

# CACHING IN ON THE GOOGLE BOOKS LIBRARY PROJECT: A NOVEL APPROACH TO THE FAIR USE DEFENSE AND THE DMCA CACHING SAFE HARBORS

JESSE S. BENNETT\*

I.	INTRODUCTION.....	1003
II.	CACHING IN.....	1007
	A. <i>Cache (pronounced 'kash,' like 'cash'). Function: noun. Definition: a hiding place for treasure or provisions.</i> .....	1007
	B. <i>The Nuts and Bolts</i> .....	1008
	C. <i>The Theory</i> .....	1010
	D. <i>Who is Caching In?</i> .....	1010
	1. <i>Personal Computers</i> .....	1011
	2. <i>Internet Service Providers</i> .....	1013
	3. <i>Search Engines</i> .....	1015
	4. <i>The Domain Name System</i> .....	1016
	5. <i>Caching Services (Google, Yahoo! &amp; Archive.org)</i> .....	1018
	(a) <i>Google</i> .....	1019
	(b) <i>Yahoo!</i> .....	1020
	(c) <i>The Internet Archive</i> .....	1020
	6. <i>Attorneys &amp; Law Enforcement</i> .....	1021
III.	THE GOOGLE BOOK SEARCH PROJECT AND THE GOOGLE LIBRARY PROJECT .....	1022
	A. <i>The Projects</i> .....	1022
	B. <i>The Ensuing Litigation and the Fair Use Defense</i> .....	1025
	1. <i>Kelly v. Arriba Soft (Kelly II)</i> .....	1027
	2. <i>Perfect 10 v. Google, Inc.</i> .....	1028
IV.	CACHING AND ISP RELATED COPYRIGHT LAW .....	1029
	A. <i>Copyright Law as it Relates to Caching</i> .....	1030
	B. <i>Copyright Law as it Relates to ISPs</i> .....	1031
V.	GOOGLE'S MATRIX V. THE COPYRIGHT ACT .....	1034
VI.	CONCLUSION .....	1038

## I. INTRODUCTION

Imagine a world ruled by computers, where humans are enslaved in tiny pods and used for their body heat and neural activity. In order to achieve submission, the computers use a giant matrix of images from the year 1999 to keep the humans from resisting their restraint, essentially enslaving their thoughts within the confines of the algorithm. Their memories, mixed with the recurring images, assure them that this false reality is actual reality. Fortunately for mankind, a rebellion has begun—a rogue uprising that will hopefully end the tyrannical reign of this algorithmic computer web, known as The Matrix.

---

\* Juris Doctor Candidate, May 2009, Florida State University College of Law; University of Florida, B.S. in Computer Engineering, 2006.

Welcome to the world of Warner Brothers Studios' *The Matrix*,<sup>1</sup> where Neo battles Agent Smith and the computer-controlled machines to destroy the matrix and free mankind. Where, as Morpheus states, "[t]he Matrix is a system . . . [and] . . . that system is our enemy."<sup>2</sup>

Although the man versus computer theme is popular in modern film culture,<sup>3</sup> it is unlikely to be part of humanity's near future. However, matrices of the magnitude envisioned in *The Matrix* are a reality. Mathematicians and computer scientists have estimated the World Wide Web (Web)<sup>4</sup> to be a matrix of at least order 12 billion.<sup>5</sup> To give some perspective, a matrix of order three is a tic-tac-toe board; now multiply that by four billion.

Among other things, computer algorithms utilizing matrix computations are extremely powerful as search engine tools.<sup>6</sup> For this reason, many search engines employ linear algebra techniques to construct matrices of immense sizes.<sup>7</sup> For example, Google's PageRank<sup>8</sup> algorithm employs the "world's largest matrix computation: order 10 billion."<sup>9</sup> Yet Google's matrix is no more than a giant index of Web content.<sup>10</sup>

1. THE MATRIX (Warner Brothers Studios 1999).

2. *Id.*

3. *See id.*; THE TERMINATOR (Orion Pictures 1984); *see also* Man vs. Machine Movies, <http://www.boxofficemojo.com/genres/chart/?id=manvsmachine.htm> (last visited Aug. 25, 2008).

4. The Web is "a part of the Internet accessed through a graphical user interface and containing documents often connected by hyperlinks . . . ." Merriam-Webster Online, <http://merriam-webster.com/dictionary/world%20wide%20web> (last visited Aug. 25, 2008). To avoid confusion, for the remainder of this Comment, the terms Web and Internet will be used interchangeably.

5. Bo Kågström, Search Engine Rankings: Using the Link Structure of the Web – Google's PageRank and Similar Approaches 4, Lecture at Umea University (Oct. 4, 2007) (transcript available at [http://www.cs.umu.se/kurser/5DA002/HT07/lectures/MBT\\_C10\\_PageRank\\_071004\\_eng\\_4p.pdf](http://www.cs.umu.se/kurser/5DA002/HT07/lectures/MBT_C10_PageRank_071004_eng_4p.pdf)) (discussing the size of the Web in terms of the order of the matrix it represents).

6. Bo Kågström, Matrix Computations and Applications - An Introduction, Lecture at Umea University (Sept. 3, 2007) (transcript available at [http://www.cs.umu.se/kurser/5DA002/HT07/lectures/MBT\\_C1\\_intro-070903eng.pdf](http://www.cs.umu.se/kurser/5DA002/HT07/lectures/MBT_C1_intro-070903eng.pdf)) (discussing several useful applications of matrix computations).

7. DAVID POOLE, LINEAR ALGEBRA: A MODERN INTRODUCTION 354 (2d ed. 2005); Amy Langville, *The Linear Algebra Behind Search Engines*, 5 J. ONLINE MATHEMATICS & ITS APPLICATIONS (2005), available at <http://mathdl.maa.org/mathDL/4/?pa=content&sa=viewDocument&nodeId=636>.

8. Phil Craven, *Pagerank Explained—Google's Pagerank and How to Make the Most of It*, WEBWORKSHOP.NET, <http://www.webworkshop.net/pagerank.html> (last visited Aug. 25, 2008).

9. Nigel Buttimore, Markov Chains for Biosequences and Google Searches, at 10, Lecture at Trinity College, Dublin (Jan. 21, 2008), available at <http://www.maths.tcd.ie/~nhb/talks/Lucent.pdf>.

10. Video: Is Google Book Search Fair Use? (Lawrence Lessig, Jan. 8, 2006) (on file at <http://video.google.com/videoplay?docid=-7256091247456149593&q=lawrence+lessig>) [hereinafter Lessig Video].

However fictional *The Matrix* may be, it is evident that many see Google's monstrous matrix as the proverbial end. It may not be the end of mankind, but at least the end of copyright protection. More specifically, the Authors Guild (AG)<sup>11</sup> and the Association of American Publishers (AAP)<sup>12</sup> are battling Google in a war over intellectual property rights.<sup>13</sup>

To be fair, authors and publishers are not disturbed by Google or its matrix per se; it is Google's use of the matrix that is of concern. The cause of the current battle is the Google Book Search Project.<sup>14</sup> More specifically, it is the Google Books Library Project's<sup>15</sup> unauthorized copying and searching of copyrighted materials that has concerned the AG and the AAP. Hence, as authors and publishers thrust their sword of copyright infringement claims, Google defends with the oft-used yet unpredictable shield of fair use.<sup>16</sup>

---

11. The Authors Guild is an advocacy group for published authors. TAD CRAWFORD & KAY MURRAY, *THE WRITER'S LEGAL GUIDE: AN AUTHORS GUILD DESK REFERENCE* 4 (Allworth Press 2002). It offers individual business and legal advice to members, as well as guide books and a quarterly bulletin. *Id.* In addition, the Guild frequently lobbies for author-favorable legislation. *Id.*

12. The Association of American Publishers is the principal trade association for American book publishers. About the Association of American Publishers, [http://www.publishers.org/main/AboutAAP/about\\_00.htm](http://www.publishers.org/main/AboutAAP/about_00.htm) (last visited Aug. 25, 2008). The Association focuses on protecting the intellectual property rights of publishers, especially copyright, and dealing with digital issues of concern to publishers. *Id.*

13. Both organizations filed suit in the Southern District of New York. Complaint, *McGraw Hill Co. v. Google, Inc.*, No. 05 CV 8881 (S.D.N.Y. Oct. 19, 2005), available at [http://www.publishers.org/main/PressCenter/Archives/2005\\_Oct/attachments/40\\_McGraw-Hill\\_v.\\_Google.pdf](http://www.publishers.org/main/PressCenter/Archives/2005_Oct/attachments/40_McGraw-Hill_v._Google.pdf) [hereinafter Complaint of McGraw Hill Co.]; Complaint, *Author's Guild v. Google, Inc.*, No. 05 CV 8136 (S.D.N.Y. Sept. 20, 2005), available at <http://pub.bna.com/eclr/05cv8136comp.pdf> [hereinafter Complaint of the Authors Guild].

14. On December 14, 2004, Google, Inc. announced the ambitious "Google Book Search" program (previously "Google Print"), making known its intent to scan materials from five major libraries and make the resources searchable online. See Press Release, Google, Inc., *Google Checks Out Library Books*, (Dec. 14, 2004), available at [http://www.google.com/press/pressrel/print\\_library.html](http://www.google.com/press/pressrel/print_library.html). "The project involves two separate programs, the 'Partner Program' (formerly the 'Publisher Program') and the 'Library Project.'" Cameron W. Westin, *Is Kelly Shifting Under Google's Feet? New Ninth Circuit Impact on the Google Library Project Litigation*, 2007 DUKE L. & TECH. REV. 2, 3 (2007) (discussing the effects on the Library Project litigation of two previous cases involving Google).

15. The Library Project is the source of the litigation because it is the subset of the Google Book Search Program that intends to copy library books regardless of whether the content is in-copyright or whether permission is given. Jonathan Band, *The Google Library Project: Both Sides of the Story*, 2 PLAGIARY: CROSS-DISCIPLINARY STUDIES IN PLAGIARISM, FABRICATION, & FALSIFICATION 1, 2 (2006), <http://www.plagiarism.org/Google-Library-Project.pdf>.

16. The fair use doctrine sets out several instances when copying of in-copyright material is allowed without the permission of the copyright holder. See 17 U.S.C. § 107 (2006). The fair use doctrine is considered unpredictable because it is not a "bright-line rule." See *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 448 n.31 (1984) (noting that Congress had "eschewed a rigid, bright line approach to fair use"). The *Sony* opinion has been consistently endorsed in this respect by the Supreme Court's decisions pertaining to fair use. See, e.g., *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 577 (1994) (pointing out

The topic of Google's fair use defense is not altogether novel. Many commentators have discussed the Google Book Search Project, the ensuing litigation, and Google's affirmative defense of fair use.<sup>17</sup> However, little has been detailed about Google's use of caching in terms of the Book Search Project, and the relevant literature contains no analysis thus far about Google's inherent similarities to Internet Service Providers (ISPs).<sup>18</sup> Google's copying of in-copyright library books may not be deemed a fair use; however, its use of caching to display the digital works may be. Unlike Agent Smith and *The Matrix*, Google may not be the enemy after all.

Naturally, Google's use of caching for the Book Search Project requires the copying of traditionally printed book content.<sup>19</sup> An immense index of this content is created and Google's cache is what makes the indexed content available on the Internet.<sup>20</sup> Therefore, it is not the copying, but the caching, which is the proper subject of a fair use inquiry. Fortunately for Google, courts have ruled that caching is a fair use;<sup>21</sup> thus, Google's use of caching within the Book Search Project should also be allowed as a fair use.<sup>22</sup> In any event, Google may find protection under the Digital Millennium Copyright Act (DMCA).<sup>23</sup>

---

that "[t]he task [of section 107] is not to be simplified with bright-line rules . . ."; Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 588 (1985); MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT, § 13.05[A] (arguing that nothing in section 107 provides "a rule that may automatically be applied in deciding whether any particular use is 'fair'").

17. For discussions of Google's fair use defense, see generally Corinna Baksik, *Fair Use or Exploitation? The Google Book Search Controversy*, 6 PORTAL: LIBRARIES & THE ACADEMY 399 (Oct. 2006), available at [http://muse.jhu.edu/journals/portal\\_libraries\\_and\\_the\\_academy/v006/6.4baksik.pdf](http://muse.jhu.edu/journals/portal_libraries_and_the_academy/v006/6.4baksik.pdf) (discussing Google's fair use defense generally); David Kohler, *This Town Ain't Big Enough For the Both Of Us—Or Is It? Reflections on Copyright, The First Amendment and Google's Use of Others' Content*, 2007 DUKE L. & TECH. REV. 5 (2007); Westin, *supra* note 14; Brian Sites, *Google The Gozerian and Fair Use Slimed: Copyright Again in the Technocrat's Den*, 47 JURIMETRICS J. 31 (2006).

18. As used herein, ISP includes Internet Service Providers, On-line Service Providers, Internet Access Providers, and communications companies that provide Internet access, among these, telephone line service providers such as AOL, digital cable service providers such as COX and Comcast, and broadband service providers such as Verizon and Embarq.

19. See Lawrence Lessig, Lessig Blog, Google Sued (Sept. 22, 2005), [http://lessig.org/blog/2005/09/google\\_sued.html](http://lessig.org/blog/2005/09/google_sued.html) [hereinafter Lessig Blog]; see also Lessig Video, *supra* note 10.

20. See Lessig Video, *supra* note 10.

21. See, e.g., Field v. Google, Inc., 412 F. Supp. 2d 1106, 1110-11 (D. Nev. 2006) (holding that Google's use of cache was a fair use).

