

SEARCH ENGINE OPTIMIZATION: WHAT WE SEE AND WHY WE SEE IT

Joseph Baillargeon *

CITE AS: 4 GEO. L. TECH. REV. 299 (2019)

TABLE OF CONTENTS

I. INTRODUCTION: WHAT IS SEARCH ENGINE OPTIMIZATION AND WHY IS IT IMPORTANT.....	299
II. CRAWLER SEARCH ENGINES: THE MECHANICS OF “GOOGLING”	301
III. GETTING RESULTS: ON AND OFF-PAGE SEO.....	303
IV. CONCLUSION	305

I. INTRODUCTION: WHAT IS SEARCH ENGINE OPTIMIZATION AND WHY IS IT IMPORTANT

Search Engine Optimization (SEO) is a technique to improve the ranking of a website on the results page of a search engine for the purpose of increasing the traffic to that site.¹ This technique has become crucial in the modern Internet era, where standing out amongst the over 1.7 billion websites can be a challenge.² This challenge derives from the fact that many of these websites are competing for the top spots of the same search queries. While such a competition may not seem that important, its outcome can actually be pivotal to the success of an online business. The unfortunate reality is that only about thirty-seven percent of online shoppers look past the top three results.³ These top three search positions have come to be known as “the golden triangle” and are highly sought after.⁴

* Georgetown University Law Center, J.D. Candidate 2022; Virginia Polytechnic Institute, B.S. Engineering Science and Mechanics, B.A. Physics. Thank you to the wonderful editors of GLTR, without whom this piece would not have been possible.

¹ Dushyant Sharma et al., *A Brief Review on Search Engine Optimization, in CONFLUENCE 2019: 9TH INT’L CONF. ON CLOUD COMPUT., DATA SCI. & ENG’G* 687, 687 (2019).

² INTERNET LIVE STATS, <https://www.internetlivestats.com/> (website live tracks total number of websites and constantly updates total) (accessed Nov. 22, 2019) [<https://perma.cc/ZR7A-BA8L>].

³ See Sharma et al., *supra* note 1, at 689.

⁴ *Id.*

The importance of this behavioral phenomenon is only increasing with time as e-commerce continues to play a larger and larger role in the economy. Studies have shown that in today's marketplace sixty-one percent of Internet users research products online and forty-four percent of online shoppers use a search engine to begin the search for a seller.⁵ Furthermore, of that forty-four percent of online shoppers, seventy-five percent never even click past the first Search Engine Results Page (SERP).⁶ Companies like Google do provide services, such as AdWords, which allow a website to be placed at the top of certain relevant searches without having to worry about SEO.⁷ However, studies have shown that seventy percent of the links that online shoppers click on are organic—meaning the links were recommended by Google's algorithm.⁸

The behavioral impact of search engine rankings is significant and certainly not limited to economics thereto. A study in India used a mock search engine to intentionally return biased results to its participants, such that the first SERP only exposed them to the positive news articles of a target candidate when that candidate was searched for.⁹ The study concluded it would be relatively easy to persuade around twenty percent of undecided voters to a target candidate, merely by changing what the voter saw in his or her search results.¹⁰

What an individual sees on the Internet is significantly determinative of what that individual will buy and how the individual will think. It has therefore become imperative to understand the algorithms that dictate what we see every time we “Google” something so we can appreciate their fallibility. Part II of this paper will help accomplish this by outlining the basics of how a search engine works. Subsequently, Part III will focus on some of the ways webmasters use search engines to gain a competitive edge through SEO. Finally, Part IV will conclude this paper by reiterating the importance of understanding these topics in a modern context.

⁵ Venkat N. Gudivada et al., *Understanding Search Engine Optimization*, COMPUTER, Oct. 2015, at 43, 43 (2015).

⁶ *Id.*

⁷ Aranyak Mehta et al., *AdWords and Generalized On-Line Matching*, in 46TH ANNUAL IEEE SYMPOSIUM ON FOUNDATION OF COMPUTER SCIENCE 1, 1 (2005).

⁸ Gudivada et al., *supra* note 5, at 43.

⁹ Robert Epstein & Ronald E. Robertson, *The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections*, 112 PNAS 1, 6 (2015).

¹⁰ *Id.* at 9.

