Internet Search Engine Copying: Fair Use Defense to Copyright Infringement

Daniel Ovanezian

Follow this and additional works at: http://digitalcommons.law.scu.edu/chtlj

Part of the Law Commons

Recommended Citation

Available at: http://digitalcommons.law.scu.edu/chtlj/vol14/iss1/5
INTERNET SEARCH ENGINE COPYING: FAIR USE DEFENSE TO COPYRIGHT INFRINGEMENT*

Daniel Ovanezian†

TABLE OF CONTENTS

I. INTRODUCTION .................................................. 268
II. SEARCH ENGINES: INTERNET TOOLS .......................... 269
   A. The Internet and the Web ...................................... 269
   B. Search Engines ................................................. 270
III. COPYRIGHT PROTECTION OF WEB SITES .......................... 274
   A. What Does Copyright Protect and Why? ....................... 274
   B. Web Site Copyrights ........................................... 275
IV. WHY WEB SITE OWNERS MIGHT ENFORCE THEIR COPYRIGHTS AGAINST SEARCH ENGINE COMPANIES .......................... 275
   A. Search Engines May Dissuade Users from Going to a Web Site or Hinder Access to a Web Site ..................... 276
   B. Web Site Abstracts May Be Displayed Along with Undesired Advertisements or Abstracts of Other Sites .......... 277
   C. Licensing Agreements Are Impractical ........................ 278
V. DO SEARCH ENGINES INFRINGE COPYRIGHTS? ..................... 280
   A. What Constitutes Copyright Infringement ..................... 280
   B. Substantial Similarity .......................................... 284
      1. Quantitative Analysis ....................................... 286
      2. Similarity to Unprotected Material ......................... 287
      3. Qualitative Analysis: Audience Test ........................ 288
   C. Infringement Conclusion ....................................... 288
VI. DOES FAIR USE APPLY TO SEARCH ENGINES? ...................... 289
   A. Four Factors of Fair Use ....................................... 289
   B. Judicial Interpretations of the Four Factors .................. 290
      1. Effect on the Potential Market .............................. 290
      2. Purpose and Character of the Use ............................ 294
      3. Nature of the Copyrighted Work ............................. 297
      4. Amount and Substantiality Used ............................. 298

* Copyright © 1997, Daniel Ovanezian.
† B.S. Electrical Engineering, Rensselaer Polytechnic Institute; Candidate for J.D., Santa Clara University School of Law, 1998.
C. Fair Use Conclusion ....................................................... 299

VII. Conclusion ........................................................................ 300

I. INTRODUCTION

Search engines have done for the Internet what Windows has done for the personal computer. Both have allowed information to be easily accessed by the inexperienced computer user. For search engines, this benefit may come at the price of violating the copyrights of thousands of Web site creators. However, would these copyright owners actually enforce their copyrights considering that search engines were created in order to facilitate access to these Web sites?

Consider the following hypothetical. During the process of cataloging the Internet, a new search engine visits a Web site that promotes a particular commercial product. The search engine copies the Web site into its database and, in the process, the search engine crashes the site's server making that site inaccessible to other Internet users. Later, a person uses the search engine to locate information about the same commercial product the search engine catalogued. Along with the Web site address for the particular commercial product mentioned above, the search engine displays two other items: a different Web site that critically reviews the product and an advertisement for a competitor's product. The owner of the crashed Web site contacts the search engine company to complain and is informed the following: the site crash was due to the site owner's improper set-up of the site's parameters; the competitor's advertisement and the critical review will continue to be displayed along with his Web site listing; the Web site owner's only remedy is to alter their site's software to block future search engines from cataloging their site.

Web site owners, however, may have an alternate recourse. They may bring an action against the search engine for copyright infringement as a means to persuade search engines to be more responsive to the above mentioned issues.

1. Windows is a computer graphical interface made by Microsoft that uses visually recognizable elements, such as icons and menus, to carry out computing functions.

2. The hypothetical is based on a search I performed using Excite's (http://www.excite.com) search engine. The search parameters used were "car," "suzuki," and "report." The information displayed included the Web site for Suzuki (an automotive company) United Kingdom, a Web site abstract concerning a report in Consumer Reports about the Suzuki Samurai sports utility vehicle in which Consumer Reports called the vehicle unsafe, and an advertisement for a different automotive company, Acura.
This article analyzes whether search engine activity constitutes copyright infringement and, if so, whether fair use would qualify as a defense to these activities. Section II provides brief descriptions of the Internet (specifically the Web) and Search Engines, respectively. Section III outlines the Web site materials that can be protected by copyright. Section IV explores in greater detail why Web site owners would choose to enforce their copyrights against search engines even though search engines are intended to facilitate access to Web sites. Section V argues that search engines do infringe Web site copyrights. Section VI analyzes the judicial interpretation of the Copyright Act and compares search engine use of copyrighted works to fair use in analogous cases. Section VII concludes that search engine copying should fall under the fair use doctrine, and comments on an alternative defense to copyright infringement — estoppel.

II. SEARCH ENGINES: INTERNET TOOLS

A. The Internet and the Web

The Internet is a decentralized network of millions of computers located around the world. There is neither a central hub that routes information nor a central governing body. The Internet began as a military research network and was ultimately opened to civilian use in 1984. Since it has been open to the public, the Internet has evolved into an international electronic information system connecting universities, governments, commercial enterprises, and individuals in order to provide access to educational, entertainment, and business resources.

Information stored on the Internet is displayed in multiple formats. One such method is an electronic display of pages which look like printed pages filled with graphics and text when displayed on computer screens. These electronic pages are linked to each other by

their developers and form a web of information spread across the Internet known as the World Wide Web (also known as WWW, or simply the Web). The Web provides a visually attractive and simple method for tying together the information contained on these electronic pages which Internet users can view by using computer software known as "browsers." With a computer, software, and a modem, it is now possible for people to browse information stored in libraries, universities, government offices, businesses, and privately maintained Web sites. What began as a means for scientists and researchers to share information with each other has turned into a "communications revolution." In addition, the Web has become a widely used medium for publishing, far exceeding the printed press as a means to publish news, information, and views in the public arena.

B. Search Engines

The growing number of Web users has resulted in a major commercialization of the Web for profit. One such area for commercialization of the Web is in providing a means for sifting through the vast quantities of information which constitutes the Web. Before search engines were available, a user interested in locating information on a specific topic would be limited to accessing Web sites that the user learned about through word of mouth or from reading publications referencing those Web sites. This method of locating information on the Web proved to be inefficient and ineffective. Soon, computer programs were created that provided users with a tool to search the Web for specific information. These tools allowed com-

8. Id.
9. Id. Two of the more popular browsers are Netscape Navigator and Microsoft Internet Explorer.
10. Id. ("Virtually anyone with access to the Internet can now develop and maintain a Web site for a modest investment").
12. Rose, supra note 7.
puter users to quickly locate Web sites in the vast Internet using different methods.\(^4\)

One of the earliest search methods involved creating Web site indices that listed Web sites by subject headings.\(^5\) This enabled a computer user to select Web site listings from the relevant subject headings. For example, to find sites on rock climbing, users would browse a list of subject headings until they found the heading "sports." Under this heading would be various subheadings, including "rock climbing."

Under that subheading would be a list of Web site addresses that contain information on rock climbing.\(^6\) Unfortunately, this search method is limited by the finite number of categories listed in the indices.

