mar. 12, 2006) (listing detailed information about how a webmaster can program HTML code into his web site to have the search engine robots ignore and not copy his web site).

[59] See Field, 412 F. Supp. 2d at 1115.

[60] See, e.g., http://www.google.com/search?hl=en&q=caching&btnG=Google+Search (last visited Mar. 16, 2006) (noting the "cached" link at the end of each search result).

[61] See 17 U.S.C. § 501(a) (2000).

[62] See Robots Exclusion, http://www.robotstxt.org/wc/exclusion.html (last visited Mar. 12, 2006) (listing detailed information about what the Robot Exclusion Standard is, how the standard is implemented, and how to make robots perform or not perform various functions based on code entered by the webmaster).

[63] Id.

[64] Id.

[65] Caching Tutorial for Web Authors and Webmasters, http://www.mnot.net/cache_docs/ (last visited Mar. 10, 2006); See also The Web Robots FAQ, http://www.robotstxt.org/wc/faq.html (last visited Apr. 27, 2006). As an example, a webmaster could program the following to not allow robots to access their site: "User-agent: *Disallow:/." Id.

assumption that webmasters would like search engine users to be able to find their web site through the search engine results.[66]

3. Benefits and Drawbacks of Caching

Generally, copying, indexing, and caching links have many benefits for ISPs and Internet users. First, search engine copying and indexing increases the Internet's performance, because search engines are able to retrieve web sites at a faster rate as a result of the web sites being stored on the search engine servers. Second, cached links are beneficial for archival purposes,[67] or for accessing information that has been removed or is unavailable because of server problems.[68] Third, cached links are helpful for web site comparisons[69] and good for allowing users to see their search terms highlighted on the web site they select.[70]

Additional arguments in favor of allowing search engine indexing, caching, and displaying cached links relate to the advantages of the Robot Exclusion Standard. According to some industry experts, the Robot Exclusion Standard is well known to web site programmers and easy to implement.[71] There is also significant public pressure for the robots to comply with webmasters' directions and not copy the web site.[72] Because the Robot Exclusion Standard allows web site owners an easy way to revoke permission, some argue this is an effective way to protect copyright and web site owners' interests.[73] However, there are conflicting views of the implications, advantages, and drawbacks of the Robot Exclusion Standard.

In contrast, copying, indexing, and caching links also have several disadvantages. First, cached links make it easy for hackers to bypass registration for web sites that require a username and password or a subscription.[74] Second, the

---

[66] *See* Posting of Danny Sullivan to Search Engine Watch Blog, *Indexing Versus Caching & How Google Print Doesn't Reprint,* http://blog.searchenginewatch.com/blog/051021-113341 (Oct. 21, 2005) [hereinafter Danny Sullivan Posting] ("[Opt-out] systems ha[ve] worked on the web, because . . . site owners want traffic.").

[67] *See* Peter Lyman, *Archiving the World Wide Web,* http://www.clir.org/pubs/reports/pub106/web.html (last visited Apr. 27, 2006) (discussing the complexities and importance of archiving the Internet); *see also* Field v. Google Inc., 412 F. Supp. 2d 1106, 1111-12 (D. Nev. 2006) (listing two of the three purposes of cached links as being good for archival purposes).

[68] Sullivan, *supra* note 3 (a cached link "helps you see a site that may no longer be online (for a short time at least)").

[69] *Field,* 412 F. Supp. 2d at 1112 (discussing the benefits of a user being able to compare the most recent version of a web site with a version that is stored in Google's cache).

[70] *Id.* The search terms that a search engine user enters into the search box appear highlighted on the cached version of the web site and therefore are easily identified when the search engine user scans the cached version.

[71] *See* Martijn Koster, *Robots in the Web: Threat or Treat?,* Apr. 1995, http://www.robotstxt.org/wc/threat-or-treat.html (last visited Apr. 27, 2006).

[72] *Id.*

[73] Email Interview with Danny Sullivan, Founder and Editor-in-Chief, Search Engine Watch, (Oct. 9, 2006) [hereinafter Sullivan Interview].

[74] *See* Celeste Biever, *Hackers Turn to Google to Find Weakest Links,* NEW SCIENTIST, Aug. 2003, http://www.newscientist.com/article.ns?id=dn4002 (last visited Apr. 27, 2006) (listing the different uses for cached links for hacking into web sites); *see also* Wiki How, http://wiki.ehow.com/Bypass-Registration-on-Web sites-using-the-Google-Cache (last visited Mar. 12, 2006) (giving instructions for hackers on how to bypass web site registration or subscriptions).

Robot Exclusion Standard is optional,[75] complicated,[76] and limited.[77] The Robot Exclusion Standard is limiting because many web site owners, such as bloggers, are unable to manually insert HTML code into their web site.[78] It places the onus on web site owners to know HTML code and to revoke permission with the specific commands in their websites.[79] Third, web site content that has been removed may remain in the search engine cache for months,[80] years,[81] or indefinitely.[82] This usurps the copyright/web site owner's exclusive right of distribution because web site owners cannot control what happens to copies of their web site. Finally, cached links mask the origin of the user who is visiting the web site.[83] This is important because webmasters are then unable to compile accurate data as to who is visiting their web sites.[84] This data is often used in revenue producing advertisements and the web site owners lose advertising revenue when this data is skewed.[85] Additionally, cached links provide a way for search engine users to never visit the source web site.[86] Even though search engine copying, and caching links have real world advantages and disadvantages, courts have hardly addressed the legal implications of this process.

C. The Legal Intersection of Copyright and Caching

At this time, no case has straightforwardly addressed the issue of direct infringement during the initial "copying" step in the search engine caching process,

---

[75] See A Standard for Robot Exclusion, http://www.robotstxt.org/wc/norobots.html#code (last visited Apr. 27, 2006) ("It is not an official standard backed by a standards body, or owned by any commercial organization. It is not enforced by anybody, and there no guarantee that all current and future robots will use it.").

[76] See generally id. (giving examples of the different code that is used to make the robots perform different functions).

[77] Posting of Russ Jones to The Search Engine Journal, http://www.searchenginejournal.com/?p=2855 (Feb. 6, 2006, 12:23EST) (standing for the proposition that many web site operators, including bloggers, are unable to change their HTML code to control search engine indexing and archiving).

[78] Id.

[79] See generally A Standard for Robot Exclusion, supra note 75 (explaining how a web site owner can opt-out).

[80] See Search Engine Roundtable, Brett Tabke Interviewed on Bot Banning, Nov. 28, 2005, http://www.seroundtable.com/archives/002863.html (last visited Apr. 27, 2006) (listing that a web site which recently revoked permission from the search engine robots remained in the various search engine caches for an extended amount of time).

[81] See The Google Cache, http://www.thegooglecache.com/ (last visited Mar. 12, 2006) ("The cache extends access to removed content, often for months if not years at a time.").

[82] See Internet Archive Frequently Asked Questions, http://www.archive.org/about/faqs.php#The_Wayback_Machine (last visited Mar. 16, 2006) ("The Internet Archive Wayback Machine is a service that allows people to visit archived versions of Web sites.").

