

Massachusetts Institute of Technology

Updating Copyright Laws to Address Concerns of Google's Cached Page Service

**6.806
Ethics and Law on the Electronic Frontier
Professor Hal Abelson**

**By
Parul Deora**

December 10, 2003

Table of Contents

<u>1.0 Introduction</u>	1
<u>2.0 Current Copyright Laws</u>	1
<u>2.1 United States Code Title 17</u>	1
<u>2.2 The Digital Millennium Copyright Act (DMCA)</u>	1
<u>2.3 Online Copyright Infringement Liability Limitation Act (OCILLA)</u>	2
<u>3.0 Google</u>	2
<u>3.1 Google’s Goals and Technology</u>	2
<u>3.1.2 High Precision</u>	2
<u>3.1.2 Academic Development</u>	2
<u>3.1.3 Usability</u>	2
<u>3.1.4 Support Novel Research</u>	3
<u>3.2 Cached Page Service</u>	3
<u>3.3 Copyright Concerns</u>	3
<u>3.4 Google and Copyright Laws</u>	4
<u>3.4.1 Fair Use</u>	4
<u>3.4.1.1 Purpose and Character of the Use</u>	4
<u>3.4.1.2 Nature of the Work</u>	5
<u>3.4.1.3 Amount and Substantiality of the Portion Used</u>	5
<u>3.4.1.4 Effect of the Use on the Work’s Potential Market or Value</u>	5
<u>3.4.2 Safe Harbors</u>	5
<u>3.4.2.1 Transitory Digital Network Communications</u>	5
<u>3.4.2.2 System Caching</u>	6
<u>3.4.2.3 Information Residing on Systems or Networks & Information Location Tools</u>	6
<u>4.0 Comparable Issues</u>	7
<u>4.1 Kelly v. Arriba Soft Corporation 336 F.3d 811 (9th Cir. 2003)</u>	7
<u>4.1.2 Arriba Soft’s Legal Challenges</u>	7
<u>4.1.2 Arriba Soft and Google</u>	7
<u>4.2 Internet Archive’s Wayback Machine</u>	8
<u>4.2.2 Internet Archive’s Legal Challenges</u>	8
<u>4.2.2 Internet Archive and Google</u>	9
<u>5.0 Suggested Update of Copyright Law</u>	9
<u>6.0 Conclusion</u>	10
<u>7.0 Endnotes</u>	12

1.0 Introduction

The existence of Google, possibly the world's most valued and preferred search engine, may be threatened due to its susceptibility to copyright infringement charges. As an increasing number of individuals are turning to the Internet as not only their primary source of information, but their only source, we have become more concerned of what information is available to us and how accessible it is online. One of Google's features designed to make knowledge easily available is its cached page service. By taking snapshots of web pages in their entirety, and then making those cached copies available to users, Google is providing Internet researchers and web-surfers the ability to access materials that are not currently available or may no longer be online. Google, however, is facing copyright concerns for its cached page service. By copying entire web pages, and then making them available on the Internet to its millions of users, Google may be violating copyright laws that specify reproduction and distribution as the exclusive rights of copyright owners.

The United States Code Title 17 enumerates what rights belong to copyright owners and what exemptions can be made for fair use purposes. An analysis of Google and the factors that determine whether an application constitutes fair use have led me to believe that the cached page service does not qualify as fair use and is susceptible to copyright laws. After reviewing the safe harbors enumerated in Section 512 of the Digital Millennium Copyright Act, I have concluded that it is ambiguous whether Google's cached page service would be protected from copyright charges. Other online services such as the Internet Archive, which faced similar legal challenges, and the case of *Kelly v. Arriba Soft*, 77 F. Supp. 2d 1116 (1999), are unlikely to serve as precedents for Google if it is brought to court for copyright infringement. Its means of caching websites to provide users with better performance do not match up with the characteristics of the Internet Archive or the aforementioned case. Only by updating digital copyright law to protect online services that cache for the purpose of providing better service, while allowing opt-out options for publishers, will we be able to maintain the existence of Google and its cached page feature.

2.0 Current Copyright Laws

Comprehending the complexity of current copyright laws is the first step to understanding copyright infringement and assessing the legality of Google's cached page service. The traditional copyrights and fair use exemptions included in Title 17 of the United States Code, and the Digital Millennium Copyright Act, specifically its provision titled the Online Copyright Infringement Liability Limitation Act, are the main elements of copyright law that need to be understood for this issue. Only by realizing the meaning of these components will we be able to properly update the laws to unambiguously exempt features such as Google's cached page service from copyright infringement.

2.1 United States Code Title 17

The copyright laws in Title 17 of the United States Code enumerate what can be copyrighted, the exclusive rights belonging to copyright owners, and the fair use limitations on those exclusive rights. Of importance in this report are the latter two categories: the exclusive rights and the fair use limitations. Section 106 of Chapter 1 states the following as what the copyright owner has the exclusive right to do and authorize: (1) reproduce, (2) prepare derivative works based upon the copyrighted material, (3) distribute, and in some cases (4) perform or display the copyrighted work.^[1] What further complicates the copyright laws are the exceptions to these rights, the fair use limitations.

Individuals may exercise what is normally considered an exclusive right, without the permission of the copyright owner, if the work is being used for purposes such as “criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research.”^[2] The factors that determine whether the employment of a copyrighted work is fair use include the following:

- (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.^[3]

Back when making a copy required a deliberate effort, it was much easier to identify copyright infringement. The vast developments in technology since Title 17 was initially put into effect, however, have obscured what constitutes fair use and have led to progressive updates in copyright law.

2.2 The Digital Millennium Copyright Act (DMCA)

The Digital Millennium Copyright Act of 1998, the DMCA, is an update to copyright law due to the increasing availability of the Internet and “the ease with which digital works can be copied and distributed worldwide virtually instantaneously.”^[4] Copyright owners sought protection against the substantial piracy that could take place if they published their works on the Internet. The DMCA is the result of content providers advocating better protection of their “copyrights in the digital world.”^[5]

Not fully satisfied, online service providers lobbied for further protection from claims of copyright infringement, leading to additional provisions for the DMCA.

2.3 Online Copyright Infringement Liability Limitation Act (OCILLA)

One of the provisions added to the DMCA was Section 512, the Online Copyright Infringement Liability Limitation Act (OCILLA). Online service providers received four safe harbors to limit their liability for copyright infringement. Though not a comprehensive list of all defenses available to online service providers, Congress provided the following four safe harbors:

- (a) Transitory digital network communications
- (b) System caching
- (c) Information residing on systems or networks at the direction of users
- (d) Information location tools^[6]

While the safe harbors do not free service providers from copyright infringement liability, they do limit the monetary penalty one would pay if found guilty of infringement. Service providers can qualify for the aforementioned limitations by: (1) adopting and reasonably implementing “a policy of terminating in appropriate circumstances the accounts of subscribers who are repeat infringers;”^[7] and (2) accommodating and not interfering with standard technical measures as defined in the text of Section 512.

