

NOTE

LEGAL RAMIFICATIONS OF “BLACK HAT” SEARCH ENGINE OPTIMIZATION

Young J. Yoon

TABLE OF CONTENTS

I. INTRODUCTION	160
II. WHAT IS “BLACK HAT” SEARCH ENGINE OPTIMIZATION?	162
A. <i>Content Scraping</i>	163
B. <i>Link Spamming</i>	164
C. <i>Other Techniques</i>	165
1. <i>Keyword Stuffing</i>	165
2. <i>Link Farming</i>	166
3. <i>Cloaking and Doorway Pages</i>	167
III. RECENT DEVELOPMENTS AGAINST COPYRIGHT INFRINGEMENT AND SPAM BY “BLACK HAT” SEARCH ENGINE OPTIMIZATION	168
IV. SO IS IT LEGAL? POTENTIAL LEGAL RAMIFICATIONS OF “BLACK HAT” SEARCH ENGINE OPTIMIZATION	169
A. <i>Copyright Infringement under the DMCA</i>	169
1. <i>Current State of the Law</i>	169
2. <i>Remedies</i>	170
3. <i>Defenses</i>	170
4. <i>Application to Content Scraping</i>	171
B. <i>Trespass to Chattel</i>	173
1. <i>Current State of the Law</i>	173
2. <i>Application to Content Scraping and Link Spamming</i>	174
C. <i>The CAN-SPAM Act of 2003</i>	175
1. <i>Current State of the Law</i>	175
2. <i>Application to Link Spamming</i>	176
V. WHO CARES? POTENTIAL PARTIES WHO MAY PURSUE LEGAL ACTION AGAINST “BLACK HAT” SEARCH ENGINE OPTIMIZATION	177
A. <i>Website Owners</i>	177
B. <i>Search Engines</i>	178
C. <i>The Federal Trade Commission (“FTC”)</i>	179

VI. GOOGLE UPDATES AND THEIR POTENTIAL FOR ABUSE.....	180
VII. PROPOSAL	182
A. <i>Important Questions to Consider in Search Engine Dispute Resolution</i>	182
B. <i>Administrative Agency (“Federal Search Commission”) Regulation: A Less-than-ideal Solution</i>	183
C. <i>Creation of a Common Law Tribunal and Industry Panel.....</i>	185
VII. CONCLUSION	186

I. INTRODUCTION

Scholars have extolled that “[t]he architecture of the Internet, as it is right now, is perhaps the most important model of free speech since the founding [of the Republic].”¹ Indeed, the Internet has become today’s most important communication tool, providing anyone with means to publish their content and reach a worldwide audience.² Search engines, the most notable of which is Google, collect data from web pages to help users reach data or information of interest.³ According to a July 2009 study, 81 percent of all Internet users enter the web environment via search engines.⁴ Google handles 7.2 billion page views every day, amounting to 87.8 billion monthly search queries worldwide.⁵ “Most websites rely on the search engine[s] for half of their traffic,”⁶ and revenue generated through Google’s advertising network “either supplements or provides full-time income to many site owners.”⁷

The popularity of using search engines in reaching Internet content has shifted the focus of web marketing to achieving high rankings in search results. Search engine optimization (“SEO”) refers to techniques used in website design and development to improve a website’s search rankings and various

¹ LAWRENCE LESSIG, CODE 237 (2d ed. 2006).

² Nursel Yalçın & Utku Köse, *What is Search Engine Optimization: SEO?*, 9 PROCEDIA SOC. & BEHAV. SCI. 487, 487 (2010).

³ *Id.*

⁴ *Id.*

⁵ *Google Facts and Figures (Massive Infographic)*, PINGDOM (Nov. 4, 2012, 7:10 PM), <http://royal.pingdom.com/2010/02/24/google-facts-and-figures-massive-infographic/>. 620 million daily visitors account for these searches. *Id.*

⁶ Editorial, *The Google Algorithm*, N.Y. TIMES, July 14, 2010, <http://www.nytimes.com/2010/07/15/opinion/15thu3.html>.

⁷ Nathania Johnson, *Google Responds to FTC’s Self-Regulatory Principles*, SEARCH ENGINE WATCH (Apr. 8, 2008), <http://searchenginewatch.com/article/2054454/Google-Responds-to-FTCs-Self-Regulatory-Principles>.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

methods have developed in the field in the new millennium.⁸ SEO techniques fall into two broad categories: “white hat,” which search engines consider ethical and best practices, and “black hat,” which include techniques that “cheat” their way around search algorithms and have questionable ethical and legal character, although they may be highly effective in the short term. Many U.S. courts take a “dim view” of these forms of “black hat” SEO, branding them as “infringement . . . techniques leading to initial interest confusion.”⁹

In August 2012, Google announced the launch of a new search algorithm “to take into account the number of valid copyright removal notices it receives for any site”¹⁰ under the Digital Millennium Copyright Act (DMCA). Though the algorithm appeared to be a friendly nod in the direction of copyright holders of web content – most notably the Motion Picture Association of America (MPAA) and the Recording Industry Association of America (RIAA) – it signaled significant consequences in the SEO industry by punishing “black hat” practices.¹¹ Google continues to battle against “black hat” SEO techniques with its May 2013 algorithm update, codenamed “Penguin 2.0,” designed to eradicate “black hat” web spam.¹²

Nevertheless, private copyright holders on the Internet may not have enough resources or interest to pursue legal action against unethical SEO practices.¹³ Because Google does not fully disclose the mechanisms behind its search engine algorithms, Internet professionals have also expressed concern about possible abuse by which the new algorithm could “punish innocent businesses [that] receive unwarranted notices.”¹⁴

⁸ See *The Time Has Come to Regulate Search Engine Marketing and SEO*, TECHCRUNCH (July 13, 2009), <http://techcrunch.com/2009/07/13/the-time-has-come-to-regulate-search-engine-marketing-and-seo/> [hereinafter *TechCrunch*]. Because most search engines closely guard their algorithms as trade secrets, some believe that search engine optimization techniques are “more voodoo than science.” *Id.*; Jeffrey D. Goldman & Eric J. German, *Pollution in the Blogosphere*, L.A. Law, June 2007, at 35.

⁹ Goldman & German, *supra* note 8, at 35.

¹⁰ Cyrus Farivar, *Google to drop search rankings of sites with many takedown notices*, ARS TECHNICA (Aug. 10, 2012, 3:58 PM), <http://arstechnica.com/tech-policy/2012/08/google-to-drop-search-rankings-of-sites-with-many-takedown-notices/>.

¹¹ *Id.*

¹² Matt Cutts, *What Should We Expect in the Next Few Months in Terms of SEO for Google?*, YOUTUBE (May 13, 2013), <http://youtu.be/xQmQeKU25zg>. “Penguin” refers to the aspects of the Google algorithm that analyze link structures for potential spam; “Panda” refers to the aspects that analyze site content. Nick Stamoulis, *Google Panda Update vs. Google Penguin Updates*, BRICK MARKETING BLOG (last visited Aug. 24, 2013), <http://www.brickmarketing.com/blog/panda-penguin-updates.htm>.

¹³ Goldman & German, *supra* note 8, at 34.

¹⁴ Deanne Katz, *Google’s New Search Rankings Penalize Copyright Infringers*, FINDLAW (Aug. 16, 2012, 5:46 AM),

In Part II, this Note surveys various “black hat” SEO techniques such as content scraping, link spamming, keyword stuffing and link farming, and considers how they differ from the industry-accepted “white hat” techniques. Part III provides a brief overview of provisions of the DMCA as well as Google’s recent efforts targeting infringers and other noncompliant, potentially unethical practices on the Internet. Part IV of this Note analyzes the current state of U.S. copyright laws, trespass to chattel provisions, and anti-spamming laws and their potential application to “black hat” SEO. This Note also explores the adequacy of these laws by considering what type of parties can bring suit or regulate “black hat” SEO (Part V), and addressing the problem of Google’s potential abuse of innocent businesses (Part VI). Finally, Part VII emphasizes the need for a more coherent, centralized regulation of web search, outlines the key points to consider for effective execution, analyzes possible solutions including those devised outside the United States, and proposes that the creation of a common law tribunal solely for search engine and online disputes is the ideal approach to provide adequate protection against “black hat” SEO.

II. WHAT IS “BLACK HAT” SEARCH ENGINE OPTIMIZATION?

A search engine is software that collects various website data, including the uniform resource locator (URL), keywords, and keyword groups that “define the content of the website, the code structure that forms the [displayed] web page, and . . . links provided on the website.”¹⁵ Special programs, commonly known as crawlers, spiders, or bots, collect such website data by “using the hyperlink structure of the web, . . . navigat[ing] through web pages periodically and captur[ing] changes . . . since [the] last navigation.”¹⁶ A search engine then indexes and stores the collected data in a database.¹⁷

Search engine optimization enables a website to appear among the top results of a search engine for certain keywords.¹⁸ To optimize a website for search performance, Google recommends various “white hat” SEO techniques including, but not limited to: creating “unique [and] accurate page titles;” making use of a “description meta tag;” improving the structure of the website’s URLs; making the website easier to navigate; optimizing displayed images; using “heading tags appropriately;” and offering “quality content and services.”¹⁹

http://blogs.findlaw.com/free_enterprise/2012/08/googles-new-search-rankings-penalize-copyright-infringers.html.

¹⁵ Yalçın & Köse, *supra* note 2, at 488.

¹⁶ *Id.*

¹⁷ *Id.*

¹⁸ *Id.*

¹⁹ GOOGLE, SEARCH ENGINE OPTIMIZATION STARTER GUIDE 4-20 (2010), *available at*

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

In contrast, “black hat” SEO techniques are “cheats” that take advantage of search algorithms. Though highly effective in the short term, search engines consider “black hat” SEO techniques unethical and often punish websites using these strategies by demoting their search rankings or, in extreme cases, banning a website outright from all search results.²⁰ Those who engage in “black hat” practices often consider search engines as enemies while the “white hats” strive to “deploy sound website-building tactics . . . [and] view [search engines] as friends.”²¹ Popular “black hat” SEO techniques include content scraping, link spamming, cloaking, doorway pages, keyword stuffing, and link farming.²²

A. Content Scraping

Websites that offer “quality content and services” attain higher rankings in search results.²³ Unfortunately, creating quality content is “time-consuming and . . . requires old-fashioned hard work,” while “purchasing interesting or

<http://www.google.co.jp/intl/en/webmasters/docs/search-engine-optimization-starter-guide.pdf> [hereinafter GOOGLE]. The guide also provides more in-depth discussions of “white hat” techniques. *Id.* Google also outlines various tools it offers to maximize effects of SEO strategies, including Google Analytics and Google Webmaster Tools. *Id.* at 31.

