A Comparative Analysis of Google Web Page Ranking Algorithms

Kundan Kumar Jha - 22MSM40052

22SDT–606: DESIGN AND ANALYSIS OF ALGORITHMS

Branch: M.Sc. Data Science

Department of Mathematics

Chandigarh University, Mohali, Punjab

Abstract

One of the most popular search engines is Google Search, although it is uncertain what determines how items are ranked. In this term paper, we discuss the outcomes of our initial efforts to reverse the ranking algorithm used by Google. There are outcomes: The ranking algorithm used by Google Search gives the most weight to citation counts. As a result, papers with a high citation count are found at higher positions far more frequently than those with a lower citation count. The web is growing every day, and most people use search engines to browse it. In this case, it is the responsibility of the service provider to respond to the internet user's search engine query with accurate, pertinent, and high-quality information. Using the contents of online pages and links between web pages, service providers must work to deliver accurate, pertinent, and high-quality information to internet users. In order to determine their benefits and drawbacks for the ranking of online pages, this study compares and analyses various web page ranking algorithms. Comparison research is conducted based on the investigation of several web page ranking algorithms to determine.

Introduction

Since the advent of the internet and particularly since the advent of Google, the world of marketing has undergone significant change. Early in the new millennium, online marketing had a growth spurt; since then, it has advanced quickly and has continued to do so. Companies must be informed of the most recent changes and what is to come in the near future if they want to succeed in the realm of online marketing. My thesis is to enlighten readers about the fundamentals of algorithms and how they relate to SEO. I'll go through what algorithms are and how they impact SEO. Do marketers need to understand algorithms, and if so, why? The main focus of this thesis is on SEO and Google's algorithm-based ranking methods. I'll start by defining algorithms and explaining their role in online marketing. The largest algorithmic changes in Google's history will be discussed, along with some information about how they impacted online marketing.

Google

The history of Google began in 1995 when two men met at Stanford University. According to the Google Company home page, the pair had developed a new search engine called BackRub by 1996 that used links "to estimate the value of particular web pages." Google's purpose is to organise the world's information and make it widely accessible and valuable. The two men then set about building the concepts that would eventually become Google. Larry Page and Sergey Brin were them.

On September 4, 1998, in California, the United States of America, Google was established. The corporation explains that the name "Google" is a wordplay from the original word "googol," a mathematical term, on their official website (Google "Google's objective is..."). The wordplay of the phrase was chosen by Larry Page and Sergey Brim as their company name because it alludes to infinity and fits the duo's goal of organising an unlimited amount of information on the internet.

According to Google, its goal is "to organise the world's information and make it generally useful" (Google Company). Google's mission statement is. They set out to provide the finest search engine user experience possible right from the start. This applies to everything Google does, not just search; when it offers new tools, it wants them to be simple to use and reliable. Google has made a point of emphasising that every design decision is made with the user in mind (Google "Ten things...").

Google Search

Google thinks it is best to specialise in one area well. Search is the one area they want to concentrate on. They have made investments in sizable research teams that work to improve search by addressing its issues. According to Google's list of "ten things," they are committed to providing a "seamless experience for millions of people." Google wants to provide the consumer with as much information as it can, but they also want to do it quickly. They are continually working to speed up the search and other programmes to satisfy the user. Google does not think that the search query should yield just anything because its goal is to make the user's search experience as pleasant as possible. Google uses more than 200+ signals and approaches to evaluate which websites are the finest for a given search query in order to deliver relevant search results. The main subject of this thesis is Google's algorithms, which are among these signals and methods.

Google is adapting to the changing world. The firm Google is insatiable. They want to keep improving the search. According to Google's "Ten Things We Consider to Be True" company ethos: We view mastering a skill as a beginning rather than a finish. We pick goals that we know we can't yet achieve because we believe that by pushing ourselves to achieve them, we will advance further than we had anticipated.