22. See Lessig Blog, *supra* note 19.

23. "The DMCA was enacted both to preserve copyright enforcement on the Internet and to provide immunity to service providers from copyright infringement liability for "passive," "automatic" actions in which service provider's system engages through a technological process initiated by another without the knowledge of the service provider." *ALS Scan, Inc. v. RemarQ Cmty., Inc.*, 239 F.3d 619, 625 (4th Cir. 2001) (citing H.R. REP. NO. 105-796, at 72 (1998) (Conf. Rep.); Digital Millennium Copyright Act, Pub. L. No. 105-304,

This Comment endeavors to uncover whether Google's use of caching provides a fair use safe harbor for its Library Project and, if not, whether it should be exempt based on the DMCA's caching safe harbors. Part II details the use of caching, both locally and on the Internet, discussing the technical details and its great public utility. Part III moves on to briefly outline the Google Book Search Project (including the Library Project), the related litigation, and the conventional fair use analysis. Part IV delves into the current state of copyright law as it relates to caching and ISPs, pointing to the DMCA's safe harbors. Finally, Part V examines the Copyright Act, Congress's reaction to technological advances and, in addition, suggests a novel use of the fair use doctrine and a possible modification to the DMCA.

## II. CACHING IN

### A. *Cache* (pronounced 'kash,' like 'cash'). *Function: noun. Definition: a hiding place for treasure or provisions.*<sup>24</sup>

In computer science, "caching" refers to the temporary storage of duplicated data or instructions, in a place where it can be easily and quickly accessed for future use.<sup>25</sup> For example, imagine your home computer as an office that contains a filing cabinet, a desk, and a bulletin board. The storage space on your hard drive is the filing cabinet where work is stored when not in use. The functional memory, or Random Access Memory (RAM), is the desk where current work is open and accessible. Finally, the cache memory is the bulletin board where frequently used items can be placed for even quicker access.<sup>26</sup>

Using the analogy above, the filing cabinet provides high-capacity storage, but access times are longer; the desk provides a location for current work, yet there is less storage space. As you go back and forth, taking files out of storage in the filing cabinet, working on

112 Stat. 2860 (1989); *see also* H.R. REP. NO. 94-1476, at 47 (1976) (discussing that Congress' intent when drafting the DMCA was, inter alia, to protect innovation and technology).

24. OXFORD AMERICAN DICTIONARY AND LANGUAGE GUIDE 128 (1999). Although the pronunciation "ka-shā," like cashay, is sometimes heard in English, it is a mispronunciation of the French word, *cache*—"to hide." Dictionary.com, <http://dictionary.reference.com/search?q=cache> (last visited Aug. 25, 2008). In addition, the Free On-line Dictionary of Computing defines "cache" as: "[a] small fast memory holding recently accessed data, designed to speed up subsequent access to the same data." Free On-line Dictionary of Computing, <http://foldoc.org/index.cgi?query=cache&action=Search> (last visited Aug. 25, 2008) ("Most often applied to processor-memory access but also used for a local copy of data accessible over a network etc.").

25. LINDA NULL & JULIA LOBUR, *THE ESSENTIALS OF COMPUTER ORGANIZATION AND ARCHITECTURE* 237-39 (Jones and Bartlett 2003).

26. Don5408's Unofficial Aptiva Support Site, Drivespace FAQ – Memory vs. Drive Space: The Difference Between RAM and Storage Space, [http://members.aol.com/don5408/drivespace/mem\\_v\\_dspace.html](http://members.aol.com/don5408/drivespace/mem_v_dspace.html) (last visited Aug. 25, 2008) (explaining the use of cache with a similar analogy).

them at your desk and returning them, you may choose to stick some of the information you are repeatedly referring to, or a copy thereof, on the bulletin board. Next time you need that information, rather than going all the way to the filing cabinet, all you have to do is look up on the bulletin board. The more often you use the information on your bulletin board, the more time you save by not digging through the filing cabinet.

### B. *The Nuts and Bolts*

The technical details of how and why the cache system works are rather simple. Computers work in binary, a language consisting of sequential ones and zeros.<sup>27</sup> Computers are machines with prescribed procedures set by their code; binary sequences consist of signals sent over wires in the form of differing voltages.<sup>28</sup> Typically, a five volt signal represents a “one” and a zero volt signal represents a “zero.”<sup>29</sup> Resistors, transistors, and capacitors<sup>30</sup> are used by microprocessors<sup>31</sup> to store the values of the signals being transmitted through the wires. Although all three are essential for microprocessor operations, the capacitor is especially important within the context of this discussion.<sup>32</sup> The capacitor is responsible for storing the signal, either “one” or “zero,” and makes it possible for data storage and manipulation—the essence of computer processing.<sup>33</sup>

However, not all signals can be maintained in capacitors without the constant supply of electricity.<sup>34</sup> Consequently, two types of mem-

27. NULL & LOBUR, *supra* note 25, at 38.

28. *Id.* at 68.

29. *Id.*

30. “Resistors provide resistance, transistors perform switching, [and] capacitors store charge.” Edward D. Manzo et al., *A Panel Discussion on Obviousness in Patent Litigation: KSR International v. Teleflex*, 6 J. MARSHALL REV. INTELL. PROP. L. 595, 619 (2007). Although all three components are essential for microprocessor operations, the capacitor is responsible for the storage of “state.” See GARY DUNNING, INTRODUCTION TO PROGRAMMABLE LOGIC CONTROLLERS 295 (Thomson Delmar Learning 2001).

31. Generally, microprocessors are the programmable digital components of a computer’s Central Processing Unit (CPU). See WILLIAM STALLINGS, COMPUTER ORGANIZATION AND ARCHITECTURE: DESIGNING FOR PERFORMANCE 37 (5th ed., Prentice Hall) (2000). Several microprocessors will make up one CPU. *Id.* The microprocessor was developed by Intel in 1971. *Id.* By 2003, nearly \$43 billion worth of microprocessors were manufactured and sold. Press Release, World Semiconductor Trade Statistics, WSTS Semiconductor Market Forecast (Oct. 28, 2003), available at <http://wsts.www5.kcom.at/public/pressrelease/pr03-10.pdf> (last visited Aug. 25, 2008).

32. In fact, many scientists consider the transistor to be the greatest invention of the twentieth century. DENNIS F. HERRICK, MEDIA MANAGEMENT IN THE AGE OF GIANTS: BUSINESS DYNAMICS OF JOURNALISM 383 (2003). However, the capacitor’s ability to store electric charges has led to the ability of computers to store data. *Id.* at 312; see also David Bondurant & Fred Gnadinger, *Ferroelectrics for Nonvolatile RAMs*, IEEE SPECTRUM, July 1989, at 30.

33. See STALLINGS, *supra* note 31, at 147.

34. NULL & LOBUR, *supra* note 25, at 234.

ory storage have been developed: volatile and non-volatile memory. Non-volatile memory can be maintained without constant electricity, while volatile memory is erased with the loss of electricity.<sup>35</sup> Hard drive memory, which acts as storage, is non-volatile.<sup>36</sup> When the computer is turned off, the state<sup>37</sup> of the capacitors is maintained and memory is not lost. This explains why one can reboot or shut down a computer for the night without losing information saved on the hard drive. However, RAM and cache are volatile memory; when the computer reboots, the cache and RAM are both reset, and the capacitors return with no state information.<sup>38</sup> When a computer is turned off, the data in both RAM and cache are erased, leaving these memory locations available for new data.<sup>39</sup>

Since capacitor states are determined by electrical signals, they are therefore subject to natural physical limitations. Computer speeds are limited to signal speeds which, due to the laws of physics, cannot exceed the speed of light (approximately 186,300 miles per second).<sup>40</sup> Hence, larger memory storage devices take longer to access information than do smaller devices because the signals must travel farther. For example, data access times for a 120-gigabyte storage device (hard drive) are much greater than for one-gigabyte storage devices (RAM), which in turn, are greater than for two-megabyte storage devices (cache). The trichotomy of information storage (made up of hard drives, RAM, and cache) has led to the development of computer systems that rely on the hierarchy of memory.<sup>41</sup> The analogy in the previous section is helpful in understanding this hierarchy. The filing cabinet takes the longest to retrieve documents, then the desk, followed by the bulletin board.

---

35. Both volatile and non-volatile memory are used in microprocessors as well as in Programmable Logic Controllers (PLCs), which are generally used for large industrial processes. See DUNNING, *supra* note 30, at 295 (discussing the use of volatile and non-volatile memory in the functioning of PLCs).

36. See NULL & LOBUR, *supra* note 25, at 623 (defining “ROM,” or “Read-Only Memory”).

37. “State” is a term of art in computer science and electrical engineering. See STALLINGS, *supra* note 31, at 243. “State” refers to whether the capacitor (or memory location) is a “one” bit, a “zero” bit, or empty (not set). *Id.*

38. NULL & LOBUR, *supra* note 25, at 234.

39. *Id.*

40. The speed of light is 299,792,458 meters/second, which is approximately 186,300 miles per second. National Institute of Standards and Technology, CODATA Value: Speed of Light in a Vacuum, <http://physics.nist.gov/cgi-bin/cuu/Value?c> (last visited Aug. 25, 2008).

41. See NULL & LOBUR, *supra* note 25, at 236.

### C. *The Theory*

Caching is based on the hierarchy of memory storage devices and their relative access speeds.<sup>42</sup> There are different types of RAM in a computer; both static RAM (SRAM) and dynamic RAM (DRAM)<sup>43</sup> exist in one system.<sup>44</sup> The latter is much slower but provides the same functionality. Why is not all RAM in a machine SRAM? The answer is simple: cost. DRAM is much less expensive to produce and use than SRAM.<sup>45</sup>

Therefore, computer engineers employ the use of SRAM for making cache copies of data while DRAM is used for most other RAM functions.<sup>46</sup> The implementation of the cache system has proven extremely beneficial in the overall efficiency of computer systems, as can be seen by a simple computation:

Assume access to main memory takes 200 cycles and access to the cache memory take[s] 15 cycles. Then code using 100 data elements 100 times each will spend 2,000,000 cycles on memory operations if there is no cache and only 168,500 if all data can be cached. That is an improvement of 91.5%.<sup>47</sup>

### D. *Who is Caching In?*

Caching is employed to allow personal computers to more efficiently and effectively handle data internally.<sup>48</sup> However, personal computers are not the only systems that can benefit from caching. ISPs, search engines, the Domain Name System (DNS),<sup>49</sup> dedicated

42. *Id.*

43. Although synchronous dynamic random access memory (SDRAM) and asynchronous dynamic random access memory (DRAM) are similar in functionality, they are not synonymous nor are they the same. Ulrich Drepper, *What Every Programmer Should Know About Memory, Part I*, LINUX WKLY. NEWS, Sept. 21, 2007, <http://lwn.net/Articles/250967/>.

44. *Id.*

45. *Id.*

46. Ulrich Drepper, *Memory Part 2: CPU Caches*, LINUX WKLY. NEWS, Oct. 1, 2007, <http://lwn.net/Articles/252125/>.

47. *Id.*

48. See NULL & LOBUR, *supra* note 25, at 236.

49. The Domain Name System serves as the Internet's "phone book" by translating Uniform Resource Locators (URLs), or more simply "domain names," into Internet Protocol (IP) addresses. See generally ZYTRAX, INC., DNS FOR ROCKET SCIENTISTS (Dec. 1, 2007), <http://www.zytrax.com/books/dns/> (discussing, in detail, DNS concepts and technical information). For example, when you type [www.google.com](http://www.google.com) (the URL) into your Web browser, the Domain Name System translates the text into 209.85.165.104 (the IP address). *Id.*; see also Daniel Karrenberg, DNS Root Name Servers Explained for Non-Experts (Sept. 2007), <http://www.isoc.org/briefings/019/briefing19.pdf>; Tim Berners-Lee, Uniform Resource Identifier (URI): Generic Syntax (Jan. 2005), <http://www.ietf.org/rfc/rfc3986.txt>. To find the IP address of any Web site while using Microsoft Windows XP, simply type "ping" followed by the URL (e.g., ping [www.google.com](http://www.google.com)) at the command prompt. To access the command prompt follow: Start Menu > All Programs > Accessories > Command Prompt, or follow: Start Menu > Run and type "cmd" in the text field.

caching services and even attorneys utilize cache systems to maximize efficiency as well.

### 1. *Personal Computers*

Personal computers always use caching.<sup>50</sup> In fact, caching was designed specifically for use with personal computers.<sup>51</sup> The ways in which personal computers employ caching can be broken into two main categories: Operating System (OS) caching and Web browser caching.<sup>52</sup>

OS caching entails the use of both the SRAM and the DRAM.<sup>53</sup> Among other things, the DRAM stores the software which is currently running on the computer.<sup>54</sup> This includes software such as Operating Systems,<sup>55</sup> word processors,<sup>56</sup> and Web browsers.<sup>57</sup> The use of the DRAM is much like the use of the desk in the earlier analogy. The DRAM's access time is much faster than the hard drive's; therefore, the Central Processing Unit (CPU)<sup>58</sup> can run the software instructions faster from that location. The use of SRAM in the personal computer can be analogized to the bulletin board from earlier. When

50. See NULL & LOBUR, *supra* note 25, at 250. For evidence of personal computer caching, both OS caching and Web browser caching, type "about:cache" in the address bar of a Web browser. The screen will display a list of cached Web sites made by the computer. One list displays the Web sites cached by the Web browser in a folder located on the hard drive. The other list displays the Web sites cached by the computer in SRAM.

51. G.C. Stierhoff & A.G. Davis, *A History of the IBM Systems Journal*, 20:1 IEEE ANNALS HIST. COMPUTING 31 (Jan. 1998).

52. As the name signifies, the operating system is the software that operates every personal computer, e.g., Microsoft Windows, Mac OS, and LINUX. NELL B. DALE, *COMPUTER SCIENCE ILLUMINATED* 320 (Jones and Bartlett 2006). A web browser is generally the software that translates html code and allows users to access the Internet, e.g., Mozilla Firefox, Internet Explorer, and Netscape Navigator. *Id.* at 481.

53. See NULL & LOBUR, *supra* note 25, at 250.

54. *Id.*

55. Technically, only one operating system can be running at a time, although many personal computers have multiple operating systems installed. For example, as of 2006, the hard drives of Apple's MacBook Notebook computers are capable of being partitioned with Microsoft Windows XP or Vista installed on one partition and OS X installed on the other partition. See Apple.com, <http://www.apple.com/getamac/windows.html> (last visited Aug. 25, 2008).