²⁰ See David Segal, *The Dirty Little Secrets of Search*, N.Y. TIMES, Feb. 12, 2011, <http://www.nytimes.com/2011/02/13/business/13search.html>. In 2011, Google demoted J.C. Penney for link spamming. *Id.*; see also Loren Baker, *Google Bans BMW for Search Spamming*, SEARCH ENGINE J. (Feb. 6, 2006), <http://www.searchenginejournal.com/bmw-booted-from-google-for-spamming/2886/>. Even prominent business corporations have fallen victim to such punishment: in February 2006, Google’s web spam team discovered deceptive SEO practices on BMW’s German website (bmw.de), which presented different content for search engine crawlers than it did to human users. *See id.* Following a complete de-listing from Google’s index, BMW responded quickly by removing the page and contacting Google for re-inclusion. Matt Cutts, *Recent Reinclusions*, MATTCUTTS.COM (Feb. 7, 2006), <http://www.matcutts.com/blog/recent-reinclusions/> (Matt Cutts is the head of Google’s web spam team).

²¹ Andrew Goodman, *Search Engine Showdown: Black Hats vs. White Hats at SES*, SEARCH ENGINE WATCH (Feb. 16, 2005), <http://searchenginewatch.com/article/2066090/Search-Engine-Showdown-Black-Hats-vs.-White-Hats-at-SES>. See SEO Black Hat, <http://seoblackhat.com> (last visited Nov. 9, 2013), for a forum of “black hat” SEO professionals and examples of animosity among the “black hat” community against search engines.

²² “Framing,” or “stealing bandwidth” by taking existing content from a website and reproducing it on a differently styled stage, is also an unethical practice. Goldman & German, *supra* note 8, at 34. However, this Note does not discuss “framing” as it is outside of the technical scope of SEO.

²³ GOOGLE, *supra* note 19, at 14.

attractive original content is also extremely expensive.”²⁴ “As a result, much of what appears on [the Internet] today is recycled, or simply stolen [material] from established competitors that spend money to create appealing content.”²⁵

Content scrapers are websites (usually blogs) that “post[] verbatim copies of content that is produced by other, more popular” websites.²⁶ This serves the purpose of “draw[ing] in users who are searching for the website [with the original content], tricking them into visiting the scraper site.”²⁷ A variation on the content scraper model is a “splog,” or a spam site that “mimics the look of a blog by lifting chunks of text from [various] other sites and pasting them together, often with the aid of automated software and usually resulting in haphazard, nonsensical prose.”²⁸

The short-term benefits of content scraping are easy to see in context of online advertising. By using pay-per-click advertising servers such as Google’s AdSense and Yahoo’s Overture, “any site can populate itself with advertisements that coincide with the subject matter covered by the site.”²⁹ Advertisers using such services pay website owners a small amount of money for every time a user visits a page on which the solicited ads are visible.³⁰ Thus, if a content scraper website “can trick a large number of people into visiting [it even for] a short period of time,” it can generate a considerable amount of ad revenue with minimal work.³¹ Content scraper sites, because they copy the content of other websites, are not only unethical but also clearly engage in copyright infringement.³²

²⁴ Goldman & German, *supra* note 8, at 32.

²⁵ *Id.* at 34.

²⁶ John Richards, *We All Know That Black Hat SEO Is Unethical and Ineffective, but It May Also Be Illegal*, OPPORTUNITIES PLANET (Feb. 17, 2012), <http://www.opportunitiesplanet.com/seo-2/black-hat-seo-could-be-illegal/>.

²⁷ *Id.*

²⁸ Goldman & German, *supra* note 8, at 36.

²⁹ *Id.* at 34.

³⁰ Richards, *supra* note 26.

³¹ *Id.* Leaders of various industries spend tens of millions of dollars for the highly targeted advertising that pay-per-click technology offers. See Rick Spence, *The Surprising Economics of Pay-per-click*, FINANCIAL POST (Jan. 27, 2012), <http://business.financialpost.com/2012/01/27/the-surprising-economics-of-pay-per-click/>.

³² John Richards, *When Engaged in SEO, Be Aware of These Legal Issues*, BAY OF BLOG (Apr. 19, 2011), <http://www.bayofblog.com/2011/04/when-engaged-in-seo-be-aware-of-these-legal-issues.html>. The Copyright Act protects an author’s exclusive right to reproduce his work in copies, and reproducing web content is no exception. 17 U.S.C. § 106(1) (2006). A website’s original authorship can be copyrighted. *What Does Copyright Protect?*, Copyright.gov, <http://www.copyright.gov/help/faq/faq-protect.html>.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

B. Link Spamming

“Most search engines look favorably upon pages that have a large number of inbound links on other websites.”³³ Link spamming is a “black hat” SEO technique that takes advantage of this algorithm by going to websites that allow the public to post visible content, “such as the comments sections of blogs and news websites,” and posting links to another website that the owner is trying to promote.³⁴ Similar to content scraping, link-spamming sites generate revenue through pay-per-click advertising.³⁵ To avoid detection and automatic shutdown by Internet service providers, link spammers often make use of “open proxies,” or machines which, by accident or design, allow anyone to access another website through them.³⁶ Spamming violates the policy of use of almost all Internet service providers,³⁷ but has questionable legal implications. Even a “semi-competent programmer” can write programs that will “link-spam vulnerable websites and blogs” with relative ease.³⁸

C. Other Techniques

Keyword stuffing, link farming, and cloaking (including doorway pages) are older, cruder techniques of “black hat” SEO. These techniques are growing increasingly obsolete as search engines become more sophisticated, and are unlikely subjects of legal disputes. Thus, this section will supply only a brief survey.

1. Keyword Stuffing

Keyword stuffing consists of “loading a web page with keywords in an attempt to manipulate a site’s search ranking[s].”³⁹ This was an easy and

³³ Richards, *supra* note 26. A website’s “inbound links” refer to links found elsewhere on the Internet that direct to said website. *What Are Inbound Links? What is their Importance in Terms of SEO?*, SEO MARKETING WORLD, <http://www.seomarketingworld.com/seo-faq/inbound-links.php> (last visited Jan. 8, 2013). Inbound links are different from reciprocal links, which occur when someone links to a website and the website links back to them in return. *Id.* Unlike reciprocal links, inbound links increase your page rank and traffic to a website. *Id.*

³⁴ Richards, *supra* note 26.

³⁵ See Charles Arthur, *Interview with a Link Spammer*, THE REGISTER (Jan. 31, 2005), http://www.theregister.co.uk/2005/01/31/link_spamer_interview/. One successful link spammer boasts that he generates revenues of £100,000 to £200,000 each month. *Id.*

³⁶ *Id.*

³⁷ *Legality of Spam - Your legal rights with respect to internet and email spam law*, SPAM ALERT, <http://www.spamalert.org/articles/legal.php> (last visited Nov. 4, 2012).

³⁸ Arthur, *supra* note 35.

³⁹ Chris Crum, *Google Penguin Update Recovery: Getting Better at Keywords*, WEBPRO NEWS (May 2, 2012), <http://www.webpronews.com/google-penguin-update-recovery->

effective way to improve rankings in the early years of search engines, but led to an undesirable user experience as meaningless keyword repetitions replaced relevant content.⁴⁰ In an April 2012 update of its algorithm, Google specifically named keyword stuffing as a target for eradication,⁴¹ labeling it a strictly “black hat” tactic.⁴² Keyword-stuffed sites, if detected, will suffer demotions in rankings or even removal from Google’s index altogether.⁴³

2. Link Farming

A link farm is “a group of web pages that all link to every other page in the group.”⁴⁴ Link farming was a highly effective SEO technique because pages with inbound links within a “tightly-knit community” would get high scores.⁴⁵ Breaking up these links into categories would further enhance rankings for member sites because they would appear more relevant to search terms.⁴⁶ Search engines can now easily detect link farms based on the distinct patterns automated link farms generate.⁴⁷ Google now specifically prohibits link farms and warns that link farming can result in de-indexing of websites or failure to pass TrustRank.⁴⁸ Search engines have also countered link farming by

getting-better-ad-keywords-2012-05. Keyword stuffing is the oldest SEO technique and became the first target for eradication. *Some Common Black Hat SEO Tricks and Their Bad Effects*, SOCIALTECHNOW (May 15, 2011), <http://www.socialtechnow.com/black-hat-seo-bad-effects/> [hereinafter *SocialTechNow*].

⁴⁰ Megan Marrs, *The Dangers of SEO Keyword Stuffing*, WORDSTREAM (Mar. 21, 2012), <http://www.wordstream.com/blog/ws/2012/03/21/dangers-of-keyword-stuffing>.

⁴¹ Crum, *supra* note 39; *see also* Rick Burgess, *Google Tweaks Algorithm to Thwart Keyword Stuffing and Link Schemes*, TECHSPOT (Apr. 25, 2012, 1:00 PM), <http://www.techspot.com/news/48338-keyword-stuffing-and-link-schemes.html> (discussing methods by which Google’s April update targeted “black hat” SEO).

⁴² Marrs, *supra* note 40.

⁴³ *Id.*

⁴⁴ Terri Wells, *Link Farming: No Good Harvest*, SEO CHAT, <http://www.seochat.com/c/a/link-trading-help/link-farming-no-good-harvest/> (last visited Dec. 28, 2012).

⁴⁵ Saptarshi Ghosh et al., *Understanding and Combating Link Farming in the Twitter Social Network*, in WWW 2012 – SESSION: SECURITY AND FRAUD IN SOCIAL NETWORKS 61, 62 (2012). Starting a link farm is an easy process, only requiring the registration of a site, uploading a design template, and adding links to it. *SocialTechNow*, *supra* note 39.

⁴⁶ Burgess, *supra* note 41.

⁴⁷ Wells, *supra* note 44.

⁴⁸ *SocialTechNow*, *supra* note 39. TrustRank is a classic anti-spamming concept that involves having an expert manually evaluate “a small set of seed pages,” identifying the reputable ones from spam, and using the “link structure of the web to discover other pages that are likely to be good.” Zoltán Gyöngyi et al., *Combating Web Spam with TrustRank*, in 30TH INT’L CONFERENCE ON VERY LARGE DATA BASES 576, 576 (2004), available at

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

tweaking their algorithms to take into account “the *content* of the web pages along with the link structure,”⁴⁹ undermining leverage of rankings by mere numbers of inbound links.