The copyright laws detailed in the previous sections do not make it clear whether Google’s cached page service would be exempt from copyright liability. A background on Google and an understanding of its cache feature will clarify why this is so.

3.0 Google

Sergey Brin and Larry Page launched Google in the Fall of 1998 with one specific mission: “to organize the world’s information and make it universally accessible and useful.”^[8] By 2003, Wired Magazine stated that Google.com entertained more than 28 million visitors each month, and that four out of five web searches occurred on Google or other sites that license its technology.^[9] Google currently receives more than 200 million search queries per day, more than half which come from outside of the United States.^[10]

The great popularity of Google is attributable to PageRank™, its system that ranks web pages.^[11] In addition to this unique feature, however, Google offers many more services and tools to its users. One such technology is its cached page service. By crawling the web and taking snapshots of web pages, Google is able to offer users links to cached sites in case the original website is unavailable. Despite its many benefits, Google is facing possible copyright concerns for its cache technology. By gathering snapshots of web pages in their entirety, and then making them available to its millions of users on the Internet, Google may be violating copyright laws that specify reproduction and distribution as the exclusive rights of copyright owners. An assessment of Google’s cache technology, its potential copyright concerns, and its relation to current copyright laws, will help us better understand its legality.

3.1 Google's Goals and Technology

In a document written by the founders of Google themselves, back when they were first introducing it, Brin and Page specified their four intended goals.^[12]

3.1.2 High Precision

Their primary goal was to improve the quality of web search engines. As the number of pages on the Internet increase, while the number of search results users view remains relatively constant, high precision becomes extremely important. Google differentiates itself from other search engines that mostly rely on manually maintained lists of popular topics or keyword matching, by utilizing the “additional structure present in hypertext to provide much higher quality search results.”^[13] At the heart of Google's software is its PageRank™ algorithm. Brin and Page defined a web page's PageRank as an “objective measure of its citation importance that corresponds well with people's subjective idea of importance.”^[14] This system allows search results to be prioritized to the query's keyword searches, allowing for greater precision.

3.1.2 Academic Development

When Google was first launched, there was still very little academic research on search engines available. Brin and Page were the first to publish a detailed description of a large-scale web search engine. They hoped that Google would provide a greater push for more development and academic understanding of search engines.

3.1.3 Usability

Another key goal was to build a search engine that many people could actually use. This not only included ease of use, but also the issue of scalability. As the Internet expands to include millions upon millions of web pages, search engines must be capable of scaling to the increasing number of resources available on the web. Google's index is currently comprised of more than three billion web pages,^[15] and is surely scalable to many more.

3.1.4 Support Novel Research

Brin and Page's final goal was to support novel research activities on large-scale web-based information. Google pursues this goal by storing the Internet's data in a compressed form so that researchers can quickly process a large amount of online data, and produce results that would have otherwise been much more difficult to generate.

Often times during research or general web searching, however, users encounter links that lead them to “Page Not Found” sites. The web pages are no longer online or their site's servers are unavailable. Google provides users the ability to still access such links via their cached page service.

3.2 Cached Page Service

Google's cached page service enables users to view snapshots of web pages from their cache, appearing as they looked when they were crawled by the system for indexing.

While crawling the web, Google downloads each and every page and analyzes it to determine the page's relevance for its PageRank™ feature. By taking a snapshot of the page, Google captures the state of the site at that moment in time. While many programs, such as browsers, cache recently accessed sites to minimize retrieval times in the near future, Google makes backups of *every* page, before it has even been requested. It then makes these snapshots available to users as cached links when that site is returned as the result of a search query. The cached page can be accessed by clicking on the “Cached” hyperlink near each search result, as indicated in Figure 1.

[Google cache raises copyright concerns | CNET News.com](#)
Google cache raises **copyright** concerns | Google's caching feature allows for quick snapshots of others' web pages. But does it infringe ...
[news.com.com/2100-1038_3-1024234.html](#) - 40k - Dec 4, 2003 - [Cached](#) - [Similar pages](#)

Figure 1: Cached Link. The cached page of a website can be accessed by clicking on the “Cached” hyperlink located near the bottom of most sites returned as search results.

Such a service is useful to users when: (1) the original site is unavailable, (2) they want to narrow in on the part of the page relevant to their search query, and (3) they lose the code to their own web pages. When an original site is unavailable due to Internet congestion or server problems, the user can still view the page's data via the cached link. Such a feature is very useful in regards to retrieving outdated online magazine articles that no longer exist. This allows users to continue their research or web surfing in a more time-efficient manner. An additional benefit, as a Google spokesman noted, the cached pages have the search query terms highlighted “to make it easier for users to find relevant information.”^[16] There have been several documented cases where website publishers accidentally deleted their own code or index files off their computers and used Google's cache to retrieve their web page. Dylan Tweney discovered Google's web page recovery service when he accidentally deleted his index.html file by mistake, causing his home page to appear as a bare list of files. When he realized he did not have a local copy, and searching for a backup would have taken ages, he Googled his home page's URL and found the cached old home page. By viewing the source of the cached version, he was able to find the original code, paste it into a new document, and restore his original home page without starting from scratch.^[17]

Despite the many benefits of Google's cache technology, it has a few snags too. Though web pages that are unavailable can be viewed through the cached link, the cached page may not have the most up-to-date information. The executive producer for ABCNews.com, Randy Stearns, is concerned that readers may access information that is not up-to-date, and may include errors that had been fixed on the original site, but were not on the archived pages.^[18] Michael Godwin, staff counsel for the Electronic Frontier Foundation, believes those are risks publishers take when placing information on the Internet. “By putting something on the Web, you're authorizing the world to look at it. By taking it down, you're taking the risk that someone might use the old data.”^[19] Google, however, took the safer route; they have a header appear at the top of every cached page to remind users that it may not be the most recent version of the page. This way, users are made aware that they may not be viewing the most current information and are not deceived into mistakenly collecting inaccurate data. Additionally, the cached

page is likely to contain very useful information, and is, thus, still a better option than the user not being able to access the information at all. Other potential drawbacks of Google's cache fall in the category of copyright infringement.

3.3 Copyright Concerns

By taking snapshots of entire web pages and then making links to those snapshots available to users without the permission of the copyright owners, Google has been treading the fine line of copyright infringement. Though users have been enjoying the service for years, recent copyright complaints by web publishers have brought the legality of Google's cached page service into question. In March of 2003, Microsoft Corporation submitted an official takedown notice to Google due to their product key being available on a site in Google's cache.^[20] Similarly, in August of 2003, CNET submitted an official notice and takedown request regarding Google's copying and use of copyrighted content available from CNET's website.^[21] Though Google ultimately removed the copyrighted materials from their cache, who's responsibility is it to assure that such copyrights are not infringed? Should copyright owners be forced to police the Internet to assure that their rights are not being infringed, or is it Google's responsibility to not involve itself in copyright violations? Above all else, is Google's method of copying and providing information truly even copyright infringement?