56. Word processors, formerly known as document preparation systems, are software applications such as Microsoft Word, Corel WordPerfect, and OpenOffice Writer that perform, inter alia, text editing and text formatting. George Rotsky, *The Word Processor: Cumbersome, but Great*, EE TIMES ONLINE, [http://www.eetimes.com/special/special\\_issues/millennium/milestones/berezin.html](http://www.eetimes.com/special/special_issues/millennium/milestones/berezin.html) (last visited Aug. 25, 2008).

57. Web browsers are software tools that allow users to access the Internet. DALE, *supra* note 52, at 481 ("A [Web] browser is a software tool that issues the request for the Web page we want and displays it when it arrives.").

58. The CPU, or sometimes simply "processor," is the central component of the computer capable of and responsible for executing programs as well as maintaining priority among software applications. Gary D. Knott, *A Proposal for Certain Process Management and Intercommunication Primitives*, 8:4 ACM SIGOPS OPERATING SYSTEMS REV. 8 (Oct. 1974). Inter alia, the CPU interprets program instructions and processes data. *Id.*

you know you will be using something often, you put the information, or a copy of it, on the bulletin board to save time. Similarly, the CPU and the software work together to determine locality of reference<sup>59</sup> and copy frequently used instructions or data in the SRAM for quicker access.<sup>60</sup>

Web browser caching,<sup>61</sup> although similar in principle to OS caching, is slightly different in its application. Web browser caching, often called client caching,<sup>62</sup> occurs when a user attempts to access a Web site<sup>63</sup> on the Internet. While attempting access, the Web browser will quickly search the user's computer to determine if a copy of the Web site exists locally.<sup>64</sup> If a copy exists (for example, if the user recently accessed that same site), the Web browser will display the local

59. Locality of reference is a general term in computer science referring to whether data is located close together in either space or time. NULL & LOBUR, *supra* note 25, at 237. This Comment primarily deals with data that is accessed close together in time (temporal locality) rather than data located physically close to each other (spatial locality) or data located in a sequence (sequential locality). *Id.*

60. NULL & LOBUR, *supra* note 25, at 238.

61. Some Computer Scientists consider Web browser caching a subset of Internet caching. S.V. NAGARAJ, WEB CACHING AND ITS APPLICATIONS 3-6 (2004). However, for purposes of this Comment, Web browser caching will be considered personal computer caching because its function is dependent on the personal computer's local RAM memory storage.

62. *See, e.g.*, Tamber Christian, *Internet Caching: Something to Think About*, 67 UMKC L. REV. 477, 477 (1999) (discussing caching as it relates to ISPs' liability for "unauthorized transmissions or displays of copyrighted materials on their networks").

63. There are several different spellings for this term. Although the terms "website" and "web site" are commonly used, The Chicago Manual of Style, The New Yorker, and dictionaries such as Merriam-Webster use the two-word, initially capitalized spelling "Web site." This is because "Web" is not a general term but a shortened version of "World Wide Web." *See* The Chicago Manual of Style Online, [http://www.chicagomanualofstyle.org/CMS\\_FAQ/InternetWebandOtherPost-WatergateConcerns/InternetWebandOtherPost-WatergateConcerns14.html](http://www.chicagomanualofstyle.org/CMS_FAQ/InternetWebandOtherPost-WatergateConcerns/InternetWebandOtherPost-WatergateConcerns14.html) (last visited Aug. 25, 2008) (explaining the formal usage of the word "Web site" and its subsequent use by The New Yorker); *see also* Merriam-Webster Online Dictionary, <http://www.m-w.com/dictionary/website> (last visited Aug. 25, 2008) (defining "Web site" as "a group of World Wide Web pages usually containing hyperlinks to each other and made available online by an individual, company, educational institution, government, or organization"). In addition, Bill Walsh, the copy chief of The Washington Post's national desk and one of American English's foremost grammarians, argues for the two-word spelling with the capital "W." BILL WALSH, LAPSING INTO A COMMA: A CURMUDGEON'S GUIDE TO THE MANY THINGS THAT CAN GO WRONG IN PRINT – AND HOW TO AVOID THEM 14-15 (2000) ("[I]f the Internet is replaced by direct-broadcast-to-brain technology tomorrow, *website* will soon look as silly as *draftdodging* and *braburning* and *goldfishswallowing*."); *see also* BILL WALSH, THE ELEPHANTS OF STYLE: A TRUNKLOAD OF TIPS ON THE BIG ISSUES AND GRAY AREAS OF CONTEMPORARY AMERICAN ENGLISH 213-14 (2004) (implying that "website" is not a legitimate compound form and that "site" is not a legitimate suffix, yet "web-crawler" is a legitimate compound); Bill Walsh, *Sharp Points: Here We Go Again–Eeee!*, THE SLOT.COM, <http://www.theslot.com/email.html> (last visited Aug. 25, 2008) (criticizing authors who attempt to compound the term "Web site" into the improper, lowercased, single word, "website").

64. *See* NULL & LOBUR, *supra* note 25, at 282; *see also* Microsoft Corp., Internet Explorer 6: How and Why to Clear Your Cache, <http://www.microsoft.com/windows/ie/using/howto/customizing/clearcache.mspx> (last visited Aug. 25, 2008) [hereinafter Microsoft, How and Why].

copy.<sup>65</sup> This will take a fraction of the time required to retrieve the Web site from the original location on the Internet. However, if a local copy does not exist on the computer, the Web browser will retrieve the site from the Internet and simultaneously store a copy in RAM.<sup>66</sup> The simultaneous copying to RAM ensures that the next time an attempt is made to access that particular Web site during the same session, the cached copy will be available for immediate retrieval.<sup>67</sup>

## 2. Internet Service Providers

ISPs routinely use caching for two reasons: to provide more rapid data retrieval for its users and to provide streamlined access to popular Web sites.<sup>68</sup> By providing customers with rapid data retrieval, ISPs are able to charge competitive rates.<sup>69</sup> By providing streamlined access, they are able to reduce the chances of bandwidth overflow.<sup>70</sup> Obviously, both reasons are economic since Internet users usually have multiple ISPs to choose from. An ISP that does not use caching consequently risks losing customers to one that does.

ISP caching, also known as proxy caching,<sup>71</sup> occurs when a network server stores a copy of a Web site on its own network.<sup>72</sup> Much like Web browser caching, when a user attempts to connect to a certain Web site, the network will first check its own network server to see if a copy of the Web site exists locally.<sup>73</sup> If so, the server will display the cached version.<sup>74</sup> This will be substantially faster than accessing the Web site from its location on the Internet, especially if the Web site's server is far away. However, if the requested site does not exist on the network, the ISP will access the Web site on the Internet and save a copy in its cache.<sup>75</sup> Much like the client caching

---

65. See Microsoft, How and Why, *supra* note 64.

66. The differences between SRAM and DRAM are only significant when discussing personal computer caching; therefore, both SRAM and DRAM will be hereinafter called RAM.

67. See NAGARAJ, *supra* note 61, at 3-6.

68. *In re Inquiry Concerning High-Speed Access to the Internet and over Cable and Other Facilities*, 17 F.C.C.R. 4798, 4810 n.76 (2002).

69. See Hal R. Varian, Local Exchange Congestion and Internet Service Providers (Mar. 1997), <http://people.ischool.berkeley.edu/~hal/Papers/phone.html>.

70. Bandwidth overflow, or congestion, refers to having more Web traffic on a physical line than bandwidth. See *id.* This can happen whenever there is excessive traffic, such as when a specific Web site gains popularity very quickly. See *id.* Bandwidth overflow commonly results in Web site unresponsiveness and even data loss. See *id.*

71. Christian, *supra* note 62, at 478.

72. RON WHITE & TIMOTHY EDWARD DOWNS, HOW COMPUTERS WORK 339 (9th ed. 2007).

73. *Id.*

74. *Id.*

75. *Id.* at 340.

scenario, the next time the Web site is requested by any user, it will be available locally.

When proxy caching is used, users get their copies from the proxy server instead of the Web site's server.<sup>76</sup> Thanks to this use of caching, the Internet is able to run faster and smoother.<sup>77</sup> However, Web pages can change frequently. News Web sites, for example, often change every few minutes.<sup>78</sup> Therefore, network servers must regularly search the Internet and continuously update the cached copies.<sup>79</sup>

It is fairly obvious how beneficial this system is, especially considering the quantity of hits received by some Web sites each day.<sup>80</sup> For example, MyDeathSpace.com,<sup>81</sup> a semi-spoof Web site devoted to "connecting" the deaths of MySpace.com<sup>82</sup> users, reportedly receives more than 100,000 hits per day.<sup>83</sup> Even if ISP caching only saved, on average, 1/10th of a second in accessing MyDeathSpace.com, the ISP would net a savings of two hours and forty-five minutes per day just for this one Web site.<sup>84</sup> Compare this with nearly a 10,000-fold savings on MySpace.com itself, where estimates are near the one billion hit per day mark, and the great benefit of ISP caching becomes relatively clear.<sup>85</sup>

76. NAGARAJ, *supra* note 61, at 3.

77. See Nat'l Cable & Telecomms. Ass'n v. Brand X Internet Servs., 545 U.S. 967, 999-1000 (2005) (noting that "[c]acheing obviates the need for the end user to download anew information from third-party Web sites each time the consumer attempts to access them, thereby increasing the speed of information retrieval").

78. See Matthew Fagan, "Can You Do a Wayback on That?" *The Legal Community's Use of Cached Web Pages In and Out of Trial*, 13 B.U. J. SCI. & TECH. L. 46, 50 (2007).

79. See Richard S. Vermut, *File Caching on the Internet: Technical Infringement or Safeguard for Efficient Network Operation?*, 4 J. INTELL. PROP. L. 273 (1997) (discussing, *inter alia*, the frequency of ISP caching).

80. Web site "hits" refer to the number of times a Web site is accessed, usually at a daily rate. See ZDNet, Hits: Definition and Additional Resources, <http://dictionary.zdnet.com/definition/hits.html> (last visited Aug. 25, 2008).

81. MyDeathSpace.com is a California based Web site that "collects the MySpace profiles of dead people and links them to news stories, obituaries or blog posts that detail their lives and deaths." Paul Sand, *MyDeathSpace.com Memorializes Youths*, NEWS TRIB. (Tacoma, WA), Apr. 17, 2006, available at <http://dwb.thenewstribune.com/news/local/story/5668473p-5083642c.html>.

82. MySpace.com is a social networking Web site where users can connect with friends and other users with similar interests. MySpace.com is currently the world's sixth most popular Web site. See Alexa, MySpace.com - Traffic Details, [http://www.alexa.com/data/details/traffic\\_details/myspace.com](http://www.alexa.com/data/details/traffic_details/myspace.com) (last visited Aug. 25, 2008).

83. See Associated Press, *Site Archives Dead MySpace Members*, Aug. 3, 2007, available at <http://www.cnn.com/2007/TECH/07/30/dying.on.the.web.ap/index.html>.

84. This 1/10th of a second savings is complete conjecture. It would be nearly impossible to calculate how much time would be saved for any particular Web site given the number of possible variables, the only constant being the number of hits per day and the distance between the network and publishing server.

85. In fact, as of April 2007, it was reported that MySpace.com was receiving between 39 and 45 billion hits per month. Scott Elkin: Myspace Statistics (May 11, 2007),

### 3. Search Engines

Search engines, such as Google, Yahoo!, and Lycos, cache Web sites as well. Most search engines function by employing software called a “web spider” or “web crawler”<sup>86</sup> which visits every Web site it can find, indexing the information it finds into a matrix of data. Indices rank from very small lists of specific keywords to very large lists consisting of every word on the site.<sup>87</sup> When a user makes a query, the search engine checks the search text against the keywords in the database and returns a list of Web pages containing some or all of the search text.<sup>88</sup> The search engine employs an algorithm to determine the display order of the Web pages found; for example, pages with more occurrences of search terms may display above pages with fewer search terms present.<sup>89</sup> In this way, more relevant Web pages are more likely to be listed first.<sup>90</sup>

Some search engines, such as Google, copy the entire Web page instead of simply a list of keywords.<sup>91</sup> Google’s use of expansive indices and its PageRank algorithm has continued to provide searches relevant to users. Many believe this has made Google the search engine of choice.<sup>92</sup> Furthermore, by storing cached copies of entire Web pages, Google is able to more quickly and efficiently make the neces-

---

<http://scottelkin.com/archive/2007/05/11/Myspace-Statistics.aspx>. That equates to roughly between 1.3 and 1.5 billion hits per day. *Id.*

86. See *Perfect 10 v. Google, Inc.*, 416 F. Supp. 2d 828, 832 (C.D. Cal. 2006) (“Google’s search engine indexes websites on the internet via a web ‘crawler,’ *i.e.*, software that automatically scans and stores the content of each website into an easily-searchable catalog.”); see also *eBay, Inc. v. Bidder’s Edge, Inc.*, 100 F. Supp. 2d 1058, 1061 n.2 (N.D. Cal. 2000) (“Programs 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.”). For the remainder of this Comment, “web crawlers” will be referred to as “webcrawlers.” See WALSH, *THE ELEPHANTS OF STYLE*, *supra* note 63, at 213. However, this should not be confused with the search engine called WebCrawler, which is a “metasearch” engine that compiles the results of all the major search engines for each particular search query. See Webcrawler.com, About WebCrawler, [http://www.webcrawler.com/webcrawler/ws/about/?\\_iceUrlFlag=11?\\_IceUrl=true](http://www.webcrawler.com/webcrawler/ws/about/?_iceUrlFlag=11?_IceUrl=true) (last visited Aug. 25, 2008) (“Using metasearch technology, WebCrawler takes results from the leading search engines (Yahoo Search!, Google, MSN, Ask), eliminates the duplicates and delivers you the most comprehensive set of results.”).

87. See Danny Sullivan, *How Search Engines Work*, SEARCH ENGINE WATCH, Mar. 14, 2007, <http://searchenginewatch.com/showPage.html?page=2168031>.

88. See *id.*

89. See *id.*

90. See *id.*

91. See Googleguide.com, *How Google Works*, [http://www.googleguide.com/google\\_works.html](http://www.googleguide.com/google_works.html) (last visited Aug. 25, 2008) [hereinafter GoogleGuide].