“Like farming,” which uses social media to leverage page rankings in a similar way link farms leverage links, has recently been replacing link farming, its evolutionary predecessor.⁵⁰ Search engines have not addressed “like farming,” perhaps because link structure on social networks is unique: the “links are not between web pages, but between users.”⁵¹ However, this new breed of link farming on social networks seems less likely to pollute the web than traditional link farming because top social network link farmers are active contributors rather than spammers or “black hat” SEO professionals.⁵²

3. Cloaking and Doorway Pages

Cloaking and doorway pages are more examples of “black hat” SEO techniques search engines have expunged from the Internet through sophisticated algorithms. Cloaking involves “presenting . . . different page[s] to human visitors and] to search engine spiders or bots for the same URL.”⁵³ A cloaked website distinguishes between a human visitor and a search engine spider through two popular methods: (1) analyzing the “user agent” text string a browser sends when the user agent requests access to the website, or (2) checking the browser’s Internet Protocol (IP) address against “a database of known IP addresses of specific search engine spiders.”⁵⁴ This ability to present different versions of a page to human and non-human visitors, combined with keyword stuffing or link farming mentioned above, can “trick” search engines into indexing highly optimized pages while human visitors never see them.⁵⁵

<http://ilpubs.stanford.edu:8090/770/1/2004-52.pdf>.

⁴⁹ Ghosh et al., *supra* note 45, at 62.

⁵⁰ Burgess, *supra* note 41.

⁵¹ Ghosh et al., *supra* note 45, at 62. For example, spammers can attempt to enhance the ranking of their tweets by acquiring links in Twitter. *Id.* at 61. “The more followers a user has, the more likely his tweets [will] rank[] highly.” *Id.* Social media link farmers believed that they, often using fake accounts created for link farming, could target (i.e. “follow”) “lay Twitter users with very few followers, who [would] then reciprocate out of social etiquette.” *Id.* at 62.

⁵² *Id.* at 66.

⁵³ *Cloaking – Black Hat?, SEOMIZATION*, <http://seomization.blogspot.com/2007/10/cloaking-black-hat.html> (last visited Apr. 7, 2013) [hereinafter *Seomization*]; see also Joerg Weishaupt, *Examining Black Hat SEO – Cloaking (Part 1)*, JOERGWEISHAUPT.COM, <http://www.joergweishaupt.com/seo/black-hat/examining-black-hat-seo-cloaking-part-1.html> (last visited Apr. 7, 2013).

⁵⁴ *Seomization*, *supra* note 53.

⁵⁵ Weishaupt, *supra* note 53.

Doorway pages achieve similar goals by sending visitors to websites they did not click.⁵⁶ Doorway pages contain seemingly relevant content, but immediately redirect arriving visitors to a completely different page that often employs keyword stuffing or link farming.⁵⁷ This ability allows website owners to “draw people in” for a particular subject then send the visitors to a page they never would have clicked if they knew what the page was truly about.⁵⁸

Today, search engines not only detect cloaking by sending previously unused spiders, but have also become very diligent in algorithmically punishing sites that contain doorway pages.⁵⁹

III. RECENT DEVELOPMENTS AGAINST COPYRIGHT INFRINGEMENT AND SPAM BY “BLACK HAT” SEARCH ENGINE OPTIMIZATION

The 1998 Digital Millennium Copyright Act has had a greater impact on the Internet than any other piece of legislation and is largely responsible for what users see on the Internet today.⁶⁰ In addition to exempting conduits, caching services, web hosts and information location tools (including search engines) in a “safe harbor” from copyright infringement liability, the DMCA established a notice-and-takedown system that allows copyright owners to request removal of infringing works from the Internet.⁶¹

A takedown notice under the DMCA must contain: (1) a physical or electronic signature of the copyright owner, (2) identification of the copyrighted work or a representative list of multiple copyrighted works, (3) identifying information on the infringing material, (4) sufficient contact information for the complaining party, (5) a statement of good-faith belief of infringement, and (6) a statement that the information in the notice is

⁵⁶ *SocialTechNow*, *supra* note 39. Though many SEO firms use them routinely, doorway pages in fact are not very effective in generating traffic. *What are Doorway Pages, and Should I Use Doorway Pages?* SEO LOGIC, <http://www.seologic.com/faq/doorway-pages> (last visited Apr. 7, 2013) [hereinafter *SEO Logic*].

⁵⁷ *SEO Logic*, *supra* note 56. Similar to cloaking, doorway pages also use the “user agent” comparison or IP address lookup methods to trick search engines. *What Are Doorway Pages?* SEARCH ENGINE WATCH (Mar. 1, 2007), <http://searchenginewatch.com/article/2048653/What-Are-Doorway-Pages>.

⁵⁸ *SocialTechNow*, *supra* note 39.

⁵⁹ *Seomization*, *supra* note 53; *see also SocialTechNow*, *supra* note 39.

⁶⁰ *DMCA Takedown 101*, BRAINZ, <http://brainz.org/dmca-takedown-101/> (last visited Nov. 4, 2012). The Computer Fraud and Abuse Act of 1986 (CFAA), the principal “catch-all” legislation for computer-related fraud, does not seem to create an applicable private cause of action for copyright infringement. *See* 18 U.S.C. § 1030(g) (2012).

⁶¹ *DMCA Takedown 101*, *supra* note 60.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

accurate.⁶²

In August 2012, Google announced that it would “start taking the number of valid copyright removal notices it receives for a site into account when ranking” websites.⁶³ The obvious purpose of the change was to punish copyright infringers and protect the rights of content creators such as the MPAA and RIAA, both of which applauded the move.⁶⁴ Copyright infringers in general, including those engaging in “black hat” content scraping, have a cause for concern under this regime.⁶⁵

Google continues to punish unethical SEO techniques with the recent update to its Penguin algorithm, dedicated to detect “black hat” web spam.⁶⁶ Matt Cutts, the current head of Google’s Webspam team, stated that the May 2013 update is more comprehensive in finding web spam and more directly targets link spam in particular.⁶⁷ Though Google has stated that it does not “retroactively apply spam penalties to sites for links that *become* ‘black hat’ after a guideline change,”⁶⁸ the new Penguin algorithm is likely to deter link spamming and other “black hat” practices.

IV. SO IS IT LEGAL? POTENTIAL LEGAL RAMIFICATIONS OF “BLACK HAT” SEARCH ENGINE OPTIMIZATION

Of the “black hat” SEO techniques mentioned above, content scraping and link spamming may be illegal under three legal frameworks: copyright infringement under the DMCA, trespass to chattel, and the CAN-SPAM Act of 2003.

⁶² 17 U.S.C. § 512(c)(3)(A) (2012).

⁶³ Chris Crum, *New Google Algorithm Immediately Raises Concerns*, WEBPRO NEWS (Aug. 11, 2012), <http://www.webpronews.com/new-google-algorithm-change-immediately-raises-concerns-2012-08>.

⁶⁴ Katz, *supra* note 14; Farivar, *supra* note 10. The MPAA and RIAA have praised the new move by Google as a “potentially significant announcement.” *Id.*

⁶⁵ This is especially so in light of the previous updates to Google’s algorithm. The “Penguin” update in April 2012 explicitly targeted sites engaging in “black hat” SEO tactics and webspam, adding a host of anti-spam measures and declaring that the search engine will “reward those who utilize their SEO powers for good.” Burgess, *supra* note 41.

⁶⁶ Cutts, *supra* note 12.

⁶⁷ *Id.* Google expects “Penguin 4” to noticeably affect 2.3 percent of English queries. Barry Schwartz, *Penguin 4, with Penguin 2.0 Generation Spam-Fighting, is Now Live*, SEARCH ENGINE LAND (May 22, 2013), <http://searchengineland.com/penguin-4-with-penguin-2-0-generation-spam-fighting-is-now-live-160544>.

⁶⁸ Chris Liversidge, *Big Brand SEO & Penguin 2.0*, SEARCH ENGINE LAND (June 3, 2013), <http://searchengineland.com/big-brand-seo-penguin-2-0-159410>.

A. *Copyright Infringement under the DMCA*

1. Current State of the Law

The Copyright Clause of the U.S. Constitution states that “Congress shall have Power . . . [t]o promote the Progress of Science and the useful Arts . . . , by securing for limited Times to Authors . . . the exclusive Right to their . . . Writings.”⁶⁹ The Copyright Act allows authors to “assert a right to exclusive use of material published” on their web pages and prevent unauthorized copying.⁷⁰ “To establish a prima facie case for copyright infringement, a plaintiff must show (1) valid ownership of a copyright, and . . . (2) that the alleged infringers violate at least one exclusive right granted to copyright holders under 17 U.S.C. § 106.”⁷¹ The DMCA allows copyright holders and their agents to demand removal of allegedly infringing content.⁷² In the context of content scraping and archiving, courts have also considered whether the infringer acted in good faith,⁷³ whether the infringement intended to divert users looking for the original content,⁷⁴ and whether the infringement has potential to cause customer confusion.⁷⁵

2. Remedies

Remedies for copyright infringement include: “(1) injunctions; (2) impounding or destruction of copies;” (3) damages for the profits lost to the copyright owner and attributable to the infringement; (4) “statutory damages of a minimum of [\$750] to a maximum of [\$30,000];” and (5) “full costs and reasonable attorney’s fees.”⁷⁶ Furthermore, a person who engages in “willful” copyright infringement for “commercial advantage or private financial gain may be subject” not only to criminal penalties, such as fines or

⁶⁹ U.S. Const. art. 1, § 8, cl. 8.

⁷⁰ Kinari Patel, “*Authors v Internet Archives*”: *The Copyright Infringement Battle Over WEB Pages*, 89 J. PAT. & TRADEMARK OFF. SOC’Y 410, 411 (2007).

⁷¹ *McNamara v. Universal Commercial Servs.*, No. 07-6079-TC, 2008 WL 4367831, at *2 (D. Or. Sept. 16, 2008); *see also* Patel, *supra* note 70, at 415.

⁷² *See* 17 U.S.C. § 512(c)(3) (2012).

⁷³ Patel, *supra* note 69, at 419. “Cache” archiving by search engine services are likely defensible under the fair use doctrine because it is in good faith. *Id.*

⁷⁴ Goldman & German, *supra* note 8, at 35.

⁷⁵ *McNamara*, 2008 WL 4367831, at *3. Although not in direct competition, because both parties were using the copyrighted article for advertising purposes, there was “potential for customer confusion.” *Id.*; *see also* *Hard Rock Café Int’l Inc. v. Morton*, No. 97 Civ. 9483 RPP, 1999 WL 701388, at *4, *5 (S.D.N.Y. Sept. 9, 1999).

