Google Page Rank Algorithm and It's Updates

Mr. Anuj Joshi

Department of Information Technology, Chandubhai S.Patel institute of Technology

Charusat University, Petlad, Anand

Prof. Priyanka Patel

Department of Information Technology, Chandubhai S.Patel institute of Technology, CHARUSAT

ABSTRACT

Google's search engine i.e Google's PageRank is one of the biggest search engines with 67% market worldwide but its algorithm is unknown. Though, Bing, Yahoo also holds good amount in search engine market but are far behind Google. Similarly, in China Baidu holds 59% of search engine market. There are billions of websites on internet while millions of websites add up every day. But in this huge dump yard of websites, every website wants to rank up in search engines search result. Question remains same as how to score up in any search result query. So, all search engines more or less have similar algorithm, but in this paper, we'll be focusing on Google's Page Rank algorithm. Google being prominent player in search engine holds a prime role for affecting your quality of website. Content can be in any form on web like email, talk, private stuffs, messages, blogs, images etc. So, there are smart crawlers which are bots or algorithm who fetch data from web and are added to databases in pretty formatted manner. On, later stage link is identified and a tree like structures are made which then describes a page or domain strength on web.

Keywords: Google, Search Engine, PageRank, Domain Authority, Page Rank Algorithm, Search Result, Page Ranking, Google Search Engine, Web Page, Keyword Stuffing, Update Google, Search Query, Exact Match, Web Site, Keyword Occurrence.

INTRODUCTION&OTHER HEADINGS

With the increase in the use of search engines, it becomes more important for web developers and SEO marketers to have their website score well in rankings in these various search engines in order to reach to their desired audiences. To optimize websites and webpages for search engines, such as Google or Bing, knowledge about the page ranking algorithms is essential. For instance, if a search engine considers how often search terms occur on a webs full text, web content should use the most relevant keywords in their web pages whenever possible to get their webpage in a top position.

For users of search engines, data regarding ranking algorithms is additionally essential so as to judge the hardiness of the results. As noticed, web site house owners do have Associate in nursing interest in having their web site displayed in high positions by search engines. Consequently, users ought to understand the algorithms so as to estimate the hardiness and so the trustiness of the search engines' results.

Understanding of any ranking algorithm will help web owners in estimating the utility of results in case of their search motive. For any instance, user trying to search for the latest match score should be showed result in preference to date while any user searching for weather forecast should be shown result based on nearby locations and not random locations. Similarly, any user searching for a product should get a result relevant to him, else the overall user experiences a search engine will get can be divesting for a company's value.

This paper is designed as follows. First, and briefing of common ranking algorithms of search engines. Then, the updates which were released time to time on order to improve the search result competition as well as user experience. And finally, all the results are described. In the conclusion section our observations and interpretation of the results is presented and an overview towards future research is suggested. [1]

RESEARCH OBJECTIVE

The research objective was to understand factors that are accounted by Google Page Ranking algorithm and how important is any factor and in how much weightage. The following research is for understanding only and didn't assure to give a final answer to any or all questions thoroughly, but is just a motivation to show in which direction an user should further go for search engine optimization.The following of the factors were analyzed:

(1)   Website's backlinks count

(2)   Website's age

(3)   Website's content and data relevancy

(4)   Website's traffic

(5)   Keyword occurrence on website.

(6)   Website's domain authority.

So, our target is here to understand how these factors matter a website and in what weightage they affect a website and its ranking. Since search engines work on basic modes, namely a page rank or a hypertext induced topic search (HITS). The research paper objective was to analyses whether Google search engine crawls text from any

figures, tables, videos or paragraphs used on any webpage.Google Page Rank was developed on a thesis that: A web page is important if it is pointed to by other important web pages. While, HITS work on a thesis that: A page is a good hub (and therefore deserves a high hub score) if it points to good authorities, and a page is a good authority if it is pointed to by good hubs[2]. The crawler plays a vital role in this process, as its prime role is to investigate all the information and recover it from the entire web. Crawler works in two phases, First, it finds a site or say discovery. Second, it investigates on the data available on site.To achieve more correct results for adrew in crawl, Smart Crawler positions locales to sort out extremely correlated site for a given subject. In the second stage, Smart Crawler achieves fast in-site arranging to unearth most noteworthy associations with an adaptable association situating.The original PageRank algorithm was described by Sergey Brin and Lawrence Page in several publications. It is given by