92. See, e.g., Richard MacManus, *Google Continues Search Engine Dominance—Even in Verticals!*, READWRITEWEB.COM, Apr. 11, 2007, [http://www.readwriteweb.com/archives/google\\_continues\\_search\\_dominance\\_incl\\_verticals.php](http://www.readwriteweb.com/archives/google_continues_search_dominance_incl_verticals.php); Kimberly Powell, *Googling Genealogy Style: 12 Google Search Tips for Genealogists*, <http://genealogy.about.com/library/weekly/aa052902a.htm> (last visited Aug. 25, 2008).

sary calculations for the PageRank.<sup>93</sup> Otherwise, each search query would require Google to “crawl” the entire Web.

In terms of search engines, particularly Google, caching is used twice.<sup>94</sup> When the web crawler copies Web pages, the pages are stored in cache.<sup>95</sup> Although the pages are saved on a server, which serves as more of a hard drive than a cache memory device, the data is only stored temporarily.<sup>96</sup> Upon the next “crawl” of the Web, the pages are replaced with newer versions.<sup>97</sup> In addition, the “crawl” is done automatically.<sup>98</sup> This type of use (automatic and temporary) has been defined as caching.<sup>99</sup> The Web pages stored on the search engine’s servers are indexed in a giant matrix for faster and easier access by the search engine.<sup>100</sup>

Similarly, when a user makes a search request, the search engine again employs caching.<sup>101</sup> The search engine makes a temporary copy of its own data in case the user requests a similar search query.<sup>102</sup> Although this automatic storage of data is more temporary than the previous example, both forms of data storage are considered caching.<sup>103</sup> Search engine effectiveness is inextricably linked to this form of caching.<sup>104</sup> Thus, caching enables search engines to provide an incredibly effective and socially useful tool for information access.

#### 4. *The Domain Name System*

The Domain Name System (DNS) is responsible for translating alphanumeric domain names (URLs) into Internet Protocol (IP) addresses.<sup>105</sup> Hence, Internet users need not memorize eight to twelve digit numeric Web site addresses. Can you imagine having to type

93. See Craven, *supra* note 8.

94. See Tiziano Fagni & Fabrizio Silvestri, *Hybrid Caching of Search Engine Results*, 52 ERCIM NEWS (Jan. 2003), [http://www.ercim.org/publication/Ercim\\_News/enw52/silvestri.html](http://www.ercim.org/publication/Ercim_News/enw52/silvestri.html); see also Evangelos P. Markatos, *On Caching Search Engine Query Results*, PROC. OF THE 5TH INT’L WEB CACHING AND WEB DELIVERY WORKSHOP (May 2000), <http://www.ics.forth.gr/carv/r-d-activities/wwwPerf/TR241/paper.html>.

95. See Andrei Z. Broder et al., *Efficient URL Caching for World Wide Web Crawling*, PROC. OF THE 12TH INT’L WORLD WIDE WEB CONF. (May 2003), available at <http://www2003.org/cdrom/papers/refereed/p096/p96-broder.html>.

96. *Id.*

97. *Id.*

98. *Id.*

99. Section 512(b) of the DMCA provides ISP safe harbors for “system caching.” 17 U.S.C. § 512(b) (2006) (setting out the safe harbor exceptions for Internet service providers which utilize system caching that is, inter alia, automatic and temporary).

100. See Fagni & Silvestri, *supra* note 94.

101. See Broder et al., *supra* note 95.

102. See NULL & LOBUR, *supra* note 25, at 237.

103. See, e.g., *id.*; Broder et al., *supra* note 95; Fagni & Silvestri, *supra* note 94.

104. Broder et al., *supra* note 95.

105. See Zytrax.com, Chapter 2 The DNS Context, <http://www.zytrax.com/books/dns/ch2/> (last visited Aug. 25, 2008); see also Karrenberg, *supra* note 49.

63.111.69.121 every time you want to check the weather online? Instead, you can easily remember [www.weather.com](http://www.weather.com).<sup>106</sup> In this way, the DNS is much like the Internet's "phone book." Due to the ease of use and resulting popularity of search engines, the DNS may not be as valuable as it once was;<sup>107</sup> however, the DNS is still an example of the social benefit of caching.

Much like the personal computer, the DNS's structure is hierarchical in nature.<sup>108</sup> This ensures that each domain name remains unique.<sup>109</sup> At the top of the hierarchy are 258 top-level domains (TLDs).<sup>110</sup> TLDs come in three types:<sup>111</sup> generic top-level domains (gTLDs) such as [.com](http://www.com) and [.org](http://www.org), which are not associated with any region or country;<sup>112</sup> country specific top-level domains (ccTLDs), such as [.uk](http://www.uk) (United Kingdom), [.au](http://www.au) (Australia), or [.jp](http://www.jp) (Japan);<sup>113</sup> and infrastructure related top-level domains which are not relevant for the typical user.<sup>114</sup> These different types of TLDs provide a mechanism for name servers to recognize Web sites that may be requested by Internet users.

Recall that domain names are alphanumeric representations of associated IP addresses. When a domain name is typed into a Web browser, either a local name server "translates" the domain name into its associated IP address or a name server higher up the chain will complete the translation.<sup>115</sup> Much like proxy (or ISP) caching, the

---

106. By typing the IP address into a Web browser's address bar, a user will be taken to the associated Web site. In this particular example, typing 65.212.118.121 will take the user to <http://www.weather.com>. Today, most Web browsers will accept [www.weather.com](http://www.weather.com), or even simply [weather.com](http://weather.com). In addition, this process can be reversed; one can look up the associated IP address of a known URL. See Self SEO.com, Find IP Address of a Website – Server IP Lookup, [http://www.selfseo.com/find\\_ip\\_address\\_of\\_a\\_website.php](http://www.selfseo.com/find_ip_address_of_a_website.php) (last visited Aug. 25, 2008).

107. See STEPHEN M. MCJOHN, *INTELLECTUAL PROPERTY: EXAMPLES AND EXPLANATIONS* 339 (2006) (stating that "as domain names become less important as identifiers (instead acting simply as addresses). . . . [a] sophisticated Internet user looking for the Web site of the Acme Potato Co. is now less likely to simply type [www.acmepotatocompany.com](http://www.acmepotatocompany.com) into the browser's window for addresses. She is more likely to use a search engine to find the Web site.").

108. See DNS FOR ROCKET SCIENTISTS, *supra* note 49; see also Karrenberg, *supra* note 49.

109. See DNS FOR ROCKET SCIENTISTS, *supra* note 49.

110. Kim G. von Arx & Gregory R. Hagen, *Sovereign Domains: A Declaration of Independence of ccTLDs from Foreign Control*, 9 RICH. J.L. & TECH. ¶ 12 (2002), available at <http://law.richmond.edu/jolt/v9i1/Article4.html#H2>.

111. *Id.*

112. *Id.*; see also Internet Assigned Numbers Authority, Root Zone Database, <http://www.iana.org/gtld/gtld.htm> (last visited Aug. 25, 2008) (listing of generic TLDs).

113. von Arx & Hagen, *supra* note 110, at 12; see also Internet Assigned Numbers Authority, *supra* note 112 (listing of country specific TLDs).

114. von Arx & Hagen, *supra* note 110; see also Internet Assigned Numbers Authority, ARPA Zone Management, <http://www.iana.org/domains/arpa/> (last visited Aug. 25, 2008).

115. Techterms.com, Name Server Definition, [http://www.techterms.com/definition/name\\_server](http://www.techterms.com/definition/name_server) (last visited Aug. 25, 2008) ("A name server translates domain names into IP addresses.

Web browser will first check with the local name server because it is closer and therefore faster.<sup>116</sup> Each local name server contains physical cache memory locations for storing translated addresses.<sup>117</sup> If the local name server does not have the associated IP address stored in cache, it will forward the request up the chain until a name server responds with the translation.<sup>118</sup> When this occurs, the local name server will send the user to the Web site and store a copy of the translation in its cache memory.<sup>119</sup> Often the local name server already knows which IP address corresponds to the entered domain name, for it is stored in cache, and the user is connected to the Web site much more quickly.<sup>120</sup>

### 5. Caching Services (Google, Yahoo! & Archive.org)

Cached Web pages are made available to end users by several companies for various purposes. When users access a popular Web page, they are typically viewing their ISPs, some other ISPs, or their own Web browser's cached copy.<sup>121</sup> However, these cached copies do not last long: Web pages typically tell the proxy server how long to store a cached version before that version becomes "stale" and requires reloading.<sup>122</sup> This allows for some cache individualization.<sup>123</sup> For example, a monthly blog<sup>124</sup> would only need to be re-cached once a month, while a highly dynamic site like The New York Times' homepage<sup>125</sup> might need to be updated every few minutes. The more often a page is re-cached, the greater the chance that when a user

---

This makes it possible for a user to access a website by typing in the domain name instead of the website's actual IP address. For example, when you type in 'www.microsoft.com,' the request gets sent to Microsoft's name server which returns the IP address of the Microsoft website.").

116. von Arx & Hagen, *supra* note 110, at 14.

117. *Id.* at 79.

118. *Id.* at 14, 79.

119. *Id.*

120. *Id.*

121. See RON WHITE, HOW COMPUTERS WORK 339-41 (2004).

122. Mark Nottingham, Caching Tutorial for Web Authors and Webmasters, [http://www.mnot.net/cache\\_docs/](http://www.mnot.net/cache_docs/) (last visited Aug. 25, 2008) [hereinafter Caching Tutorial].

123. See Fagan, *supra* note 78, at 51-52 (discussing caching services in general).

124. The term "blog" is a variation from the original "Web logs" that were combinations of collected links of personal commentary found around the Web. See David Gulbransen, *Welcome to the Blawgosphere*, CBA REC., Apr. 2006, at 3, 37 ("On the surface, a blog is nothing more than a Web site that is updated frequently and offers different mechanisms for reading the content other than a traditional web browser. The 'blogger' writes a short article or blurb using blog software to 'post' the entry to a Web site.").

125. The *New York Times* online edition, NYTimes.com, was the most popular online newspaper site in 2005, reaching 11,405 unique viewers per day, making it both dynamic and popular. See ZDNet.com, Most Popular Newspaper Sites: NY Times, USA Today, Washington Post (Nov. 16, 2005), <http://blogs.zdnet.com/ITFacts/?p=9508>. If it were not re-cached regularly, viewers would not be reading the most recent news stories and would likely turn elsewhere for their news. See Caching Tutorial, *supra* note 122.

requests a page, that cached page will actually be the most recent version.<sup>126</sup>

On the other hand, search engines only update their cached copies when they crawl the Web.<sup>127</sup> This happens far less often than ISPs update their cached copies.<sup>128</sup> Therefore, a cached version of a page found through a search engine looks further back into a Web page's "past" than does a cached version by an ISP. A user viewing a search engine's cached copy<sup>129</sup> will likely see the Web page as it existed some time in the past, rather than the way it looks at the time of the search. These "old" versions of Web pages are particularly useful to those seeking to view Web pages that have changed or no longer exist. Understanding the utility of such "old" cached copies, some search engines have made them available to the searching public in addition to the most "current" cached copies.

Three primary sources for viewing cached Web pages exist: Google, Yahoo!, and the Internet Archive.

(a) *Google*

In 1997, Google introduced a feature that allowed users to access copies of almost any Web page within Google's index in the form it was in upon Google's last Web crawl.<sup>130</sup> The page accessed by the user could be anywhere from minutes to months old, depending on when Google last "crawled" that particular site.<sup>131</sup> For the first time, users could gain access to a Web page through a search engine even after it had been removed from the Internet.<sup>132</sup>

When a user performs a Google search, along with the results of the query comes a "cached" link below the results. Following that link leads the user to the cached snapshot stored on Google's server. Although popular, Google is not the only mainstream search engine to provide caching services.

---

126. *See id.*

127. *See* Google Cached Pages: What Are Cached Pages, [http://www.googleguide.com/cached\\_pages.html](http://www.googleguide.com/cached_pages.html).

128. *See* Vermut, *supra* note 79, at 308.

129. A search engine's cached copy is the internal copy that a search engine has stored on its server, which it uses to determine the relevance of a page, and is not the page that is retrieved when a user links to the actual search result. *See* Field v. Google, Inc., 412 F. Supp. 2d 1106, 1110-11 (D. Nev. 2006) ("When clicked, the 'Cached' link directs an Internet user to the archival copy of a Web page stored in Google's system cache, rather than to the original Web site for that page.") The former copies are the primary topic of this subsection, while the latter copies are the ISP, or browser, cached sites referred to previously.

130. Stefanie Olsen, *Google Cache Raises Copyright Concerns*, CNET NEWS, July 9, 2003, [http://www.news.com/2100-1032\\_3-1024234.html](http://www.news.com/2100-1032_3-1024234.html).

131. *Id.*

132. *Id.*

*(b) Yahoo!*

Yahoo! offers a similar service to Google called “My Web.”<sup>133</sup> Much like Google’s caching service, when a user searches the Web, a “cached” link is placed below the site’s description and takes the user to the page saved during Yahoo!’s last Web crawl.<sup>134</sup> However, Yahoo! provides its users with an additional feature. Upon performing a search, a user is presented with the search results, the appropriate “cached” link, as well as a “save” link.<sup>135</sup> Users can click the “save” link and take their own snapshot of the page. Next time they view this cached snapshot, the page will appear as it did when the user last clicked the “save” link.<sup>136</sup>

Essentially, MyWeb users can save Web pages to prevent them from being lost or overwritten during the next Web crawl.<sup>137</sup> However, users must log in to Yahoo! and manually save a Web page to prevent its loss; there is no way for users to view unsaved Web pages if the cached snapshot has already been replaced.<sup>138</sup>

*(c) The Internet Archive*

The “Wayback Machine” (named after Mr. Peabody’s time machine from *The Rocky and Bullwinkle Show*)<sup>139</sup> can be found on the Internet Archive’s Web site.<sup>140</sup> The service stores every snapshot taken from Web sites crawled by the Alexa webcrawler since 1996.<sup>141</sup> For example, the Wayback Machine contains 2734 versions of “www.cnn.com” from 2001, when it took snapshots as often as several times a day.<sup>142</sup> These snapshots are open to the public; anyone with an Internet connection and a Web browser can view them.