⁷⁶ Patel, *supra* note 69, at 415; *McNamara*, 2008 WL 4367831, at *4 (quoting 17 U.S.C. § 504(c)(1)-(2)).

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

imprisonment,⁷⁷ but also to increased damages up to \$150,000.⁷⁸ An infringement is willful if the infringer had actual awareness or should have known that her actions constituted infringement, or “acted with reckless disregard for, or willful blindness to, the rights of the copyright” holder.⁷⁹

3. Defenses

Several defenses are available to copyright infringers under the doctrines of fair use, innocent infringement, and implied license. A copyright infringer may still collect and present copyrighted content as long as such activity constitutes fair use.⁸⁰ Courts consider four factors to determine whether an activity is fair use:

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.⁸¹

The fair use doctrine has indemnified search engines in copyright claims over potential infringement in “caching” services.⁸² Search engines such as Google periodically “cache” websites in order to provide the “pages to users more quickly and . . . in case the originals become unavailable.”⁸³ When the use of copied material is a commercial one, however, the infringer is liable under a presumption of likely future harm to the potential market.⁸⁴

A copyright infringer may also escape liability if the infringement was “innocent.”⁸⁵ An infringement is innocent when the infringer “proves by a

⁷⁷ Patel, *supra* note 69, at 415.

⁷⁸ *McNamara*, 2008 WL 4367831, at *4 (quoting 17 U.S.C. § 504(c)(1)-(2)).

⁷⁹ *Id.* (quoting *Island Software & Computer Serv., Inc. v. Microsoft Corp.*, 413 F.3d 257, 263 (2d Cir. 2005)).

⁸⁰ 17 U.S.C. § 107 (2012). For example, use of copyrighted content for “criticism, comment, news reporting, teaching (including multiple copies for classroom use), [and] scholarship or research” constitutes fair use. *Id.*

⁸¹ *Id.*

⁸² See *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006) (finding “caching” constitutes fair use and granting summary judgment).

⁸³ Patel, *supra* note 69, at 424.

⁸⁴ *McNamara*, 2008 WL 4367831, at *3 (quoting *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (1994)). This is true unless the infringed work is “transformative of the original.” *Id.*

⁸⁵ *Id.*; see also 17 U.S.C. § 504(c)(2) (2012).

preponderance of evidence that she was unaware that her actions constituted infringement, and she had no reason to believe that her acts constituted infringement.”⁸⁶

Lastly, the implied license defense is available to the copyright infringer.⁸⁷ “[A] copyright owner may grant a nonexclusive license expressly or impliedly through her conduct.”⁸⁸ A court may find an implied license “where the copyright holder engages in conduct from which the other party may properly infer that the owner consents to his use.”⁸⁹ “Silence or lack of objection may . . . be the equivalent of a nonexclusive license.”⁹⁰ In the search engine caching cases, courts have considered the availability of “opt-out” protocols to presume implied license.⁹¹

4. Application to Content Scraping

A copyright owner who suspects another site is scraping her original content can easily establish a prima facie case of infringement. The original content is “fixed” on the Internet, a “tangible medium of expression,” giving rise to valid copyright protection,⁹² and verbatim copying clearly violates exclusive rights to reproduce and distribute such copies to the public.⁹³ Those who engage in content scraping for “black hat” SEO purposes do not act in good faith because their primary purpose is to confuse, trick, and divert users from finding the original content in order to generate online advertising revenue.⁹⁴

Some copyright infringers may attempt to “cloak their [misappropriation] in the language of fair use, . . . presenting their sites as commentaries” on the original content.⁹⁵ This provides an infringer with room to purport that the infringement was for the “purpose of engaging in legitimate criticism or comparative advertising,” especially because search engines cannot discern the

⁸⁶ *McNamara*, 2008 WL 4367831, at *3 (quoting 17 U.S.C. § 504(c)(2)).

⁸⁷ *Parker v. Yahoo!, Inc.*, No. 07-2757, 2008 WL 4410095, at *3 (E.D. Pa. Sept. 25, 2008) (citing *MacLean Assocs., Inc. v. Wm. M. Mercer-Meidinger-Hansen, Inc.*, 952 F.2d 769, 778-79 (3d Cir. 1991)).

⁸⁸ *Id.*

⁸⁹ *Id.* (quoting *Field*, 412 F. Supp. 2d at 1116).

⁹⁰ *Id.* (citing *Field*, 412 F. Supp. 2d at 1116).

⁹¹ *Parker*, 2008 WL 4410095, at *3-4 (allowing search engine’s “caching” of literary work because there was an implied license). “Opt-out mega-tag [is] a ‘well-known industry standard’” that allows a site owner to block search engines from indexing a site in search results or providing access to the site through cached links. *Id.* at *3. An implied license exist if the site owner does not exercise this “opt-out” option. *Id.*

⁹² See 17 U.S.C. § 101 (2012).

⁹³ 17 U.S.C. § 106(1, 3) (2012).

⁹⁴ See *Richards*, *supra* note 26.

⁹⁵ *Goldman & German*, *supra* note 8, at 35-36.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

difference between copyright infringement and commentary as protected by the First Amendment.⁹⁶ Despite this caveat, the commercial purpose behind “black hat” SEO is likely to defeat any fair use defense.⁹⁷ For similar reasons, content scraping for SEO purposes is unlikely “innocent” infringement.

No implied license exists for content scraping because there are no effective “opt-out” measures. Content scraping as a “black hat” SEO technique is much more difficult to prevent than caching by search engines because it does not honor “no-archive” HTML meta-tags or “robots.txt” instructions.⁹⁸ Various services can help a website owner monitor content scraping, including: Copyscape, which finds reproductions of copyrighted content on other websites; Google Alerts, which alerts the content owner each time a duplicate title shows up on Google; and Google Webmaster Tools, which allow monitoring of all inbound links to the original website.⁹⁹ These services are only useful in monitoring content scraping, however, and do not provide any viable “opt-out” measures to prevent infringers from copying content. Similarly, cautionary measures such as CAPTCHA validation (security measure against machine-generated posts), pinging (notifying search engines of each legitimate update), and inclusion of canonical links (labeling content with identifying information) are passive techniques that do not actively deny access to content scrapers.¹⁰⁰ Thus, content scraping as a “black hat” SEO technique is *prima facie* copyright infringement and is actionable under the DMCA.

B. Trespass to Chattel

1. Current State of the Law

The Restatement (Second) of Torts imposes tort liability on a trespasser to

⁹⁶ *Id.*

⁹⁷ See *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 562 (1985) (rejecting fair use defense because the unauthorized reproduction of memoir quotes had commercial motives); see also Goldman & German, *supra* note 8, at 36. In *Live Nation Motor Sports*, the plaintiff claimed that the link impeded the plaintiff’s ability to “sell sponsorships [and] advertisements on its own website as the ‘exclusive source’ of the [appropriated] webcasts.” *Live Nation Motor Sports, Inc. v. Davis*, No. 3:06-CV-276-L, 2006 WL 3616983, at *1 (N.D. Tex. Dec. 12, 2006); see also Goldman & German, *supra* note 8, at 36.

⁹⁸ *Field v. Google, Inc.*, 412 F. Supp. 2d 1106, 1113 (D. Nev. 2006). HTML meta-tags and “robots.txt” instructions are industry-standard protocols that prohibit search engine spiders from crawling one’s website. *Id.*

⁹⁹ Amanda DiSilvestro, *How to Prevent Content Scraping Before It Starts*, FAMOUS BLOGGERS (May 15, 2012), <http://www.famousbloggers.net/prevent-content-scraping.html>.

¹⁰⁰ *Id.*

chattel if and only if:

(a) he dispossesses the [possessor] of the chattel; or (b) the chattel is impaired as to its condition, quality, or value; or (c) the possessor is deprived of the use of the chattel for a substantial time; or (d) bodily harm is caused to the possessor, or harm is caused to some person or thing in which the possessor has a legally protected interest.¹⁰¹

For Internet chattels, different states have required varying degrees of interference to constitute trespass. In *eBay. v. Bidder's Edge*, the court extended the traditional notion of trespass to chattel to web pages, stating that “[c]onduct that does not amount to a substantial interference with possession, but which consists of intermeddling with or use of another’s personal property, is sufficient to establish a cause of action for trespass to chattel.”¹⁰²

Some courts have required more than “intermeddling” to find trespass to chattel by focusing on consumption of web server capacity and denial of access to other potential visitors.¹⁰³ In *Register.com v. Verio*, the Second Circuit applied New York law to find that the defendant likely committed a trespass because “an unauthorized consumption of computer systems’ capacity depletes the capacity available at a given time for authorized end-users, which may diminish the condition, quality, or value of the systems.”¹⁰⁴ In *Snap-On Business Solutions. v. O’Neil & Associates*, the court held that even intangible damage, such as temporary crashes and slowing down of computer servers, may constitute a trespass to chattels.¹⁰⁵ Some states do not recognize trespass to intangible property.¹⁰⁶

¹⁰¹ RESTATEMENT (SECOND) OF TORTS § 218 (1965).

¹⁰² *eBay, Inc. v. Bidder’s Edge, Inc.*, 100 F. Supp. 2d 1058, 1070 (N.D. Cal. 2000). However, the court also noted that there is “some uncertainty as to the precise level of possessory interference required to constitute an intermeddling.” *Id.*

¹⁰³ *Snap-On Bus. Solutions, Inc. v. O’Neil & Associates, Inc.*, 708 F. Supp. 2d 669, 678 (N.D. Ohio 2010). “The interference has to amount to more than intermeddling.” *Id.* (quoting *Dryden v. Cincinnati Bell Tel. Co.*, 734 N.E.2d 409, 416 (Ohio Ct. App. 1999)).

¹⁰⁴ *Register.com, Inc. v. Verio, Inc.*, 356 F.3d 393, 438 (2d Cir. 2004) (quoting *eBay v. Bidder’s Edge*, 100 F. Supp. 2d 1058, 1071) (internal quotations omitted). The court applied the Restatement definition of trespass to chattel: “when a party intentionally damages or interferes with the use of property belonging to another, [where] interference may be accomplished by dispossessing another of the chattel” (which does not require a showing of actual damages), “or using or intermeddling with a chattel in possession of another” (which requires a showing of actual damages). *Id.* at 437 (quoting RESTATEMENT (SECOND) OF TORTS §§ 217-21 (1965)) (internal quotations omitted).