Originally, the purpose of copyright law was to protect those who invent or develop tangible works, so to ultimately promote the sciences and useful arts. This would imply that the protection of works is the primary goal, and that copyright owners should have the choice to share their exclusive rights rather than be forced to go about protecting them. In the reasoning in *Whelan v. Jaslow*, 797 F.2d 1222 (1986), the judge stated that, "We must remember that the purpose of the copyright law is to create the most efficient and productive balance between protection (incentive) and dissemination of information, to promote learning, culture, and development."^[22] As technologies have developed, however, it has become increasingly difficult to balance the benefits of new technological services with the protection of copyrights. Godwin believes that once an individual puts information up on the Web he or she is implicitly authorizing its reproduction, since the Internet functions by the use of copies.^[23] Publishers, however, are likely to disagree and would prefer an opt-in policy, where they choose to be included in Google's cache or not. Since Google's technology automatically caches every web page though, seeking permission from individual sites that contain copyrighted information is not possible. Sacrificing Google's cache feature in the name of copyright protection, however, would eliminate the several advantages of the tool described in Section 3.2.

There are, however, several opt-out options available to publishers who do not want their sites cached. They may notify Google of the infringement activity, and Google will remove the link as requested, as in the aforementioned cases of Microsoft and CNET's complaints. Additionally, Google respects the several mechanisms available through HTTP that disable a site's ability to be cached. Webmasters can prevent the caching of their sites by placing a "noarchive" meta in the header of each page. By encoding the robots.txt file into one's website, a publisher can be assured that Google's robots will not be able to crawl and cache their site.

Publishers who may be especially interested in utilizing opt-out options are registration-only sites such as many online newspapers. Why would individuals pay to access online news articles when they could read them for free off of Google's cache instead? This complicates the entire matter because now the cache is negatively affecting the site's market value. Google, however, is more than willing to work with such site owners to address the issue. Christine Mohan, a spokeswoman at the publisher of NYTimes.com, announced that, "We are working with Google to fix that problem—we're going to close it so when you click on a link it will take you to a registration page. We have established these archived links and want to maintain a consistency across all these access points."^[24]

The available opt-out options and Google's openness to work with publishers have prevented the eruption of any major lawsuits so far, but some believe it is only a matter of time. Danny Sullivan, editor of Search Engine Watch, states that "It's very much an issue that has yet to be tested, and I fully expect that it will be."^[25]

A closer look at Google's cached page service reveals a serious legal dilemma: Is Google allowed to make copies of web pages and then make those copies available to the public without the copyright owners' permission?

3.4 Google and Copyright Laws

Fred von Lohman, an attorney at the Electronic Frontier Foundation stated that, "Many of us copyright lawyers have been waiting for this issue to come up: Google is making copies of all the Web sites they index and they're not asking permission. From a strict copyright standpoint, it violates copyright."^[26] A Google spokesman, however, disagrees and stated, "We've evaluated this from a legal perspective, including copyright law, and have determined that Google's cached page service complies with the law."^[27] Google declined my request for further comment on how it complies with copyright law,^[28] and after analyzing the copyright laws detailed in Section 2.0 myself, I concluded that it is ambiguous whether current copyright laws would protect Google's cached page service.

Google clearly violates two of the four exclusive rights outlined in Section 2.1, a copyright owner's right to reproduction and distribution of the copyrighted work. By crawling the Internet and taking full snapshots of each web page, it is reproducing the copyrighted work. By then making that cached page available to search engine users via a link on the search results page, Google is distributing the copyrighted work. The only way Google could exercise these otherwise exclusive rights is if it qualifies as fair use, or is otherwise exempt due to the DMCA safe harbors.

3.4.1 Fair Use

Section 2.1 outlines the following as the factors that determine whether an act constitutes fair use: (1) the purpose and character of the use, (2) the nature of the work, (3) the amount and substantiality of the portion used, and (4) the effect of the use on the work's market or value.

3.4.1.1 Purpose and Character of the Use

Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994), states that the purpose and character of the use involves a consideration of whether

“the new work merely supersedes the objects of the original creation, or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is transformative.” The more transformative the new work, the less will be the significance of other factors, like commercialism, that may weight against a finding of fair use.”^[29]

The founders of Google claimed many noble design goals for their search engine, such as research and academic development as detailed in Section 3.1. While maintaining their mission of collecting the world’s information and making it easily available, Brin and Page have rejected deals with great monetary benefit for the sake of Google’s integrity.^[30] Nonetheless, on the highest level, Google is a business and the commercial nature of its cached page service works against its ability to be considered fair use. Though some could argue that the cache’s purpose has an educational element due to its promotion of research, this aspect of its intent is unlikely to outweigh its overriding commercial function. Jason Schultz, staff attorney at the Electronic Frontier Foundation, agrees, “Many of the pages in Google's cache are wholesale copies of the original pages. Google is clearly using these pages to boost its business. And many sites offer archives for subscribers only. These factors could weigh against Google in a law suit.”^[31]

According to *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 579 (1994), above, however, the transformative element of the use’s character weighs more heavily than its commercialistic purpose. Unfortunately for Google, its cache adds little, if any, new expression, meaning, or message to its cached web pages. By simply taking snapshots of the original sites and making those very snapshots available to the public, it is simply reproducing and redistributing the original copyrighted work with no additional information in or analysis of the actual content on the page. Therefore, both the purpose and character of Google’s cached page service weigh against a finding of fair use.

3.4.1.2 Nature of the Work

Determining fair use requires an assessment of the nature of the original work, acknowledging “that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied.”^[32] Google’s cache takes snapshots of all web pages that contain a variety of copyrighted original works. Such content could range from factual reports and biographies to artistic poems, stories, and photos; published or unpublished. According to June M. Besek, author of “Copyright: What Makes a Use ‘Fair’,” “In general, the law is more sympathetic to copying a fact-based work—a history or biography, for example—than it is to copying a fanciful work such as the latest Harry Potter book.”^[33] Therefore, the artistic works are closer to that core of intended protection. Since Google’s cache copies both the factual and artistic materials in their entirety anyways, the nature of the cache factors unambiguously against fair use.

3.4.1.3 Amount and Substantiality of the Portion Used

Google's cached page service requires that the content of web pages be copied in their entirety for the purpose of the tool to be achieved. Besek says "Generally, the more that is taken, the less likely it is to be fair use..."^[34] Nonetheless, if the cached links are made available so that the information can still be accessed when the original site is down, they must contain the entire site's contents to be of use. If only a small portion of an article was viewable in the cached link, for example, one would still need to access the original site to read the entire article and understand it to completion. The cache would be useless at that point since the user would still need to wait for the original site to be online again. Hence, the amount copied is reasonable in relation to the purpose for copying. In Section 3.4.1.1, however, the purpose for copying itself was found to weigh against fair use. We can therefore infer that the amount and substantiality of the portion used also factors against fair use.