A more extensive method involves the use of "search engines," computer programs that search through databases for Web sites that contain information suggested by the user.\(^7\) A common misconception is that search engines traverse the Web’s structure only after they receive a request from a user.\(^8\) In fact, another type of program, called "robots," "spiders," or "crawlers" continually rambles the Web to create a database of Web sites.\(^9\) A search engine then searches through this database to return a list of all the sites that match a search request entered by the user.\(^10\)

Which Web sites a robot decides to search varies; each robot uses a different strategy depending on how it is programmed. In general, a robot starts from a previously generated list of the most popular Web site addresses. These Web sites contain links that will lead the robot to other sites on the Web.\(^11\) A robot will "parse," or copy, information from the sites it visits. Some robots copy just the site's

---

15. One of the first topical search indices was Yahoo! (http://yahoo.com), started by two Stanford University students. Johnson-Laird, supra note 4, at 23.
20. Id.
21. Id.
address while others will copy words, sentences or even the entire contents of a site in order to generate a summary, or abstract, of the site. Despite the possibility of perpetuating the above mentioned misconception, the term "search engine" will, hereafter, encompass the functions performed by both robots and search engines because independent analysis is not required due to their common ownership.

Search engines can return a list of upwards of tens of thousands of site addresses containing information matching a search request. The site addresses, or URLs, are displayed to the user in different highlighted colors. This allows the user to click the computer's mouse on a highlighted site address in order to connect the user to that new address, also known as hyperlinking or simply linking. Recent claims by search engine companies have placed the number of Internet addresses listed by major search engine providers to be over 50 million, with the number rising by millions each month. In contrast to the index method, the drawback to this search method is that the quantity of information retrieved on many search requests is too large to be of practical use. The ambiguity of the query language makes it difficult to narrow search results to relevant material. How-

22. There are some problems with this practice:
   By the time a spider completes its run through the Web, many links are already outdated, the newest pages aren't indexed at all, and sites that require passwords aren't included. Excite, with 50 million Web pages, claims to have one of the largest libraries, but it ignores 100 million others.

Ziegler, supra note 18.

23. For example, a search I performed with the WebCrawler search engine (http://webcrawler.com) with the search term "car" returned a list of 60,000 sites as of September 30, 1997.

24. URL is an abbreviation for Uniform Resource Locator. It is an addressing system used on the Internet to find information. URLs are in the format of service://hostname/path, so a URL of http://techlaw.scu.edu/links.html shows you how the Internet finds that document. The service (http, in this case) portion shows that it is part of the Web. The host name (techlaw.scu.edu) shows (extremely roughly, since virtual domain names exist) the host computer where the document is located. The path (links.html) shows more specifically where the document is related on the host. G. Burgess Allison, THE LAWYERS QUICK GUIDE TO NETSCAPE NAVIGATOR 15 (1997).


ever, by adding additional key words it is possible to pull more relevant material to the top of the search results list.

The early search engines were created by graduate students who placed them on the Internet for others to freely access.\textsuperscript{27} Search engines quickly became a success and their use grew so large (over a million inquiries a day) that their creators saw a potential for profit. Some creators sold their search engines to commercial Internet service providers who charged users an access fee.\textsuperscript{28} Other creators continued to allow free use of their search engines, but sold advertising space on their search engine's Web site. Currently, both fee-for-access and free access search engines generate substantial income from sales of advertising space.\textsuperscript{29} The advertising space appears at the top of almost every retrieval page. New advertisements appear on a user's computer screen each time a search query is completed. In television broadcasting, advertisers can buy time slots associated with a particular television program. Similarly with search engines, advertisers can purchase advertising space from search engine providers associated with a particular search parameter. Users who search for the keyword "printers," for example, would see an advertisement from a printer company, whereas an automobile advertisement might appear during a search for automobile sites.\textsuperscript{30}

The main function of a search engine is to provide users with a list of Web site links. These standard links, which send users to pages at other Web sites when the links are activated, do not raise copyright problems as long as the destination sites are themselves Web sites open to the general public. The very essence of the Web is linked site pages, so everyone who runs a Web site implicitly agrees to allow others to link to their site.\textsuperscript{31} In addition to these links, some search engines also create abstracts of Web sites by copying words

\textsuperscript{27} One of the first large scale search engines was WebCrawler, located at http://webcrawler.com. It was created in early 1994 by Brian Pinkerton, a graduate student at the University of Washington, and sold to America Online in 1995. Lycos, http://www.lycos.com, is a search engine developed in 1994 under the direction of Michael Mauldin at Carnegie Mellon University. Hakala, supra note 14.

\textsuperscript{28} See id.

\textsuperscript{29} Interview with Mark A. Van Heren, Software Engineer and founder of Excite, in Mountain View, Cal. (Dec. 20, 1996).

\textsuperscript{30} Jane Hodges, Sponsoring Search Terms Becomes a Popular Option, ADVERTISING AGE, Jan. 15, 1996, at 38. Search engine providers typically sell words for 1.5 cents to 10 cents per impression depending on the ad category, or $300 to $750 per month. An impression is recorded when a user downloads a page with an advertisement on it. Id.

\textsuperscript{31} Rose, supra note 7.
and sentences from Web pages. It is the creation and display of these
abstracts that may expose search engine companies to copyright in-
fringement liability.

III. COPYRIGHT PROTECTION OF WEB SITES

A. What Does Copyright Protect and Why?

Copyright protects original works of authorship fixed in a tangi-
ble medium from being copied without permission from the copy-
right holder.\textsuperscript{32} A result of this protection is that copyright holders re-
ceive compensation for use of their works. However, the primary
purpose of copyright is \textquoteleft[t]o promote the Progress of Science and the
useful Arts, by securing for limited Times to Authors and Inventors
the exclusive Right to their respective Writings and Discoveries.\textsuperscript{33}
This protection serves as an incentive to authors to create works that
will benefit the public.

However, existing copyright laws may not be able to provide ef-
fective protection for certain works of authorship, such as the digital
transmission of works on the Internet, that were not envisioned when
the existing copyright laws were enacted. Some commentators be-
lieve that \textquoteleft[i]ntellectual property law cannot be patched, retrofitted,
or expanded to contain digitized expression any more than real estate
law might be revised to cover the allocation of broadcasting spec-
trum. We will need to develop an entirely new set of methods as be-
fits this entirely new set of circumstances.\textsuperscript{34} These commentators
advocate the creation of new types of copyright protection, such as
the right to cache files\textsuperscript{35} or the right of net transmission.\textsuperscript{36}

Other commentators believe that existing copyright laws can
and must be applied to original works on the Internet.\textsuperscript{37} In fact, the

\begin{footnotesize}
\begin{enumerate}
\item U.S. CONST. art. I, § 8, cl. 8.
\item John Perry Barlow, \textit{The Economy of Ideas}, WIRED, Mar. 1994, at 84-85.
\item David Post, \textit{Bargaining in the Shadow of the Code: File Caching, Copyright, and
Contracts Evolving in Cyberspace}, \textit{in Symposium, supra} note 4, at chs. 7, 9; \textit{see infra}
note 59 for definitions of caching and mirroring.
\item Mark A. Lemley, \textit{Dealing With Overlapping Copyrights on the Internet, In
Symposium, supra} note 4, at chs. 5, 43-46.
\item Id. at 3; Larson, \textit{supra} note 6, \textit{citing The Working Group on Intellectual Property
Rights}.
\end{enumerate}
\end{footnotesize}
House Report on the Copyright Act\textsuperscript{38} supports this view. The Report states that "there is no disposition to freeze the doctrine in the statute, especially during a period of rapid technological change."\textsuperscript{39} Aside from some minor changes to the Copyright Act, recommended by the Clinton Administration’s Information Infrastructure Task Force, dramatic changes to current copyright laws are not being considered.\textsuperscript{40} Therefore, in the absence of any proposed legislation, activities on the Internet must be evaluated in the context of current copyright laws.