¹⁰⁵ *Snap-On*, 708 F. Supp. 2d at 679.

¹⁰⁶ *See, e.g., Inventory Locator Serv., LLC v. Partsbase, Inc.*, No. 02-2695 MA/V, 2005 WL 2179185, at *11-12 (W.D. Tenn. Sept. 6, 2005) (focusing on the physical nature of a

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

Other courts have required a showing of actual damages or diminished value to establish trespass to chattel. In *CompuServe v. Cyber Promotions*, the defendant directed spam e-mails to the plaintiff’s subscribers, inducing many of them to terminate their subscriptions.¹⁰⁷ The damage the plaintiff sustained as a result was sufficient to establish the defendant’s behavior as trespass to plaintiff’s chattel.¹⁰⁸ Under California law, “an electronic communication that neither damages the recipient’s computer system nor impairs its functioning” does not constitute trespass to chattels.¹⁰⁹ In *Intel Corp. v. Hamidi*, the court required a showing of intermeddling harmful to a materially valuable interest rather than mere interference that does not amount to dispossession.¹¹⁰ Furthermore, the court asserted that intermeddling is actionable only if it impairs the “condition, quality, or value” of a chattel or deprives the possessor’s use of the chattel for a time, amounting to interference “so substantial that it is possible to estimate the loss caused thereby.”¹¹¹ Several other courts have followed this approach in the determination of trespass to chattel.¹¹²

2. Applications to Content Scraping and Link Spamming

Depending on the jurisdiction, a copyright owner may successfully raise a trespass to chattel claim against content scraping as a “black hat” SEO technique. In *eBay*, the defendant’s website generated its auction database by scraping content from various popular sites, including eBay.com.¹¹³ The court found a trespass based on the defendant’s intermeddling with the plaintiff’s personal property.¹¹⁴ In *Snap-On*, the court refused to grant summary judgment to the defendant who ran scraping programs to recover data licensed to the

web server); *see also* Univ. Tube & Rollform Equip. Corp. v. YouTube, Inc., 504 F. Supp. 2d 260, 269 (N.D. Ohio 2007) (refusing to find trespass to chattels because defendant “did not make physical contact with the computers hosting the [plaintiff’s] website”).

¹⁰⁷ *CompuServe Inc. v. Cyber Promotions, Inc.*, 962 F. Supp. 1015, 1023 (S.D. Ohio 1997).

¹⁰⁸ *Id.*

¹⁰⁹ IAN C. BALLON, 1 E-COMMERCE AND INTERNET LAW: TREATISE WITH FORMS § 5.05(1) (2d ed. 2001), *available at* Westlaw ECOMMINTLAW.

¹¹⁰ *Intel Corp. v. Hamidi*, 30 Cal. 4th 1342, 1351 (2003).

¹¹¹ *Id.* at 1357.

¹¹² *See, e.g.*, *Jedson Eng’g, Inc. v. Spirit Constr. Servs, Inc.*, 720 F. Supp. 2d 904, 926 (S.D. Ohio 2010) (finding trespass because plaintiff “suffered harm in terms of the diminished value of servers as a safe, secure location for project files”); *A.V. v. iParadigms, LLC*, 562 F.3d 630, 647 (4th Cir. 2009) (requiring a showing of a causal link between electronic transmission and actual impairment).

¹¹³ *eBay, Inc. v. Bidder’s Edge, Inc.*, 100 F. Supp. 2d 1058, 1062 (N.D. Cal. 2000).

¹¹⁴ *Id.* at 1070.

plaintiff, causing unusually large spikes in website traffic and inducing several server crashes.¹¹⁵ However, some jurisdictions may require a showing of physical intrusion to servers or a causal link between the content scraping and a loss of materially valuable interest.¹¹⁶ Such loss of materially valuable interest usually stems from a denial of service to other potential visitors and can be especially difficult to prove in the case of larger-scale websites, which often prepare their servers for unusual spikes in traffic in order to prevent crashes.¹¹⁷

Link spamming is less likely to cause server problems because posting a link in a public forum does not require much bandwidth, unless frequent and automated spamming rises to the level of threatening server crashes.¹¹⁸ However, if a website owner can prove that extensive link spamming has deterred patrons from visiting the site or encouraged them to terminate their accounts, the owner has a potential claim under the *CompuServe* analysis of trespass to chattel.¹¹⁹

C. The CAN-SPAM Act of 2003

1. Current State of the Law

On December 16, 2003, President George W. Bush signed into law the Controlling the Assault of Non-Solicited Pornography and Marketing Act (CAN-SPAM Act).¹²⁰ The Act in effect has made spamming legally permissible in the United States, as long as it follows certain criteria: “a truthful subject line; no false information in the . . . headers or sender ‘information’; ‘conspicuous’ display of [sender information], and other minor requirements.”¹²¹ A key feature of the CAN-SPAM Act is its requirement of “opt-out” information in all messages that spammers send.¹²² If a spammer

¹¹⁵ *Snap-On*, 708 F. Supp. 2d at 675, 679.

¹¹⁶ *See, e.g.*, *Intel Corp. v. Hamidi*, 30 Cal. 4th 1342, 1351 (2003); *see also* BALLON, *supra* note 109.

¹¹⁷ *See generally* BALLON, *supra* note 109. For examples of how large-scale websites prepare for such unusual spikes in traffic, see Elizabeth Millard, *The Enterprise Guide to Web Traffic Spikes*, E-COMMERCE TIMES (Sept. 24, 2003), <http://www.ecommercetimes.com/story/31644.html>.

¹¹⁸ For estimates of how much bandwidth a public blog requires to handle organic visitors, see Jim Makos, *How Much Bandwidth Does Your Blog Use?*, JIMMAKOS.COM (Feb. 23, 2012), <http://jimmakos.com/2012/02/how-much-bandwidth-does-your-blog-use/>.

¹¹⁹ *CompuServe*, 962 F. Supp. at 1023.

¹²⁰ Amit Asaravala, *With this Law, You Can Spam*, WIRED (Jan. 23, 2004), <http://www.wired.com/techbiz/media/news/2004/01/62020>.

¹²¹ *Legality of Spam*, *supra* note 37.

¹²² Asaravala, *supra* note 120.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

harvests user information without compliance to these provisions, aggravated or accelerated penalties may apply.¹²³ Governmental entities, including the Federal Trade Commission and states, are the primary enforcers of the Act.¹²⁴ The Act also created a cause of action for Internet service providers.¹²⁵

2. Application to Link Spamming

Despite the legislative intent behind the CAN-SPAM Act, the legislation has been ineffective or counterproductive in curbing spam on the Internet.¹²⁶ “Opting out” is unwise for the user because it merely gives the spammer a confirmation that a message reached an active e-mail account,¹²⁷ and any recipient who is aware of potential risks of opting out is unlikely to exercise the option.¹²⁸ Although “proponents of the CAN-SPAM Act argue[d] that a federal law will . . . bolster state efforts to find [the] spammers,” the Act in effect took away the rights for citizens to sue spammers directly or pursue a class action suit.¹²⁹

The CAN-SPAM Act is hardly applicable to link spamming because the legislation’s primary target is e-mail spamming. Aware of the loopholes, link spammers differentiate themselves from e-mail spammers.¹³⁰ There is no arguable distinction “between a person using a computer to post a comment, and a person using a computer to use a computer to post a comment.”¹³¹ Exploiting open proxy servers and comment functions open to the public are perfectly legal. Furthermore, most blogging platforms and web frameworks, such as WordPress, often offer features that require administrative approval of user comments,¹³² making application of the CAN-SPAM Act to link spamming unlikely.

¹²³ *Legality of Spam*, *supra* note 37; *see also* 15 U.S.C. § 7706(f)(3)(C) (2012).

¹²⁴ *See* 15 U.S.C. § 7706(b)-(f).

¹²⁵ 15 U.S.C. § 7706(g).

¹²⁶ *Legality of Spam*, *supra* note 37. Professor Lessig of Stanford Law School has called the Act “an abomination at the federal level . . . [it is] ineffective and . . . affirmatively harmful because it preempts state legislation.” Asaravala, *supra* note 120.

¹²⁷ *Legality of Spam*, *supra* note 37.

¹²⁸ *Id.*

¹²⁹ Asaravala, *supra* note 120.

¹³⁰ Arthur, *supra* note 35.

¹³¹ *Id.* Experienced spammers can even simulate human response in automated authentication. *Id.*

¹³² *See* [Comment Moderation](http://codex.wordpress.org/Comment_Moderation), WORDPRESS CODEX, http://codex.wordpress.org/Comment_Moderation (last visited Apr. 7, 2013), for default comment moderation features on WordPress, a popular blogging platform. *See How it Works – Find Out How Akismet Catches Spam*, AKISMET.COM, <http://akismet.com/how/#gs1> (last visited Apr. 7, 2013), for an example of other anti-spamming software.

V. WHO CARES? POTENTIAL PARTIES WHO MAY PURSUE
LEGAL ACTION AGAINST “BLACK HAT” SEARCH ENGINE OPTIMIZATION

Though “black hat” SEO techniques pollute the Internet community through deceptive practices and creation of user confusion, not many parties are likely to pursue legal action. The few entities that might have incentives to sue or regulate “black hats” are: website owners, search engines, and government entities, specifically the Federal Trade Commission.

A. Website Owners

Content scraping is not always bad for copyright owners. If a scraper website attributes credit to the original content, the infringing site’s “black hat” SEO efforts can “actually help drive traffic to [the original] website and help [it] gain visibility.”¹³³ Thus, litigation is rare¹³⁴ and website owners may often lack incentives to bring suit against the infringers. “In many cases, . . . finding the contact information” of the infringing website and asking the infringer to take down the content is enough.¹³⁵ If the copyright infringer or link spammer is not in direct competition with the website from which it appropriates content, copyright owners have little potential to recover actual damages.¹³⁶

Furthermore, few website owners “will have the resources or interest to test the issue.”¹³⁷ Though they may view the conduct of “black hats” as irritating, website owners “may be reluctant to litigate what appear to be close calls.”¹³⁸ The difference between a victim’s loss and litigation cost determine the private incentive to bring suit.¹³⁹ A website owner’s loss from the conduct of “black hats” is often minimal.

Nevertheless, as the SEO services industry expands and awareness of SEO’s marketing potential heightens, more website owners are likely to consider legal action against the “black hats.”¹⁴⁰ In *Paul Jones v. BullsEye Event Group*, the

¹³³ DiSilvestro, *supra* note 99.