3.4.1.4 Effect of the Use on the Work's Potential Market or Value

For the cached page service to negatively effect a site's potential market or value, users must be accessing the page via the cached link more often than they are viewing the page via the original site. Google, however, has admitted that there are very few clicks on the cached links, and that most users go directly to the original sites linked in their search result pages. For those that do access the cached link, Google includes a header that states the URL to the page's original site and encourages users to go to that site for the most up-to-date information. Sites who would be adversely affected by the cache can also utilize the opt-out options mentioned in Section 3.3. Judith Jennison, the defense lawyer for a search engine titled Arriba Soft, says the following about Google, "The fact that the search site has an opt-out program would likely illustrate that the market for original copyrighted works can be protected, which is a significant factor in fair-use analysis."^[35] Additionally, as shown in Figure 1, the link to the cached page is very discrete, so users would be drawn to the original site's link first. Therefore, Google's cached page service has little effect on the original site's market or value, and thus favors fair use.

With the fourth factor being the only one that supports fair use, Google's cached page service does not qualify for fair use based solely on United States Code Title 17. An assessment of the DMCA and its safe harbors may provide the answer to whether Google's cache is legal or not.

3.4.2 Safe Harbors

Section 512 of the DMCA, described in Section 2.3 above, lists four safe harbors that apply to online service providers: (1) transitory digital network communications, (2) system caching, (3) information residing on systems or networks at the direction of users, and (4) information location tools. Whether these safe harbors apply to Google depends on whether the search engine can be defined as a service provider according to the DMCA. Section 512 defines a server provider as

“an entity offering transmission, routing, or providing connections for digital online communications, between or among points specified by a user, of material of the user’s choosing, without modification to the content of the material as sent or received” or “a provider of online services or network access, or the operator of facilities thereof.”^[36]

Some interpret this definition to apply only to Internet service providers (ISPs), such as AOL, who offer people access to the Internet without modifying the actual content on the web. According to Jonathan Bick, Professor of Internet Law at Pace Law School and Rutgers Law School, search engines may register with the Copyright Office as an ISP.^[37] My inquiry to Google’s Help Team regarding whether they consider themselves an ISP received the following response: “Thank you for your note. We are not an internet service provider as we do not offer any hosting or email services. We are a search engine only.”^[38] Therefore, according to this interpretation, the DMCA safe harbors do not apply to Google. Others, however, believe Section 512’s definition of a service provider applies more broadly to include “Internet service providers (ISPs), search engines, bulletin board system operators, and even auction web sites.”^[39] With such an interpretation, the safe harbors have the potential to apply to Google’s cached page service.

Assuming search engines, and therefore Google, do qualify as service providers, whether its cached page service qualifies as one of the four safe harbors requires a separate assessment.

3.4.2.1 Transitory Digital Network Communications

When the provider acts as a data conduit, transmitting digital information from one point on a network to another at someone else’s request; the transmission, routing, or providing of connections for the information, including transient copies that are made automatically in the operation of the network, are covered by the first safe harbor.^[40] Google could be considered to behave as a data conduit since it transmits links as results to users who request sites related to specific search query terms. Its cached pages are the copies that “are made automatically in the operation of the network,” since all pages are backed-up with no discrimination of the page’s contents.

The Copyright Office Summary of Section 512 of the DMCA lists several conditions the provider must meet to qualify for this safe harbor. The first requirement is that a person other than the provider must initiate the transmission. From the perspective of Google as a search engine, this condition is met because the cached links are only transmitted if a user queries terms related to the contents of that page. From the perspective of the cached page service, however, this condition is not met because the system automatically crawls and caches every page, before a user has even requested them. According to the executive summary of the DMCA by the law offices of Lutzker & Lutzker LLP, the service provider cannot place material online.^[41] Google’s cached page service, however, makes their copies available online via their cached links. The second requirement is that the transmission, routing, provision of connections, or copying must be executed by an automatic process without selection of material by the service provider. Both Google and its cached page service meet the copying aspect of this requirement since pages are

automatically copies regardless of their content. Which links are actually outputted to users as search results, however, are determined by Google's PageRankTM system, which could be considered to be a "selection of material by the service provider." Google and its cache service meet the third requirement that the provider cannot determine the recipients of the material since they must output results to any user of the search engine. It is ambiguous whether the cached page service meets the fourth requirement that states that any intermediate copies must not be retained for longer than reasonably necessary. What time span constitutes "reasonably necessary" is not defined. The cached page service maintains copies of pages anywhere from a few days to a few months depending on when the system crawls that site again. Schultz believes it is ambiguous whether such durations are considered reasonably necessary. "The DMCA safe harbor might protect Google, although caching has to be "intermediate" and temporary", so there's some question as to whether they meet that standard".^[42] Google and its cached page service both satisfy the last condition requiring that no modification is made to the content. The pages are simply backed-up in their entirety, and distributed in their entirety; the content is not edited or modified in any way to potentially change the gist of the content.

The analysis of the different requirements above leads us to conclude that it is uncertain whether Google's cached page service would be protected by the limitation for transitory communications.

3.4.2.2 System Caching

The safe harbor for system caching limits the liability of service providers who retain copies of material for a limited time so that they can be quickly retrieved the next time they are requested by the user. Though Google and its cached page service do abide by many of the system caching limitation's requirements, such as not modifying the retained material and removing pages they have been notified to eliminate, they do not qualify for the system caching limitation.

The wording of the limitation implies that its core is to protect providers such as browsers who cache pages recently visited by web-surfers so that they can quickly reload that page when requested by the user again. Google's cached page service, however, is not caching for that reason. Its cache serves to provide users with cached links of pages if the original site is unavailable. The Legal Protection of Digital Information specifies that the "Section 512 provides safe harbors only to service providers, and then only when the alleged infringing material is not supplied or used by the service provider or its employees."^[43] In this case, however, Google, the supposed service provider, is both supplying and using the copies it makes as a service to users, in addition to its core business as a search engine. Nevertheless, Godwin stated that since the DMCA does not make a distinction between browser caches and search engine caches as its written, it is unlikely that a judge would find a search engine cache illegal.^[44] Conversely, Congress did not specify applying the caching exemption to search engines, and may have had a different intent in mind.

Even if one believed that Google did achieve this limitation's core goal, it still would not qualify for protection under this safe harbor because it does not meet at least two of the

limitation's requirements. To meet this limitation, the "provider must limit users' access to the material in accordance with conditions on access (e.g. password protection) imposed by the person who posted the material."^[45] According to the Legal Protection of Digital Information, "A cache should not provide a way of bypassing an access control system for the material."^[46] As proven by the NYTimes.com situation described in Section 2.3, where users were able to view cached versions of articles that were supposed to be available only to subscribers who had paid for the service, Google is not limiting users' retrieval to the proper access conditions required by the site owners. Secondly, Google is not abiding by the requirement that the "provider must comply with rules about "refreshing" material—replacing retained copies of material with material from the original location..."^[47] When a user is viewing one of Google's cached pages and attempts to refresh the page, the user is not directed to the material's original location, the same cached page is reloaded instead.