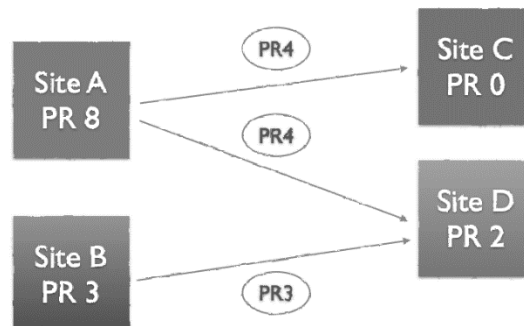PR(A) = (1-d) + d (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))

Where ,

PR(A) is the PageRank of page A,

PR(Ti) is the PageRank of pages Ti which connect to page A,

C(Ti) is the quantity of outbound connections on page Ti and

d is a damping factor which can be set in the vicinity of 0 and 1

Along these lines, above all else, we see that PageRank does not rank sites all in all, but rather is resolved for each page independently. Further, the PageRank of page is recursively characterized by the PageRanks of those pages which connect to page A.

The PageRank of pages Ti which connect to page A does not impact the PageRank of page A consistently. Inside the PageRank calculation, the PageRank of a page T is constantly weighted by the quantity of outbound connections C(T) on page T. This implies the more outbound connections a page T has, the less will page. An advantage from a connection to it on page T.

The weighted PageRank of pages Ti is then included. The result of this is an extra inbound connection for page A will dependably build page A's PageRank.

At long last, the total of the weighted PageRank's of all pages Ti is increased with a damping factor d which can be set in the vicinity of 0 and 1. Along these lines, the stretch out of PageRank advantage for a page by another page connecting to it is diminished.

In light of the measure of the real web, the Google web index utilizes an approximate, iterative calculation of PageRank esteems. This implies each page is allocated an

underlying beginning quality and the PageRanks of all pages are then ascertained in a few calculation hovers in view of the conditions controlled by the PageRank calculation.

So, in simple words. The calculations don't work if they're performed just the once. Correct values are obtained through several iterations. Suppose we've a pair of pages, A and B, that link to every alternative, and neither have the other links of any kind. Page Rank of A depends on Page Rank worth of B and Page Rank of B depends on Page Rank worth of A. we will not calculate A's Page Rank till we all know B's Page Rank, and that we cannot calculate B's Page Rank till we all know A's Page Rank. However, playing a lot of iterations will bring the values to such a stage wherever the Page Rank values don't amendment. Therefore, a lot of iterations are necessary whereas shrewd Page Ranks.

So, it was briefing about how the Page Rank algorithm works, after acquiring data from websites and calculating various backlinks on various sites, let us see how various factors affects and play a prominent role in this algorithm.

METHODOLOGY

A.   Impact of Backlinks Counts

Google regularly iterates the above algorithm to update its database and collect new data from various websites. Similarly, crawler regularly visit a website and updates and stores all the information it gets regarding the linking between various pages. To see the effect of backlinks or say in-bound links and out-bound links on various pages Search Engine Optimization (SEO), We created backlinks for 3 different websites, Website A with 100 links to various weak domain authority i.e. Page Ranked websites while Website B with only 50 links to a strong domain authority and similarly Website C with 100 links to a strong domain authority page.
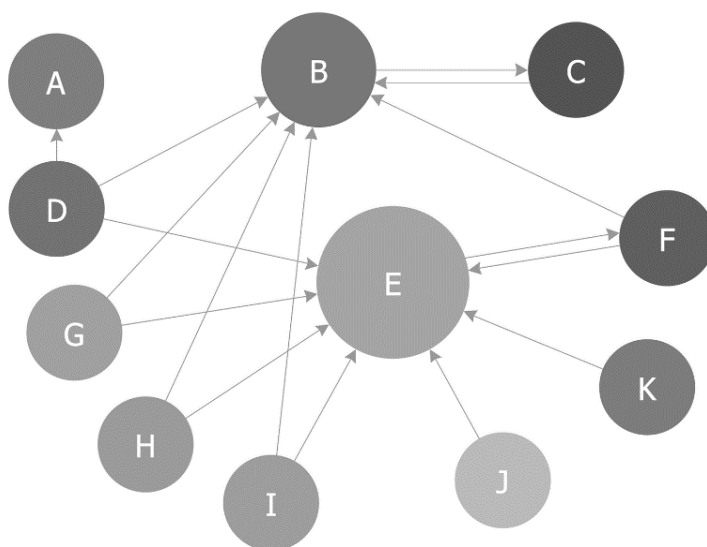
B.   Impact of Website's Age

To understand effect that a website's age has upon its ranking, we executed few search queries for same keyword. Even with different keywords but for same category search queries were executed to draw the relevancy and understand the importance of age on a website's page ranking on web.