[83] The Google Cache, http://www.thegooglecache.com/ (last visited Mar. 12, 2006) ("The cache allows Google to serve site content anonymously. [If you] [d]on't want the owner of a site to know you are looking at their goods . . . just watch the cache instead.").

[84] See Eric Schlachter, Cache-22, http://www.ericgoldman.org/Articles/cache22article.htm (last visited Oct. 30, 2006) (explaining the various ways cached links cause web site owners to lose advertising revenue).

[85] See id.

[86] See Sullivan Interview, supra note 73.

or the legality of cached links, but three recent cases illustrate the interplay between the search engine caching process and copyright. Interestingly, the three cases discussed in this section implicate Google, but unfortunately for Google, it only won two of the three cases.

1. Field v. Google

One of the first cases highlighting the issues surrounding indexing and caching was *Field v. Google*.[87] In this case, Field argued that Google infringed his rights when a search engine user clicked on the cached link to Field's writings, which were available for free on his web site.[88] However, the *Field* court specifically made an extra effort to discuss that Field was not claiming infringement during the initial scan and copy by the "googlebot."[89] This distinguishes *Field* from a situation in which a web site owner sues a search engine for the initial copying of his web site. The *Field* court indicated that the result may have been different if Field would have claimed infringement during the initial copying step.[90]

Ultimately, the *Field* court held there was no direct infringement by Google when a search engine user clicked on the cached web site link.[91] In addition to the copyright issues in the case, the *Field* court indicated that it was punishing Field for manufacturing a claim against Google because of his bad faith prior to the lawsuit.[92] The *Field* court also applied several defenses that precluded the finding of liability including: implied license,[93] estoppel,[94] fair use,[95] and DMCA safe harbor.[96]

---

[87] 412 F. Supp. 2d 1106 (D. Nev. 2006).

[88] *See id.* at 1115.

[89] *See id.*

[90] *See generally id.* (explaining "Field does *not* allege that Google committed infringement when its Googlebot . . . made the initial copy of the Web pages containing his copyrighted works and stores those copies in the Google cache."). Although the court points out what Field did not allege, they do not discuss the potential outcome of such a cause of action.

[91] *See id.* at 1114 (granting Google's motion for summary judgment that "by operating its cache and presenting 'Cached' links to works within it, Google does not directly infringe Field's copyrighted works."). Additionally, the court notes that Field did not claim "Google was liable for indirect infringement (contributory or vicarious liability)." *Id.*

[92] *See id.* at 1123 (stating "Field's own conduct stands in marked contrast to Google's good faith").

[93] *Id.* at 1116 ("A web site publisher can instruct a search engine not to cache the publisher's web site by using a "no-archive" meta-tag . . . . Despite this knowledge, Field chose not to include the no-archive meta-tag on the pages of his site.").

[94] *See id.* "A plaintiff is estopped from asserting a copyright claim 'if he has aided the defendant in infringing or otherwise induced it to infringe or has committed covert acts such as holding out . . . by silence or inaction.'" *Id.* To prevail on an estoppel defense, a defendant must prove four elements: (1) the plaintiff knew of the defendant's allegedly infringing conduct, (2) the plaintiff intended that the defendant rely upon his conduct or acted so that the defendant had a right to believe it was intended, (3) the defendant was ignorant of the true facts, and (4) the defendant detrimentally relied on the plaintiff's conduct. *See id.* Because the court found that all four estoppel factors were present, the court granted Google's motion for summary judgment on the estoppel defense. *Id.*

[95] *Id.* at 1123 (explaining why fair use weighs in favor of Google).

[96] *Id.* at 1125 (granting Google's motion for partial summary judgment that it qualifies for the DMCA safe harbor provisions).

2. *Parker v. Google*

Parker v. Google[97] illustrates a direct copyright infringement claim where the court barely addresses the search engine process and copyright. Parker, an author, claimed direct copyright infringement when Google automatically archived a posting he put on USENET, an online bulletin board.[98] Parker further claimed direct copyright infringement when Google produced a list of links in response to a user's search query with excerpts of his website within the list of links.[99] The *Parker* court addressed the direct infringement claim of the archived USENET postings by considering Google an ISP without discussion, and dismissed the complaint, following the same reasoning of the *Costar Group v. Loopnet, Inc.* court.[100] The *Parker* court dismissed Parker's claim, in part, because Google did not have the requisite volitional conduct to satisfy a claim for direct infringement.[101]

Parker's complaint regarding Google's direct infringement via Google's process of indexing and caching websites was also dismissed for failure to state a claim on which relief can be granted.[102] The court devoted one paragraph of analysis and relied on *Field v. Google* and the DMCA safe harbor to relieve Google of liability with little explanation.[103] Parker's claims of Google's contributory and vicarious liability were dismissed as well.[104]

3. *Perfect 10 v. Google*

Another illustrative case dealing with copyright infringement and search engines is *Perfect 10 v. Google, Inc.*[105] At issue in *Perfect 10*, was whether Perfect 10's copyrights were infringed when Google displayed Perfect 10's fee-based photographs in its image search.[106] The Google image search works the same way the text search works in that Google sends robots to make copies of photographs. After the initial copying, Google displays the full images through its image search in thumbnail form.[107] The *Perfect 10* court held that Google directly infringed the copyrights and that the fair use defense did not apply.[108]

---

[97] 422 F. Supp. 2d 492 (E.D. Pa. 2006).

[98] *See id.* at 496.

[99] *Id.*

[100] *See id.* at 497–98; *see also* Costar Group, Inc. v. Loopnet, Inc., 373 F.3d 544, 550 (4th Cir. 2004) (holding volitional conduct is a required element to prove direct infringement).

[101] *Parker*, 422 F. Supp. 2d at 497.

[102] *See id.* at 498 (explaining automatic caching does not constitute direct infringement).

[103] *See id.*

[104] *Id.* at 498–500.

[105] 416 F. Supp. 2d 828 (C.D. Cal. Feb. 17, 2006).

[106] *See id.* at 838.

[107] *See id.* at 832–33 (explaining the thumbnail images).

[108] *Id.* at 851.

> The first, second, and fourth fair use factors weigh slightly in favor of P10. The third factor weighs in neither party's favor. Accordingly, the court concludes that Google's creation of thumbnails of P10's copyrighted full-size images, and the subsequent display of those thumbnails as Google Image Search results, likely do not fall within

II. ANALYSIS

Three topics of discussion are generated in light of the brief history of the intersection of copyright and cached links. First, Section A analyzes the differences in the outcome of the Google cases. Second, Section B discusses the policies in favor of and against the search engine's initial copying of web sites and displaying cached links. Third, Section C applies direct infringement and possible defenses to a hypothetical claim against a search engine for the copying of a web site and displaying a cached link.