An analysis of the conditions required for the system caching limitation leads us to conclude that Google's cached page service would unlikely qualify for the system caching safe harbor. Attorney von Lohman agrees and stated that, "Most people agree that the caching exception in the DMCA is obsolete. I don't think it would cover Google's cache."^[48]

3.4.2.3 Information Residing on Systems or Networks & Information Location Tools

These safe harbors intend to limit the liability service providers could face for infringing material that exists on their servers or that their links point to, respectively. A limitation that allows the caching of sites to benefit users by making the information available when the original site is down is what Google needs. These safe harbors, however, protect service providers who are worried about being liable for material that they host on their systems or link to without knowing whether that content is legal or not. Google's cached page service could not be exempt from copyright law due to these limitations since these safe harbors are addressing different issues.

After analyzing the current copyright laws, both the Title 17 fair use exemptions and the DMCA's safe harbors, it is still ambiguous whether Google's cached page service would be ruled legal if brought to court.

4.0 Comparable Issues

Others have faced similar legal challenges, but it is unclear whether the decisions made in those cases would be the same as in Google's situation. In *Kelly v. Arriba Soft Corporation*, 336 F.3d 811 (9th Cir. 2003), a search engine's reproduction and display of thumbnail images was found to be fair use, but its display of full-sized images has yet to be decided. An organization called the Internet Archive faced legal challenges similar to Google since it copies web pages with the aim of creating an online library of all web pages, but it received full exemption from the Copyright Office.

4.1 Kelly v. Arriba Soft Corporation 336 F.3d 811 (9th Cir. 2003)

When Leslie Kelly, a professional photographer, found reproductions of his images being displayed as thumbnail and full-sized images on Arriba Soft's results page, he charged the search engine with copyright infringement. The United States District Court for the Central District of California ruled that though Kelly had sufficiently proven that Arriba had made unauthorized reproductions and displays of his works, those reproductions and displays constituted fair use under Section 107 of the Copyright Act. Upon appeal, the United States Court of Appeals for the 9th Circuit affirmed in part, reversed in part, and remanded in part the district court's ruling. The Court of Appeals found the search engine's use of thumbnails to be fair use, but remanded the decision regarding the full-image displays to further proceedings since the district court was not to have ruled on that issue.

4.1.2 Arriba Soft's Legal Challenges

Arriba Soft was a search engine unlike most others; rather than outputting text, the results appeared in the form of small images. From July 1999 to near August 2000, two links accompanied the thumbnails in the results page: Source and Details. Clicking on the Source link, or the thumbnail itself, two new windows would pop up. While one of these two windows was the home page of the images original site, the other was a full-sized version of the original image.

Leslie Kelly filed copyright infringement charges against Arriba Soft when he discovered that the search engine was reproducing and displaying his images as thumbnails. Arriba conceded to the violation of reproduction and display rights for the thumbnails only, and thus, argued that the thumbnail images constituted fair use.

The district court, however, ruled on both the thumbnails *and* the full-sized images, and found Arriba's use of the photos to be fair use. This decision broadened the scope of Kelly's original motion to include the full-sized images, and it also extended Arriba's concession to cover the full-sized images. The court found Arriba's use of the images to be sufficiently transformative and harmless to the value of Kelly's works, and therefore, fair use of Kelly's images.

Kelly appealed the decision to the United States Court of Appeals for the 9th Circuit. The resulting holding was that the district court ruled correctly in finding the search engine's use of the thumbnail images to be fair use. The district courts ruling regarding the full-sized images, however, was reversed and remanded since "neither party moved for summary judgment as to the full-size images and Arriba's response to Kelly's summary judgment motion did not concede the prima facie case for infringement as to those images."^[49]

4.1.2 Arriba Soft and Google

As one of the few cases involving a search engine's copyright liability, *Kelly v. Arriba Soft Corporation*, 336 F.3d 811 (9th Cir. 2003), may provide insight into what Google can or cannot do. Analyzing the reasoning behind why Arriba Soft constitutes fair use will allow us to determine whether this case would serve as a solid precedent for Google's cached page service.

The circuit court found that the purpose and character of Arriba Soft's use of the images as thumbnails weighed in favor of fair use. Though the search engine clearly had a commercial purpose, it was not using Kelly's images to promote its website, nor was it trying to sell the images. "Instead, Kelly's images were among thousands of images in Arriba's search engine data-base. Because the use of Kelly's images was not highly exploitative, the commercial nature of the use weighs only slightly against a finding of fair use."^[50] I agree with this reasoning because a user cannot simply save a thumbnail image instead of buying the actual one. Attempts to enlarge the thumbnail image would diminish its quality, and become worthless to the user. If similar reasoning were applied to Google's cache, it too would have only a slight weighing against fair use because, though it is also of a commercial purpose, each of its snapshots are among millions of cached copies automatically downloaded into Google's database. Regarding the character of the search engine, Arriba Soft was found to be sufficiently transformative. "Arriba's use of Kelly's images in the thumbnails is unrelated to any aesthetic purpose. Arriba's search engine functions as a tool to help index and improve access to images on the internet and their related web sites."^[51] The factor of purpose and character was found to favor Arriba "due to the public benefit of the search engine and the minimal loss of integrity to Kelly's images."^[52] This analysis does not apply to Google because its cached page service actually takes snapshots of entire web pages, and displays the entire web page. Users could actually view the cached link and never need to see the original site; individuals who wanted an image, however, would still need to retrieve a full-sized image if they wanted to use it, the thumbnail would not suffice. Google, therefore, is of a less transformative nature than Arriba Soft, and its purpose and character of use would factor against fair use. This analysis coincides with my conclusion in Section 3.4.1.1.

Regarding the fair use factor of the nature of Arriba Soft, the circuit court found it to slightly favor Kelly due to the artistic nature of the images, and since many of them were published prior to appearing on the search engine's result pages. Google's cached pages consist of many artistic and published works so the court's finding in Arriba Soft would imply a weighing against fair use for Google, as reasoned in Section 3.4.1.2.

The circuit court found the factor of amount and substantiality of use to favor neither party. The reasoning was that Arriba Soft had reproduced and displayed the entire work, but that amount was reasonable according to its purpose of use. Had the search engine only displayed a portion of the image, the photo would be harder to identify, and the results page would be useless. In Section 3.4.1.3, I had rationalized that Google needed to copy the web page in its entirety to serve the cache's purpose of making the page available even if the site is down. Had Google only copied and displayed a portion of the page, the user would still need to access the original site to view the remaining portions.

At the point where the user would still have to wait for the original site to come back online, the benefits of the cached page service are lost. The difference between Arriba Soft and Google, however, is that Arriba's purpose was found to be legitimate, whereas I had determined that Google's purpose factored against fair use. Therefore, even though the amount Google copied is reasonable for its objective, since its purpose weighs against fair use, its nature factors against fair use too.