\textit{B. Web Site Copyrights}

Web sites often resemble the pages of printed books or magazines.\textsuperscript{41} The copyrights in these magazine-like Web pages will be very similar to traditional print media copyrights. In addition to protecting the literary work appearing on these printed pages, the Copyright Act also protects the rights of authors in their graphic works.\textsuperscript{42} As such there is a copyright in all original text and images on Web sites. Furthermore, there may be a copyright in the arrangement of the text and graphics as laid out on a page, depending on the degree of originality in page design and use of original graphic elements. There will also be a collective work copyright\textsuperscript{43} in the Web site as a whole, consisting of the collection of separately copyrighted text, images and other works which make up the contents of the site. Finally, when a Web site contains much information that is not itself copyrighted, such as statistics gathered from various public sources, there may be a copyright protecting the original manner in which the site gathers, arranges, and presents that information to users at the site.\textsuperscript{44}

\textbf{IV. Why Web Site Owners Might Enforce Their Copyrights}

\textsuperscript{39} H.R. Rep. No. 94-1476, at 66.
\textsuperscript{40} See infra note 52.
\textsuperscript{41} Rose, \textit{supra} note 7 ("Indeed, many Web sites function as 'magic magazines,' where the user can either read the pages in sequence or jump around to other pages at the same site or at other sites instantly").
\textsuperscript{43} 17 U.S.C. § 101 (1994) (defined as "a work, such as a periodical issue, anthology, or encyclopedia, in which a number of contributions, constituting separate and independent works in themselves, are assembled into a collective whole").
\textsuperscript{44} Id.
AGAINST SEARCH ENGINE COMPANIES

Although search engine companies profit from the labors of copyright owners, search engines are not the scourge of the Internet to copyright owners who display their creations on the Web. In fact, search engines provide copyright owners with useful tools to search the Web for infringement of their copyrights. However, there are circumstances in which owners may try to enforce their copyrights against search engine providers. Such circumstances include situations where search engines dissuade or hinder access to a site, undesired advertisements are displayed along with a site address, and licensing agreements are not feasible.

A. Search Engines May Dissuade Users from Going to a Web Site or Hinder Access to a Web Site

People who create Web sites normally want them to be found by others. However, search engines may actually reduce the number of people that access a particular site by providing an inaccurate abstract, supplanting the site with the abstract, or overloading the site’s server.

An incomplete or inaccurate site abstract may actually dissuade users from accessing a particular Web site. Although the abstracts displayed by search engines are intended to be solely informational, rather than endorsements, they might influence user decisions. These abstracts may be all that are read by users in order to help them select from among the sites returned by a query. Web site creators have little control over the content of these abstracts because most abstracts are automatically generated by a search engine’s computer program rather than written by the Web site owners themselves.45 In addition, these programs generate abstracts from information contained on the introductory pages of Web sites rather than from the deeper pages containing more substantive material.46 These practices of creating summaries often result in inaccurate or incomplete abstracts of Web

45. NetFirst Will Help Tame the Internet, THE INFO. ADVISOR, May 1995. Most search engines, including InfoSeek, excerpt the first couple of lines while others, such as NetFirst, actually have a human read a site and create an abstract.

46. Joseph Heifer, Commercialization of the Internet: A Webmaster’s Perspective, SEARCHER, Oct. 1995, at 22 (while site creators want to be found on the Web as easily as possible, some, such as corporations, “want to embed the information content of their Web sites within processes and algorithms of an interactive character to create a customer or add value to an existing customer”).
sites. A search engine user who would otherwise consider a site relevant may choose not to visit a site after reading an inaccurate or incomplete abstract. Therefore, since search engines might actually reduce access to some sites, Web site creators may want to enforce their copyrights as a means of persuading Web search providers to alter information contained within the search engine abstracts.

In other cases, search engine abstracts may supplant Web sites. A reproduction from a small Web site may serve the same informational function as the actual Web site. An abstract containing the essence of a Web page may serve to fulfill the needs of users who would have otherwise read the actual Web site, so that the user might not thereafter wish to see the original Web site contents. A Web site creator might care whether his material is read on his site rather than on the search engine site for different reasons. For example, a particular Web site may be new and expanding in size. A user reading a search engine abstract may not be aware of this and therefore may not be inclined to access the site at a later time. Alternatively, a Web site may contain graphics or advertisements that would otherwise not be seen when reading a search engine abstract.

Finally, certain search engine robots have overloaded networks and servers when they have logged onto a site. The resulting problems manifest themselves in refused connections, system performance slowdowns, and in extreme cases a system crash. If a search engine continually hinders access to a site, the site owner may want to enforce his rights under copyright as a means of persuading the search engine provider to reconfigure its robot.

All of the above factors operate to decrease the number of users accessing a site, which in turn results in lost readership and possibly lost revenues.

**B. Web Site Abstracts May Be Displayed Along with Undesired Advertisements or Abstracts of Other Sites**

Web search providers merely facilitate access to material on the Web without endorsing the material to which they point users. However, Web search providers do have the power to decide what advertisements get displayed on their search engine’s Web site. In contrast, Web site owners whose addresses are returned as a result of a search have no control over the advertisements that are displayed

47. Koster, supra note 19.
along with their address. Therefore, while search engine providers can decline to accept advertisements that they consider to be offensive or politically incorrect, Web site owners have no such ability.

Furthermore, when a broad search is performed for a certain product or service, the Web sites retrieved may include sites that criticize the product or service. Thus, abstracts critical of a product or service may be displayed along with the Web site for that product or service. For example, if a search for a particular model car is performed, a search engine may return a list of Web sites containing critical reviews of the car along with the Web site of the car’s manufacturer. Needless to say, the car manufacturer would not be pleased to have abstracts criticizing its car displayed alongside the abstract of their product.

Therefore, certain Web site owners may choose to enforce their copyrights as a way of preventing search engines from associating certain advertisements with their sites or displaying critical reviews together with their sites.

C. Licensing Agreements Are Impractical

"Media experts predict that the Internet and the World Wide Web will become the greatest advertising medium since television." Although Web site creators presumably want their sites advertised and accessed, some do not want search engines to exclusively profit from providing access to their copyrighted works. Web site owners may want search engine companies to pay for licensing rights. However, when search engines link to millions of Web sites, paying a reasonable licensing fee to even a fraction of those site owners would not be practical because search engine providers would only have limited funds available to license rights. Also, they would have to divide such funds among all the Web sites that want to license their copyrights. This could result in per-site licensing funds that would be too low to satisfy Web site owners.

---

48. Van Heren, supra note 29.
50. Rose, supra note 7, at 59 ("The publisher [or search engine provider] has to divide the money available for rights licensing among all works used at the site, but this may result in a per-work license payment that would strike the owners of the works used as insultingly small").
A possible legislative solution could be to impose compulsory licensing on the industry. Congress enacted a compulsory licensing scheme in the cable industry in order to reduce the need for negotiations among thousands of program copyright owners and cable operators for the right to rebroadcast copyrighted programs.\textsuperscript{51} Similarly, Congress could enact compulsory licensing for the Internet in order to reduce the need for negotiations between tens of thousands of Web site copyright owners and search engine companies for the right to reproduce copyrighted Web sites for the purpose of indexing the Web.

However, current sentiment disfavors compulsory licensing of copyrights. The Working Group on Intellectual Property Rights (a task force committee formed by President Clinton to make recommendations regarding changes to intellectual property law and policy) stated that "in certain circumstances, particularly where transactions costs are believed to dwarf per-transaction royalties, Congress has found it necessary to provide for compulsory licenses, ... [T]he Working Group finds that compulsory licensing of intellectual property rights is neither necessary nor desirable."\textsuperscript{52} The reasons given by the Working Group are that "[c]ompulsory licensing disregards marketplace forces, ... treats all works alike, even though their value in a competitive marketplace would likely vary dramatically ... [and] also treats all users alike."\textsuperscript{53} These same concerns are readily applicable to the Internet. As such, Congress may disfavor compulsory licensing on the Internet for the same reasons stated by the Working Group.

Therefore, confronted with the prospect of millions of Web site owners enforcing their copyrights and no legislative solutions on the


\textsuperscript{53} Id.
horizon, search engine providers will likely assert that their actions either do not constitute copyright infringement or are a fair use.