Algorithms

What exactly are algorithms, and how do they relate to online marketing? In response to the question of what algorithms are, Google states unequivocally: "You want the solution, not billions of web pages. Computer programmes called algorithms search for hints to provide you with the information you need. Using Google's internal search algorithms Since there are now billions of web pages on the internet, there are probably hundreds or even more that contain useful information. But how can we ever discover the information we need from all of those pages if we have so many of them? No one will take the time to browse thousands of pages. Because of this, Larry Page and Sergey Brin began creating algorithms, which are computer operations and formulas that facilitate search.

Google's algorithms have advanced significantly, and today they focus on more than 200 distinct signals to help consumers discover the content they're looking for online. These signals, or at least a portion of them, are crucial for marketers to understand and be aware of because Google does not make all of its signals available to the general public. These signals are what the algorithms use to determine whether or not to display the page in the top results. Examples of the signals include keywords on websites, location, the timeliness of site information, and page rankings in internet searches.

Crawling and classifying pages serve as the beginning of the search process. Search engines such as Google have created custom site crawling robots for this purpose. These are occasionally referred to as spiders. "Googlebot" is the name of Google's web-crawling robot. These crawling robots scan each page as they move from one to the next, analysing the information to decide if the page is useful or not.

Google algorithm’s major changes

One of the most widely known improvements to Google's web search capabilities is the addition of PageRank. Google claims that PageRank calculates a website's relevance by counting the quantity and calibre of links pointing to a page. The relevance of a website is assumed based on the notion that a website with more links from many other websites is more important. In the section that follows, I'll go into further detail on PageRank.

Google Panda, an algorithm that concentrated on eliminating low-quality websites from the search results, and Google Penguin an algorithm that concentrates on lowering spammy websites, are two more significant developments in Google's history. In the sections that follow, I'll go into more depth on the Panda and Penguin updates.

PageRank

PageRank is the first algorithm which comes to mind when the topic of algorithms is brought up. What exactly is it? What does it do, or why is PageRank important for marketing staff to understand? What benefit does a marketer gain from understanding PageRank? These are significant inquiries that demand responses.

Websites are ranked using the Google search engine's PageRank algorithm in the search results. The more important a page is to search engines, the higher its PageRank. The creators of Google, Larry Page and Sergey Brin, created PageRank. Google Company’s "Our goods and services" states that Page once described the supposedly ideal search engine as one that "understands precisely what you mean and gets you back what exactly you want." These words have guided Google's development of its search engines, which aim to simplify user access to information as much as possible. Despite this, Google has risen to the challenge of developing faster and smarter search engines.

The definition of PageRank by **Marziah Karch** in her web post "What Is PageRank and How Do I Use It?" in About.com Guide states that the algorithm is patented by Stanford and gets its name from Larry Page. In contrast to earlier search engines, PageRank places less emphasis on term density. Karch claims that with earlier search engines, it was feasible to trick the system by cramming the page with keywords. PageRank, on the other hand, values pages that have a lot of links pointing to them. Links are viewed by PageRank as "votes," and whenever one website is connected to another, that other page receives a vote.

Karch draws attention to the fact that PageRank is unique in that it also rates the pages that link to another page. Vote value is lower for pages having low PageRank compared to pages with greater PageRank. Linking doesn't have a quantity problem. What counts is the calibre of the pages.

Even if PageRank seems great, Karch points out in her article that it is not perfect. By manipulating the PageRank algorithm, so-called "Black Hat" webmasters or advertisers have discovered certain strategies to improve their page ranking. There are several strategies to try to trick PageRank, one of which is "link farming," which refers to the practice of using links as the content of a website in order to improve ranking. Google has taken action against this type of behaviour, and as a result, pages with an excessive number of low-quality links have lost some of their page rank value. Although linking to these sites from your website won't necessarily have an impact on it, you shouldn't link to them in the first place.

In this paper, Marziah Karch describes how PageRank is calculated using a scale that ranges from 1 to 10. The ranking is given to specific website pages, not to the entire website. Obtaining relevant links for your website is necessary if you want to improve your page rating. Trading ties, for instance with business partners, is one opportunity.