The final factor for determining fair use, market effect, was found to favor Arriba Soft. Since the thumbnail images included the original sites address, interested buyers were actually being driven to Kelly's site rather than away from it. Even if users only wanted more information regarding the photo, or wanted to see a full-sized image of the photo, they would go to the original site. The market or values of Kelly's images were not being negatively impacted by Arriba Soft's search engine. Though Google's cached link pages also include the original sites URL, the same need to visit the original site does not exist as with Arriba Soft. Had the court been ruling on Arriba Soft's display of full-sized images, there would be more of a similarity to Google's situation. In Section 3.4.1.4, I had concluded that the market effect favors fair use for Google, but Arriba Soft's case does not set a precedent for why that is true for the cached page service.

Jennison, Arriba Soft's defense lawyer, believes this case could serve as a precedent for Google. "In Google's case, the result would likely be the same, because the temporary caching for indexing purposes would be fair use per *Kelly v. Arriba soft*."^[53] I disagree, however, because Google goes beyond simply temporarily "caching for indexing purposes." Google's cached page service may hold backup copies anywhere from days to months, whereas Arriba Soft only holds it for 24 hours. Additionally, Google's feature is not meant for indexing, but is presented as an alternative to viewing the original site, especially when the original website is unavailable. The full-sized images would be more comparable to the snapshots of pages taken by Google's crawler because both contain the work in its entirety. With the focus of *Kelly v. Arriba Soft Corporation*, 336 F.3d 811 (9th Cir. 2003), being on the thumbnail images rather than the full-sized images, it does not serve as a proper precedent for Google's cached page service.

4.2 Internet Archive's Wayback Machine

Brewster Kahle, inventor of the Wide Area Information Servers (WAIS) system, founder of WAIS Inc., and 1982 MIT graduate, founded the Internet Archive in 1996. His goal was to construct a digital library that would preserve the Internet's contents.^[54] To collect the web's information, the organization programmed computers to crawl the Internet by downloading a web page, and then downloading the graphics and other pages it links to. As the process continues throughout each page, more and more of the Web gets stored into the Internet Archive's databases. To serve as a library, the organization developed the Wayback Machine, a program that "organizes the billions of pages and allows anyone online to look up the contents of the archive."^[55] Kahle told CNET News.com that his archive is often used for patent research, and by designers and students who wish to understand the evolution of the Web's design and display.

4.2.2 Internet Archive's Legal Challenges

A program that copies web pages to archive them and make them available to others is surely going to face legal challenges, and the Internet Archive confronted several. Aside from the legal and social issues involving privacy, import/export restrictions, and possession of social property,^[56] the Internet Archive's greatest legal challenge was copyright protection. Certain provisions of the DMCA, such as Section 1201 that prohibits the circumvention of access controls, threatened the Internet Archive's ability to archive software titles. Kahle, therefore, petitioned the Copyright Office for an exemption from Section 1201 to be granted for "Literary and audiovisual works embodied in software whose access control systems prohibit access to replicas of the works."^[57] In late October of 2003, the Internet Archive was granted exemptions to reproducing the following four categories of works:

- (1) Compilations consisting of lists of Internet locations blocked by commercially marketed filtering software applications that are intended to prevent access to domains, websites or portions of websites, but not including lists of Internet locations blocked by software applications that operate exclusively to protect against damage to a computer or computer network or lists of Internet locations blocked by software applications that operate exclusively to prevent receipt of email.
- (2) Computer programs protected by dongles that prevent access due to malfunction or damage and which are obsolete.
- (3) Computer programs and video games distributed in formats that have become obsolete and which require the original media or hardware as a condition of access.
- (4) Literary works distributed in ebook format when all existing ebook editions of the work (including digital text editions made available by authorized entities) contain access controls that prevent the enabling of the ebook's read-aloud function and that prevent the enabling of screen readers to render the text into a specialized format.^[58]

Though these exemptions will only last for three years, at which time the Internet Archive will need to petition once again, they will aid the organization with achieving its goal until then.

4.2.2 Internet Archive and Google

The Internet Archive's technological process is similar to that of Google's cached page service. Both download copies of pages while crawling the web, and later make those copies available to users. The key distinction, however, is that the Wayback Machine was built for non-profit, preservation purposes, whereas Google's cache is a commercial tool meant for the convenience of web surfers. This difference is sufficient reason for why the Internet Archive's legal challenges would not set precedence for Google because Section 108 and 117 of the Copyright Act, which govern archiving, specifically exempt the Internet Archive's reproduction of web pages. Even without these sections, copying for archival purposes is protected by fair use exemptions. As the House of Representatives reported when passing the Copyright Act of 1976,

"The efforts of the Library of Congress, the American Film Institute, and other organizations to rescue and preserve this irreplaceable contribution to our cultural life are to be applauded, and the making of duplicate copies for purposes of archival preservation certainly falls within the scope of 'fair use.'"^[59]

With Google specifically saying that its “cache feature does not attempt to create a permanent historical record of the Web,”^[60] it would not qualify for fair use under Section 108 or 117, or by any exemptions received by the Internet Archive.

5.0 Suggested Update of Copyright Law

All the analysis thus far indicates that if Google were currently brought to court with copyright infringement charges against its cached page service, it is unclear whether they would be liable or not. To promote the availability of research and ensure the preservation of Google’s cached page service despite potential copyright charges, it is vital that the copyright laws be updated to allow for such features.

Chapter 4 of the Digital Dilemma outlines several reasons that could justify updating copyright law to cover Google’s cached page service, specifically seven categories “into which exceptions and limitations to copyright owners’ rights seem generally to fall.”^[61] Google’s cached page service could fall into three of those categories: (1) those that are based on public interest grounds, (2) those that promote flexible adaptation of the law to new circumstances, and (3) those that cover situations in which uses or copying of protected works are incidental to otherwise legitimate activities, or implicitly lawful given the totality of circumstances.^[62] Laws that permit libraries and archives to make copies for preservation purposes are examples of exceptions that fall into the first category. One could argue that the value Google’s cache feature adds to research and the proliferation of information is of public interest, and should consequently also be exempt. As technologies rapidly change, the courts must often apply copyright laws to situations that were never intended by the original writers of the law. Such a case includes *Sony v. Universal City Studios*, 464 U.S. 417, 104 S. Ct. 774, 78 L. Ed. 2d 574 (1984), where the home taping of television programs for time-shifting purposes was ruled as fair use. Eventually, however, the copyright laws themselves must be adjusted to accommodate the changes in society. The DMCA was written around the same time that Google began its services on the Internet. The writers of the DMCA were most likely not taking Google’s cached page feature into consideration when determining what qualifies as an exemption, since they were not even familiar with it at the time. An update to copyright law that unambiguously legalizes search engine caching could be justified by the need to adapt laws to new circumstances. Lastly, the overall look at Google’s cached page tool in Section 3.4.1 has already concluded that the service has little if any adverse effects on other markets, and can be of great value to users. Since the end result of providing users access to pages whose original sites are down is a legitimate activity, the third category may also justify exempting the incidental use and copying of protected works that may occur. Some could turn this into a means versus ends debate, but ultimately, it is just one of many justifications for updating copyright law.