A. The Conflicting Case Holdings of Parker, Perfect 10, and Field

1. Parker and Field

Even though *Parker* and *Field* had holdings consistent with each other, the *Parker* court's reliance on *Field* was misplaced for two reasons. First, the cases were factually distinguishable because each of the plaintiffs argued copyright infringement at different stages in the search engine process.[109] Field argued infringement when a search engine user clicked on a link to his website displayed in the search engine results, not when Google was copying his website during the caching and indexing process.[110] On the other hand, Parker argued Google directly infringed during the initial copying, indexing, and caching of his posting to USENET.[111] Second, the *Parker* court did not address this critical difference even though the *Field* court indicated Google may have been found liable had Field claimed direct infringement during the initial copying, indexing, and caching process; i.e. exactly what Parker claimed in his suit. Nonetheless, the *Parker* court adopted the *Field* holding with little discussion.[112]

Interestingly, the *Parker* court also applied the DMCA safe harbor defense even though neither Parker nor Google raised the issue in their pleadings.[113] The *Parker* court listed the DMCA safe harbor provision in a footnote with no discussion of its applicability to Google.[114] Exactly like the *Field* court, the *Parker* court ignored the initial requirements for DMCA protections requiring ISPs to have "subscribers" or "account holders" as discussed previously in Section I(A)(2)(b).

---

the fair use exception. The court reaches this conclusion despite the enormous public benefit that search engines such a Google provide.
Id.

[109] *Compare* Field v. Google, Inc., 412 F. Supp. 1106, 1115 (D. Nev. Jan. 19, 2006) (stating Field was not claiming direct infringement when Google made the initial copy of his web site, instead he was claiming infringement when search engine users clicked on the cached link), *with* Parker v. Google, Inc., 422 F. Supp. 492, 498 (discussing Parker was claiming direct infringement when Google copied, indexed, and displayed a cached link of his USENET posting).

[110] *See Field*, 412 F. Supp. at 1115.

[111] *See Parker*, 422 F. Supp. at 496–97.

[112] *Id.* at 498.

[113] *See id.* at 497.

[114] *Id.*

2. Perfect 10 and Field

Similar to *Parker*, *Field* and *Perfect 10* were factually distinguishable from each other.[115] Ultimately, however, the resulting holdings were a result of many different factors. First, in *Field*, the court punished Field for his bad faith in manufacturing a claim against Google.[116] Bad faith was demonstrated because Field wrote the works displayed on his web site over a few days, knew about the Robot Exclusion Standard, and objected when Google offered to remove the content.[117] Conversely, Perfect 10 established a valid claim against Google because, in part, it suffered tangible losses to its bottom line and there was no evidence of bad faith.[118]

Second, Field was displaying his writings for free while Perfect 10 was displaying its photographs using a fee-based subscription service.[119] The *Perfect 10* court seemed to sympathize with Perfect 10 because of the huge monetary losses that Perfect 10 was suffering as a result of Google's conduct.[120] The *Field* court, on the other hand, noted that Field was displaying his works for free and, therefore, was not suffering any financial harm.[121]

Finally, both courts analyzed the commercial use factor of Google's fair use defense in radically different ways.[122] Essentially, the *Field* court minimized the financial impact that Field's web site made on Google's profits because his web site was only one of billions included in Google's search engine.[123] Accordingly, the inclusion or exclusion of Field's web site made no difference to Google or to its profits.[124] Conversely, the *Perfect 10* court found that Google benefited financially in

---

[115] *Compare* Field v. Google Inc., 412 F. Supp. 2d 1106, 1115 (D. Nev. Jan. 19, 2006) (stating Field was not claiming direct infringement when Google made the initial copy of his web site, instead he was claiming infringement when search engine users clicked on the cached link), *with* Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828, 838 (C.D. Cal. Feb. 17, 2006) (discussing Perfect 10 was claiming direct and secondary infringement when Google copied, distributed, and displayed its copyrighted images).

[116] *Compare Field*, 412 F. Supp. 2d at 1123 (discussing that Field manufactured his claims against Google), *with Perfect 10*, 416 F. Supp. 2d 832 (discussing Perfect 10 as a legitimate company that makes most of its money from the sale of its copyrighted images).

[117] *See Field*, 412 F. Supp. 2d 1123.

[118] *See Perfect 10*, 416 F. Supp. 2d at 832 (explaining Perfect 10 generates all of its income from the sale of copyrighted works, licensing, and subscriptions). When search engine users find Perfect 10's copyrighted images in the image search, Perfect 10 loses money because the search engine user can view the image for free, as opposed to paying Perfect 10 to view the image.

[119] *Compare Field*, 412 F. Supp. 2d at 1114 (discussing that Field displayed his works to the world for free), *with Perfect 10*, 416 F. Supp. 2d at 832 (discussing Perfect 10's web site subscriptions for $25.50 per month that allow user's to view Perfect 10's copyrighted images).

[120] *Perfect 10*, 416 F. Supp. 2d at 832 (explaining Perfect 10's substantial monetary losses due to Google's conduct).

[121] *Field*, 412 F. Supp. 2d at 1114 (discussing that Field initially displayed his works for free on the Internet).

[122] *Compare Field*, 412 F. Supp. 2d at 1120 (discussing that there is no evidence that Google profited from using Field's work because his works made up a small minority of billions that are stored in Google's database), *with Perfect 10*, 416 F. Supp. 2d at 846 (finding that Google derives "significant commercial benefit" from the Google Image Search).

[123] *Field*, 412 F. Supp. 2d at 1120 (listing the small financial impact that Field's web site has on Google's profits).

[124] *See id.*

several ways by including Perfect 10's images in its image search.[125] Unlike *Field,* the inclusion of Perfect 10's copyrighted images in Google's search engine image search was directly connected to Google's profits.[126]

B. Additional Policy Considerations

This section goes beyond the legal implications of the search engine caching process and explores public policy considerations relating to the advantages and disadvantages of the process. Section one explores the ways that search engines fulfill important public needs and further the goals of copyright. Section two discusses the negative policy implications associated with allowing search engines to copy web sites and to display cached links without the web site owner's express permission.

1. Benefits of Search Engines and Caching

Some of the main goals of copyright law are to promote creativity and to encourage the free flow of information to the public.[127] One view is that search engines fulfill the important goals of copyright through their services because they allow Internet users to find information on a wide variety of topics.[128] Without search engines, the only way for an Internet user to locate a web site is if the user types in the web site address directly into the browser.[129]

Beyond providing a valuable public resource, it appears that search engines comply with the Robot Exclusion Standard.[130] In addition, search engines promptly

---

[125] *See Perfect 10,* 416 F. Supp. 2d at 846.

[126] *See id.*

[127] *Compare* Whelen v. Jaslow, 797 F. 2d 1222, 1235 (3rd Cir. 1986) (explaining the ultimate goal of copyright is to balance the rights of copyright holders in such a way as "to create the most efficient and productive balance between protection (incentive) and dissemination of information, to promote learning, culture and development."), *with* NATIONAL RESEARCH COUNCIL, THE DIGITAL DILEMMA: INTELLECTUAL PROPERTY IN THE INFORMATION AGE 23 (National Academy Press) (2000) [hereinafter THE DIGITAL DILEMMA] (explaining that the Internet has a profound impact on the availability of information because of its size and depth and that it is "one of the world's largest libraries"). The idea that that the Internet is one of the world's largest libraries is directly in line with the specific goal of copyright of dissemination of information. *Id.*

[128] *See* Walker, *supra* note 43 (explaining that search engines provide a "roadmap" for Internet users to locate information and that Internet users would be severely limited in the amount of information they would be able to access without the aid of search engines).