However, the Wayback Machine does not cache a copy of a page every time the page is updated and often pages are not added to the

133. See Yahoo! Search, My Web 2.0 BETA - FAQ - Yahoo!, <http://myweb2.search.yahoo.com/myresults/faq#1> (last visited Aug. 25, 2008).

134. *Id.*

135. *Id.*

136. *Id.*

137. *Id.*

138. See *id.* With MyWeb 2.0 Beta, it may be possible to see cached Web pages that the user did not, individually, save. *Id.* With the introduction of MyWeb 2.0’s “Community” feature, it will be possible for users to search the saved, cached Web pages of others in their community. *Id.* However, at least one user must still manually save the page in order for it to not be overwritten by the webcrawler. *Id.*

139. Heather Green, *A Library as Big as the World*, BUSINESSWEEK, Feb. 28, 2002, [http://www.businessweek.com/technology/content/feb2002/tc20020228\\_1080.htm](http://www.businessweek.com/technology/content/feb2002/tc20020228_1080.htm) (attributing the name of the “Wayback Machine” to *The Rocky and Bullwinkle Show*).

140. Internet Archive, <http://www.archive.org/index.php> (last visited Aug. 25, 2008).

141. Internet Archive, Frequently Asked Questions, <http://www.archive.org/about/faqs.php> (last visited Aug. 25, 2008) [hereinafter Internet Archive FAQ #1].

142. Internet Archive, Wayback Machine, [http://web.archive.org/web\\*/http://www.cnn.com](http://web.archive.org/web*/http://www.cnn.com) (last visited Aug. 25, 2008) (listing search results for “http://www.cnn.com”).

database for up to six months.<sup>143</sup> Rather, the Wayback Machine only caches a copy every time its webcrawler actually visits that page.<sup>144</sup> In addition, users may request to have their sites removed from the Wayback archive;<sup>145</sup> however, the Wayback may not honor all requests.<sup>146</sup> These limitations aside, the Wayback Machine provides a wealth of cached snapshots and is likely the most comprehensive archive of Web history to date. Over forty billion snapshots are available from among almost two petabytes<sup>147</sup> of data and the Wayback Machine is growing at a rate of twenty terabytes<sup>148</sup> per month.<sup>149</sup>

### 6. Attorneys & Law Enforcement

Lawyers and police often use Web caching services in their research. In fact, the practice has become so common that attorneys will often ask their assistants to “‘do a Wayback on that.’”<sup>150</sup> “The archives are most attractive to specialists in intellectual-property law—in particular, areas such as domain-name battles—and have been used by companies [such as] . . . Playboy Enterprises Inc.”<sup>151</sup> For example, in cybersquatting<sup>152</sup> cases, attorneys are beginning to use the Wayback Machine “as a matter of course.”<sup>153</sup>

143. Internet Archive FAQ #1, *supra* note 141.

144. *Id.* (noting that webpages are cached when the webcrawler crawls the Web and that “about 50% of all pages on the web . . . change[] from [the] previous visit”).

145. Internet Archive, The Internet Archive’s Policies on Archival Integrity and Removal, <http://www2.sims.berkeley.edu/research/conferences/aps/removal-policy.html> (last visited Aug. 25, 2008).

146. *Id.*

147. One petabyte is equal to approximately one million gigabytes. Internet Archive: Petabox, <http://www.archive.org/web/petabox.php> (last visited Aug. 25, 2008).

148. One terabyte is equal to approximately one thousand gigabytes; therefore, in comparison, one petabyte is equal to approximately one thousand terabytes. See Merriam-Webster.com, <http://meriamwebster.com/dictionary/terabyte> (last visited Aug. 25, 2008).

149. Internet Archive FAQ #1, *supra* note 141 (stating that “[the] Wayback Machine contains almost 2 petabytes of data and is currently growing at a rate of 20 terabytes per month”).

150. To “do a Wayback on that” refers to using the Wayback Machine to research past versions of Web sites related to litigation. David Kesmodel, *Lawyers’ Delight: Old Web Material Doesn’t Disappear*, WALL ST. J., July 27, 2005, at A1, available at [http://online.wsj.com/article\\_email/SB112242983960797010-H9je4Nglad4o52tbXmIbq6Jm4.html](http://online.wsj.com/article_email/SB112242983960797010-H9je4Nglad4o52tbXmIbq6Jm4.html) (“In 2003, . . . the company cited the Wayback Machine during a court hearing to prove that a defendant used the term ‘sex court’ on his Web site only after Playboy aired a TV show with the same name. In his defense, the site operator asserted he had been using the name months before. The case was settled midtrial.”).

151. *Id.*

152. Cybersquatting involves one party buying a domain name of a well-known product, company, or trademark before the owner of the product, company, or trademark can. The cybersquatter then offers the domain for sale at an inflated price. This activity is actionable under the Anticybersquatting Consumer Protection Act. 15 U.S.C. § 1125(d)(1) (2000).

153. Kesmodel, *supra* note 150.

Similarly, law enforcement may use Web caching services to investigate cyber crimes.<sup>154</sup> Tallahassee, Florida, Police Department investigator Kent Campbell stated that his office has used cached Web pages to investigate cyber crimes and, on one occasion in particular, it was helpful in acquiring useful evidence.<sup>155</sup> Unfortunately, the question of whether this evidence is admissible at trial has yet to be decided. In two well-known cases dealing with cached Web sites, the courts discussed the merits of using such pages; however, neither came down decisively on the issue.<sup>156</sup> However, in *Commonwealth v. Diodoro*, the court ruled that child pornography found in the defendant's Internet cache was sufficient to support a sexual abuse of children conviction.<sup>157</sup>

### III. THE GOOGLE BOOK SEARCH PROJECT AND THE GOOGLE LIBRARY PROJECT

#### A. *The Projects*

Since its humble beginnings as a research project in 1996,<sup>158</sup> Google has become the most popular search engine in the United States<sup>159</sup> and possibly the world. In fact, recently, the word "google" was added to the dictionary as a verb.<sup>160</sup> Google's success is due, in part, to its accurate PageRank system,<sup>161</sup> as well as its abundant

154. The Computer Crime & Intellectual Property Section of the U.S. Department of Justice "is responsible for implementing the Department's national strategies in combating computer and intellectual property crimes worldwide." U.S. Department of Justice, About the Computer Crime & Intellectual Property Section, <http://www.cybercrime.gov/ccips.html> (last visited Aug. 25, 2008). "The Computer Crime Initiative is a comprehensive program designed to combat electronic penetrations, data thefts, and cyberattacks on critical information systems." *Id.*

155. Kent Campbell, Tallahassee Police Dep't, Presentation to the Florida State University College of Law Cyber Law Seminar (Oct. 30, 2007).

156. *See* *Telewizja Polska USA, Inc. v. EchoStar Satellite Corp.*, No. 02 C 3293, 2004 WL2367740, at \*5-6 (N.D. Ill. Oct. 15, 2005); *see also* Kesmodel, *supra* note 150 (discussing the eventual mistrial of a Canadian murder case due to use of Wayback Machine evidence and a cybersquatting case involving Vodafone Group and related cached Web pages).

157. *Commonwealth v. Diodoro*, 932 A.2d 172, 174-75 (Pa. Super. Ct. 2007) (finding the defendant guilty of possession and control of child pornography based on thirty images found in the temporary Internet cache of his personal computer).

158. Google, Google Corporate Information; Google Milestones, <http://www.google.com/intl/en/corporate/history.html> (last visited Aug. 25, 2008).

159. Danny Sullivan, *Nielsen NetRatings Search Engine Ratings*, SEARCHENGINEWATCH.COM, Aug. 22, 2006, <http://searchenginewatch.com/showPage.html?page=2156451>.

160. *See* Nate Anderson, "Google" Declared a Verb, ARS TECHNICA, July 6, 2006, <http://arstechnica.com/news.ars/post/20060706-7198.html>. Google, the transitive verb with inflected forms: googled and googling, means "to use the Google search engine to obtain information about (as a person) on the World Wide Web." Merriam-Webster.com, <http://merriamwebster.com/dictionary/google> (last visited Aug. 25, 2008).

161. *See, e.g.*, Constantin Daniela, *Why is Google So Popular?*, HELIUM, <http://www.helium.com/tm/501146/google-moment-number-search> (last visited Aug. 25, 2008).

advertising revenues.<sup>162</sup> Refusing to relinquish its stronghold on the market, Google pursues its self-proclaimed mission “to organize [all of] the world’s information and make it universally accessible and useful.”<sup>163</sup> With this in mind, in October 2004, Google announced the development of the Book Search Project and invited major publishers to participate.<sup>164</sup> Google designed the Book Search Project to allow users to search the actual text of all books in the Project’s database. Google would no longer be used just for searching Web sites. Books would be scanned into the database and included in the Book Search Project upon submission by publishers. Much like its Web search, the PageRank algorithm would display books, based on relevance, which contained the users’ search terms. The amount of text which could be viewed by users would be determined by the publishers.<sup>165</sup> Because all books in this project were submitted by publishers, there would be no copyright issues.

This all changed in November 2006 when Google announced it would go forward with the Library Project regardless of any agreements, or lack thereof, with the publishers.<sup>166</sup> This variation on the Book Search Project entailed creating a massive searchable index<sup>167</sup> of every book located in each of several large libraries across the world.<sup>168</sup> The Library Project partners Google with the University of Michigan, Harvard University, The New York Public Library, Stanford University, and Oxford University.<sup>169</sup> Each library will receive a digitized copy of each book scanned in exchange for giving Google ac-

---

162. Google reported revenues of \$4.32 billion for the third quarter of 2007, a 57% increase from the third quarter of the previous year. Press Release, Google, Inc., Google Announces Third Quarter 2007 Results (Oct. 18, 2007), *available at* [http://www.google.com/press/pressrel/2007Q3\\_earnings\\_google.pdf](http://www.google.com/press/pressrel/2007Q3_earnings_google.pdf) [hereinafter Google Third Quarter Results].

163. Google, Google Corporate Information: Company Overview, <http://www.google.com/corporate/> (last visited Aug. 25, 2008).

164. Google Book Search, History of Google Book Search, <http://books.google.com/googlebooks/newsviews/history.html> (last visited Aug. 25, 2008). Google Book Search was previously named “Google Print.” *Id.*

165. Google Book Search, About Google Book Search, <http://books.google.com/googlebooks/about.html> (last visited Aug. 25, 2008).

166. *See generally* Posting of Nathan Naze to The Official Google Blog, <http://googleblog.blogspot.com/2006/11/new-way-to-browse-books.html> (Nov. 21, 2006, 18:51 EST).

167. In the context of Google’s use of others’ content, Google uses a database of information and an index of terms within that content. *See* Posting of Tim O’Reilly to O’Reilly Radar, [http://radar.oreilly.com/archives/2005/08/google\\_library.html](http://radar.oreilly.com/archives/2005/08/google_library.html) (Aug. 12, 2005). This distinction is important, and although the terms “database” and “index” are often used interchangeably, they are two distinct elements.

168. Google Book Search, Google Book Search Publisher Questions, [http://books.google.com/googleprint/publisher\\_library.html](http://books.google.com/googleprint/publisher_library.html) (last visited Aug. 25, 2008).

169. Google Book Search, Google Book Search Library Partners, <http://books.google.com/googlebooks/partners.html> (last visited Aug. 25, 2008).

cess to the book along with agreeing not to allow any other commercial search engines access to the digital version.<sup>170</sup>

The main difference between the Book Search Project and the Library Project is that the Library Project proposes to digitize all books in each library regardless of publisher permission.<sup>171</sup> In other words, many in-copyright books are being copied without the express permission of the copyright holders.<sup>172</sup> To many authors and publishers, this is a real problem.<sup>173</sup>

In response to heavy criticism by the AG and the AAP, Google suspended the copying of books for the Library Project from August until November 2005<sup>174</sup> to allow for the implementation of an “opt-out” program, in which authors and publishers<sup>175</sup> could request that their books not be included in the Project.<sup>176</sup> Thus:

[The owner] can participate in the Partner Program, in which case it would share in revenue derived from the display of pages from the work in response to user queries; it can let Google scan the book under the Library Project and display snippets in response to user queries; or it can opt-out of the Library Project, in which case Google will not scan its book.<sup>177</sup>

Several other online giants are involved in similar endeavors, yet in less controversial manners. For example, Yahoo! and Microsoft MSN are working with Internet Archive to digitize public domain works or works authorized by the copyright owners.<sup>178</sup> Project Gutenberg<sup>179</sup> has made public domain works available online for years.<sup>180</sup>

170. See Katie Hafner, *Libraries Shun Deals to Place Books on Web*, N.Y. TIMES, Oct. 22, 2007, at A1, available at <http://www.nytimes.com/2007/10/22/technology/22library.html?pagewanted=1&r=1>.

171. See Band, *supra* note 15, at 1; see also Allan R. Adler, *The Google Library Project* 9 (Sept. 2006), available at [http://www.publishers.org/main/Copyright/attachments/ARA\\_paper.doc](http://www.publishers.org/main/Copyright/attachments/ARA_paper.doc).

172. See Band, *supra* note 15, at 1; see also Adler, *supra* note 171, at 12.

173. See Complaint of McGraw Hill Co., *supra* note 13, at 4; see also Complaint of the Authors Guild, *supra* note 13, at 2.

174. JOHNATHAN BAND, AMERICAN LIBRARY ASSOCIATION, *THE GOOGLE LIBRARY PROJECT: THE COPYRIGHT DEBATE 2* (Jan. 2006), <http://www.ala.org/ala/washoff/oitp/googlepaprfnl.pdf>.

175. The “opt-out” program allows for the owner of the copyright to request that their copyrighted works not be included in the Library Project. See Google Book Search Help Center, *What if I Find One of My Books in Google Book Search and I Would Like It Removed?*, <http://books.google.com/support/bin/answer.py?answer=43756&topic=9011> (last visited Aug. 25, 2008) [hereinafter Google Book Search Help Center]. If the author no longer owns the rights to the copyright, then the publisher would be required to “opt-out” by following Google’s prescribed steps. *Id.*

176. Band, *supra* note 15, at 2.

177. *Id.*

178. See, e.g., Elinor Mills, *Microsoft to Offer Book Search*, CNET NEWS.COM, Oct. 26, 2005, [http://www.news.com/2102-1025\\_3-5913711.html](http://www.news.com/2102-1025_3-5913711.html).