Google Panda

Now that we are more familiar with PageRank, we can discuss the upcoming significant algorithm shift in Google history. The Google Panda algorithm is the other one, and it drew a lot of attention when it was introduced in 2011 because it damaged so many websites. Google Panda's primary goal, according to Danny Sullivan in his article "Google Forecloses On Content Farms With "Panda" Algorithm Update" (2011) in Search Engine Land, was to identify "content farms" and prevent them from appearing in Google search results. Google Panda was introduced in February 2011. Sullivan points out that, while Google does not explicitly state that the upgrade targets content farms, Matt Cutts has made references to these websites. Sites having "shallow or low-quality material," as Google explains them in its official blog article "Google search and search engine spam," are referred to as "content farms" (2011). In the same post referenced above, Sullivan explains that Google Panda also intended to take action against scraper sites, which, as their name suggests, scrape content from other websites in an effort to rank well. The websites with the original content may not appear on Google's first page as a result, which infuriates the webmasters as well as marketers of the original content.

Google Penguin

Penguin, a Google algorithm designed to combat webspam, was introduced in April 2012. The Penguin will affect around 3% of search queries, according to Google's Matt Cutts in the post "Another step to encourage high-quality sites" (2012) on the official Google Developer Central Blog. Compared to languages like Polish, which have been found to have greater spam, this percentage only applies to English, German, Chinese, and Arabic.

Google declared that the objective of this algorithm was to lower rankings for websites that disregarded Google's quality standards. In Danny Sullivan's article "Google Launches 'Penguin Update' Targeting Webspam In Search Results" (2012), he lists the main offences as keyword stuffing, link scheme, cloaking, and intentional duplicate content.

What does keyword stuffing entail? Keyword stuffing refers to a practice where excessive keyword use is done to alter a website's position in search results. Google believes that overusing keywords on web pages (sometimes known as "keyword stuffing") detracts from the user experience. Because keywords are crammed into the writing as much as possible, the user is given a bad experience because the language is confusing and frequently repetitious. In other words, the material offers the reader nothing pertinent. The lack of original, pertinent, and understandable content is taken into account as a signal to lower the page's rating. Google thus advises employing keywords wisely in context to promote a user-friendly reading and browsing experience. Google has illustrated the kind of text that should not be written and would be considered to be keyword stuffing.

Google Hummingbird

In 2013 Google released a new algorithm upgrade. Danny Sullivan from Search Engine Land was one of the first people to submit all pieces of information on this algorithm adjustment. Sullivan explains that Google has created a new search algorithm named Hummingbird in his post "FAQ: All About the New Google "Hummingbird" Algorithm" (2013). Google began utilising Hummingbird in September, although the move was not officially announced until October 2013. Being "precise and quick," like hummingbirds, is where the name comes from. One of the most significant shifts in Google history is Hummingbird.

According to reports, Hummingbird is more of a "new engine" than an actual update. According to Sullivan, the Hummingbird engine is entirely new but incorporates elements of the Panda and Penguin engine systems. Despite being "a new engine," Hummingbird's introduction does not forbid "existing engines" like PageRank. Sullivan claims that Hummingbird takes advantage of PageRank as well as other elements. That algorithm has no impact on SEO. Hummingbird does not alter the result as Panda and Penguin did; rather, it merely enables Google to interpret the signals in a novel manner.

Why do marketers need to know about algorithms?

It's time to consider why, having covered the most significant and well-known algorithmic modifications, I covered them all. What was the reading's actual purpose? I sincerely believe that better exploitation of these algorithms is achievable when marketing personnel is aware of and understands them better, leading to improved positioning in search engines. Understanding algorithms is essential for working well in an internet marketing platform.

Understanding algorithms and how they evolve gives you a solid foundation for SEO. Understanding the algorithms that drive SEO makes it simpler for marketers to adhere to its regulations if they are aware of its rationale. Understanding algorithms and staying current on their changes, in my opinion, helps one stay current on developments in internet marketing and the business sector while also giving them a deeper understanding of SEO.