[129] *See generally id.*

[130] *See* Field v. Google Inc., 412 F. Supp. 2d 1106, 1117 (D. Nev. Jan. 19, 2006) (stating that Google complies with the Robot Exclusion Standard and that Google removes web sites at the request of the owner). *But see* Search Engine Roundtable, *supra* note 80. Brett Tabke, the owner of Webmaster World (one of the largest web sites for webmasters), was interviewed after he decided to ban all web robots from visiting his web site. *Id.* When asked about why he thought that it would take a long time for search engines to remove his web site from the search engine results, he presented data to prove that his web site was still showing up in all search engines except one. *Id.* Specifically, his web site remained in Gigablast for 180 days, in Jeeves for 120 days, in MSN for 90 days, and in Slurp and Google for 60 days after he changed the Robot Exclusion Standard on his web site to disallow all robots. *Id.*

respond to take down notices if webmasters do not want their web site indexed or archived.[131] This compliance shows the search engines' good faith efforts to comply with industry standards, and shows that the search engines respect webmasters wishes.[132]

Lastly, archiving the Internet should be an important goal of the public,[133] similar to the archival purpose of the Library of Congress.[134] Archiving the Internet is essential and search engines are the obvious candidates to carry out this function because they are already performing these duties to some extent,[135] and already have the capabilities and resources to archive.[136]

2. Drawbacks of Search Engines and Caching

While there are many advantages to search engines and links to cached web sites, there are just as many arguments against search engines copying web sites and using cached links without the web site owner's express permission.

First, the fact that an industry has been utilizing something for many years does not necessarily mean the process or application is legal or it complies with the goals of copyright law.[137] Just because search engines rely on caching to provide fast service and accurate data, does not render the legal implications of their actions irrelevant.

Second, while some search engines respect the Robot Exclusion Standard, others do not.[138] Taking a web site down after it has been copied and displayed as a cached link is a solution that exists if the web site owner disapproves of the copying and

---

[131] *Field,* 412 F. Supp. 2d at 1122 (discussing that Google was willing to remove Field's web site from the search engine index and remove the web site from the cache after it had learned that Field did not want his web site to be included).

[132] *See, e.g., id.* at 1122–1123.

[133] *See* THE DIGITAL DILEMMA, *supra* note 127, at 98 (recognizing that preservation provides innumerable advantages to society).

[134] *See* The Library of Congress, http://www.loc.gov/about/ (last visited Apr. 6, 2006) (stating that the purpose of the library of Congress is "[t]o make its resources available and useful to the Congress and the American people and to sustain and preserve a universal collection of knowledge and creativity for future generations.").

[135] *See* Lyman, *supra* note 67 (standing for the proposition that search engines already are archiving the web for commercial purposes).

[136] *See* Ignatius, *supra* note 47, at 49. "Google maintains tens of thousands of servers to store all those cached web sites it searches." *Id.*

[137] *See, e.g.,* Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 125 S. Ct. 2764 (2005) (giving an example of a technology (peer to peer file sharing) that had been used for many years to enable others to illegally download music).

[138] *See* Search Engine Roundtable, *supra* note 80 (discussing the reasons why search engines have "fostered an era of robots.txt disrespect" and discussing the problems with "rogue bots" (robots that do not follow the Robot Exclusive Standard)).

caching.[139] This solution is plausible, except the removal does not impede the resulting damage caused from unwanted and rapid dissemination.[140]

Third, although information generally appears for free on the Internet, there are some authors who post on their web site information for limited amounts of time for a fee or subscription.[141] Cached links to a web site defeat the author's control of the work, and in some cases, defeat the author's control over the fees for the work.[142] Moreover, search engines could be perceived as helping hackers circumvent standard technical measures through cached links, which is illegal.[143] Also, a presumption can be made that if a web master changes something on a web site or removes the web site, he did this because he no longer wants people to see the information.[144] Existence of the web site in the search engine cache defeats this purpose and the web site owner's exclusive control over the copyrighted material.

Equally important, in *Field,* Google argued two of the three stated purposes for cached links were for archiving the web.[145] However, Google argued in the very same case that web site storage is sufficiently "intermediate and temporary" so that Google qualifies for DMCA safe harbor protection.[146] It is contradictory to argue that cached links serve archival purposes, but then only retain these archival copies for fourteen days.[147] Archiving the Internet is an important goal, but economic questions, technical questions, and legal questions as to who is in the best position to archive the Internet remain unanswered.[148]

---

[139] Field v. Google Inc., 412 F. Supp. 2d 1106, 1117 (D. Nev. Jan. 19, 2006) (explaining the process of removing a web site from the search results if the web site owner disagrees with the web site copying and caching).

[140] *See* THE DIGITAL DILEMMA, *supra* note 127, at 38 (explaining the differences between traditional publishing and Internet publishing of information and the speed at which information travels over the Internet).

[141] *See, e.g.,* Geek News, *Stephen King Goes Online, Again* http://www.geek.com/news/geeknews/q22000/pda200721001924 (last visited May 14, 2006) (Jul. 21, 2000) (giving an example of a well-known author who posts parts of his books on his web site in a pay per read program that requires readers to be honest about their reading time and to donate the correct amount to his web site).

[142] *See* Olsen, *supra* note 42.

[143] 17 U.S.C. § 1201 (2000).

[144] *See, e.g.,* David Whelan, *Google Me Not,* FORBES.COM (Aug. 16, 2004) http://www.forbes.com/forbes/2004/0816/102_print.html (discussing sensitive information that appears online and how this information is included in search engines, even after it has been removed).

[145] Field v. Google, Inc., 412 F. Supp. 2d 1106, 1112 (D. Nev. 2006). Google's reasons for allowing access to web sites through cached links included: Archival copies that are available when a search engine can not access the actual web site and archival copies that allow people to examine how a web site has been altered over time. *Id.*

[146] *Id.* at 1124 (accepting the Google argument that the timeframe of fourteen to twenty days for web site storage as sufficiently intermediate to qualify Google for the DMCA safe harbor provision).

[147] *See* Richard Koman, *How the Wayback Machine Works,* O'REILLY WEBSERVICES.XML.COM, Jan. 21, 2002, http://www.xml.com/pub/a/ws/2002/01/18/brewster.html. The Wayback Machine could make this argument much more than Google because The Wayback Machine actually archives copies of web sites for an unlimited amount of time and allows Internet users to compare any number of copies of the web site, not just one copy from the previous 14 days like Google. *Id.*

[148] *See* THE DIGITAL DILEMMA, *supra* note 127, at 208–09 (explaining the associated problems with archiving the Internet and suggesting that a task form should be created to address these problems and decide the details of archiving the Internet).