V. DO SEARCH ENGINES INFRINGE COPYRIGHTS?

A. What Constitutes Copyright Infringement

Subject to certain statutory limitations, the Copyright Act of 1976 gives copyright owners the exclusive rights in their copyrighted works to: reproduce the copyrighted work, prepare derivative works\(^5\) based on copyrighted works, distribute copies of the copyrighted work to the public, and display the copyrighted work publicly.\(^5\) Anyone who engages in any of these activities with a copyrighted work, without the permission of the copyright owner, is liable for copyright infringement. "Copies" are defined as "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."\(^5\) Any form of dissemination in which a "material object" does not change hands is not a reproduction, no matter how many people are exposed to the work.\(^5\) Courts, however, have generally accepted the premise that electronic transmissions on the Internet are material objects within the definition of "copies."\(^\) As such, search engine activity is subject to existing copyright law.

As previously discussed, there are two main types of search engines: those which merely provide address links to Web sites and those which, in addition to the links, provide abstracts of Web sites.\(^\)

\(^5\) A derivative work is defined as one that is based upon preexisting works, such as an abridgment or a condensation, or any form in which a work may be recast or transformed. 17 U.S.C. § 101 (1994).


\(^5\) “Recently, some Internet service providers (as distinguished from search engines) have also been copying, otherwise known as “mirroring” or “caching,” popular Web sites to
However, even the search engines that do not provide abstracts still copy material from sites. When a search engine's robot locates a Web site, a copy of the site is held in the search engine's Random Access Memory ("RAM") in order to catalogue the site into the search engine's database. Although a copyrighted work has not been displayed to the user, loading the Web site into the search engine's RAM may constitute a reproduction, or copy, of a copyrighted work which is a potential copyright infringement.

Information that is loaded into a computer's RAM exists only as long as the computer is kept on and is erased when a computer is turned off. Yet, under § 101 of the Copyright Act, this information can still be "perceived, reproduced, or otherwise communicated" with the aid of a computer because data loaded into RAM can be viewed on a computer screen, saved on a disk, printed out, or transmitted to other computers via a network. The issue is whether "copies" that exist in a computer's RAM are "fixed" such that they qualify as reproductions under the Copyright Act. Section 101 defines "fixed" as "sufficiently permanent or stable . . . for a period of more than a transitory duration." At least two courts have held that fixation of a computer file in RAM constitutes copying within the meaning of § 106. However, these court holdings appear to be

---

their own locations so that users can browse those sites within the provider's computer systems instead of calling out" on the Internet. Rose, supra note 7, at S9. Rose explains further that:

Their purpose is probably just to reduce the burden of redundant traffic on their Internet connections. . . . While mirroring may be an attractive solution to the problem of managing the Internet providers' computer resources, unless permission is obtained from the owner of the Web sites, such wholesale copying would still seem to be copyright infringement and possibly trademark infringement as well.

Rose, supra note 7 at S9. There are two kinds of caching Web servers: pull-caches and push-caches. Pull-caches respond to actual demand for Web pages by storing the most frequently accessed Web pages. Push-caches work by receiving pages from Web sites in anticipation of those pages being required. Johnson-Laird, supra note 4, at 25.

61. Lemley, supra note 36, at 4.
contrary to the legislative history of the Copyright Act and have been extensively criticized.

A non-displayed copy of a Web site in a search engine's RAM is at least arguably a violation of the Copyright Act. However, for the purposes of this section, such actions are presumed to be justified through the "fair use" defense to copyright infringement.

The other type of search engine compiles abstracts from information contained within Web sites, in addition to address links. Within this category, some search engine providers use computer programs to automatically create these abstracts. To create these abstracts, a robot program will first search the Internet for new Web pages and then temporarily copy the textual information found within these pages onto either a hard disk drive or into RAM. Once the page is in temporary memory, the abstract can be automatically computer generated by copying words and sentences from the information stored on the hard drive or in RAM. The words and sentences selected for the abstract vary depending on the search engine. For example, Excite's search engine copies three sentences from a Web page to create an abstract. The sentences selected are intended to be most representative of the material contained on the Web page.

---

64. H.R. Rep. No. 94-1476, at 52-53 (1976) ("[T]he definition of 'fixation' would exclude from the concept purely evanescent or transient reproductions such as those projected briefly on a screen, shown electronically on a television or other cathode ray tube, or captured momentarily in the memory of a computer") (emphasis added).

65. See Lemley, supra note 36, at 5 ("There are a number of problems with the MAI opinion — it does not refer to the legislative history, it does not discuss the 'transitory duration' prong of the fixation test, and the sources it cites are generally inapposite"), Niva Elkin-Koren, Copyright Law and Social Dialogue on the Information Superhighway: The Case Against Copyright Liability of Bulletin Board Providers, 13 CARDOZO ARTS & ENT. L.J. 345, 381-382 (1995).

66. See infra Part VI.

67. See supra Part II.

68. Van Heren, supra note 29.

69. Id.

70. Id. See also Laurie Flynn, Businesses Spring Up to Help With Web Searches, AUSTIN AM.-STATESMAN, Oct. 9, 1996, at D6.

71. See Van Heren, supra note 29 (the Web page is computer modeled to create vector representations of the sentences in a multidimensional space. The vectors are compared to identify three vectors which are closest to the center of the space defined by all the vectors. The center vectors comprise the sentences that contain the highest frequency of words found in the greatest number of sentences. For example, if vector 1 contains the words \{X,X,X,X,X,Y,A\}; vector 2 \{X,Y,A,A\}; and vector 3 \{Y,A,B\}, then vector 2 would be closest to the center of the vector space. Although word X has the highest frequency, it does...
After a Web site is copied, the address and an abstract of the site is compiled into a database. Following the creation of an abstract, the original information copied from a Web site is deleted to make memory space available for creation of new abstracts from other Web sites. Other search engine providers, such as NetFirst, create their databases by having humans read Web sites and then write abstracts about those sites. The legal issue presented by the creation of "abstracts" is whether such search engine activity constitutes copyright infringement.

When a search engine user inputs a search request, the search engine will then scan its own database, rather than the Internet, and display relevant links and abstracts to the user. As will be discussed further, both computer and human generated abstracts qualify as "copies" or "derivative works." Thus, search engine providers might be precluded from creating these abstracts and displaying them to the public unless authorized by the owner of the copyrighted material contained in the site.

In order to establish a cause of action for copyright infringement a plaintiff must prove two elements: ownership of a valid copyright and unauthorized copying by the defendant of constituent elements of the work that are original. Copying of constituent elements of the work is based on the substantial similarity between the original work and the copy. Damage or harm to a plaintiff is not an element of the cause of action. Therefore, a Web site owner does not have...
to suffer any actual harm, such as financial loss, in order to enforce his copyrights against a Web search provider.

The first element, ownership, would most likely not be an issue in an infringement action against a Web search provider. As such, this article only examines whether Web search providers may be liable for copyright infringement by copying Web site original works.