Algorithms’ impact on SEO

Well, knowing the algorithms is certainly helpful for SEO. Online marketers frequently rely on the expertise of search algorithms to know the tiny secrets for getting the finest outcomes. According to Eric Enge, Stephan Spencer, Rand Fishkin, and Jessie Stricchiola in "Art of SEO: Mastering Search Engine Optimization," understanding crawling and indexing is crucial for SEO practitioners as it aids them in making decisions about how to proceed in order to achieve their objectives (2009, 30). In a YouTube video titled "What are the main three-five SEO areas where webmasters do the most mistakes?" Matt Cutts expresses the view that marketers should have a thorough understanding of how search engines operate. (2013).

Based on Matt Cutts' video on web marketers' most common SEO blunders, Barry Schwartz prepared a brief post. Matt Cutts' top five SEO blunders are stated by Schwartz in the article "The Top 5 SEO Mistakes as Per Google's Matt Cutts." "Not utilising webmaster resources and understanding how Google works and what SEO is about" is one of these prevalent errors.

Since algorithms are the "behind the scenes" element of Google search, I believe this is a compelling argument for internet marketers to learn at least the basics of them. Additionally, without algorithms, spam sites could receive more attention than really informative sites, making SEO ineffective.

CONCLUSION

I've introduced algorithms and provided some fundamental information about them in this essay. The algorithms are the "behind the scenes" of the search, as I've explained, thus it's critical to comprehend them and how they operate. As I've already mentioned, search algorithms organise all of the online data so they can respond to search queries with the most relevant information possible. In order to give the searcher relevant results, algorithms rank the pages in addition to organising the content. This implies that the algorithms scan the pages for information and decide whether it is relevant—in which case the ranking rises—or irrelevant. Naturally, as marketers, we want our page to rank higher, thus we must be aware of the elements that influence Google's perception of the page.

In other words, algorithms serve as the foundation for SEO. The criteria used by algorithms to determine if a webpage has good or terrible attributes form the basis for SEO regulations. When the algorithms are already known and provide support for the marketer's understanding of the online environment, it is simpler for them to adhere to the SEO guidelines and comprehend how well SEO is performed. Additionally, despite claims that SEO will change and that specific SEO techniques would improve rankings, a marketer with a solid understanding of the algorithms will be able to differentiate between what is only a rumour and what information needs to be taken seriously.

References

1) Amie Baumwell. *Panda 2.5 Does Some Damage.*

2) Barry Schwartz. Search Engine Land. *The Top Five SEO Mistakes According To Google's Matt Cutts*.

3) Enge E., Spencer S., Fishkin R., Stricciola J.S. *The Art of SEO; Mastering Search Engine Optimization*, United States of America: O'Reilly Media Inc. 2010.

4) John J. Meier and Thomas W. Conkling, "Google Scholar's Coverage of the Engineering Literature: An Empirical Study," *The Journal of Academic Librarianship*, vol. 34, 2008.

5) Judit Bar-Ilan, "Which h-index? — A comparison of WoS, Scopus and Google Scholar," *Scientometrics*, vol. 74, 2007, 257-271.

6) Péter Jacsó, "Google Scholar: the pros and the cons," *Online Information Review*, vol. 29, 2005, 208-214.

7) White, Bruce, "Examining the claims of Google Scholar as a serious information source," *New Zealand Library & Information Management Journal*, vol. 50, 11-24.

8) Beel, Jöran and Gipp, Bela "Collaborative Document Evaluation: An Alternative Approach to Classic Peer Review," *Proceedings of World Academy of Science, Engineering and Technology*, Vienna: 2008, vol. 31, 410-413. Available at: www.scienstein.org.

9) S. Pal, V. Talwar, and P. Mitra, "Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions: In IEEE Trans. Neural Networks, 13(5), PP.1163–1177,2002.