Finally, Google also argued in *Field* an additional advantage of cached links is that the search terms the user entered to find the web site show up highlighted on the cached version of the web site.[149] Because the same function can be performed in most web browsers using the "find" function, its value to Internet users is debatable at best.[150]

C. How Would a Court Address Caching?

No court has addressed the exact question of whether copying and caching of web sites by search engines is legal. Accordingly, this section presents a hypothetical situation as to how a court might address this problem. Assume the facts of *Field v. Google* are changed so that the plaintiff, Arthur ("Art") Author, is claiming direct infringement on the initial copy of his writings and photographs on his web site.[151] Art's web site has been visited by search engine spiders and is listed in the search results with a cached link. Moreover, Art requires a subscription to his web site to view chapters of his latest book. This section will analyze the requirements for direct copyright infringement and take a closer look at the defenses, as applied to Art's situation.

1. Direct Copyright Infringement

First, assume that Art can prove he owns the copyright to his writings and photographs, satisfying the first requirement to prove copyright infringement.[152] Next, regarding the second requirement, even if the search engines say their robots are only taking "snapshots," the end result is that the robot is making a copy of the entire web site. This copying appears to satisfy the second requirement for copyright infringement.[153] However, it is unclear if the volitional conduct aspect of the copyright infringement claim is satisfied.[154] Because the courts have found that the copying process cannot be automatic, it remains unclear if the conduct of the robots in the hypothetical is automatic, thus shielding the search engine from liability.[155]

There are three views courts can take when analyzing the volitional conduct requirement. First, a court could decide the volitional conduct requirement is

---

[149] *See Field,* 412 F. Supp. 2d at 1112 (discussing the third purpose of cached links as good for allowing fast identification of search query terms).

[150] *See e.g.,* Firefox Help, http://www.mozilla.org/support/firefox/menu (last visited Apr. 10, 2006) (giving an example of how to find search terms on a web site through a web browser through a non-cached version of the web site).

[151] This hypothetical presents a claim of direct copyright infringement during the initial copy of Art's web site as opposed to the claim made in Field v. Google, Inc., 412 F. Supp. 2d 1106, 1115 (D. Nev. 2006), when a search engine user clicked on a cached link to his writings.

[152] *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991). Recall the two requirements to prove direct infringement are ownership of a valid copyright and copying by the defendant of the main elements of the work that are original. *Id.*

[153] *See id.* (discussing the second prong as copying by the defendant).

[154] *See generally* Religious Tech. Ctr. v. Netcom On-Line Commc'n Servs., Inc., 907 F. Supp. 1361, 1367–69 (N.D. Cal. 1995).

[155] *Id.*

inapplicable to the hypothetical scenario because it was initially implemented to protect innocent service providers.[156] Examples of innocent service providers include copy machine manufacturers that provide the machine to do the infringing or ISPs that provide the software that automatically forwards messages to bulletin board subscribers.[157] As applied to Art's situation, the search engine is most analogous to the person using the copy machine to copy a book as opposed to an innocent service provider. Therefore, under this example, an active infringer would not be able to use volitional conduct as a shield from liability.

Second, the court could apply the volitional conduct requirement and find it is satisfied because the robots' conduct is automatic. Along the same reasoning, the robot programming could still qualify as a volitional act because people program the robots to copy. Using a different example, assume that a search engine programmed a robot to search the Internet and make copies of copyrighted songs or programmed its robots to download music illegally. Surely liability has not vanished simply because the search engine programmed a robot to do the dirty work for it, even though the robots illegal music downloading was automatic.

Finally, the court could apply the volitional conduct requirement and find it is not satisfied. The *Parker* and *Field* courts adopted this view and, accordingly, the search engines did not satisfy this requirement and were not liable for direct copyright infringement.

The best view, however, is to apply the volitional conduct requirement and find it is satisfied by virtue of the back end programming by the search engines. When someone controls a software program, or in this case, a robot, liability should not be lost. As a result, search engines should be held responsible for the actions of their robots because they directed the robots to perform the web site copying.

2. Defenses

This section assumes direct copyright infringement has been established on the initial copying of the web site and examines some possible defenses.

a. DMCA

Even though both the *Parker* and *Field* courts applied the DMCA to Google, applying the DMCA safe harbor provisions to search providers is difficult for a

---

[156] *See id.*

[157] *See, e.g., id.* The volitional conduct in *Religious Technologies* is distinguishable from the hypothetical because the court was reviewing a process where "[i]ncidental copies automatically made on [the ISPs] computers using [the ISP's] software as part of a process initiated by a third party . . . . [the defendant] did not take any affirmative action that directly resulted in copying of plaintiffs' works other than by installing and maintaining a system whereby software automatically forwards messages . . . and temporarily stores copies on its system." *Id.* at 1368.

variety of reasons.[158] First, the stated legislative history makes it clear that the DMCA safe harbors are to shield ISPs from liability based on infringing conduct of third parties.[159] The goal of Congress was to curb copyright infringement through cooperation of ISPs and copyright owners to prevent such conduct.[160] Unfortunately for search engines, the DMCA legislative history shows that the DMCA safe harbors were meant to protect ISPs from infringement committed by third parties, and not to protect them from claims for direct infringement.[161] It appears that Congress did not contemplate that the ISPs themselves would face claims of copyright infringement.[162]

Second, the threshold requirements to qualify for DMCA safe harbor protection mention "subscribers" and "account holders," neither of which seem to apply to search engines.[163] One interpretation suggests search engines are not eligible for DMCA safe harbor protection because they cannot meet the DMCA threshold requirements.[164] On the other hand, at least one commentator has suggested that search engines are covered by the DMCA safe harbors and simply do not have to meet the threshold requirements for eligibility.[165] Ultimately, qualifying for the DMCA safe harbors is problematic for search engines, and its applicability to the situation presented in this section is questionable.[166]

b. Fair Use

While the application of the DMCA safe harbor provision to search engines is tenuous, the fair use defense may be more appropriate.[167] The four fair use factors are: (1) the nature of the work, (2) the amount and substantiality of the work used, (3) the effect on the market for the value of the copyrighted work, and (4) the commercial nature of the use.[168] As to the first factor of fair use, the nature of Art's

---

[158] *See generally* Walker, *supra* note 43.

[159] *See* Ellison v. Robertson, 357 F.3d 1072, 1076 (9th Cir. 2004) (explaining Congress's intent behind the DMCA as a way to promote the cooperation between ISP's and copyright holders and to shield innocent ISPs from liability).

[160] *Id.*

[161] *See* Walker, *supra* note 43 (listing the DMCA safe harbors as a shield from claims of infringement by third parties).

[162] *See generally* DMCA History, *supra* note 5 (noting the absence of any discussion of search engines, themselves, being shielded from liability for claims of direct infringement against them).

[163] *See* Walker, *supra* note 43, at 40 (explaining that search engines do not have subscribers or account holders and that search engine users do not have to pay for the service).