An additional provision should be added to the DMCA that protects search engines that cache for the purpose of providing better service to users. Inspired by existing copyright law, the following is a suggested provision to be added to Section 512 of the DMCA.

Limitations on cached page services for search engines

(1) LIMITATION ON LIABILITY. — A search engine shall not be liable for monetary relief, or for injunctive or other equitable relief, for infringement of copyright by reason of the caching and display on a system or network controlled or operated by or for the search engine in a case in which —

- (A) the caching is carried out through an automatic technical process without selection of the material by the search engine for the purpose of making the material available to users of the system or network
- (B) the search engine does not select the recipients of the material except as an automatic response to the request or query of that user
- (C) no copy of the material made by the search engine in the course of such storage is maintained on the system or network in a manner ordinarily accessible to anyone other than anticipated recipients
- (D) the material is transmitted through the system or network without modification of its content
- (E) the address of the original website is clearly indicated on the cached page
- (F) the cached page indicates that it may not be the most up-to-date version
- (G) the link to the original website is more predominantly visible on search result pages than the cached page link, and
- (H) the search engine does not receive a financial benefit directly attributable to the cached material, in a case in which the search engine has the right and ability to control such activity

(2) CONDITIONS. — The conditions required are that —

- (A) if the cached website has in effect a condition that a person must meet prior to having access to the material, such as a condition based on payment of a fee or provision of a password or other information, the search engine permits access to the stored material in significant part only to users of its system or network that have met those conditions and only in accordance with those conditions; and
- (B) if the owner of a copyrighted work requests for the removal of his or her work from the cache, the search engine must respond expeditiously to remove, or disable access to, that material upon notification as described in subsection (3).

(3) DESIGNATED AGENT. — The limitations on liability established in this subsection apply to a search engine only if the it has designated an agent to receive notifications of requests for material removal described in subsection (2)(B), by making available through its service, including on its website in a location accessible to the public, and by providing to the Copyright Office, substantially the following information:

- (A) the name, address, phone number, and electronic mail address of the agent.
- (B) other contact information which the Register of Copyrights may deem appropriate.

The Register of Copyrights shall maintain a current directory of agents available to the public for inspection, including through the Internet, in both electronic and hard copy formats, and may require payment of a fee by the search engines to cover the costs of maintaining the directory.

(4) ELEMENTS OF NOTIFICATION. —

To be effective under this subsection, a request for material removal must be a written communication provided to the designated agent of a search engine that includes substantially the following:

- (A) a physical or electronic signature of a person authorized to act on behalf of the material's owner.
- (B) identification of the material that is to be removed, and information reasonably sufficient to permit the search engine to locate the material.

(C) information reasonably sufficient to permit the search engine to contact the complaining party, such as an address, telephone number, and, if available, an electronic mail address at which the complaining party may be contacted.

(D) a statement that the complaining party has a good faith belief that use of the material in the manner complained of is not authorized by the copyright owner, its agent, or the law.

(5) CONDITIONS FOR ELIGIBILITY. —

ACCOMMODATION OF TECHNOLOGY. — The limitations on liability established by this section shall apply to a search engine only if it accommodates and does not interfere with standard technical measures for publishers to opt-out of caching such as robots.txt files and meta tags.

DEFINITION. — As used in this subsection, the term “standard technical measures” means technical measures that —

(A) have been developed pursuant to a broad consensus of copyright owners and service providers in an open, fair, voluntary, multi-industry standards process;

(B) are available to any person on reasonable and nondiscriminatory terms; and

(C) do not impose substantial costs on search engines or substantial burdens on their systems or networks.

(6) DEFINITIONS. —

(A) SEARCH ENGINE. — As used in this limitation, the term “search engine” means a computer program that retrieves documents, files, or other information from a database or network, or the operator of such a program.

(B) MONETARY RELIEF. — As used in this section, the term “monetary relief” means damages, costs, attorneys' fees, and any other form of monetary payment.

The aforementioned suggested update to copyright law would allow Google to continue operation of its cached page service liability free, while still allowing publishers the option of not being copied or of requesting their material to be removed from the cache.

6.0 Conclusion

With the Internet expanding at phenomenal rates, Google's cached page service holds great values for those turning to the Web for information. The recently raised copyright concerns have stirred debate on the legality of Google's cache. Current copyright laws do not exempt the cache as fair use, and it is ambiguous whether it would be protected by the DMCA safe harbors. There is confusion whether the limitations apply to search engines, and even if they do, it is unclear whether they would specifically cover Google's cached page service. There are few examples to turn to for insight on what would happen if Google were brought to court. One of the few known cases of a search engine charged with copyright infringement does not serve as a solid precedent since the case involved thumbnail images, whereas Google's cache returns web pages in their entirety. That same case, however, will soon have a proceeding regarding the search engine's display of full-sized images, which may be more comparable to Google's cache. Other

organizations that also reproduce and display web pages in their entirety are often times not analogous because they conduct such activities for archival purposes, which are exempt from copyright law. Only by adding a provision to the DMCA that protects search engines who partake in cached page services, while still allowing publishers opt-out options, can we bring copyright law up to par with the rapid changes in technology and preserve a most valued service.

7.0 Endnotes

[¹] 17 U.S.C.S. § 106 (2003).

[²] US Code Collection, “Sec. 107. – Limitations on exclusive rights: Fair use;” available from <http://www4.law.cornell.edu/uscode/17/107.html>; Internet; accessed November 22, 2003

[³] 17 U.S.C.S. § 107 (2003).

[⁴] Legal Protection of Digital Information, “Chapter 3: Copyright of Digital Information;” available from <http://www.digital-law-online.com/lpdi1.0/treatise33.html>; Internet; accessed November 29, 2003

[⁵] Ibid.

[⁶] Ibid.

[⁷] “The Digital Millennium Copyright Act of 1998: U.S. Copyright Office Summary,” December 1998, Page 9

[⁸] Google’s Corporate Information, “Google History;” available from <http://www.google.com/corporate/history.html>; Internet; accessed December 4, 2003

[⁹] Wired Magazine, “Google vs. Evil;” available from http://www.wired.com/wired/archive/11.01/google_pr.html; Internet; December 4, 2003

[¹⁰] Google Press Center, “Google Fun Facts;” available from <http://www.google.com/press/funfacts.html>; Internet, accessed December 5, 2003

[¹¹] Google Technology, available from <http://www.google.com/technology/index.html>; Internet; accessed December 4, 2003

[¹²] The Anatomy of a Large-Scale Hypertextual Web Search Engine;” Brin, Sergey; Page, Lawrence; available from <http://www7.scu.edu.au/programme/fullpapers/1921/com1921.htm>; Internet; accessed December 4, 2003

[¹³] Ibid.