¹³⁴ Goldman & German, *supra* note 8, at 36. “Reported litigation against blogs is rare; reported litigation against splogs is nonexistent.” *Id.*

¹³⁵ DiSilvestro, *supra* note 99.

¹³⁶ *McNamara*, 2008 WL 4367831, at *4.

¹³⁷ Goldman & German, *supra* note 8, at 34.

¹³⁸ *Id.* at 34, 36. Website owners may be especially wary when pursuing litigation “will be characterized as trying to stifle free speech.” *Id.* at 36.

¹³⁹ Keith N. Hylton, *Litigation Costs and the Economic Theory of Tort Law*, 46 U. MIAMI L. REV. 111, 122 (1991).

¹⁴⁰ See Complaint at 4, *Schoemaker v. Cota*, No. 8:12-cv-00027 (D. Neb. Jan. 16, 2012), for a recent case that, though not directly involving “black hat” techniques, illustrates the newfound economic importance of SEO; in the complaint, the plaintiff sued the defendant for using a photograph documenting the plaintiff’s pay-per-click advertising profits to sell the defendant’s own SEO services.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

plaintiff filed a complaint against a competing website that extensively appropriated and reproduced content from the plaintiff’s website.¹⁴¹ The plaintiff alleged that the defendant infringed the plaintiff’s copyright under 17 U.S.C. § 501 et seq., and that he suffered “substantial damage to [his] business in the form of diversion of trade, loss of profits, injury to goodwill and reputation, and the dilution of the value of [his] rights.”¹⁴² The plaintiff asked for maximum statutory damages of \$150,000 by alleging willful infringement.¹⁴³

In *Nuesoft Technologies, Inc. v. John Does 1-5*, an extreme case of content scraping, the plaintiff brought a copyright infringement suit against unnamed defendants engaging in phishing scams by operating “mirror sites” of the plaintiff’s popular, original website.¹⁴⁴ Arguing that the defendants knowingly and intentionally copied website content without permission, the plaintiff asked for statutory damages under 17 U.S.C. § 504(c).¹⁴⁵ Similarly, in *Liberty Media Holdings v. Gan*, the plaintiff brought a copyright and trademark infringement suit against the defendant alleging that the defendant reproduced and profited from the plaintiff’s original web content by “site ripping,” another extreme type of content scraping that copies entire websites.¹⁴⁶

B. Search Engines

Prominent search engines offer online advertising platforms such as Google’s AdSense and Yahoo’s Overture. These services allow any website to populate itself with advertisements that are relevant to the website’s content.¹⁴⁷ Getting clicks on such ads generates revenue not only for the website owners, but also for the search engines, who take a cut of all payments made on the platform.¹⁴⁸ Websites engaging in “black hat” SEO often serve the purpose of

¹⁴¹ Complaint at 3, *Paul Jones v. BullsEye Event Grp. LLC*, No. 2:12-cv-06929 (C.D. Cal. Aug. 10, 2012). The plaintiff operates a website selling tickets for sporting events such as the Super Bowl, and the site ranks very highly in search engines. The defendant’s website, by copying the plaintiff’s content on over 45 separate pages and purporting to be “associated with, sponsored by, and/or affiliated” with the plaintiff, exploited the SEO efforts the plaintiff had paid for. *Id.* at 3, 8.

¹⁴² *Id.* at 9-10.

¹⁴³ *Id.* at 10.

¹⁴⁴ Amended Complaint at 1-3, *Nuesoft Techs., Inc. v. John Does 1-5*, No. 1:08-cv-03912 (N.D. Ga. Dec. 31, 2008).

¹⁴⁵ *Id.* at 7-8.

¹⁴⁶ Complaint at 1, *Liberty Media Holdings, LLC v. Gan*, No. 1:11-cv-02754 (D. Colo. Oct. 21, 2011). Plaintiff demanded statutory damages, actual damages, legal fees and injunctive relief. *Id.* at 2.

¹⁴⁷ Goldman & German, *supra* note 8, at 34.

¹⁴⁸ *Id.*

generating revenue by exploiting this system at the expense of content creators who have achieved higher search rankings through legitimate “white hat” techniques.¹⁴⁹ Seeking legal assistance from the “leading search engines may offer a remedy,” but the search engines are likely to have “conflicting motives” because they also benefit from the “placement of contextual ads.”¹⁵⁰ Search engines have appeased aggrieved users by punishing “black hats” through mere demotion of search rankings,¹⁵¹ and are not likely to take part in any legal initiatives.

C. The Federal Trade Commission (FTC)

Regulation of “black hat” SEO techniques by the FTC is another solution businesses have suggested in order to curb unethical practices on the Internet.¹⁵² However, any attempt to directly regulate the Internet is likely to trigger public outrage.¹⁵³ The Search Engine Marketing Professional Organization (SEMPO), a non-profit organization consisting of “thousands of marketing professionals across fifty countries, . . . calls for a ‘free market’ approach to the Internet with little or no regulation.”¹⁵⁴ In a letter to FTC Chairman Jon Leibowitz, SEMPO emphasized that “[s]earch is not a government-run utility, established by law and thus subject to bureaucratic oversight.”¹⁵⁵ SEMPO further asserted that, because consumers and businesses

¹⁴⁹ See *supra* Part II for detailed discussion on how “black hat” SEO techniques take advantage of pay-per-click advertising platforms.

¹⁵⁰ Goldman & German, *supra* note 8, at 36.

¹⁵¹ See Arthur, *supra* note 35.

¹⁵² Abby Johnson, *SEMPO: FTC Shouldn’t Regulate Google and Other Search Engines*, WEBPRO NEWS (Nov. 4, 2012, 9:21 PM), <http://www.webpronews.com/sempo-ftc-shouldnt-regulate-google-and-other-search-engines-2011-12>.

¹⁵³ *Id.* One need not look much further than to the 2012 controversy over the Stop Online Piracy Act (SOPA) and the Protect Intellectual Property Act (PIPA) for the likely public response to government regulation of the Internet. “In December [2011], a group of influential technology figures, including founders of Twitter, Google, and YouTube, published an open letter to lawmakers stating that the legislation would enable Internet regulation and censorship on par with the government regulation in China and Iran.” Edward Wyatt, *White House Says It Opposes Parts of Two Antipiracy Bills*, N.Y. TIMES, Jan. 14, 2012, <http://www.nytimes.com/2012/01/15/us/white-house-says-it-opposes-parts-of-2-antipiracy-bills.html>. Opponents of the legislation “mounted a web-wide protest that include[d] a 24-hour shutdown of the English-language Wikipedia.” Jenna Wortham, *A Political Coming of Age for the Tech Industry*, N.Y. TIMES, Jan. 17, 2012, <http://www.nytimes.com/2012/01/18/technology/web-wide-protest-over-two-antipiracy-bills.html>.

¹⁵⁴ Johnson, *supra* note 155.

¹⁵⁵ *Id.* (quoting Letter from Board of Directors, SEMPO, to Jon Leibowitz, Chairman, Federal Trade Commission (Nov. 15, 2011), *available at*

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

are the driving force behind search, each company should be “free to develop its own approach” in technologies and algorithms to fulfill the needs of its customers.¹⁵⁶

Opponents of government regulation of the Internet point out that “all businesses are subject to the same rules for both organic and paid search[es].”¹⁵⁷ They contend that “consumers feeling a [sufficient] level of trust with search engines”¹⁵⁸ is too important to sacrifice by government regulation.¹⁵⁹ Though the FTC may have more incentives and resources than website owners or search engines to combat “black hat” SEO, opponents express skepticism over government competency, advocating that the “market would solve the problem long before government regulators could figure out what was wrong or how to fix it.”¹⁶⁰

VI. GOOGLE UPDATES AND THEIR POTENTIAL FOR ABUSE

One may ask: is it not enough that Google will punish unethical behavior on the Internet, even if the law is inadequate to regulate it? Google’s 2012 decision to demote websites with a high number of DMCA takedown notices is a good way to punish companies that rely on infringement for their business, but it may also “punish innocent businesses who receive unwarranted notices.”¹⁶¹ As Google carefully maintains opacity over inner workings of its search algorithms, this potential for abuse is a source of concern for Internet users.¹⁶²

The notion that Google has amassed too much unrestrained power on the

<http://www.scribd.com/doc/72797167/Sempo-Search-Regulation-Arti>).

¹⁵⁶ *Id.*

¹⁵⁷ *Id.*

¹⁵⁸ *Id.*

¹⁵⁹ *See id.*

¹⁶⁰ Don Racheter & Daniel Oliver, *Government Shouldn’t Regulate Google Searches*, THE GAZETTE, Oct. 21 2012, <http://thegazette.com/2012/10/21/government-shouldnt-regulate-google-searches/>. Racheter and Oliver criticize the notion that “the FTC should get into the practice of evaluating the millions of services and products offered on the Internet by Google and all its competitors” as simply “preposterous.” *Id.* The authors also posit that “[e]verything we know about government regulation tells us that the quality of searches would get worse, not better, if government regulated them.” *Id.*

¹⁶¹ Katz, *supra* note 14. Overzealous enforcement of the DMCA is especially apparent on sites like YouTube. *See* Eric Limer, *Piracy Is Not Theft but Unwarranted Takedowns Are*, GEEKOSYSTEM (Apr. 20, 2012), <http://www.geekosystem.com/unwarranted-takedowns-are-theft/>.

¹⁶² *See* Farivar, *supra* note 10. “Google’s demotion of some websites may be abused, simply because they may be accused of copyright violations, rather than evaluated or even convicted.” *Id.*

Internet is not a novel one. Though “Google argues that its behavior is kept in check by [competing search engines] like Yahoo or Bing, . . . Google has become the default search engine for many Internet users.”¹⁶³ The company’s dominance in Internet search has incited criticisms that there is “little pressure from the competition that would force Google to change any of [its] practices.”¹⁶⁴ Google, in particular, has been the subject of several antitrust investigations.¹⁶⁵

The apprehension about the new Google algorithms and their abuse is similar to the widespread concerns about biased ad and search placement. The advertising methods search engines offer are frequently under attack by claims that they are forms of “stealth marketing.”¹⁶⁶ “Although anyone can open an account to buy paid search listings, the rules are arbitrary” in that search engines can terminate these accounts “at any time, without notice to the website owner.”¹⁶⁷ Furthermore, the search engine is free to choose which company “may or may not buy traffic within its index” at any time.¹⁶⁸ Google’s near-monopolizing popularity means that “any company that falls out of favor” with the search engine, whether for engaging in “bad practice or a simple disagreement, can find itself at risk of going out of business.”¹⁶⁹

The more troubling issues regarding Google’s DMCA punishment include what constitutes a “high number” of takedown notices, whether there is any chance of recourse once a website is demoted, and how to confirm the validity of accusations.¹⁷⁰ Although the DMCA has a counter-notice system to dismiss

¹⁶³ *The Google Algorithm*, *supra* note 6.