[164] *See* Walker, *supra* note 43, at 40 (discussing some of the threshold requirements that use the terms "account holders" and subscribers).

[165] *See* Walker, *supra* note 43, at 40.

[166] *See* Field v. Google, Inc., 412 F. Supp. 2d 1106, 1123 (D. Nev. 2006). The court listed reasons that Field's complaint regarding the DMCA's safe harbor provisions application to Google was not properly presented. *Id.* Furthermore, note the court's failure to address the threshold requirements for eligibility and the court moving on to specific detailed provisions of the DMCA. *Id.* *See also* Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828, 838 (C.D. Cal. 2006.) (noting that Google brings up the DMCA as a possible defense, but the court does not address it at this stage of the claim).

[167] *Compare Field,* 412 F. Supp. 2d at 1123 (finding fair use defense applied to Google), *with Perfect 10,* 416 F. Supp. 2d at 851 (finding fair use was likely to not apply to Google).

[168] *See* Harper Row, Publishers v. Nation Enters., 471 U.S. 539, 560–69 (1985) (explaining the fair use factors).

work is primarily his web site, and secondarily, his writings and works appearing on the web site. However, the scope of fair use is expanded somewhat when published works are at issue.[169] Accordingly, the first fair use factor weighs slightly in favor of the search engines because Art's works are copied after they are published.

The second fair use factor is the amount and substantiality of the work used.[170] Search engines copy web sites in their entirety.[171] As a result, this factor weighs in Art's favor.

The third fair use factor is the effect on the market for the value of the copyrighted work.[172] This factor is difficult to assess in relation to the search engine copying the web site and including it in the index because data about the effects of this on a web site is difficult to find. Most commentators believe inclusion in search engines benefits web sites, as opposed to harming the market for the web sites.[173] However, the "benefits" to be included in a search engine may be overstated, because statistics suggest only nine to thirteen percent of web site visitors come from search engines.[174] Potentially, web site owners gain benefits by being included in search engine results, and the effect on the market for the value of the works is not harmed. Accordingly, this factor weighs in favor of the search engines.

The third fair use factor, as analyzed relating to cached links, however, has a different outcome. In Art's situation, cached links allow web site visitors to bypass his registration system and therefore directly impact the market for his works. Because cached links directly affect the market for Art's work, this factor weighs in favor of Art. In sum, the third fair use factor outcome may be different depending on the step of the search engine process that the court reviews.

Finally, the last fair use factor is the commercial nature of the use. As mentioned in Part II(A), the *Field* and *Perfect 10* courts looked at this factor from extraordinarily different viewpoints.[175] A court could apply the commercial use factor as applied in *Field* and decide that Art's web site is only one in billions within the search engine and therefore has no effect on the commercial success of the search engine.[176] Alternatively, the court could apply the commercial use factor utilized in *Perfect 10* and decide that Art's web site is directly connected to the search engine's financial success.[177]

---

[169] *See Harper Row, Publishers*, 471 U.S. at 563 (explaining the scope of fair use is "narrower with respect to unpublished works.").

[170] *See id.* at 564–65.

[171] *See* Danny Sullivan Posting, *supra* note 66.

[172] *See Harper Row, Publishers*, 471 U.S. at 566.

[173] *See* Danny Sullivan Posting, *supra* note 66.

[174] *See* Sullivan Interview, *supra* note 73 ("Studies show sites often get 9–13% of their traffic from search engines. Plenty of people scream and yell if they get dropped from search results. Anecdotally, far, far more people are concerned about being in search engines than staying out of them.").

[175] *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578 (1994) (stating the first factor of fair use is "the purpose and character of the use, including whether such use is of a commercial nature, or is for non-profit educational purposes"). An additional consideration is if the new work is transformative. *Id.*

[176] *See* Field v. Google, Inc., 412 F. Supp. 2d 1106, 1120 (D. Nev. 2006). The hypothetical court would be applying the reasoning of the *Google* court in this case. *Id.*

[177] *See* Perfect 10 v. Google, Inc., 416 F. Supp. 2d at 846 (C.D. Cal. 2006). The hypothetical court would be applying the reasoning of the *Perfect 10* court in this case.

The better application of the commercial nature of the use factor is the *Perfect 10* view, because search engines' products and revenues are derived directly from copying web sites. Additionally, search engines provide cached links as an extra service to attract more users, and thus, more advertising money. The fastest search engines with the most relevant results attract more users and more advertising. Without web sites, search engines would cease to exist.

Ultimately, the fair use defense outcome is determined according to how much weight a court gives each factor. The outcome may be different depending on which step of the search engine caching process the court is evaluating. So even though the entire existence of search engines rests on copying web sites, courts must consider that search engines serve a purpose that furthers access of information and ultimately, promotes the goals of copyright. The extent search engines actually further access to web sites is a critical element in a court deciding to apply fair use. Although, web sites receiving less than fifteen percent of its traffic from search engines does not seem like a strong enough number to warrant the *carte blanche* given to search engines.

Fair use applicability to cached links is much less likely than the applicability to the initial web site copying. Importantly, cached links do not further copyright goals of providing access to information, except in limited situations where data has been removed or is unavailable from a web site. Accordingly, this example illustrates that all of the fair use factors favor Art and web site owners in general. Therefore, fair use should not be granted for search engines for cached links.

c. Implied License

An implied license[178] is another defense that could be used against a claim of copyright infringement.[179] Google successfully used the implied license defense in *Field*.[180] The court indicated that Field's knowledge of the Robot Exclusion Standard in addition to his failure to indicate a preference to be excluded in the robot searches constituted an implied license.[181] While the application of an implied license defense was straightforward in *Field*, the application is not so straightforward in Art's situation. The evidence Google presented in *Field* indicated that most web masters know about the Robot Exclusion Standard.[182] Nevertheless, a court would be unable to find an implied license if the web master did not know about the Robot Exclusion

---

[178] *See* Effects Assocs. Inc. v. Cohen, 908 F. 2d 555 (9th Cir. 1990) (explaining that a license may be granted orally or can be implied by conduct); *see also Field*, 412 F. Supp. 2d at 1116 (stating that an implied license can be assumed when the other party reasonably believes that the owner consented to the use of the copyrighted material).

[179] *Field*, 412 F. Supp. 2d at 1116 (explaining that an implied license was available as a defense because Field knew about the Robot Exclusion Standard and failed to implement it in his web site, thereby, granting Google an implied license).

[180] *Id.*

[181] *Id.*

[182] *Id.* at 1112.

Standard.[183]   Simply, because Art does not know about the Robot Exclusion Standard, the implied license defense will not suffice.

III. PROPOSAL

Given the rate that the Internet has expanded, intellectual property challenges in cyberspace have multiplied at increased at equally fast rates. This rapid expansion has left traditional copyright law far behind. This presents a new problem for copyright holders, because the scope of the Internet allows for fast dissemination of copyrighted material.[184] While some efforts have been made by Congress and volunteer standards groups, these efforts have not adequately protected copyright holders. For example, the Robot Exclusion Standard has been suggested and partially implemented, but compliance is sporadic[185] and is not overseen by any official body.[186] Additionally, this standard requires webmasters to know how to program HTML code to protect their copyrighted material.[187] To compound the problem, courts have taken divergent views in the Google cases[188] and leave no clear direction for copyright holders or search engines. It seems apparent the current structure is not the optimal situation for copyright holders and search engine users. This section proposes a unique balance between the need for search engines and the protection of copyright holders' interests.