B. Substantial Similarity

Whether search engine activity is actionable depends on whether substantial material is copied from a Web site. Over a century ago, the Supreme Court held that for an unauthorized reproduction to constitute actionable infringement of a copyright there must have been copying of a substantial portion of the copyrighted work. This rule has remained in effect since the passage of the Copyright Act of 1976. In other words, a reproduction must be substantially similar to a copyrighted work to constitute actionable infringement. Because search engine abstracts are created by directly reproducing and compiling words and sentences from a Web page, factual "copying" would most likely be conceded by Web search providers. Therefore, a finding of legal copying and thus actionable infringement would turn entirely on whether the works are substantially similar. Slight or trivial similarities are not substantial and therefore are not infringing. However, two works may not be literally identical and yet, for the purposes of copyright infringement, may be found to be substantially similar. Unfortunately, "no bright line rule exists as to what quan-

79. Stillman v. Leo Burnett Co., 720 F. Supp. 1353, 1357 (N.D. Ill. 1989) (stating that, a "defendant who has copied from a plaintiff's work as a factual matter . . . may not have copied as a legal matter").
82. A threshold element of copying is factual copying: whether the defendant in creating their work used the plaintiff's material. Because the nature of an abstract is to describe the contents of a Web site, the existence of an abstract might be considered direct evidence of copying. See Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 359-60 (1991). However, if it is not possible to establish copying as a factual matter by direct evidence, as it is rare that the plaintiff has available a witness to the physical act of copying, then copying can be established indirectly by the plaintiff's proof of "access" and substantial similarity. 4 NIMMER, supra note 75, § 13.01[B]. It would be virtually impossible for a Web search provider to deny they had access to Web sites. Therefore, it is presumed that the "access" to the plaintiff's work would be conceded by a Web search provider.
83. 4 NIMMER, supra note 75, § 13.03[A].
turn of similarity is permitted before crossing into the realm of substantial similarity."

In determining whether substantial similarity exists, treatises have distinguished two different forms of similarity which have been coined "comprehensive nonliteral similarity" and "fragmented literal similarity." Comprehensive nonliteral similarity exists where the fundamental essence or structure of a work is duplicated in another work, rather than merely a particular line or paragraph or other minor segment. In contrast, fragmented literal similarity is literal similarity (virtually, though not necessarily, completely word for word) between the plaintiff's work and defendant's work, however this is not comprehensive similarity — that is, the fundamental substance or overall scheme of the plaintiff's work has not been copied; no more than a line, or paragraph, or page of the copyrighted work has been appropriated.

Search engine abstracts that are automatically generated by compiling only words and sentences from an entire Web page can be more readily categorized as fragmented literal similarity. The leading case on when fragmented literal similarity becomes substantial enough to constitute copyright infringement is *Feder v. Videotrip Corp.* The *Feder* court, relying heavily on the treatise *Nimmer On Copyright*, used both a quantitative and qualitative analysis to determine the existence of substantial similarity. The *Feder* court, giving more weight to the qualitative analysis, stated that:

> Where there is literal similarity . . . between the plaintiff's work and defendant's work . . . [but it] is not comprehensive . . . [then] the question in [this] case is whether the similarity relates to matter which constitutes a substantial portion of plaintiff's work - not whether such material constitutes a substantial portion of defendant's work. [Although] the quantitative relation of the similar material to the total material contained in the plaintiff's work is

---

85. *4 Nimmer, supra* note 75, § 13.03[A] (In attempting to understand the manner in which courts have drawn the line of substantial similarity, it is helpful to distinguish between two quite different forms of similarity. This distinction has received almost no express judicial recognition, so that it is necessary to invent our own terminology in contrasting the two such forms). The treatise terminology is quoted in *Twin Peaks Prods., Inc. v. Publications Int'l*, Ltd., 996 F.2d 1366, 1372 n.1 (2d Cir. 1993).
86. *4 Nimmer, supra* note 75, § 13.03[A][1].
certainly of importance... no bright line rule exists as to what quantum of similarity is permitted before crossing into the realm of substantial similarity. Even if the similar material is quantitatively small, if it is qualitatively important the trier of fact may properly find substantial similarity. If, however, the similarity is only as to nonessential matters, then a finding of no substantial similarity should result. 89

While the Feder court analyzed the difference between quantitative and qualitative similarity, for the purposes of clarity it is necessary to more carefully analyze both.

1. Quantitative Analysis

Unfortunately, there is no minimum quantitative measure that, as a matter of law, would establish a finding of substantial similarity. Whereas in one case the Supreme Court held that copying 300 words out of 200,00090 (0.15% of the original work) is clearly substantial,91 other courts have held that copying 100 words out of 70,000 (0.14% of the original work) was de minimus (not substantial).92 Although in a few cases it was held that copying but a single sentence was substantial,93 there appears to be a general consensus among most circuit and district courts that the copying of only one or two sentences is de minimus.94 Although these cases do not provide a definitive rule, they help to provide a guideline for comparison

Search engines typically copy a few sentences from a Web site.95 Although the quantity of Web site contents will vary dramati-

89. Id. at 1173-1176 (D. Colo. 1988).
91. Id. at 566.
93. Dawn Assocs. v. Links, 203 U.S.P.Q. (BNA) 831, 835 (N.D. Ill. 1978) (the sentence was part of promotional advertising for a motion picture and considered, by the court, to be a prominent and integral part of the plaintiff's advertising material). Universal City Studios, Inc. v. Kamar Indus., Inc., 217 U.S.P.Q. (BNA) 1162, 1166 (S.D. Tex. 1982) (the inscription of "E.T. Phone Home" appearing on defendant's products were found to be substantially similar to key lines of dialogue from a copyrighted motion picture because they would be readily recognizable to the lay observer).
95. Flynn supra note 70.
cally, on average, three sentences comprise 60 words and constitute approximately 1% of a Web site. In comparison to the above cited cases, both in percent and number of words reproduced, search engines copy a greater amount of material than courts have held to be *de minimus*. In both *Werlin* and *Toulmin*, a one to two sentence reproduction was held to be *de minimus*. In contrast, search engines reproduce what amounts to twice as many sentences from a Web site. Furthermore, the one percent of material that search engines reproduce is approximately seven times greater than the amount held to be substantial in *Harper*. Therefore, based on a quantitative analysis, search engine abstracts could be considered substantially similar to the Web sites from which they are compiled.

2. Similarity to Unprotected Material

Often, Web sites will consist of some original material and some noncopyrightable material, such as facts. Because no copyright exists in facts per se, the copyright in a Web site dealing with factual matters cannot be infringed by an abstract that copies such facts. But, if the factual copying occurs by verbatim copying of a particular description of such facts then copyright infringement can exist, so long as there is some originality in the particular description of such facts. In *Feist Publications Inc. v. Rural Telephone Services Co.*, the court refused to find copyright infringement of a phone book listing copied by a rival directory because the factual listings of names, towns, and numbers in alphabetical order reflected an utter absence of originality. Therefore, based on the reasoning in *Feist*, a verbatim copying of Web site phrases containing facts would only constitute actionable infringement if Web site creators described these facts with some originality.

---

96. Percentage calculated from an average of ten random Web sites.
101. *Marcus v. Rowley*, 695 F.2d 1171, 1177 (9th Cir. 1983) ("Even if it were true that plaintiff's book contained only facts, this argument fails because defendant engaged in virtually verbatim copying").
103. *Id. at 348-9.*
104. *Id.*
3. Qualitative Analysis: Audience Test

When determining if substantial similarity exists, courts have given greater weight to a qualitative analysis rather than a quantitative analysis. The qualitative test for substantial similarity has traditionally been the "Audience Test," otherwise known as the "Ordinary Observer Test," however, the Supreme Court has yet to rule on the legitimacy of this test. The Audience Test has been defined by some courts to be "whether an average lay observer would recognize the alleged copy as having been appropriated from the copyrighted work." Because abstracts are verbatim copies of Web page sentences, there is little doubt that the average person would recognize the sentences as having been appropriated from corresponding Web pages. Furthermore, search engine programs are designed to copy the sentences which are most representative of the material contained on a Web page. Therefore, it would appear that the goal in creating an abstract is to copy material with the most qualitative importance. As such, search engine abstracts are, by design, intended to be qualitatively similar to Web sites.