II. CRAWLER SEARCH ENGINES: THE MECHANICS OF “GOOGLING”

Most people are accustomed to using crawler-based search engines, which include search engines like Bing, Yahoo!, and Google.¹¹ This paper will focus on Google, as Google handles around seventy-five percent of Internet searches.¹² In general, the crawler-based search engine takes four major steps to try to find and display what a user is looking for: crawling, indexing, calculating relevancy, and retrieving the results.¹³ First, the search engine must run a large number of computer programs, known as bots, that scour the Internet for as many webpages as they can find.¹⁴ Next the search engine stores a copy of the webpages into a giant document database, which Google calls Caffeine.¹⁵ Once the database of webpages has been created or updated, the contents of those websites are formatted to remove unnecessary information and keywords and phrases are extracted.¹⁶ The document database then uses these keywords and phrases to create indexes to be later queried by the user.¹⁷ Finally, when the user enters a keyword or phrase into the browser, the text string is transformed into its canonical form—a form that allows the search engine to take into account misspellings, capitalization, related words, noise words, etc.—and the Google search engine searches the indexed document database, Caffeine, for this canonical form, returning the most relevant URLs using Google’s ranking algorithms.¹⁸

The ranking algorithm is one of the most important aspects of the search engine. Since its inception, Google has been working diligently to perfect this algorithm by improving its accuracy. The original Google ranking algorithm, called PageRank, ranks webpages based on: (1) the number of outgoing links on a page, the number of backlinks, or pages that link into that

¹¹ Sharma et al., *supra* note 1, at 687.

¹² Dave Davies, *Meet the 7 Most Popular Search Engines in the World*, SEARCH ENGINE J. (Jan. 7, 2018), <https://www.searchenginejournal.com/seo-101/meet-search-engines/#close> [<https://perma.cc/8CB3-AXJ6>].

¹³ Sharma et al., *supra* note 1, 687–88.

¹⁴ Gudivada et al., *supra* note 5, at 44.

¹⁵ Sharma et al., *supra* note 1, at 687; Dave Davies, *Google’s Caffeine Update: Better Indexing & Fresher Search Results*, SEARCH ENGINE J. (Nov. 20, 2017), <https://www.searchenginejournal.com/google-algorithm-history/caffeine-update/#close> [<https://perma.cc/8CB3-AXJ6>].

¹⁶ Gudivada et al., *supra* note 5, at 44.

¹⁷ *Id.*

¹⁸ Crosby Grant, *Canonical Form: The Hidden Keywords in Paid Search*, SEARCH ENGINE LAND (Dec. 26, 2011), <https://searchengineland.com/canonical-form-the-hidden-keywords-in-paid-search-100603> [<https://perma.cc/H7JR-DUUE>]; Gudivada et al., *supra* note 65, at 44.

page; and (2) the quality of those links.¹⁹ This method allows the internet to “vote” on which webpage it deems the most relevant.²⁰ Once people realized this, however, artificial rank inflation coincided, as people exploited the algorithm by doing things like loading their websites with hidden links in text colors that matched the backgrounds.²¹ These tactics undermine the quality of the results returned by the search engine and are known as “black hat” SEO practices because they are contrary to the Google guidelines for creating websites.²² The competing interests of search engines seeking to produce high quality results and companies seeking placement at the top of the SERP have created an arms race that has proliferated the number and elevated the quality of Google’s various ranking algorithms.

Some such ranking algorithms include: Panda, Penguin, the Pirate, Payday Loan, Hummingbird, Pigeon, Possum, and Fred. Panda is designed to evaluate the quality of information on a website—such as having too little or too much information, poor format, grammar error, spelling mistakes, unreliable information, or low-quality content.²³ Penguin checks for contextual page linking (links that are surrounded by text), and determines if the links are from trustworthy sources.²⁴ The Pirate algorithm blocks or de-ranks sites that have received several reports for copyright infringement.²⁵ Payday Loan filters out pornography, casino, and high interest loan sites.²⁶ Hummingbird attempts to interpret the user’s intent when they input a keyword or phrase into the browser.²⁷ Pigeon finds the user’s location and alters ranking based on the location of companies in the user’s area.²⁸ Possum improves the ranking of websites or businesses in the top position of the search result to coincide with those nearest to the location of the searcher.²⁹ Finally,

¹⁹ Ian Rogers, *The Google Pagerank Algorithm and How it Works*, PRINCETON UNIV., <https://www.cs.princeton.edu/~chazelle/courses/BIB/pagerank.htm> [https://perma.cc/7WAK-7WQJ].

²⁰ *Id.*

²¹ Sharma et al., *supra* note 1, at 688.