A. Governing Body and Regulation

One of the major obstacles in Internet copyright protection is the lack of an Internet governing body.[189] Policies can be suggested by different volunteer

---

[183] It is impossible for a court to find an implied license if the web site owner does not know about cached links or the Robot Exclusion Standard. *See, e.g.,* Field v. Google, Inc., 412 F. Supp. 2d 1106, 1116 (D. Nev. 2006) (copyright owner knows of use and remains silent).

[184] *See* Whelan, *supra* note 144 (discussing how information can be propagated by 150 search engines and, in turn, create massive problems if the web site owner does not want this information disseminated).

[185] *See, e.g.,* Search Engine Roundtable, *supra* note 80 (discussing the problem of "rogue bots," which are robots that ignore the Robot Exclusion Standard). One of the negative effects of rogue bots is that they seriously diminish system performance. *Id.* As a result, web site owners spend unnecessary time and money dealing with these rogue bots. *Id.*

[186] *See* A Standard for Robot Exclusion, *supra* note 75 ("It is not an official standard backed by a standards body, or owned by any commercial organization. It is not enforced by anybody, and there no guarantee that all current and future robots will use it.").

[187] The Google Cache, http://www.thegooglecache.com/ (last visited Mar. 8, 2006) (discussing how the burden should be on search engines to gain permission from web site owners before the copying process as opposed to web site owners having to know HTML programming to revoke permission after the copying has already taken place).

[188] *Compare* Field v. Google, Inc., 412 F. Supp. 2d 1106, 1109 (D. Nev. 2006) (finding no liability against Google), *with* Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828, 859 (C.D. Cal. 2006) (finding liability against Google).

[189] *See* Michael J. Brady, Lawrence O. Monin, Curtis R. Tingley & Timothy L. Skelton, *The World Wide Web and the New World of Litigation: A Basic Introduction,* 66 DEF. COUNS. J. 497 (1999). Three groups set policies and technical standards for the Internet on a volunteer basis. *Id.*

organizations, but until these policies are enforced, any and all proposed standards will remain optional. The sheer scope of the Internet makes this a daunting task. Ideally, the organization or group responsible for deciding standards for the Internet should take the concerns of all parties into consideration when setting these standards. In the past, parties with the most clout and power directed and promulgated Internet standards.[190] While a decision about who should govern the Internet is outside of the scope of this comment, it is important to understand that Internet governance and enforcement is a major obstacle to any solution to the problems discussed throughout.

The Robot Exclusion Standard provides a good illustration of how a typical Internet standard is developed. Also, it shows why the need for input from all parties and enforcement of the Robot Exclusion Standard are critical components that are missing from the current structure. The Robot Exclusion Standard was developed with input from robot authors and was initially brought about because of web site owners wanted to block robots that were causing problems with their web sites.[191] It is unclear how much influence each of the parties had in the decision about the Robot Exclusion Standard, but it is clear that the standard was written to benefit robot authors.[192] It put the burden on web site owners to deny the robots access as opposed to the robots asking permission prior to copying.[193] Specifically, the Robot Exclusion Standard requires millions of web site owners to block robot access as opposed to requiring thousands of robot owners to make changes during the robot development.

Until there is an authoritative body that can enforce these standards, any solution proposed is subject to the same optional treatment that exists now. Input from all parties is crucial to insure that the standards are fair and that the standards reflect the interests of everyone. It is within this framework that we examine the other facets of copyright protection on the Internet and possible solutions.

The Internet Society coordinates International cooperation. *Id.* The Internet Architecture Board sets hardware and software standards. *Id.* Finally, the Internet Engineering Task Force facilitates communications between different network administrators and owners. *Id.* While all of these organizations are helpful to setting guidelines, none of them have the authority to enforce their recommendations. *See* THE DIGITAL DILEMMA, *supra* note 127, at 209.

[190] *See* Search Engine Roundtable, *surpra* note 80 (discussing how during the development of Internet standards, search engines have dictated the Robot Exclusion Standard in the same way that Microsoft and Netscape manipulated HTML standards to benefit themselves).

[191] *See generally* A Standard for Robot Exclusion, *supra* note 75. The Robot Exclusion Standard was initially developed to deal with rogue bots. *Id.* The paper that discusses the Standard was a compilation of comments and discussion from robot authors, people with an interest in robots, and members of the Technical World Wide Web Mailing List. *Id.* The Robot Exclusion Standard's origin seemingly came from web authors who were concerned about rogue bots, yet the discussion was influenced by robot authors. *Id.* The goal of robots is to visit and copy information from web sites, so for robot authors, developing an opt-in system would have made it much more difficult for their robots to function. *Id.*

[192] *See* A Standard for Robot Exclusion, *supra* note 75.

[193] *See* A Standard for Robot Exclusion, *supra* note 75.

B. Blanket Licenses

This proposal presents a two-fold solution to search engines copying and caching links balancing the rights of both search engines and web site owners. First, Section 1 discusses a solution to the initial search engine copying. Second, Section 2 explores a solution regarding search engines displaying cached links.

1. Search Engines Initial Web Site Copying

I propose search engines enter into aggregate blanket license agreements with the Internet Corporation for Assigned Names and Numbers ("ICANN").[194] ICANN is the appropriate organization to perform this function, because it oversees the entire domain name system.[195] As part of the binding agreement between registrars and domain name registrants, the search engine rights to initially copy a website for inclusion in search engine results would be automatically included.[196] ICANN is well suited to act as an intermediary between the search engines and website owners because all domain names are routed through ICANN.[197]

Similar blanket agreements were negotiated in an analogous situation. In *Broadcast Music Inc. v. Columbia Broadcasting System*,[198] the music industry had a problem similar to web site owners' current problems and devised a solution of blanket licensing. Because of the makeup of the music industry, there were many artists who owned copyrighted material, but the logistics of the individual licensing and enforcement made it impossible for television networks and others to work out license agreements with each individual artist.[199] As a result, two organizations formed to oversee the blanket licensing agreements.[200] Blanket licenses are appropriate when there are thousands of potential users for the copyrighted material, thousands of copyright owners, and millions of copyrighted works.[201] The plaintiffs claimed an antitrust violation based on the blanket license, but the Supreme Court determined a blanket license was appropriate and legal.[202] As part of its reasoning, the Court detailed other examples of blanket licenses, including compulsory blanket licenses for secondary transmissions by cable television systems and compulsory licenses for jukeboxes.[203] Finally, the Court recognized that

---

[194]   *See* ICANN, http://icann.org/faq (last visited Oct. 30, 2006) (explaining ICANN's responsibilities pertaining to web site domain names).