179. See Gutenberg: About, <http://www.gutenberg.org/wiki/Gutenberg:About> (last visited Aug. 25, 2008) (“Project Gutenberg is the first and largest single collection of free elec-

Moreover, in 2005, Amazon.com announced it would sell eBooks<sup>181</sup> through a search program much like Yahoo!'s.<sup>182</sup> In addition, Random House, one of the largest publishers of trade books, has claimed it has a business plan for allowing similar online viewing.<sup>183</sup> Yet, Google stands alone with its plan to scan copyrighted content without permission; these other programs either scan public domain works only or use an "opt-in" system for authors and publishers<sup>184</sup> of copyrighted works.<sup>185</sup>

### B. *The Ensuing Litigation and the Fair Use Defense*

From August until November 1, 2005, Google suspended scanning to allow for content owners to exercise their "opt-out" privileges.<sup>186</sup> However, in September of 2005, the Authors Guild's president alleged "a plain and brazen violation of copyright law," and several authors filed suit against Google for copyright infringement.<sup>187</sup> The next month, five major publishing companies, all members of the AAP, filed suit as well;<sup>188</sup> both suits were filed in the District Court for the Southern District of New York.<sup>189</sup> Google responded that its Library Project was "fully consistent with both the fair use doctrine . . . and the principles underlying copyright law itself."<sup>190</sup>

Without an understanding of the fair use doctrine and its role in copyright law, Google's response may seem like a formulaic lie to the media. However, Google's claim of legality may be legitimately based on this relatively new doctrine with historic roots. "From the infancy of copyright protection, some opportunity for fair use of copyrighted materials has been thought necessary to fulfill copyright's very pur-

---

tronic books, or eBooks. Michael Hart, founder of Project Gutenberg, invented eBooks in 1971 and continues to inspire the creation of eBooks and related technologies today.").

180. *Id.*

181. An eBook is the digital media equivalent of a conventional printed book.

182. Elinor Mills, *Amazon, Random House Throw Book at Google*, CNET NEWS.COM, Nov. 3, 2005, [http://www.news.com/2102-1025\\_3-5931569.html](http://www.news.com/2102-1025_3-5931569.html).

183. *Id.*

184. For the remainder of this Comment, authors and publishers will be referred to as content owners, each possessing the same rights as the other.

185. See Westin, *supra* note 14, at 6.

186. Band, *supra* note 15, at 2.

187. Press Release, The Authors Guild, Authors Guild Sues Google, Citing "Massive Copyright Infringement" (Sept. 20, 2005), available at <http://www.authorsguild.org/advocacy/articles/authorsguildsuesgooglecitingmassivecopyrightinfringement.html>.

188. Press Release, The Association of American Publishers, Publishers Sue Google over Plan to Digitize Books (Oct. 19, 2005) available at [http://publishers.org/main/PressCenter/Archives/2005\\_Oct/Oct\\_03.htm](http://publishers.org/main/PressCenter/Archives/2005_Oct/Oct_03.htm).

189. Complaint of McGraw Hill Co. *supra* note 13, at 1; see also Complaint of the Authors Guild, *supra* note 13, at 1.

190. Posting of Susan Wojcicki, Vice Pres. of Product Management, Google, The Official Google Blog, <http://googleblog.blogspot.com/2005/09/google-print-and-authors-guild.html> (Sept. 20, 2005, 21:04 EST).

pose, “[t]o promote the Progress of Science and useful Arts. . . .”<sup>191</sup> The use of copyrighted content that would otherwise infringe the content owner’s exclusive rights is not infringing if it is determined to be a fair use.<sup>192</sup> Section 107 of the Copyright Act enumerates four non-exclusive factors to be considered in any fair use analysis. The factors, which should be considered on a case-by-case basis, are: “(1) the purpose and character of the use . . . ; (2) the nature of the copyrighted work; (3) the amount . . . used in relation to the copyrighted work as a whole; and (4) the effect . . . upon the potential market for or value of the copyrighted work.”<sup>193</sup> Although the statute requires that all four factors be considered, the first and fourth factors have been given the greatest weight.<sup>194</sup>

While the outcome of Google’s most recent battle is not yet determined, many scholars have speculated about the court’s inevitable fair use analysis.<sup>195</sup> Although fair use is an “equitable rule of reason” to be determined on a “case-by-case basis,”<sup>196</sup> many commentators have looked to a set of seminal cases involving the application of copyright law to search engines for guidance. In particular, *Kelly v. Arriba Soft Corp. (Kelly II)*,<sup>197</sup> cited in various fair use analyses of district courts (including the Southern District of New York),<sup>198</sup> involved “the application of copyright law to the vast world of the [I]nternet and [I]nternet search engines.”<sup>199</sup> More recently, a pair of 2006 Ninth Circuit decisions have been thoroughly discussed in relation to the Library Project, namely *Field v. Google, Inc. (Field)*<sup>200</sup> and

191. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 575 (1994) (quoting U.S. CONST., art. I, § 8, cl. 8).

192. Karl Oakes, *Copyright and Intellectual Property*, 18 C.J.S. COPYRIGHTS § 113 (2008).

193. 17 U.S.C. § 107 (2000); *Campbell*, 510 U.S. at 579-90.

194. *Field v. Google, Inc.*, 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006).

195. *See, e.g.*, Baksik, *supra* note 17, at 405 (speculating that while the court’s analysis of fair use will likely lean in Google’s favor, “[t]he question will come down to ‘whether the public service will outweigh the commercial exploitation’”) (quoting Elisabeth Hanratty, *Google Library: Beyond Fair Use?*, 2005 DUKE L. & TECH. REV. 10 (2005)); Manali Shah, *Fair Use and the Google Book Search Project: The Case for Creating Digital Libraries*, 15 COMM.LAW CONCEPTUS 569 (2007) (arguing that Google is unlikely to succeed under a fair use analysis); Westin, *supra* note 14 (arguing that the court will likely find Google’s Library Project to be a fair use); Sites, *supra* note 17 (predicting a finding of fair use for Google); *see also* Lessig Video, *supra* note 10 (arguing that Google’s use is a fair use).

196. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 448 n.31 (1984) (citing H.R. REP. NO. 94-1476, at 65-66 (1976)).

197. *Kelly v. Arriba Soft Corp. (Kelly II)*, 336 F.3d 811 (9th Cir. 2003).

198. *See* *Faulkner v. Nat’l Geographic Soc’y*, 294 F. Supp. 2d 523, 547 (S.D.N.Y. 2003), *rev’d on other grounds*, 409 F.3d 26 (2d Cir. 2005) (determining that the use of magazine cover photographs arranged into a “photo montage” was transformative); *see also* *Bill Graham Archives, LLC v. Dorling Kindersley Ltd.*, 386 F. Supp. 2d 324, 333 (S.D.N.Y. 2005) (finding that the use of reduced-size images of concert posters in a documentary book was a fair use when the use was transformative and lower resolution).

199. *Kelly II*, 336 F.3d at 815.

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

*Perfect 10 v. Google, Inc. (Perfect 10)*.<sup>201</sup> A full discussion of *Field* will be saved for Part IV of this Comment.

1. Kelly v. Arriba Soft (*Kelly II*)

In *Kelly II*, after withdrawing its prior decision for procedural reasons,<sup>202</sup> the Ninth Circuit held that Arriba Soft's generation of thumbnail images<sup>203</sup> in response to user searches was a fair use.<sup>204</sup> Much like Google, defendant Arriba Soft<sup>205</sup> operated an Internet search engine.<sup>206</sup> One distinction was that Arriba's search engine displayed small "thumbnail" images, in response to a user's search terms, rather than text.<sup>207</sup> Plaintiff Kelly was a photographer who displayed copyrighted images on his own Web site and licensed his images to others.<sup>208</sup> Kelly complained to Arriba about his images being indexed by the search engine and, after Arriba removed the links to the site, Kelly sued for copyright infringement for images that were licensed to third-party Web sites.<sup>209</sup>

The Ninth Circuit analyzed all four statutory fair use considerations and determined that Arriba's use of Kelly's images was a fair use.<sup>210</sup> Regarding the first factor (purpose and character of the use), the Ninth Circuit held that Arriba's use was "more incidental and less exploitative in nature than more traditional types of commercial use."<sup>211</sup> Moreover, the court found that any commercial significance was mitigated by the transformative nature of the thumbnails.<sup>212</sup> In addition, the court gave significant weight to the great public benefit bestowed by the utility of Internet search engines.<sup>213</sup> For example,

---

201. *Perfect 10 v. Google, Inc.*, 416 F. Supp. 2d 828 (C.D. Cal. 2006).

202. *Kelly II*, 336 F.3d at 815. While Arriba conceded a prima facie case of infringement regarding only the thumbnail images, and not as to the in-line linked full-size images, the district court addressed both issues. *Id.* at 816. In the Ninth Circuit's initial ruling, *Kelly v. Arriba Soft Corp. (Kelly I)*, 280 F.3d 934 (9th Cir. 2002), *withdrawn*, 336 F.3d 811 (9th Cir. 2003), the appellate court affirmed summary judgement as to the thumbnail images, but held that in-line linking constituted a "display" for purposes of the Copyright Act, and reversed the district court's ruling of fair use as to the full-size images. *Kelly I*, 288 F.3d at 947-48. *Kelly I* was subsequently withdrawn, and, in *Kelly II*, the Ninth Circuit reaffirmed that generation and display of the thumbnail images was a fair use. *Kelly II*, 336 F.3d at 815-17.

203. Thumbnail images are reduced size, lower-resolution versions of original pictures displayed on the Internet. *Kelly II*, 336 F.3d at 815.

204. *Kelly II*, 336 F.3d at 822.

205. Although Arriba Soft changed its name to "Ditto.com" during litigation, this Comment will use the name "Arriba" to avoid confusion. *Id.* at 815 n.1.

206. *Id.* at 815.

207. *Id.*

208. *Id.* at 815.

209. *Id.* at 816.

210. *Id.* at 817-18.

211. *Id.* at 818.

212. *Id.*

213. *Id.* at 820.

Google's search engine allows users to more effectively harness the raw potential of the Internet and all it contains.<sup>214</sup> Without Google, or search engines in general, access to relevant information on the Web would certainly be more elusive.<sup>215</sup>

Turning to the second statutory factor (nature of the copyrighted work), the Ninth Circuit determined that Kelly's works were creative in nature; however, this weighed only slightly against fair use as Kelly's work had been previously published.<sup>216</sup> Under the third factor (amount of the work used), the court found that, although Kelly's works were copied in their entirety, this was reasonable based on Arriba's intended use of the copies.<sup>217</sup> Considering the fourth factor (market effects), the court held that Arriba's thumbnails would not harm the commercial value or any potential market for Kelly's images.<sup>218</sup> This was due, in part, to the fact that the picture quality of the thumbnails was inferior to that of Kelly's originals and no foreseeable market for thumbnails existed.<sup>219</sup>

Despite Arriba's incidental commercial purpose and significant copying of Kelly's creative works, the Ninth Circuit held that Arriba's generation and display of thumbnail images was a fair use under the Copyright Act.<sup>220</sup>

## 2. Perfect 10 v. Google, Inc.

In *Perfect 10*, the Central District of California held that, despite similarities to *Kelly II*, Google's use of thumbnail versions of copyrighted images was not a fair use.<sup>221</sup> The court's decision hinged on two main distinctions not present in *Kelly II*. First, the court determined that Google directed users, through the thumbnail indexes, to third-party Web sites that carried infringing copies of Perfect 10's copyrighted images.<sup>222</sup> The court stated that "Google's thumbnails lead users to sites that directly benefit Google's bottom line."<sup>223</sup> Second, the court held that Google's thumbnails, although transformative, were also consumptive due to the fact that Perfect 10 had entered into a licensing agreement with Fonestarz Media Limited for

---

214. See Kohler, *supra* note 17, at 25.

215. *Id.*

216. *Kelly II*, 336 F.3d at 820; see also *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 555 (1985) (holding that "the author's right to control the first public appearance of his undissemated expression will outweigh a claim of fair use").

217. *Kelly II*, 336 F.3d at 821.

218. *Id.* at 821-22.

219. *Id.*

220. *Id.* at 822.

221. *Perfect 10 v. Google, Inc.*, 416 F. Supp. 2d 828, 851 (C.D. Cal. 2006).

222. *Id.* at 849.

223. *Id.* at 847.

reduced-size images for download and use on cell phones.<sup>224</sup> The thumbnails were “essentially the same size and of the same quality” and could supersede the images licensed to Fonestarz.<sup>225</sup>

The second and third statutory factors were analyzed with similar results to *Kelly II* and had little influence on the fair use determination with this court.<sup>226</sup> Due primarily to the potential harm to the cell phone image download market, the district court held Google’s use of thumbnail images in this instance was likely not a fair use and issued an injunction in favor of Perfect 10.<sup>227</sup>

However, the Ninth Circuit reversed the district court’s decision, holding that *Kelly II* controlled the case.<sup>228</sup> In reversing, the court of appeals agreed with the district court’s view that there were differences between Arriba’s search engine at issue in *Kelly II* and Google’s search engine.<sup>229</sup> However, it held that those differences were not substantial enough to make Google’s use unfair.<sup>230</sup> Specifically, the Ninth Circuit stated that it “must weigh Google’s superceding and commercial uses of thumbnail images against Google’s significant transformative use, as well as the extent to which Google’s search engine promotes the purposes of copyright and serves the interest of the public.”<sup>231</sup>

The question of whether Google’s Library Project constitutes a fair use has yet to be determined. However, fair use is merely an impediment to proper analysis of the project as both caching services and ISPs have been granted a safe harbor against copyright infringement.