²² *Id.*

²³ Abuzar Khan, *Top 12 Google Algorithm Updates You Need to Know in 2019*, SEO BASICS (May 29, 2019), <https://www.seobasics.net/google-algorithm-updates> [https://perma.cc/5RP3-9XRE].

²⁴ *Id.* at §§ 6–7.

²⁵ *Id.* at § 8.

²⁶ *Id.* at §§ 10–11.

²⁷ *Id.* at §§ 11–12.

²⁸ *Id.* at §§ 13–15.

²⁹ *Id.*

Fred looks for excessive ads, low-value content, and websites that offer very little user benefit.³⁰

The aggregate effect of these algorithms makes sure that users get the highest quality information relevant to what they are looking for. These algorithms take into account over 200 factors³¹ to accomplish their respective functions, which is where legitimate, “white hat” SEO practices come into play. SEO is not about gaming the system, but rather, optimizing a website to help search engines return the most pertinent results to the user.³²

III. GETTING RESULTS: ON AND OFF-PAGE SEO

SEO is not an exact science. It requires quite a bit of guess work because Google’s ranking algorithms are constantly being updated and tested, and their exact algorithms are guarded as a trade secret.³³ Google commonly sequesters a small subsection of the Internet to test new algorithms, so it is possible for two people using Google at the same time to be using different algorithms.³⁴ The SEO community is left to rely on a mishmash of vague guidance and second hand sources to determine best practices. One of the best sources comes from Google itself, which publishes its own guidelines on how to design a website with SEO in mind.³⁵ Additionally, because Google patents some portions of its algorithms, those portions are publicly available and provide a broad overview of what is going on behind the scenes.³⁶ But outside of what Google chooses to disclose, the only other reliable SEO strategies come from trial and error and approximations (using mathematical and machine learning techniques).³⁷ The information from all these different

³⁰ Aleh Barysevich, *A Cheat Sheet to Google Algorithm Updates from 2011 to 2018*, SEARCH ENGINE WATCH (Oct. 10, 2018), <https://www.searchenginewatch.com/2018/10/10/a-cheat-sheet-to-google-algorithm-updates-from-2011-to-2018/> [<https://perma.cc/74WJ-UY6J>].

³¹ Gudivada et al., *supra* note 5, at 43.

³² *Id.* at 46.

³³ Kristine Forderer, *Trade Secrets: A Valuable Tool in Your IP Protection Strategy*, COOLEY GO (Feb. 4, 2019), <https://www.cooleygo.com/trade-secrets-a-valuable-tool-in-your-ip-protection-strategy> [<https://perma.cc/HU33-PSNP>]; Steven Levy, *Exclusive: How Google’s Algorithm Rules the Web*, WIRED (Feb. 22, 2010), https://www.wired.com/2010/02/ff_google_algorithm/ [<https://perma.cc/KU2L-XD7L>].

³⁴ Levy, *supra* note 33, at 687.

³⁵ GOOGLE, SEARCH ENGINE OPTIMIZATION STARTER GUIDE (2010), <http://static.googleusercontent.com/media/www.google.com/en/us/webmasters/docs/search-engine-optimization-starter-guide.pdf> [<https://perma.cc/RDK3-NRP4>].

³⁶ Lawrence Page, US Patent 6,285,999 B1, <https://patentimages.storage.googleapis.com/37/a9/18/d7c46ea42c4b05/US6285999.pdf> [<https://perma.cc/3M2B-NA26>].

³⁷ Hengameh Banaci & Ali Reza Honarvar, *Web Page Rank Estimation in Search Engine Based on SEO Parameters Using Machine Learning Techniques*, 17 INT’L J. COMP. SCI. &

sources has culminated into a plethora of SEO strategies that are often divided up into two distinct groups: (1) on-page, which refers to the actual content and code the makes up a person's website; and (2) off-page, which refers to all of the information about your website that is not contained on your website.³⁸ The strategies in each of these two groups can be further categorized as either a white hat or black hat strategy.³⁹