C. Infringement Conclusion

Search engine abstracts are substantially similar to the Web sites from which they are compiled. Not only do search engines copy quantitatively substantial material as compared with facts in other cases, but they are also designed to copy Web site material of the most qualitative importance. Therefore, the reproduction of such substantial material would constitute actionable copyright infringement. In any given case, the issue of substantial similarity cannot be resolved without considering the purpose for which the defendant's work is used. The purpose for which a copied work is used, along with other factors, is considered by courts in determining whether such behavior is actionable. In many cases, liability will be determined based on the doctrine of fair use, rather than merely whether substantial material is copied from an original work.

106. 4 NIMMER, supra note 75, § 13.03[E][1].
107. Ideal Toy Corp. v. Fab-Lu Ltd., 360 F.2d 1021, 1022 (2d Cir. 1966).
108. 4 NIMMER, supra note 75, § 13.03[A][2].
109. See infra Part VI[B][2].
VI. DOES FAIR USE APPLY TO SEARCH ENGINES?

A. Four Factors of Fair Use

Search engine providers may seek to avoid the cost and complexity of licensing or liability for possible copyright infringement by asserting the affirmative defense of fair use. This defense allows a defendant to avoid liability for activity which would otherwise constitute copyright infringement.110 The doctrine first appeared in 1841 in *Folsom v. Marsh*,111 although the precise term “fair use” did not make its appearance until 28 years later.112 Fair use was fashioned by judges to balance the author’s right to compensation for his work against the public’s interest in the widest possible dissemination of ideas and information.113 The doctrine was eventually codified with the enactment of the 1976 Copyright Act.114 Although § 106 of the Copyright Act confers exclusive rights to a copyright owner, § 107 of the Act provides others the privilege to make “fair use” of a copyrighted work.115 Specifically, § 107 states, in pertinent part:

Notwithstanding the provisions of section ... 106 ..., the fair use of a copyrighted work, including such use by reproduction in copies ... for purposes such as criticism, comment, news reporting, teaching..., scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include:

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) the nature of the copyrighted work;

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

---

113. Id.
(4) the effect of the use upon the potential market for or value of the copyrighted work.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.116

Because these factors are only broadly stated in § 107, courts are left with almost complete discretion in determining the extent to which any factor is present in a particular case.117

B. Judicial Interpretations of the Four Factors

The Supreme Court has held that § 107 requires a case-by-case determination whether a particular use of a copyrighted work is fair and noted that the four factors listed in the statute were only "nonexclusive" factors to be considered in determining whether there is a fair use in any particular case.118 Although § 107 provides no guidance as to the relative weight given to each factor, the Court held that "[this] last factor [effect on the market] is undoubtedly the single most important element of fair use. [Citation omitted] Fair use, when properly applied, is limited to copying by others which does not materially impair the marketability of the work which is copied."119

1. Effect on the Potential Market

The "effect of the use upon the potential market for or value of the copyrighted work" is the last factor listed in § 107.120 However, the Court has held it to be the single most important element of fair use.121 In Harper & Row, Publishers, Inc. v. Nation Enterprises,122 the Court held that a reproduction which directly competed for a share of the market for a copyrighted work would materially impair the marketability of that copyrighted work and, thus, would not be fair use within the meaning of the Copyright Act.123 In Harper, Na-

116. Id.
117. 4 Nimmer, supra note 75, § 13.05[A].
119. Id. at 566.
123. Id. at 568.
tion Magazine published a 2,250 word article, at least 300 words of which consisted of verbatim quotations of copyrighted expressions taken from an unpublished manuscript of President Ford's memoirs. The Harper Court held that the use of the quotations was not a fair use within the meaning of the Copyright Act primarily because the purpose of the use was to directly compete for a share of the market for excerpts of the manuscript; therefore, the use would materially impair the marketability of the copyrighted work.

Other courts have analyzed the effect on the market factor by balancing the benefit the public will derive if the use is permitted against the personal gain the copyright owner will receive if the use is denied. The less adverse the financial effect on the copyright owner, the less the public benefit need be shown to justify a fair use defense. However, any adverse effect on the potential market for the copyrighted work by reason of disparaging or unfavorable remarks in the allegedly infringing work is not relevant to the analysis. The adverse effect that is considered is the extent to which the copy supplants the original work in current or potential markets.

What are the adverse effects of search engines on the market for Web site copyrighted material? Some search engines merely copy a Web site into RAM or onto a hard drive and provide the search engine user with the Internet address of the site without displaying any of the copied information. If copied material is not displayed to the user then the search engine is not providing a substitute for the original. As such, there is no adverse effect on the market for the original material. On the contrary, the search engine has a beneficial effect on the market for the original work by facilitating access to that work.

Of greater concern are the search engines that display abstracts of Web sites to users. Although in theory these abstracts should stimulate interest in Web sites, an inaccurate or incomplete abstract could actually discourage interest in a Web site. A search engine

124. Id. at 545.
125. Id. at 568 (the Court considered all four factors in its decision that the reproductions did not constitute a fair use).
128. See supra Part IV[A].
user who would otherwise consider a site relevant may choose not to visit a site after reading an inaccurate or incomplete abstract.

Furthermore, an abstract that incorporates the heart of a Web site or substantially all of the material in a Web site (i.e. if the Web site contains less than a page of material) will end up serving the same function as that Web site. Such an abstract may fulfill the needs of users who would have otherwise read the actual Web site, thereby eliminating the user’s wish to see the original Web site. Therefore, search engine abstracts might adversely effect the market for some Web sites by dissuading users from accessing those sites. However, in comparison to other cases, it seems unlikely that courts would find that abstracts would realistically diminish the marketability of Web sites.

First, courts may analogize an incomplete or inaccurate abstract to that of a disparaging or unfavorable remark on a copyrighted work. As such, a search engine abstract may have an adverse effect but not one that is cognizable under the Copyright Act. Second, an incomplete abstract of a Web site could arguably cause a user to access that site for more information. Third, the chances of generating a misrepresentative abstract are small since they are created by verbatim copying. Finally, in contrast to Harper, search engines do not materially impair the marketability of Web sites because their purpose is not to directly compete for a share of the same market as the Web sites from which they copy material. Rather, search engine abstracts increase the market for Web sites by providing information about and facilitating access to those Web sites.

The effect on the “potential” market for Web site material must also be considered. A potential market has been defined as either an immediate or delayed market and includes harm to derivative works. But does this definition encompass any theoretical market no matter how unlikely that the copyright owner would engage in that market? For example, if copyrighted song lyrics have been appropriated for a 1960s trivia quiz book, could a potential market be defined

---

129. Id.
132. Cable/Home Communication Corp. v. Network Prods., Inc., 902 F.2d 829, 845 (11th Cir. 1990) (an immediate market exists concurrently with the copyrighted material’s market. A delayed market exits subsequent to and is the result of the copyrighted material’s market).
as the possibility of licensing song lyrics for quiz books? One court has favored fair use on this factor where the defendant's work fills a market niche that the plaintiff has no interest in occupying. Applying this to search engines, it could be reasonably argued that providing an indexing or search tool to the World Wide Web is a market niche that individual Web site creators would have no interest in occupying. However, when the Second Circuit applied the market niche reasoning it held that a potential market for television episodes included the market for books summarizing the episode plots. Therefore, a similar argument could be that the potential market for a publisher who owns multiple Web sites includes the market for an index that summarizes the information within those Web sites.

Therefore, a court's resolution of the effect on the potential market for an original work would lead back to a balancing of the benefit the public will derive if the use is permitted against the personal gain the copyright owner will receive if the use is denied. Because the Web comprises over 50 million Web pages and is growing by millions each month, search engines are essential for the survival and expansion of the Web. As such, permitting use of Web site material for search purposes ensures survival of an important and widespread means to publish news, information, and views to the public. Such a public benefit can readily be seen to outweigh any small pecuniary gains to Web site owners from licensing the rights to summaries of their Web sites.

Therefore, the fourth factor would seem to favor Web search engine providers, due to the substantial benefits of Web search tools to the public and the improbability that search engine abstracts would either directly compete for market share of copyrighted works, diminish market access to copyrighted works, or cause significant loss of potential revenues from licensing.