[¹⁴] Ibid.

[¹⁵] Google Press Center, “Google Fun Facts;” Olsen, Stefanie; July 9, 2003; available from <http://www.google.com/press/funfacts.html>; Internet, accessed December 5, 2003

[¹⁶] CNET News.com, “Google Cache Raises Copyright Concerns;” available from http://news.com.com/2100-1032_3-1024234.html; Internet; accessed November 5, 2003

- [17] Microdoc News, "Google Cache as Friend;" Tweney, Dylan; April 18, 2003, available from <http://microdoc-news.info/home/2003/04/22.html>; Internet; accessed December 5, 2003
- [18] CNET News.com, "Google Cache Raises Copyright Concerns;" 2003
- [19] Electronic Frontier Foundation Staff Counsel Michael Godwin, interviewed by Parul Deora, 8 December 2003, phone
- [20] Chilling Effects, "Microsoft Complains of Product Key in Google Cache;" March 12, 2003; available from <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=586>; Internet; accessed November 22, 2003
- [21] "CNET Complains of WinPro in Google Cache;" August 28, 2003; available from <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=845>; Internet; accessed November 22, 2003
- [22] *Whelan v. Jaslow*, 797 F.2d 1222 (1986).
- [23] Electronic Frontier Foundation Staff Counsel Michael Godwin, interviewed by Parul Deora, 8 December 2003, phone
- [24] CNET News.com, "Google Cache Raises Copyright Concerns;" 2003
- [25] *Ibid.*
- [26] *Ibid.*
- [27] *Ibid.*
- [28] The following is Google Team's response to my inquiry into their copyright concerns. Response received December 8, 2003, Re: [#5294358] Questions for MIT Paper, from google to pdeora

Hi Parul,

Thank you for your note. We apologize for our delayed response. At this time, the information you've read is all of the information that's publicly available. As you may imagine, because Google is a privately held company, there is some information that we're unable to provide to the public at this time.

We hope that your class assignment went well, and we'd love to field any other questions you may have in the future.

Regards,
The Google Team

- [29] *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 579 (1994).
- [30] "In May, Anita Roddick, the outspoken British founder of the Body Shop, blasted Google in her blog for yanking a text ad for her site. Google's explanation: Roddick had called actor John Malkovich a "vomitous worm" in her blog, violating a Google policy against accepting ads for sites that are "anti-" anything. After Roddick protested, Google offered to reinstate the ad in exchange for a promise from Roddick that she would remove the Malkovich reference from the first page of her site. When she refused, Brin had a decision to make: Should he give in and accept Roddick's money, or stand by his principles? He chose his principles." Wired Magazine, "Google vs. Evil;" available from http://www.wired.com/wired/archive/11.01/google_pr.html; Internet; December 4, 2003

- [31] Electronic Frontier Foundation Staff Attorney Jason Schultz, interviewed by Parul Deora, 4 December 2003, phone
- [32] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579, 127 L. Ed. 2d 500, 114 S. Ct. 1164 (1994).
- [1[33]] Educause Review, "Copyright: What Makes a Use 'Fair'?" Besek, June M.; November/December 2003; available from <https://www.educause.edu/ir/library/pdf/erm0368.pdf>; Internet; accessed December 8, 2003
- [34] *Ibid.*
- [35] CNET News.com, "Google Cache Raises Copyright Concerns;" 2003
- [36] 17 U.S.C. § 512
- [37] New Jersey Law Journal, "Exploitation of Trademarks on the Internet;" Bick, Jonathan; December 8, 2003; available from http://web.lexis-nexis.com/universe/document?_m=49f26a9782e72323b352d099df6632fc&_docnum=1&wchp=dGLbVzz-zSkVA&_md5=3df59410dfd4ae94a7156dda0454b2a0; Internet; accessed December 9, 2003
- [38] Google Help Team, questioned by Parul Deora, 9 December 2003, email
- [39] Chilling Effects, "What defines a service provider under Section 512 of the DMCA?" available from <http://www.chillingeffects.org/dmca512/notice.cgi?NoticeID=586#QID127>; Internet; accessed November 22, 2003
- [40] "The Digital Millennium Copyright Act of 1998: U.S. Copyright Office Summary," December 1998, Page 10
- [41] Lutzker & Lutzker LLP, "The Digital Millennium Copyright Act;" available from <http://www.arl.org/info/frn/copy/osp.html>; Internet; accessed November 5, 2003
- [42] Electronic Frontier Foundation Staff Attorney Jason Schultz, interviewed by Parul Deora, 4 December 2003, phone
- [43] Legal Protection of Digital Information, "Digital Law Online: Caching;" available from <http://www.digital-law-online.com/lpdi1.0/treatise36.html>; Internet; accessed November 29, 2003
- [44] Electronic Frontier Foundation Staff Counsel Michael Godwin, interviewed by Parul Deora, 8 December 2003, phone
- [45] "The Digital Millennium Copyright Act of 1998: U.S. Copyright Office Summary," Page 11
- [46] Legal Protection of Digital Information, "Digital Law Online: Caching;" available from <http://www.digital-law-online.com/lpdi1.0/treatise36.html>; Internet; accessed November 29, 2003
- [47] "The Digital Millennium Copyright Act of 1998: U.S. Copyright Office Summary," Page 11
- [48] CNET News.com, "Google Cache Raises Copyright Concerns;" 2003
- [49] *Kelly v. Arriba Soft Corporation*, 336 F.3d 811 (9th Cir. 2003).
- [50] *Ibid.*
- [51] *Ibid.*

[52] Ibid.

[53] CNET News.com, "Google Cache Raises Copyright Concerns;" 2003

[54] Internet Archive, "Archiving the Internet;" Scientific American; November 4, 1996; available from http://www.archive.org/sciam_article.html; Internet; accessed November 29, 2003

[55] Business Week Online, "A Library as Big as the World;" Green, Heather; February 28, 2002; available from http://www.businessweek.com/technology/content/feb2002/tc20020228_1080.htm; Internet; accessed November 25, 2003

[56] Internet Archive, 2003

[57] Kahle, Brewster; Macgillivray, Alexander; Lessig, Lawrence; Seltzer, Wendy; "Rm 2002-4 – 17 USC § 1201 Exemptions Notice of Inquiry," December 18, 2002

[58] "Statement of the Librarian of Congress Relating to Section 1201 Rulemaking;" available from http://www.copyright.gov/1201/docs/librarian_statement_01.html; Internet; accessed December 7, 2003

[59] H.R. REP. NO. 94-1476, at 73 (1976).

[60] CNET News.com, "Google Cache Raises Copyright Concerns;" 2003

[61] Digital Dilemma, "Chapter 4: Individual Behavior, Private Use, and Fair Use;" available from http://www.nap.edu/html/digital_dilemma/ch4.html; Internet; accessed November 22, 2003

[62] Ibid.