[195]   *See id.*

[196]   *See, e.g.,* Sallen v. Corinthians, 273 F. 3d 14, 20 (1st Cir. 2001) (explaining "registrants must accept the UDRP's terms in order to register a domain name . . . .").

[197]   *See* ICANN, http://icann.org/faq (last visited Oct. 30, 2006) (ensuring "that every address is unique and that all users of the Internet can find all valid addresses.").

[198]   441 U.S. 1 (1979).

[199]   *Id.* at 5 (explaining "as a practical matter it was impossible for the many individual copyright owners to negotiate with and license the users and to detect unauthorized uses.").

[200]   *Id.*

[201]   *Id.* at 20.

[202]   *Id.* at 24 (the "blanket license has provided an acceptable mechanism for at least a large part of the market for the performing rights to copyrighted musical compositions . . . .").

[203]   *Id.* at 15.

Congress has created explicit statutory authority for copyright owners to collect royalties through licenses.[204]

Applying the model set forth in *BMI*, when a person registers a domain name, she would have the opportunity to decide if she wanted her website included in search engines. If she decided she did not want to be included, she could opt-out and choose not to receive the license fee proceeds. ICANN could compile a list of the number of domain names to be included, and this would determine the amount of the license agreement fee. Each search engine would negotiate reasonable rates for the inclusion of all of the domain names that elected to be included. Copyright royalty judges,[205] or a similar neutral party, could oversee these negotiations to make sure both parties received a fair agreement. Once an agreed rate was determined, this money would be either distributed to each domain name owner or alternatively, would be reflected in a lower cost of keeping a domain name registered. Each website owner would be compensated for the copying of their material for inclusion into search engines. Not only would this model compensate web site owners, but blanket licensing would lower the cost of registering and maintaining web site registration. As a result, the goals of copyright would be advanced because more people would be encouraged to register web sites, and thus, create new works.

Additional support for a blanket license is bolstered by evidence that Google has already entered into agreements with some news agencies to copy and display online news articles.[206] These agreements show that Google values these contributions to their search engine and rewards the authors accordingly. Google and other search engines should value other non-news websites as much as news stories. Google has recognized the value of including these news sources in their search results and should recognize there are many other valuable web sites that should receive similar appropriate treatment and compensation.[207]

Detractors from this proposal may argue that website owners receive enough compensation in the form of benefits from search engines by driving Internet traffic to a website. As the argument goes, people would never find websites without search engines. However, this argument breaks down, because many websites are successful without people using search engines to find the website. For example, most people know www.ebay.com, www.cnn.com, www.weather.com, or www.yahoo.com without ever using a search engine.

Additionally, detractors could argue that search engines cannot afford to compensate website owners *and* stay in business at the same time. A quick look at the potential costs to search engines compared to search engine revenue show a different story though. By 2010, the search engine industry is predicted to exceed

---

[204] *Id.* at 15–16.

[205] *See* 17 U.S.C. § 118 (2000) (explaining how copyright owners may negotiate royalty fees); *see also* Copyright Royalty and Distribution Reform Act of 2004, Pub. L. No. 108–419, § 801, 118 Stat. 2341, 2342 (2004) (creating Copyright Royalty Judges).

[206] *See* Caroline McCarthy, *Google reveals payment deal with AP*, NEWS.COM (August 3, 2006) http://news.com.com/2100-1038_3-1024234.html (last visited Nov. 2, 2006) (explaining the deal between Google and the AP that allows Google to use the AP's news stories and pictures in exchange for financial compensation).

[207] *See id.* "Google has always believed that content providers and publishers should be fairly compensated for their work so they can continue producing high-quality information . . . ." According to Google's statement, wouldn't web site owners be considered "content providers"?

revenues of twenty-three billion dollars.[208] At this level of revenue, no matter what royalties were agreed upon between web site owners and search engines, search engines would be readily able to compensate website owners without great financial hardship. The ongoing expense of maintaining the license would be small compared to the profits of the search engines. Overall, the web site owners would be compensated in exchange for their permission for search engines to copy and use their web site for profit.

2. Opt-In for Cached Links

Cached links are problematic for web site owners for many reasons and provide few appreciable benefits to web site owners.[209] Consequently, there should be an opt-in standard for the search engine to display a link to a cached version of the web site.[210] Simply, the web site would initially default to opt-out for search engine to post cached links, but the web site owner could opt-in, if he approved of the search engines displaying a cached link to his web site.

Various factors support this solution. First, an opt-in system for cached links defaults towards protecting copyright as opposed to allowing hackers to usurp registration or masking the origin of visitors so web sites lose money. Cached links do little to further the access of information and there are many more disadvantages to cached links. As a result, copyright protection should be the foremost priority and the opt-in system should be implemented.

Second, Google and other search engines should not be afraid to implement the opt-in system because it is already recognized in a different context. Google Print, Google's effort to create an index of books, recognizes a partial opt-in system.[211] Web site owners opt-out for the indexing part of the process, but opt-in to have their books displayed with a cached link.[212] There should be no reason why Google treats the text of books any differently than the text of web sites. Web sites are as important as any other copyrighted work and retain all of the exclusive rights that come with copyrights.

IV. CONCLUSION

The rapid expansion of the Internet has created copyright problems that have not been addressed before. The search engine caching process involves an initial step of copying a web site without express permission from the web site owner. This copy is used to create a large index, which in turn, is used to generate search results for

---

[208] Gord Hotchkiss, *Search Revenue to Reach 23 Billion by 2010*, SEARCH ENGINE LOWDOWN.COM, http://www.searchenginelowdown.com/2005/02/search-revenue-to-reach-23-billion-by.html (last visited Nov. 1, 2006).

[209] *See* discussion *supra* Section II(B)(2) (explaining the problems with cached links).

[210] *See* Danny Sullivan Posting, *supra* note 66 (noting that the hybrid opt-in/opt-out standard is currently being utilized in the Google Print system).

[211] *Id.* (comparing Google Print's opt-in for caching to regular Google's opt-out for caching).

[212] *Id.*

millions of people.   Search engines also display cached links unless the web site owner opts out.

The assumption that these processes are legal has continued for many years with few legal challenges.   Currently, the responsibility to opt-out of the initial copying and caching links is delegated to the web site owner.   These search engine practices conflict with the copyright owner's exclusive rights.

The future is not uncertain though.   An appropriate solution to these problems presented itself in *BMI v. Columbia*.   The *BMI* factors are applicable to the Internet and provide a model with which search engines and web site owners can work together for the benefit of everyone.   Copyrights are protected and search engine results are comprehensive and fast.

Further, an opt-in model is desirable for cached links because cached links have minimal benefit and enormous cost.   Web site owners lose subscription money through registration bypass, and advertising money through origin masking.   The benefits to cached links are minimal and debatable.   Consequently, cached links should be opt-in for web site owners who truly desire their web site to be stored and displayed as a cached link.

Implementing these proposals creates a good situation for everyone.   Web site owners receive compensation for use of their copyrights and search engines comply with copyright law and are able to continue their profitable business models.