133. 4 NIMMER, supra note 75, § 13.05[A][4].
134. Twin Peaks Prods. Inc. v. Publications Int'l, Ltd., 996 F.2d 1366, 1377 (2d Cir. 1993) (fourth factor favors plaintiff television producer, which may rightfully wish to occupy market of books summarizing episode plots).
135. Id.
136. See discussion supra note 26.
137. Rose, supra note 7.
138. Id.
2. Purpose and Character of the Use

The first factor listed in § 107 is "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes." This factor does not identify which uses beyond nonprofit educational purposes will constitute a fair use. However, the preamble to § 107 lists purposes that might qualify as fair use: criticism, comment, news reporting, teaching, scholarship, and research. The uses listed in the preamble are merely illustrative and not exhaustive, therefore, types of uses beyond the six enumerated in the preamble may also be considered fair use.

The purpose and character of a search engine is to provide users with a tool to search for specific information on the Web in order to facilitate their research. The Web was initially used for scientific research but has since expanded to include academic, business, and personal uses. As previously discussed, one type of search engine merely copies a Web site into RAM or onto a hard drive in order to generate a list of Web site addresses of relevance to the user without displaying any copied information to the user.

Since the use is to create a directory tool for researching the Internet and not to display any copyrighted Web site material, courts may likely consider this activity a fair use. On the other hand, other search engines do use the copied information to construct abstracts of Web sites for display to search engine users. Although these abstracts contain copied information, they are merely intended to supplement the researching function by providing users with a brief description of the content found in the Web site. Originally, search engines were placed on the Internet for others to freely access. As nonprofit research tools, the copying performed by search engines, including the display of abstracts, might be readily identified as a fair use.

Currently, however, Web search providers are generating substantial profits by either charging usage fees or selling advertisement

142. Larson, supra note 6.
143. See supra Part V[A].
144. See American Geophysical Union v. Texaco, Inc., 37 F.3d 881, 900 (2d Cir. 1994).
space. Therefore, the use made of copyrighted material by Web search providers is of a commercial nature. The Harper Court stated that "[t]he fact that a publication [is] commercial as opposed to non-profit is a separate factor that tends to weigh against a finding of fair use." The commercial nature of a use, however, is not dispositive since many fair uses are conducted for profit. As the Campbell Court noted, "[I]f, indeed, commerciality carried presumptive force against a finding of fairness, the presumption would swallow nearly all of the illustrative uses listed in the preamble paragraph of § 107, including news reporting...[and] teaching...since these activities 'are generally conducted for profit in this country.'" As another court noted, "publishers of educational textbooks are as profit-motivated as publishers of scandalmongering tabloid newspapers." Therefore, the commercial nature of search engine use of copyrighted works would probably not be given too much weight in a fair use analysis, especially since search engines do not compete for the same market share as, or supplant, the original Web site material.

As such, the fair use inquiry will focus on the character of the use, rather than the commerciality. The Supreme Court described this inquiry as an investigation as to "whether the new work merely 'supersede[s] the objects' of the original creation, or instead adds something new, with a further purpose or different character,...it asks, in other words, whether and to what extent the new work is "transformative." The reason why the Court believed a transformative work to be a fair use is because the creation of that transformative work furthered the goal of copyright — to promote the sciences and the arts. The Campbell Court held that parody was fair use, similar to comment or criticism, because it could provide a social benefit by shedding light on an earlier work, and in the process create a new work. Similar to the reasoning in Campbell, the material that a search engine copies is not intended to supplant the original

---

148. 4 NIMER, supra note 75, § 13.05[A][1][a].
149. See supra Part VI[B][1].
150. Campbell v. Acuff-Rose Music, Inc., 510 U.S. at 579 (the material must be used in such a way that it is both productive and employed in a different manner or for a different purpose than the original material).
151. Id.
152. Id. at 579-80.
Web site. Rather, its purpose is to alert the user to the existence of that Web site. At best, no copied material is displayed to the user with non-abstract search engines. At worst, only a few sentences reproduced from material within a Web site are displayed to the user as with abstract search engines. Although a verbatim reproduction is not "transformative," it could be considered a form of neutral commentary that provides a benefit to the public: facilitating access to information.

Some courts, however, have been reluctant to afford fair use protection to reproductions that contain little criticism or comment and are used for commercial purposes. In Religious Technology Center v. Netcom On-Line Services, Inc., the court held that "the [large] percentage of the plaintiff's works copied combined with the minimal added criticism or commentary negates a finding of fair use." However, unlike Religious Technology Center, a search engine abstract does not comprise a large percentage of and is not a practical substitute for an original work that likely consists of hundreds of sentences.

In contrast, the Twin Peaks court held that although a work may be a criticism, mere summarization of a work, which does not contain transformation nor serve a legitimate purpose, will weigh against a finding of fair use. Unlike Twin Peaks, search engine abstracts, though not transformative, are summaries that serve a legitimate purpose. Search engines further the copyright goal of promoting science by providing a tool that is essential to the survival of the Internet as a state-of-the-art publishing medium. Furthermore, an earlier court in Triangle Publications, Inc., v. Knight-Ridder Newspapers, Inc. held that copying the cover page of a magazine for use in comparative advertisements was a fair use. That court determined that the first factor weighed in favor of fair use mainly because "there was no at-

154. Id. at 1249. Defendant copied all or almost all of many works, which were predominately short documents of less than three pages. Id. at 1247.
155. Id. at 1249.
156. Twin Peaks Props., Inc, v. Publications Int'l, Ltd., 996 F.2d 1366, 1374-1376 (2d Cir. 1993) (a work can be criticism notwithstanding that it aims at television lowbrow audience; but a book summarizing the precise details of eight television episode plots, without transformation, serves no legitimate purpose).
157. 626 F.2d 1171, 1172 (5th Cir. 1980).
1. INTERTAETSEARCHENVGINE COPYNG

1. TEMPT TO PALM OFF... [PLAINTIFF'S] PRODUCT AS THAT OF [DEFENDANT'S]. "

Similar to the reasoning in Triangle, the copying of a Web page for information indexing is classified under fair use because search engines do not attempt to “palm off” indexed Web pages as their own.

Therefore, courts may be more inclined to extend the scope of fair use to non-transformative reproductions, which do not supplant original works and only copy insubstantial amounts, to further scientific progress that benefits the public. As such, this factor may weigh slightly in favor of Web search engine providers.

3. Nature of the Copyrighted Work

The second factor listed in § 107 is “the nature of the copyrighted work.”159 Under this factor, the more creative a work, the more protection it should be given from copying. Whereas, the scope of fair use should not extend to more informational or functional works.160 This distinction is based on the belief that some works are closer to those Congress intended for copyright protection.161

It is difficult, if not impossible, to generalize the typical “nature” of Web sites that are copied by search engines. The content found within different Web sites varies from almost completely factual, such as a weather report Web site,162 to highly creative, such as a Web site containing poems.163 Therefore, among all the factors, the nature of the copyrighted work is the most fact specific factor subject to nearly exclusive analysis on a case-by-case basis. However, all cases share at least one fact in common: a Web site is created to facilitate the distribution of a “work” to the public. Some courts have commented that material prepared for general public distribution may be more susceptible to fair use copying.164 Yet, it must be remembered that the purpose of copyright is to encourage an author to dis-

158. Id. at 1176.
160. Diamond v. Am-Law Corp., 745 F.2d 142, 148 (2d Cir. 1984) (“informational ... works may be more freely published under Section 107 than those of a creative nature”).
seminate ideas and information to the public by giving that author exclusive rights for a limited time. If works prepared for general public distribution are readily susceptible to fair use copying, authors will be less inclined to share their works with the public, thereby defeating the purpose of copyright.