#### IV. CACHING AND ISP RELATED COPYRIGHT LAW

Outside the Google Library Project litigation lies a realm of law that has recently developed to deal with ISPs and caching. As detailed in Part II, caching is used by several online services for a variety of purposes, such as to increase overall Internet speed and decrease bandwidth congestion. Thus, caching has helped advance technology by allowing more Internet users to have quicker access to more information. A few representative cases can help shed some light on the evolution of the law in this relatively new area.

---

224. *Id.* at 831.

225. *Id.* at 849.

226. *Id.* at 849-50.

227. *Id.* at 851.

228. *Perfect 10, Inc. v. Amazon.com, Inc. (Perfect 10 II)*, 487 F.3d 701, 724-25 (9th Cir. 2007) (reversing the district court’s decision).

229. *Id.*

230. *Id.*

231. *Id.* at 722.

### A. Copyright Law as it Relates to Caching

In a fairly early case involving copyrights and computers, the Fifth Circuit noted that “the act of loading a program from a medium of storage into a computer’s memory creates a copy of the program . . . .”<sup>232</sup> Although the court in *Vault* held that there was no infringement, the case set the precedent that creating digital copies, even when not reproduced in portable medium, constitutes copying for the sake of analyzing infringement. However, the courts did not discuss the application of such a rule in the context of RAM or caching for several years. Nearly four years later, the Ninth Circuit in *MAI Systems Corp. v. Peak Computer, Inc.*<sup>233</sup> addressed the distinction, holding that a copy of software “made in RAM is ‘fixed’ and qualifies as a copy under the Copyright Act”<sup>234</sup> and is, therefore, copyright infringement. Although this decision has received unfavorable treatment, it has not been overruled.<sup>235</sup>

More recently, and more relevant to this Comment, in *Field v. Google, Inc.*,<sup>236</sup> the District Court of Nevada held that Google’s cache storage did not constitute direct copyright infringement because the display of cached versions of Web pages was a fair use.<sup>237</sup> Field was an attorney and a poet.<sup>238</sup> He objected to Google indexing some of his copyrighted poetry, which was posted on his own Web site.<sup>239</sup> Applying the four fair use factors of Section 107,<sup>240</sup> the court determined

232. *Vault Corp. v. Quaid Software, Ltd.*, 847 F.2d 255, 260 (5th Cir. 1988).

233. *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511 (9th Cir. 1993).

234. *Id.* at 519; *see also* 17 U.S.C. § 101 (2000):

“Copies” are material objects, other than phonorecords, 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. The term ‘copies’ includes the material object, other than a phonorecord, in which the work is first fixed.

...

A work is ‘fixed’ in a tangible medium of expression when its embodiment in a copy or phonorecord, by or under the authority of the author, is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration. A work consisting of sounds, images, or both, that are being transmitted, is “fixed” for purposes of this title if a fixation of the work is being made simultaneously with its transmission.

17 U.S.C. § 101.

235. *See, e.g.*, *DSC Commc’ns Corp. v. Pulse Commc’ns, Inc.* 170 F.3d 1354 (Fed. Cir. 1999) (disagreeing with *MAI Systems*); *Telecomm Technical Servs., Inc. v. Siemens Rolm Commc’ns, Inc.*, 66 F. Supp. 2d 1306 (N.D. Ga. 1998) (declining to follow *MAI Systems*); *Applied Info. Mgmt., Inc. v. Icart*, 976 F. Supp. 149 (E.D.N.Y. 1997) (declining to follow *MAI Systems*).

236. 412 F. Supp. 2d at 1123.

237. *Id.*

238. *Id.* at 1110.

239. *Id.*

240. 17 U.S.C. § 107 (2000).

that making cache copies of copyrighted material is a fair use.<sup>241</sup> Similar to *Perfect 10*, the court determined that Google's cache was highly transformative which, in conjunction with Google's great social utility, outweighed any argument of its commercial nature.<sup>242</sup> The district court also placed little emphasis on the second and third factors of fair use and found no evidence of any market for the copyrighted work, the fourth factor.<sup>243</sup> In addition, the court held that Field's claim for damages was "precluded by operation of the 'system cache' safe harbor of Section 512(b) of the Digital Millennium Copyright Act (DMCA)."<sup>244</sup>

Finally, in *Parker v. Google Inc.*, the District Court of the Eastern District of Pennsylvania dismissed Parker's claims charging Google with direct copyright infringement based on Google's use of caching and the DMCA safe harbors.<sup>245</sup> Plaintiff's claims arose as a result of both Google's archiving of Usenet postings that contained excerpts of plaintiff's copyrighted works, and its display of excerpts of plaintiff's copyrighted Web site in search results.<sup>246</sup> The district court determined that Google did not engage in the requisite volitional conduct necessary for direct copyright infringement; such copying was a natural by-product of Google's automated search engine and related technologies.<sup>247</sup>

Relying on *Field*, the district court determined that Google's acts were akin to a user's use of his or her ISP to transmit infringing material to a third party.<sup>248</sup> Such actions do not give rise to direct infringement claims against ISPs due to the DMCA safe harbors.<sup>249</sup> On appeal, the Third Circuit affirmed the dismissal of Parker's claims since Google's caching of infringing Usenet postings lacked the requisite "volitional conduct."<sup>250</sup>

### B. Copyright Law as it Relates to ISPs

Setting some early groundwork in ISP litigation, prior to the DMCA, the court in *Religious Technology Center v. Netcom On-Line*

---

241. *Field*, 412 F. Supp. 2d at 1123.

242. *Id.* at 1119-20.

243. *Id.* at 1120-21.

244. *Id.* at 1109. See generally 17 U.S.C. § 512(b) (2000) (setting out the safe harbor exceptions for service providers, and now search engines, utilizing system caching).

245. *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 504 (E.D. Pa. 2006) (dismissed because plaintiff "failed to state a claim on which relief can be granted").

246. *Id.* at 495.

247. *Id.* at 495, 497.

248. *Id.* at 497.

249. *Id.*

250. *Parker v. Google, Inc. (Parker II)*, 242 Fed. App'x 833, at \*3 (3d Cir. July 10, 2007).

*Communication Services*<sup>251</sup> held that an “Internet access provider”<sup>252</sup> was not directly liable for cached copies of copyrighted material.<sup>253</sup> In *Religious Technology*, the plaintiffs sued the operator of an Internet Bulletin Board Service (BBS)<sup>254</sup> as well as the operator’s ISP for direct infringement after copyrighted materials were posted on the BBS and cached on the ISP’s computers.<sup>255</sup> Although the district court looked to *MAI Systems*<sup>256</sup> to determine that the copies were permanently fixed in the ISP’s RAM, it refused to hold Netcom liable for its machine’s “passive” operation.<sup>257</sup>

Two years later, the District Court for the Northern District of Illinois granted a motion for summary judgment against a host computer operator<sup>258</sup> for contributory infringement, yet refused to hold the ISP responsible for direct copyright infringement.<sup>259</sup> In *Marobie-FL, Inc. v. National Association of Fire Equipment Distributors*,<sup>260</sup> the National Association of Fire Equipment Distributors (NAFED) purchased clip art<sup>261</sup> from Marobie-FL, Inc. and placed it on its Web site via the host computer administered by Northwest Nexus, Inc., NAFED’s ISP.<sup>262</sup> Marobie noticed a drop in sales of its copyrighted clip-art, determined it was due to NAFED’s unauthorized publication

251. *Religious Tech. Ctr. v. Netcom On-Line Commc’n Servs., Inc.*, 907 F. Supp. 1361 (N.D. Cal. 1995).

252. For the purpose of this Comment, “Internet access provider” and ISP are synonymous.

253. *Religious Tech.*, 907 F. Supp. at 1368, 1372 (holding that, although copies were “sufficiently fixed” on ISP’s network, the ISP was not directly liable for the copies, did not receive direct financial benefit from the infringing materials, and copyright holders were not entitled to preliminary injunction).

254. Originally, BBSs were computer systems running software that allowed users to connect to them via their own computer systems. PATRICE FLICHY, *THE INTERNET IMAGINAIRE 75* (MIT Press 2007). These BBSs allowed the outside users to perform functions such as downloading software and data, uploading data, reading news, and exchanging messages with other users. *Id.* Today, the term BBS generally refers to any online forum or message board. *Id.*

255. *Religious Tech.*, 907 F. Supp. at 1365.

256. *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 511 (9th Cir. 1993).

257. *Religious Tech.*, 907 F. Supp. at 1367.

258. A “host computer operator” is the operator, company or person, who hosts a Web site. *Marobie-FL, Inc. v. Nat’l Ass’n of Fire Equip. Distribs.*, 983 F. Supp. 1167, 1171 (N.D. Ill. 1997). The Web site itself, consisting of computer files, resides on the host computer and the files are broadcast over the Internet for users to view. *Id.* In this particular instance, however not usually, the “host computer operator” and the ISP are one and the same. *Id.*

259. *Id.* at 1181 (holding that the host computer operator, the ISP in this case, could be liable for contributory copyright infringement but not for direct copyright infringement).

260. 983 F. Supp. 1167 (N.D. Ill. 1997).

261. According to Merriam-Webster, clip art is “ready-made usually copyright-free illustrations sold in books or as part of a software package from which they may be cut and pasted or inserted as artwork.” Merriam-Webster.com, <http://merriamwebster.com/dictionary/clip%20art> (last visited Aug. 25, 2008).

262. *Marobie-FL, Inc.*, 983 F. Supp. at 1172.

on the Internet, and subsequently sued both NAFED and Northwest.<sup>263</sup> The district court commented that although

Northwest provides a service somewhat broader than the service provided by the Internet access provider in *Religious Technology Center*, the court nevertheless finds that Northwest only provided the means to copy, distribute or display plaintiff's works, much like the owner of a public copying machine used by a third party to copy protected material.<sup>264</sup>

Clearly, the court acknowledged that Northwest served as more than just a gateway to the Internet due to its operation of the host computer.<sup>265</sup> However, "like a copying machine owner, Northwest did not actually engage in any [infringing] . . . activities itself," hence it could not be held liable for direct infringement.<sup>266</sup>

In 1998, the DMCA was enacted to help preserve copyright enforcement on the Internet and "to implement the World Intellectual Property Organization Copyright Treaty and Performances and Phonograms Treaty."<sup>267</sup> In addition, Congress intended to provide immunity to ISPs from copyright infringement for automatic, or passive, actions initiated by Internet users without the ISPs' knowledge.<sup>268</sup> Section 512(b) of the DMCA also includes provisions designed to address the issues that arise in traditional system caching scenarios.<sup>269</sup>

In particular, section 512(b) provides a safe harbor for system caching, permitting the "intermediate and temporary storage" of online content.<sup>270</sup> In order for the safe harbor to apply, the content must be: (1) intermediate and temporary; (2) "made available online by a person other than the service provider;" (3) transmitted to a third person at his or her direction; and (4) stored through an automatic technical process so as to make the material available to users of the system who request access to it from the host computer.<sup>271</sup> In addition, the material must be transmitted without modification and where the content is made available online without the copyright owner's authorization, the ISP must respond "expeditiously" to remove or disable access to the allegedly infringing content.<sup>272</sup>

Clearly, Congress intended to facilitate the great social utility of ISPs, which provide Internet access to the general public, by provid-

---

263. *Id.*

264. *Id.* at 1178.

265. *Id.*

266. *Id.*

267. Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998).

268. See H.R. Rep. No. 105-796, at 72-73 (1998) (Conf. Rep.).

269. See 17 U.S.C. § 512(b) (2000).

270. *Id.*

271. 17 U.S.C. § 512(b)(1)(A)-(C).

272. 17 U.S.C. § 512(c)(1)-(3) (2006).

ing liability immunity for the copyright infringement of others.<sup>273</sup> Moreover, the legislature recognized the importance of system caching to the overall speed and efficiency of the Web.<sup>274</sup> However, it is not clear that these safe harbors will protect Google from the litigation surrounding its Library Project.

## V. GOOGLE'S MATRIX V. THE COPYRIGHT ACT

Google intends to digitize and make available nearly 20 million books through caching and indexing.<sup>275</sup> Some of these books, but not all, are currently in-copyright materials.<sup>276</sup> For those books that are in-copyright, Google offers an "opt-out" program in which copyright holders may remove their books from the list contained in Google's index.<sup>277</sup> Similarly, if a copyright holder's book has yet to be digitized, the owner may have the title removed from Google's list of books to be scanned.<sup>278</sup> Equally as important, if an in-copyright book is digitized, indexed, and available through Google's site, only snippets will be available to searchers.<sup>279</sup> As a precaution, creative searchers who attempt to display different snippets of a single book are blocked from viewing more than the original search results.<sup>280</sup> In contrast, public domain works will be displayed in full text and can be viewed and searched multiple times.<sup>281</sup>

It seems obvious upon discussing the precautionary measures taken, including the use of snippets and the "opt-out" program, that Google anticipated possible copyright issues. In fact, Google states, "[t]he Library Project's aim is simple: make it easier for people to find relevant books—specifically, books they wouldn't find any other way such as those that are out of print—while carefully respecting authors' and publishers' copyrights."<sup>282</sup> Google plans to spend up-

273. *Perfect 10, Inc. v. CCBill*, 340 F. Supp. 2d 1077, 1086 (C.D. Cal. 2004).

274. *See id.* Although the legislature has not explicitly endorsed that system caching adds efficiency and speed to the Web, it may be implied through the protection afforded to ISPs for such processes.

275. Lessig Video, *supra* note 10. In fact, some estimates put the number of books Google plans to digitize as high as 32 million. Jonathan V. Last, *Google and Its Enemies*, WEEKLY STANDARD, Dec. 10, 2007, available at <http://www.weeklystandard.com/Content/Public/Articles/000/000/014/431afuv.asp>.

276. *Id.*

277. Google Book Search Help Center, *supra* note 175.

278. Google Book Search, What About Books That I Don't Want in Google Book Search at All?, <http://books.google.com/support/bin/answer.py?answer=43755&topic=9011> (last visited Aug. 25, 2008).

279. *See* Google Book Search Library Project, <http://books.google.com/googlebooks/library.html> (last visited Aug. 25, 2008); *see also* Lessig Video, *supra* note 10.