A crucial first step in implementing white hat, on-page SEO is to conduct "keyword research."⁴⁰ Keyword research is the process of finding commonly searched phrases and topics whose Google search results would be best to appear in.⁴¹ For instance, if someone had a Mexican restaurant in Washington DC, they might want the restaurant's webpage to be the top result in a search for, "best Mexican restaurants in DC." Google provides tools to determine the competitiveness and frequency of use for these searched phrases,⁴² to help decide which searches would be best to get highly ranked for. Once a phrase has been decided on, adjustments can be made to the webpage, including: the title tag, header tags, the URL, and the content of the webpage, such that they all gravitate around the target phrase.⁴³ It is ideal for the keyword to occur around every 5 to 7 words per 100 words.⁴⁴ Other things the Google algorithms are known to look at are: organic looking anchor text (text that is also a URL), a custom 404 page (what happens when you link to a page that does not exist), and a custom privacy policy.⁴⁵

Examples of off-page SEO techniques include things like placing backlinks to your website on other high-quality websites via blog comments⁴⁶ and social media sites like LinkedIn and Twitter.⁴⁷ This strategy is referred to as Link Building.⁴⁸ Another crucial off-page SEO technique is ensuring that Google has an accurate sitemap (information about and structure of every

NETWORK SEC. 95, 96 (2017) (discussing the extensive research that he been done since 2005 to computationally approximate web ranking),

http://paper.ijcsns.org/07_book/201705/20170513.pdf [<https://perma.cc/LC7J-NL6G>].

³⁸ Sharma et al., *supra* note 1, at 688–89.

³⁹ Gudivada et al., *supra* note 5, at 44.

⁴⁰ Sharma et al., *supra* note 1, at 688.

⁴¹ *Id.*

⁴² Google, *How Keyword Planner Works*, GOOGLE ADS, https://ads.google.com/intl/en_en/home/tools/keyword-planner/ [<https://perma.cc/9G7Y-DS39>].

⁴³ Sharma et al., *supra* note 1, at 688.

⁴⁴ Muhammad Naeem Ahmed Khan & Azhar Mahmood, *A Distinctive Approach to Obtain Higher Page Rank Through Search Engine Optimization*, SĀDHANĀ 1, 3 (Mar. 2018).

⁴⁵ Gudivada et al., *supra* note 5, at 48–49.

⁴⁶ Sharma et al., *supra* note 1, at 689.

⁴⁷ Khan & Mahmood, *supra* note 44, at 8.

⁴⁸ *Id.*

webpage on a website) for the website.⁴⁹ When Google sends its crawler bot to a website to be copied, it creates a sitemap as it traverses all of the links on the page.⁵⁰ By submitting its own sitemap directly to Google, a website can ensure that all of its webpages have been properly indexed and can be found through the search engine.⁵¹

Unfortunately, as with other automated systems, once the underlying algorithms are dissected, the system can be exploited. Exploiting these ranking algorithms to raise a website's rank is known as black hat SEO. These practices include things like keyword stuffing, where the website is filled with many different, often unrelated, keywords to increase the number of search engine queries for which a website is returned.⁵² Cloaking is a method where hackers cause the crawler bots and the users to see different content, usually to hide malicious content.⁵³ Hackers also often write computer code that causes automatic webpage generation, copying content word for word from well-known sites in attempts to rank for certain keywords.⁵⁴ These underhanded methods end up costing the businesses they displace an estimated \$130 billion annually.⁵⁵

IV. CONCLUSION

SEO is something that cannot be ignored by those who have or want to have an online presence. Until human behavior changes and people begin consistently looking past the first few search results, one of the best ways to increase traffic to a website is making sure the search engine algorithms are working in the site's favor. Search engines, like Google, have been increasing their accuracy by leaps and bounds over the past decade. However, at the end of the day these algorithms are still, and probably will always be, imperfect systems that are based off of imperfect metrics. Unless a website invests, at least minimally, in SEO, it is vulnerable to outranking by savvy webmasters with lower quality content who aspire to make a quick buck. The implications of this go far beyond just e-commerce and have profound effects on how the public is informed. Misinformation and disinformation can easily be ranked higher than accurate information.⁵⁶ Therefore, it is—now more than ever—

⁴⁹ Gudivada et al., *supra* note 5, at 50.

⁵⁰ *Id.*

⁵¹ *Id.*

⁵² Sharma et al., *supra* note 1, at 688.

⁵³ Gudivada et al., *supra* note 5, at 45.

⁵⁴ Gudivada et al., *supra* note 5, at 45.

⁵⁵ *Id.* at 44.

⁵⁶ STACIE HOFFMAN ET AL., OXFORD TECH. & ELECTIONS COMM'N, THE MARKET OF DISINFORMATION 11 (2019).

crucial to inoculate the populous to these manipulations by making sure they understand the fallibility and limitations that underlie their search results.