Because this factor is heavily dependent on a fact specific analysis, a determination of which side this factor would tend to favor is not possible.

4. Amount and Substantiality Used

The third factor listed in § 107 is "the amount and substantiality of the portion used in relation to the copyrighted work as a whole." Search engines copy entire Web pages verbatim onto RAM or hard drives in order to provide users with links or to create abstracts of those Web pages. As such, this factor would tend to weigh against a finding of fair use. However, since these copies are never displayed to the user, this factor may be of less importance when weighed against the insubstantial effect of such non-displayed copies on the potential market of Web sites.

Of greater importance is the amount and substantiality used in creating abstracts displayed to the user. This analysis is very similar to that of substantial similarity for actionable copyright infringement, involving both a quantitative and qualitative analysis. The general rule applied by many courts is that the essence of a copyrighted work cannot be reproduced. In the Harper case, the Court held that although the words that were actually quoted were an insubstantial portion of the work, in view of the expressive value of the excerpts, Nation Enterprises took what was essentially the heart of the book. In contrast, the Triangle court held that reproducing the cover of a TV guide magazine was not substantial because the essence of the magazine, the program schedules and articles, was not copied.

166. See infra Part V[D].
168. Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. at 564-565 (the quoted portions were excised "precisely because they qualitatively embodied Ford's distinctive expression").
The relevant inquiry in the present case, therefore, is what is the "essence" of a Web site? As with the "nature" of a Web site, it is difficult to generalize the typical "essence" of different Web sites. Yet, as previously discussed, a search engine is designed to copy sentences which are most representative of the material contained on a Web page. As such, arguably the goal of a search engine is to capture a Web site's essence, whatever that essence may be. Whether the goal is achieved, however, can only be determined on a case-by-case basis. On the other hand, the essence of a Web site may be a reflection of its look and feel, as well as the power of the creator's words. Thus, an abstract that simply lifts the words most often used does not reproduce the "essence" of the site. Generally, though, this factor may weigh against fair use because search engines copy more material than the amount considered substantial by most courts, and, regardless of actual success, search engines are designed to reproduce the heart of a Web site.

C. Fair Use Conclusion

How do the § 107 factors balance out in the fair use analysis? Generally, the improbability that search engines would adversely effect the potential market of Web sites seems to favor a finding of fair use. Although considered the most important factor by the Supreme Court, this factor alone is not dispositive to the fair use analysis. Courts may consider Web sites, generally, to be highly creative works and search engines to serve only the pecuniary interests of their owners. As such, if the rights of Web site copyright owners are considered paramount and if abstracts are deemed merely non-transformative verbatim reproductions of a Web site's essence, then other § 107 factors may outweigh the market effect factor. However, as one commentator has noted, "whatever [outcome] is chosen, it ought to be chosen on the basis of policy concerns, not by automatic application of rules designed with another context entirely in mind." 171

170. See supra Part V[B][3].
171. The Anatomy Of the Internet Meets The Body of The Law in SYMPOSIUM, supra note 4, at 46.
VII. Conclusion

Whether search engine abstracts violate a Web site creator's copyrights depends both on the quantitative and qualitative substantiality of material copied from the Web site. Web sites vary dramatically in both content and size. Due to this variation, a determination of the quantitative similarities between an abstract and its respective Web site would only be feasible on a case-by-case basis. Search engines, however, are designed to copy sentences which are most representative of a Web site's content, regardless of what that content may be. Assuming search engines perform as designed, the resulting abstracts could readily be seen as qualitatively similar to the Web sites from which they are generated. Since greater weight is given to a qualitative analysis, based on qualitative similarity alone, courts should consider that the compilation of search engine abstracts constitutes an infringement of the Web site owner's copyright.

However, a holding of copyright infringement would not be the end determination of liability. Search engine companies would argue that although substantial material is copied from an original work, the defense of fair use makes such copying nonactionable. The doctrine of fair use requires courts to consider factors such as the effect of the copying on the potential market of the copyrighted work, the nature and purpose of use of the copyrighted work, and the amount and substantiality of the copyrighted work that is used. The fact that search engine abstracts would not likely adversely effect the potential market of Web sites would tend to favor a finding of fair use. Whereas, if abstracts are deemed to be merely non-transformative, verbatim reproductions of the essence of highly creative Web sites, then the insubstantial effect these abstracts have on the Web site's market may be outweighed, and the fair use defense will be defeated.

An analysis of fair use, however, is not limited to the four factors enumerated in the Copyright Act. When evaluating fair use factors, an extremely important consideration by most courts has been the public interest served by the use of the copied material and by copying the work itself. Courts have come to recognize that there are situations in which the copyright holder's interest in a maximum financial return must be subordinated to the greater public interest in

---

the development of art, science, and industry. The Internet is a major technological achievement of science itself and, more importantly, enables the progress of other sciences and arts. The public interest in the Internet is not only the development of the publishing industry, but also the development of what has been called a communications revolution. Search engines are an essential tool for the survival of the Internet as it expands. Therefore, in light of the benefit provided to the public and the improbability of supplanting copyright works, search engine copying of Web sites should be considered a fair use. A contrary ruling would severely limit the effectiveness of the Internet as a practical tool for information gathering and for the dissemination of information. Such a ruling could possibly eliminate what has become one of this century’s most effective means of “promot[ing] the Progress of Science and the useful Arts.”

While not discussed in this article, if search engines are faced with claims of copyright infringement, an alternate defense could be that the Web site owners acquiesced to the copying of their work. Such a defense is based on the doctrine of estoppel. A Web site owner’s long-standing acquiescence to the copying of his work may

---

174. Meeropol v. Nizer, 417 F. Supp. 1201, 1207 (S.D.N.Y. 1976); Rosemont Enters., Inc. v. Random House, Inc., 366 F.2d 303, 307 (2d Cir. 1966). See also Iowa State Univ. Research Found. v. American Broad. Cos., 651 F.2d 57, 60 (2d Cir. 1980) (fair use “permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which the law is designed to foster”); Howard C. Anawalt, Nine Guidelines and a Reflection on Internet Copyright Practice, in SYMPOSIUM, supra note 4, at chs. 2, 10 (the concept of fair use allows copying for certain socially useful purposes).

175. See supra Part II[A].

176. Lemley, supra note 36, at 38 (“Either copyright owners will literally shut down the Net by enforcing the rights the courts have declared they have, or those ‘rights’ will continue to be ignored in the interests of day-to-day living . . . ”).


178. As stated by the Ninth Circuit, the estoppel defense has four elements:

1. the party to be estopped must know the facts of the defendant’s infringing conduct;
2. he must intend that his conduct shall be acted on or must so act that the party asserting the estoppel has the right to believe it is so intended;
3. the latter [infringer] must be ignorant of the true facts; and
4. he must rely on the former’s conduct to his injury.

Hampton v. Paramount Pictures Corp., 279 F.2d 100, 104 (9th Cir. 1960). These elements have been adopted by numerous district and circuit courts. 4 NIMMER, supra note 75, § 13.07 n.3 (1997).
be sufficient to prove one of four estoppel elements.\textsuperscript{179} Web site owners have the means to prevent selective search engine robots from visiting their site by placing a simple file onto their server.\textsuperscript{180} As such, acquiescence could be implied from a Web site owner's knowledge of the copying performed by search engines coupled with the failure to initiate preventive measures. However, a search engine's greatest difficulty in proving an estoppel defense would be in proving another estoppel element, ignorance of the facts. The appearance of a copyright notice on copies of a work has been held to be sufficient to give an infringer the requisite notice to prevent claims of ignorance.\textsuperscript{181} As such, merely posting a copyright notice on a Web site page may be all that is needed to defeat an estoppel claim.


\textsuperscript{180} Koster, \textit{supra} note 19.

\textsuperscript{181} Hampton v. Paramount Pictures Corp., 279 F.2d at 104.