¹⁶⁴ James Gelfer, *Do Search Engines Need Government Regulation?*, IEPLEXUS (July 19, 2010), <http://www.ieplexus.com/industry-news/5209-do-search-engines-need-government-regulation/>.

¹⁶⁵ *The Google Algorithm*, *supra* note 6. See Sam Gustin, *U.S. Google Antitrust Probe Spurs Internet-Regulation Debate*, TIME, Oct. 15, 2012, <http://business.time.com/2012/10/15/ftc-antitrust-probe-against-google-sets-up-internet-regulation-clash/>, for a further look into Google’s recent involvement in antitrust issues.

¹⁶⁶ Viva R. Moffat, *Regulating Search*, 22 HARV. J. L. & TECH. 475, 488 (2009). In 2002, “the FTC issued a letter responding to a complaint about the inclusion of paid placement ads in search engine results.” *Id.* at 485. The letter recommended “clear and conspicuous disclosure of paid placement” and inclusion on advertising platforms. *Id.*

¹⁶⁷ *TechCrunch*, *supra* note 8.

¹⁶⁸ *Id.*

¹⁶⁹ *Id.* Some hypothesize that the fear of losing Google’s favor causes businesses to incur additional costs in hiring web traffic agencies and brokers, diminishing chances for small businesses to “actually compete and gain search market share.” *Id.*

¹⁷⁰ Farivar, *supra* note 10; *see also* Crum, *supra* note 63 (“No court or other umpire confirms that the accusations are valid, although copyright owners can be liable for bad-faith accusations.”).

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

frivolous accusations,¹⁷¹ website owners “may not know about, or have the ability to easily challenge, notices sent to Google. . . . Sites that host a lot of content, or are very popular, may receive a disproportionate number of notices . . . without being disproportionately infringing.”¹⁷² Thus, user-generated content sites are likely to experience harm from this change, despite the DMCA’s intent to protect them.¹⁷³ Though the Penguin update of May 2013 seems less likely to raise abuse concerns, metrics show that the update unexpectedly victimized several large websites such as the Salvation Army’s.¹⁷⁴

VII. PROPOSAL

As described above, “black hat” SEO techniques pollute the Internet and harm creators of original web content. However, the current state of the law does not provide adequate regulatory protection against these unethical practices nor incentivizes website owners to pursue legal action. Many web professionals, though wary of the government’s potential to abuse power, believe that the solution lies within regulation of search engine companies and their protocols, combined with laws that define legal uses of search technologies.¹⁷⁵

Such centralized regulation of search engines is possible through two solutions: establishment of an administrative agency to oversee search operations, or creation of a common law tribunal combined with an industry panel to resolve search engine disputes. However, the general antipathy towards governmental control over the Internet presents a major obstacle to search engine regulation, and the United States has been the most prominent supporter of an “open” Internet on the global stage.¹⁷⁶ This section first

¹⁷¹ 17 U.S.C. § 512(g)(3) (2012).

¹⁷² Crum, *supra* note 63.

¹⁷³ *Id.*

¹⁷⁴ See Matt McGee, *Penguin 2.0 Losers: Porn Sites, Game Sites, & Big Brands Like Dish.com & the Salvation Army*, SEARCH ENGINE LAND (May 23, 2013, 4:14 PM), <http://searchengineland.com/penguin-2-0-losers-porn-sites-game-sites-big-brands-like-dish-com-the-salvation-army-160744>.

¹⁷⁵ Interview with Le Zhang, Founding Manager, Rapid Labs, in Bos., Mass. (Mar. 12, 2013). The debate among web scholars is fierce and polarized, with one side “offering arguments for agency regulation and others urging a free-market approach.” Moffat, *supra* note 166, at 477.

¹⁷⁶ Joseph Menn, *Internet Regulation Seen at National Level as Treaty Talks Fail*, REUTERS (Dec. 16, 2012), <http://www.reuters.com/article/2012/12/17/net-us-telecoms-treaty-fail-idUSBRE8BG00K20121217>. Many nations, including the United States – which made a powerful opposing statement – “refused to sign a [2012] global telecommunications treaty that opponents feared could lead to greater government control over online content

inquires into the types of problems an ideal solution must address to effectively resolve search engine disputes while minimizing controversy over Internet regulation. An analysis of the two possible solutions under these issues reveals that agency regulation of search is far from an ideal solution. Therefore, the government should establish a common law tribunal for search engine dispute resolution and create a panel of diverse interest groups to develop industry guidelines.

A. Important Questions to Consider in Search Engine Dispute Resolution

“Search engines sit at the nexus of a variety of information flows . . . between users, content providers, and advertisers.”¹⁷⁷ This variety gives rise to competing policy concerns when resolving search engine disputes, requiring trade-offs between “privacy and access, transparency and efficiency, and being found and remaining hidden.”¹⁷⁸ Therefore, regulation of search must have as its foundation fair information practices and must maintain an ongoing dialogue with input from consumer advocates, the government, and other stakeholders.¹⁷⁹

Regulation of search must address several key conundrums. First is the question of how the government will decide if search results are “fair.”¹⁸⁰ Each search engine dispute is likely to be different, and developing a set of durable and objective legal criteria would be difficult because cyberspace creates a “quicksilver technological environment that might make yesterday’s regulation superfluous tomorrow.”¹⁸¹ Second, a uniform regulation “will make results from all search engines standardized.”¹⁸² This may undermine competition and experimentation, hindering growth and innovation.¹⁸³ Third, regulation would require transparency, because the “algorithms of search engines are generally trade secrets and [may] therefore result in undetected, inherent biases.”¹⁸⁴ However, “if search engines have to disclose their proprietary algorithms,” manipulating the system also becomes easier.¹⁸⁵ Lastly, regulation of search must not encroach on informational autonomy.¹⁸⁶ The elements of

and communications.” *Id.*

¹⁷⁷ Moffat, *supra* note 166, at 479.

¹⁷⁸ *Id.* at 477.

¹⁷⁹ See Johnson, *supra* note 155.

¹⁸⁰ Gelfer, *supra* note 164.

¹⁸¹ Urs Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead*, 8 YALE J. L. & TECH. 201, 224 (2006).

¹⁸² Gelfer, *supra* note 164.

¹⁸³ *Id.*

¹⁸⁴ Gasser, *supra* note 181, at 232-33.

¹⁸⁵ Gelfer, *supra* note 164.

¹⁸⁶ Gasser, *supra* note 181, at 227-228.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

informational autonomy – “freedom to make choices among alternative sets of information, . . . right to express [one’s] own beliefs and opinions, . . . [and every user’s right to] participat[e] in the creation of information, knowledge and entertainment”¹⁸⁷ – are the most crucial virtues that shape the Internet. Therefore, frameworks for search engine governance “should aim to maximize access to search engines both for users and content providers on non-discriminatory terms.”¹⁸⁸

B. Administrative Agency (“Federal Search Commission”) Regulation: A Less-than-ideal Solution

Some scholars have called for the creation of a “Federal Search Commission” to directly oversee operation of search engines.¹⁸⁹ The proponents of this approach liken search engines to “utilities, essential facilities, and common carriers,” concluding that the government should regulate search engines “similarly to telecommunications firms or airlines,” offering their services on a non-discriminatory basis.¹⁹⁰ Google’s dominance over the industry further justifies this approach because in a world where Google did not control such a “significant portion of consumer behavior, this type of regulation [would not] be necessary.”¹⁹¹ Transparency is a key value that the establishment of an administrative agency would achieve, as “traffic generators that use rule-based algorithms to determine result sets [would have to] disclose their methodologies.”¹⁹²

This approach, however, has more drawbacks than benefits. First, direct regulation by an agency does not guarantee objectiveness. “It is difficult to imagine that a federal agency, or anyone else, [could devise] an objective set of criteria for evaluating the propriety of search engine results.”¹⁹³ Furthermore, an agency is unlikely to do a substantially better job of controlling or responding to claims of improper bias in the search rankings than search engines currently do themselves.¹⁹⁴

A cause for greater concern is that “federal agency regulation can be a particularly slow-moving process” that is unsuitable for application in search.¹⁹⁵ Search results are dynamic: “search technology, . . . related business

¹⁸⁷ *Id.* at 227-28.

¹⁸⁸ *Id.* at 231.

¹⁸⁹ Moffat, *supra* note 166, at 488.

¹⁹⁰ *Id.* at 487-89. Opponents debate this point vigorously, arguing that “traditional justifications for agency regulation simply are not present.” *Id.* at 478.

¹⁹¹ *TechCrunch*, *supra* note 8.

¹⁹² *Id.*

¹⁹³ Moffat, *supra* note 166, at 493.

¹⁹⁴ *Id.*

¹⁹⁵ *Id.* at 492.

models, and consumer behavior are . . . changing so rapidly.”¹⁹⁶ Thus, a federal agency cannot “effectively regulate the industry or the technology.”¹⁹⁷ The sluggishness of agency regulation may also “lock in” standards or technologies that are less than optimal.¹⁹⁸

Lastly, the public may view direct regulation by an administrative agency as an encroachment on informational autonomy. A recent survey shows that seventy-seven percent of adults expressed the view that “the government has no business regulating search engine results.”¹⁹⁹

Therefore, creation of a “Federal Search Commission” is not an ideal solution to effectively regulate search protocols. Nevertheless, concerns over providing adequate protection against unethical Internet practices such as “black hat” SEO offer a strong argument for “substantially more intrusive and centralized intervention.”²⁰⁰

C. Creation of a Common Law Tribunal and Industry Panel

Though agency regulation is inappropriate for the search industry, “a more coherent and centralized approach” than currently implemented is still necessary.²⁰¹ The alternative, and better, solution is for the “federal courts to take on . . . the task of regulating search engines.”²⁰² Creation of a common law tribunal that specializes in Internet search issues will give the government the appropriate power to centralize regulation without hindering the industry.