280. Lessig Video, *supra* note 10.

281. Google Book Search Library Project, *supra* note 279; Lessig Video, *supra* note 10.

282. Google Book Search Library Project, *supra* note 279. Google's ultimate goal is "to work with publishers and libraries to create a comprehensive, searchable, virtual card

wards of \$800 million to complete the project.<sup>283</sup> Apparently, its attorneys feel safe resting on the fair use defense in court. However, the affirmative defense of fair use may actually impede Google's plans to digitize books. In fact, Google's best defense may have already been addressed by several courts and the DMCA.

After *Parker*, it is clear that the caching of copyrighted material is a fair use, providing that the copier lacks volitional conduct.<sup>284</sup> However, Google intends to make wholesale copies of books for its index without necessarily receiving permission from content owners.<sup>285</sup> This will likely be seen as volitional by the courts because Google will be physically scanning and copying the books itself. Although the court in *Kelly II* allowed the volitional copying of entire photographs because it was reasonable for the intended use,<sup>286</sup> similar courts may not agree. Therefore, Google's best defense may be found under the DMCA.

Recall that section 512(b) of the DMCA provides a safe harbor for system caching as long as the content is: (1) intermediate and temporary; (2) made available by another; (3) requested by a third person; and (4) stored through an automatic technical process.<sup>287</sup> Google's Library Project can meet all but the second element. As to the first element, Google's copying is intermediate and temporary,<sup>288</sup> although the index may be permanent. After the initial copying of the book, the text is indexed in Google's matrix and caching is employed for snippet retrieval.<sup>289</sup> Addressing the third element, all book searches will be made by Internet users from remote host computers<sup>290</sup> wherever Internet access is available. As to the fourth element, Google's PageRank algorithm and its use of cache memory is completely automatic.<sup>291</sup>

---

catalog of all books in all languages that helps users discover new books and publishers discover new readers." *Id.*

283. Last, *supra* note 275.

284. *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 497 (E.D. Pa. 2006) (stating that "a plaintiff must also show volitional conduct on the part of the defendant in order to support a finding of direct copyright infringement" (quoting *Field v. Google, Inc.*, 412 F. Supp. 2d 1106, 1114-15 (D. Nev. 2006))).

285. See Lessig Blog, *supra* note 19; see also Lessig Video, *supra* note 10.

286. *Kelly II*, 336 F.3d 811, 820-21 (9th Cir. 2003) (stating, however, that "[w]hile wholesale copying does not preclude fair use per se, copying an entire work militates against a finding of fair use" (quoting *Worldwide Church of God v. Phila. Church of God, Inc.*, 227 F.3d 1110, 1118 (9th Cir. 2000))).

287. 17 U.S.C. § 512(b) (2000).

288. See *Field v. Google, Inc.*, 412 F. Supp. 2d 1106, 1114-15 (D. Nev. 2006) (holding that Google's use of cache was intermediate and temporary).

289. Google Book Search Library Project, *supra* note 279; Lessig Video, *supra* note 10.

290. This element is inherent in almost all Web searching.

291. See GoogleGuide, *supra* note 91.

Section 512(b)(2) of the DMCA also requires that the content must be transmitted without modification.<sup>292</sup> Additionally, where the content is made available online without the copyright owner's authorization, the ISP must respond "expeditiously" to remove or disable access to the content.<sup>293</sup> Again, Google's Library Project falls within these provisions. Once the index is created, Google's algorithm and use of cache ensures that the content presented to searchers is unchanged from its original version.<sup>294</sup> In addition, Google's "opt-out" program allows content owners to require that their in-copyrighted materials be removed from Google's index altogether.<sup>295</sup> In fact, Google as a content provider almost directly mirrors that of an ISP—the very function Congress intended to protect.<sup>296</sup>

The DMCA was created, in part, to protect the great social utility of information access provided by ISPs.<sup>297</sup> Congress saw fit to use caching as the catalyst for this protection because of its automated process and its great public benefit.<sup>298</sup> Correspondingly, caching has proven beneficial in many areas, including personal computers, the DNS, and Web browsers.<sup>299</sup> In addition, the advent of caching has led to the advancement of search engines like Google, which has improved information access and retrieval as well as overall Internet efficiency.<sup>300</sup> Even the court in *Kelly II* recognized the great public benefit bestowed by the utility of search engines via cache,<sup>301</sup> even so, it is debatable whether the DMCA should cover these types of technologies as well.

When discussing the implications of the modern copyright system, it is important to point out the intrinsic conflicts inherent in it, particularly in terms of technological advances. The intent of the Copyright Act is to protect the limited rights of creators and at the same time enhance the technological advancement of society.<sup>302</sup> In particular, the purpose is to "promote the Progress of Science and useful Arts."<sup>303</sup> The conflict is evident when a company like Google creates a technological advancement that also stands to be very commercially valuable. Whose rights are more important? Is it society's right to

292. 17 U.S.C. § 512(b)(2)(A) (2000).

293. *Id.* § 512(b)(2)(E).

294. Google Book Search Help Center, *supra* note 175; Lessig Video, *supra* note 10.

295. Google Book Search Help Center, *supra* note 175; Lessig Video, *supra* note 10.

296. *See* ALS Scan, Inc. v. RemarQ Cmty., Inc., 239 F.3d 619, 625 (4th Cir. 2001); Perfect 10, Inc. v. CCBill, 340 F. Supp. 2d 1077, 1086 (C.D. Cal. 2004).

297. *See* cases cited *supra* note 296.

298. *Id.*

299. *See supra* Part II.

300. *See supra* Part II.D.3.

301. *See Kelly II*, 336 F.3d 811, 820 (9th Cir. 2003).

302. H.R. REP. NO. 94-1476, at 47 (1976).

303. *Id.*

foster innovation for the greater good, or the copyright holder's right to protect its monopoly?

Recall that Google is considered the world's most popular search engine.<sup>304</sup> With worldwide and U.S. Google searches reaching nearly 200 million<sup>305</sup> and 95 million<sup>306</sup> per day respectively, there is not much doubt of Google's presence. Moreover, the "Google [Library Project] could be the most important contribution to the spread of knowledge since Jefferson dreamed of national libraries."<sup>307</sup> Recently, in May 2007, California officially recognized the Internet Archive as a library,<sup>308</sup> and one can only imagine that Google's Library Project is next.

It is difficult to calculate how many people actually use Google<sup>309</sup> and, more importantly, how many people search rather than type URLs into their Web browser.<sup>310</sup> However, it is easy to see that the ease of searching, using Google's matrix and PageRank algorithm, makes Google, and other search engines, nearly as important as ISPs. Both ISPs and search engines provide online content, valuable information, and research capabilities to Internet users. Both ISPs and search engines utilize proxy caching to increase Internet access speeds, decrease bandwidth overflow rates, and improve the overall efficiency of the Web. In addition, an often overlooked similarity is that both ISPs and search engines are for-profit entities whose bottom lines are affected by popularity and copyright infringement litigation.

---

304. David Bowen, *Drowning in Information*, FIN. TIMES, Mar. 20, 2003, available at <http://search.ft.com/ftArticle?%20queryText=Drowning+In+Information&y=0&aje=true&x=0&id=030320009486&ct=0>; Thomas L. Friedman, *Is Google God?*, N.Y. TIMES, June 29, 2003, available at <http://query.nytimes.com/gst/fullpage.html?res=9C05E0D8163AF93AA15755C0A9659C8B63>.

305. Friedman, *supra* note 304.

306. One study reported that of the nearly 5.6 billion U.S. originated Web searches per month in July 2006, 49.2% were Google searches. Sullivan, *Nielsen NetRatings*, *supra* note 159. Therefore, it can be approximated that nearly 2.8 billion Google searches originated from U.S. searchers in July 2006; that is nearly 95 million per day. *Id.*

307. See Lessig Blog, *supra* note 19.

308. Adrian McCoy, *The Internet Gives Birth to an 'Official' Online Library*, PITTSBURGH POST-GAZETTE, June 24, 2007, available at <http://www.post-gazette.com/pg/07175/796164-96.stm>.

309. See Sullivan, *Nielsen NetRatings*, *supra* note 159. For an explanation of how extrapolations are made to calculate the number of users, see *supra* note 306.

310. This number would be nearly impossible to calculate; however, the popularity of Google itself tends to imply that many people are simply searching for Web sites, rather than typing in URLs. See Sullivan, *Nielsen NetRatings*, *supra* note 159; see also MCJOHN, *supra* note 107, at 339 (stating that sophisticated Web users have turned away from using URLs in favor of search engines).

## VI. CONCLUSION

Google has been called the most popular Web search engine in the world<sup>311</sup> and is also a powerful corporation.<sup>312</sup> However, Google is not only popular and powerful, it is also extremely useful. In fact, Google has patented both its searching algorithm<sup>313</sup> and its use of caching<sup>314</sup>—evidence of its utility, novelty, and non-obviousness.<sup>315</sup> Moreover, several courts have commented on Google's utility in terms of the great benefit it conveys to society; that benefit being information access on a grand scale rivaling the world's greatest libraries.<sup>316</sup>

However, Google is not competing with conventional libraries.<sup>317</sup> On the contrary, Google is a service that can be, and should be, effectively used by libraries and librarians alike.<sup>318</sup> The Google Library Project, in particular, aids in directing researchers to locate relevant library books.<sup>319</sup> Yet, the success of the Library Project and that of Google in general is predicated on its use of caching. Coincidentally, like ISPs, in order to cache, Google must copy first.

Recall that Google's use of caching is what makes the copied content available on the Web.<sup>320</sup> The copying itself is simply the means of getting the content into Google's matrix. Therefore, it is not the copying, but the caching, which is the proper subject of a fair use inquiry. Fortunately for Google and society, courts have ruled that caching is

311. In fact, as of the writing of this Comment, Google.com was the most popular Web site in the United States, and the fourth most popular Web site in the world. Google.com: Traffic Details from Alexa, [http://www.alexa.com/data/details/traffic\\_details/google.com](http://www.alexa.com/data/details/traffic_details/google.com) (last visited Aug. 25, 2008).

312. Google is an American corporation, traded publicly on the National Association of Securities Dealers Automated Quotations (NASDAQ) stock exchange, with 15,916 full-time employees as of September 30, 2007. *See* Google Third Quarter Results, *supra* note 162. In addition, Google recently reported third quarter revenues of \$4.32 billion. *Id.*

313. *See* System & Method for Searching an Extended Database, U.S. Patent No. 7,174,346 (filed Sept. 30, 2003) (issued Feb. 6, 2007).

314. *See* Accelerating User Interfaces by Predicting User Actions, U.S. Patent Application Publication No. 20060047804 (filed June 30, 2004) (published Apr. 13, 2006).

315. *See* 35 U.S.C. § 101-103 (2000). In order for a U.S. patent to issue, the invention must be "new and useful." *Id.* § 101. The invention must also be novel. *Id.* § 102. In addition, the invention must be non-obvious. *Id.* § 103.

316. *See, e.g.,* Perfect 10 v. Google, Inc., 416 F. Supp. 2d 828, 848-49 (C.D. Cal. 2006) ("It is by now a truism that search engines such as Google Image Search provide great value to the public."); Field v. Google, Inc., 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006) (commenting that "[t]he Internet is replete with references from academics, researchers, journalists, and site owners praising Google's cache"); *Kelly II*, 336 F.3d 811, 820 (9th Cir. 2003) (finding search engine's use of copyrighted material transformative in part because it "benefit[ted] the public by enhancing information-gathering techniques on the internet").

317. *See* Robert J. Lackie, *Google's Print and Scholar Initiatives: The Value of and Impact on Libraries and Information Services*, in LIBRARIES AND GOOGLE 57, 63-69 (William Miller & Rita M. Pellen eds., 2005).

318. *See id.*

319. *See id.*

320. *See supra* Part II.D.3.

a fair use.<sup>321</sup> In addition, the DMCA has set out caching safe harbors within which Google almost squarely fits. Moreover, the statutory limitations of the DMCA show that Congress is willing to modify copyright law as technology changes.<sup>322</sup> Thus, Google's caching should be allowed as a fair use; however, if not, the DMCA should be amended to afford Google the same safe harbors as ISPs.

One suggestion would be to eliminate the second element of Section 512(b),<sup>323</sup> essentially transforming the DMCA into a caching safe harbor for ISPs and search engines. This would effectively fit the Library Project squarely within the DMCA without altering its intent. Complying with the Act requires a function for content removal;<sup>324</sup> therefore, removing 512(b)(2), the source element, should not alter the function of the DMCA. For Google, the Library Project's "opt-out" program provides the necessary content removal function.

Finally, this Comment has compared Google to both the human-enslaving machines of *The Matrix* and the human-empowering services we call Internet Service Providers. Hopefully, it is clear which comparison is more accurate. Google is new, Google is useful, and most of all, Google is innovative—the very things the Copyright Act intends to protect.<sup>325</sup> It should not be thought of as a menace to society, stealing the rights of authors and publishers. On the contrary, Google is Neo,<sup>326</sup> freeing the rights of literary content from the monopolistic hold of its owners. This Comment should not be read to imply that content owners, the Copyright Act, the Attorney General, or the American Association of Publishers are the enemy; however, it certainly *is* implying that Google is not.

---

321. See, e.g., *Field v. Google, Inc.*, 412 F. Supp. 2d 1106, 1110-11 (D. Nev. 2006) (holding that Google's use of cache was a fair use).

322. See Pamela Samuelson, *The Generativity of Sony v. Universal: The Intellectual Property Legacy of Justice Stevens*, 74 *FORDHAM L. REV.* 1831, 1861-62 (2006). Samuelson claims that though the DMCA expands the exclusive rights of copyright holders to a degree, these provisions should be construed narrowly so as not to stifle innovation. *Id.* Samuelson also claims that the existence of the safe harbor provisions supports this contention. *Id.*

323. 17 U.S.C. § 512(b)(2) (2000).

324. See 17 U.S.C. § 512(c) (2000).

325. See H.R. REP. NO. 94-1476, at 47 (1976).

326. Recall that Neo was the hero in the fictional movie introduced in Part I. See *THE MATRIX*, *supra* note 1. In *The Matrix*, Neo destroyed the enemy (the Matrix) and freed mankind from slavery. *Id.*