Technology changes in unforeseeable ways, and “controlling [it] is difficult, if not impossible.”²⁰³ Thoughts on any given technology will not remain static, and the governmental body that oversees that technology must be able to adapt to the mercurial changes.²⁰⁴ Common law development is “iterative [and] fact-intensive” in nature.²⁰⁵ Common law may also “be more adaptable in the face of rapid change than a strict statutory structure or a set of agency regulations

¹⁹⁶ *Id.*

¹⁹⁷ *Id.*

¹⁹⁸ *Id.*

¹⁹⁹ Danny Goodwin, *77% of Americans Oppose Government Regulation of Search Engines*, SEARCH ENGINE WATCH (Jan. 11, 2011), <http://searchenginewatch.com/article/2050190/77-of-Americans-Oppose-Government-Regulation-of-Search-Engines-Survey>. With twelve percent undecided on the issue, only eleven percent agreed government regulation of search is necessary. *Id.*

²⁰⁰ Moffat, *supra* note 166, at 489.

²⁰¹ *Id.* at 478.

²⁰² *Id.* Such forum should also be responsible for resolving search engine disputes. *Id.* at 499.

²⁰³ *Id.* at 500.

²⁰⁴ *Id.*

²⁰⁵ *Id.* at 501.

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

formulated in response to an earlier issue or previous-generation technology.”²⁰⁶ In other words, “the flexibility of common law renders it less likely to inhibit innovation or lock in [outdated] standards.”²⁰⁷ Several federal statutes “already govern certain aspects of search engine operation,” and the federal court system is certainly capable of performing “ongoing updates and . . . filling in the gaps between [the] statutes.”²⁰⁸

In contrast to agency regulation, this common-law approach is advantageous in its non-intrusiveness.²⁰⁹ “Under a common law system, the applicable legal principles tend to be more general,”²¹⁰ therefore less likely to infringe informational autonomy on the Internet. Outside the United States, industry leaders have called for precisely this approach. In Australia, the primary telecommunications industry body proposed creation of an independent judicial body to resolve online copyright disputes.²¹¹ The proposal further recommends establishing a panel of Internet service providers and rights holders who “will agree to cooperatively establish and jointly fund a Copyright Industry Panel.”²¹² The Panel’s main responsibilities include “prepar[ing] and disseminat[ing] educational material including but not limited to copyright issues, the availability of legal content online, how to make [I]nternet services secure from unauthorized use, and how consumers can avoid infringing activity.”²¹³ The addition of an industry panel is an important dimension in the common-law tribunal approach, helping represent various groups’ interests in the operation of the tribunal.

Creation of a common law forum specializing in search engine and other online disputes is superior to a direct regulation by an administrative agency. To provide better, more centralized protection against copyright infringement and web spamming through “black hat” SEO techniques, the government should adopt this approach to regulate search engines and their protocols.

²⁰⁶ *Id.*

²⁰⁷ *Id.* at 504.

²⁰⁸ *Id.* at 503.

²⁰⁹ *Id.* at 502.

²¹⁰ *Id.*

²¹¹ COMMUNICATIONS ALLIANCE, LTD., AUSTRALIAN INTERNET SERVICE PROVIDER PROPOSAL: A SCHEME TO ADDRESS ONLINE COPYRIGHT INFRINGEMENT 8-9 (2011), available at http://commsalliance.com.au/__data/assets/pdf_file/0019/32293/Copyright-Industry-Scheme-Proposal-Final.pdf [hereinafter *ISP Proposal*].

²¹² *Id.*

²¹³ *Id.* In other words, the panel will issue guidelines on “best practices” on the Internet. The proposal further suggests that the industry panel take part in operations of an appeals process for those “who receive notices (similar to DMCA takedown notices) and “believe that they are not responsible for any infringing activity.” *Id.* at 9.

VIII. CONCLUSION

A substantial majority of Internet users rely on search engines to reach information of interest on the Internet.²¹⁴ As search engines became more popular, various optimization techniques have emerged to exploit search algorithms to improve rankings of websites on search result pages.²¹⁵ However, search engine companies view the more extreme of these techniques as unethical, dubbing them “black hat” search engine optimization and punishing websites that use them through demotion or de-indexing.²¹⁶ Google, who has fought a long battle against unethical practices on the web, has recently modified its search algorithm to consider the number of takedown notices²¹⁷ under the DMCA, and to directly target websites engaged in web spamming.²¹⁸ Though the algorithm updates primarily protect major copyright holders on the web such as the MPAA and the RIAA, it is likely to affect the SEO industry because “black hat” techniques often engage in copyright infringement as well as spamming and potential trespass to chattels.²¹⁹

Evolution of search engine technology has eradicated abated many “black hat” SEO techniques, such as keyword stuffing, link farming, cloaking, and doorway pages.²²⁰ However, two techniques – content scraping and link spamming – still survive and continue to pollute the Internet with practices of questionable legal character.

A content scraper posts verbatim copies of content that other, more popular websites have published on the web.²²¹ Content scraping often involves pay-per-click advertising models, generating revenue by tricking users into visiting a site that mimics the original content for which the users searched.²²² This practice violates the original author’s exclusive right to reproduce his work in copies, constituting copyright infringement.²²³ Furthermore, overloading a web server with excessive content scraping may constitute trespass to chattel.²²⁴

Link spamming involves abusing other websites, especially those open to

²¹⁴ See Yalçın & Köse, *supra* note 2, at 487.

²¹⁵ See *supra* Part II.

²¹⁶ Marrs, *supra* note 40; *SocialTechNow*, *supra* note 39.

²¹⁷ See Farivar, *supra* note 10.

²¹⁸ Cutts, *What Should We Expect*, *supra* note 12.

²¹⁹ See *supra* Part IV.

²²⁰ See *SocialTechNow*, *supra* note 39.

²²¹ Richards, *supra* note 26.

²²² See *id.*

²²³ See 17 U.S.C. § 106(1) (2012).

²²⁴ See *eBay*, 100 F.Supp.2d at 1070 (finding that overloading of servers is sufficient to establish a strong likelihood of proving intentional, unauthorized interference with plaintiff’s possessory interest in the computer system, and to proximately result in damage to plaintiff).

2014] “BLACK HAT” SEARCH ENGINE OPTIMIZATION

posts by the public, to litter the Internet with inbound links.²²⁵ The technique can improve a website’s search rankings because search engine algorithms reward sites with a high number of inbound links.²²⁶ Though this technique does not infringe any copyright, a victim of link pollution may have a claim under the principles of trespass to chattel²²⁷ or the CAN-SPAM Act of 2003.

Nevertheless, current law is largely inadequate to provide appropriate protection against “black hat” SEO offenses. Website owners are likely to be reluctant to pursue legal action against content scrapers or link spammers because the loss from their conduct is often minimal, especially on issues that appear to be close calls when framed within free speech issues.²²⁸ Few cases involving search engine disputes are currently on the docket.²²⁹ Though users expect search engines to diligently rid the Internet of deceitful practices, search engines generate a majority of their revenue through their online advertising platforms and may have conflicting incentives to pursue any legal initiatives against “black hat” SEO.²³⁰ The government – the FTC in particular – has appropriate resources and incentives to centralize regulation of the Internet,²³¹ but faces public opposition and skepticism over its competency in providing such regulation.²³²

The inadequacy of the current legal system in the protection of rights holders must not persist. Self-regulation by Google is not sufficient to keep the “black hats” in check, nor trustworthy. Google may have become too powerful in the recent years for market forces to check, and its dominance has raised many concerns over abuse of its power, biases in search result presentation, and even antitrust violations.²³³ Though there is a chance in the long run that continued “black hat” abuse will motivate content-generators to put more pressure on the search engines to pursue eventual legal actions, such resolution in the near future is difficult to imagine. Therefore, a more coherent and centralized approach through government regulation is risky but necessary.

Scholars have proposed two potential solutions to centralize regulation of search engines: establishment of an administrative agency and creation of a common law tribunal. The ideal approach must guarantee that such regulation maintains fairness in search result ranking criteria, does not hinder growth and

²²⁵ Richards, *supra* note 26.

²²⁶ Richards, *supra* note 2626.

²²⁷ See *eBay*, 100 F. Supp. 2d at 1070.

²²⁸ Goldman & German, *supra* note 8, at 36.

²²⁹ See, e.g., Complaint at 3-4, *Boston Pads, LLC v. The Velvet Pillow.com*, No. SUCV2014-00197 (Mass. Super. Ct. Jan. 17, 2014).

²³⁰ Goldman & German, *supra* note 8, at 34.

²³¹ See Johnson, *supra* note 7.

²³² See Racheter & Oliver, *supra* note 160.

²³³ Gelfer, *supra* note 164; see also *The Google Algorithm*, *supra* note 6.

innovation, balances transparency against potential abuse, and does not encroach informational autonomy.²³⁴ Direct regulation of search by an administrative agency – a “Federal Search Commission” – does not satisfy many of these qualifications. An agency is unlikely to excel in controlling or responding to “claims of improper bias in the search rankings.”²³⁵ Due to its slow-moving nature, federal agency regulation is not only unsuitable for regulating the ever-changing search industry but may also result in the “locking in” of “sub-optimal standards or technologies.”²³⁶ Furthermore, such agency regulation is likely to severely infringe informational autonomy.²³⁷

The alternative and best solution to centralize regulation of search technologies is to create a common law tribunal for search engine disputes with a complementary industry panel. Because common law development is “iterative [and] fact-intensive [in] nature,” it is much more adaptable to rapid and unforeseeable changes than a strict statutory structure or agency regulations.²³⁸ The federal court system is much more capable of performing “ongoing updates” and “filling in the gaps between [existing] statutes.”²³⁹ Common law principles are also more general than statutes and regulations,²⁴⁰ suggesting that this approach is less likely to encroach on informational autonomy on the Internet. Finally, an industry panel to publish guidelines for best practice, offer expert knowledge in appeals, and represent interests of various groups completes an ideal system for regulation of search.²⁴¹

To improve on the current state of legal protection against copyright infringement and web spamming through “black hat” SEO techniques, the government should establish a common law tribunal specializing in search engine disputes and complemented by an industry panel. Such centralized regulation is the first step toward a web professional’s dream: a truly organic, black-hat-free Internet.

²³⁴ Gelfer, *supra* note 164. *See also* Gasser, *supra* note 181, at 227-28, for a discussion on the importance of informational autonomy on the Internet.

²³⁵ Moffat, *supra* note 166, at 493.

²³⁶ *Id.* at 492.

²³⁷ Gasser, *supra* note 181, at 230.

²³⁸ Moffat, *supra* note 166, at 501.

²³⁹ *Id.* at 503.

²⁴⁰ *Id.* at 502.

²⁴¹ *ISP Proposal*, *supra* note 3, at 8-9.