C.   *Impact of Content and Data relevancy*

We wanted to know how a website showing content relevant to searched query and data on it affects Page Rank, whether a keyword must occur in a document or if Google considers synonyms as well. To accomplish this, the results of various search engines were observed, and search engines usually highlights the keywords that you have searched to show the relevancy between searched terms and showed results. On further reading, we found that Google released various updates which are mentioned in further section of this paper, that how it considers the relation between website type and content on it.

D.   *Impact of Traffic on Website*

It is a strange observation that first goal of SEO is to get traffic on your website, while traffic on your websites ultimately drives your Page rank and SEO. So, it's the goal which affects itself in a recursive manner. As number of clicks on a website are recorded and crawler visits that website more and improves it's ranking.

To observe the same, we searched various websites and observed its bounce rate and estimate visit through a tool called Similar Web, which helped in understanding the effect of traffic on any page authority.

E.   *Impact of Keyword Occurrence on Website*

To analyze whether or not the prevalence of a research term i.e. keyword in an exceedingly page's title has a control on its ranking, results of few search queries were analyzed. It absolutely was compared however usually the highest results contained the search term within the title and the way usually the last 10 results contained the search term within the title.

F.   *Impact of Domain Authority*

The plausible way to find out the role of Domain Authority on various Page's Rank was to find their result in search engine. Also, to clarify this observation, we used MozRank and Similar Web Tools to find the relevancy between both.

UPDATES IN PAGE RANK'S ALGORITHM

Google releases regular updates and modifies its search algorithm, sometimes these updates are announced officially while sometimes they are done confidentially. These modifications no doubt gives SEO experts a challenge but at the same time it improves the competition and helps a website to rank up. Basically, these updates and modifications are for improving user experience and shows more relevant data.

Although, ranking up is definitely difficult but surely it is not impossible, you just need to follow and update your website according to the factors that affect your website. There are hundreds of updates few of them were unconfirmed while few were official and are confirmed. But here we are not going to mention all the updates. Rather we will put lights on some important updates that actually affected the quality of websites and changed the SEO techniques.

1)   *Boston in February,2003*

It was a small but among the first update which was experienced by users. In this update, main concept was to introduced monthly updates but later that idea failed. Here few modifications in algorithm were made and indexing was improved.

2)   Florida in Novermber,2003

This update was made primarily to curb the problem of keyword stuffing. Many websites lost their ranking because of this update. SEO word received this first jerk through this update.

3)   Nofollow in January,2005

This update was released to combat the spams and unnecessary outbound links. Though there was not an algorithmic change but nofollow attribute was introduced for websites to control spammy big comments on websites to get unnecessary link.

4)   Panda in February,2011

It was one of the revolutionary update in Google's history. A major change in algorithm took place and it cracked down the sites with low content, keyword stuffing issues, duplicated, plagiarized or thin content and user spammed content. It worked as a filter, which assigned a "quality score" to web pages.

Google frequently rollouts Panda updates which has caused penalties and recoveries much faster. Also, these updates focus on No-content copy, too many outbound links are bad with keywords, no multiple pages with same keywords.

5)   Penguin in April,2012

Google moved the Penguin Update in April 2012 to get the sites that are spamming Google's indexed lists. This refresh is basically gone for diminishing hunt rankings of sites that disregard Google's Webmaster Guidelines and utilize dark cap SEO methods to falsely expand positioning of their sites, by acquiring or purchasing joins through some wrong practices.

Penguin regular release focuses on reducing practice of improper link building, links with over optimized anchor text and keyword stuffing. In case of Penalization, it reduces the rank of a page instead of penalizing entire website.

6)   Hummingbird in August,2013

Hummingbird enables Google to better translate look inquiries and give comes about that match searcher expectation (rather than the individual terms inside the question). While watchwords keep on being essential, Hummingbird algorithm makes it feasible for a webpage to score good and rank for an inquiry regardless of whether it does not carry the correct words the searcher has entered and this is accomplished with the assistance of normal dialect handling that depends on inactive semantic ordering, co-happening terms and equivalent words.

It considered some 200+ factors that can affect a websites' ranking and search result. It also focused on mobile marketing. Hummingbird was a term coined to describe "fast and precise", that it can focus on meaning of a phrase or a keyword rather than an individual keyword. Use of knowledge graph was found here for better understanding of meaning and ensuring better results.

7)   Pigeon in July, 2014

In the year 2014, Google shook everyone especially the local SEO world with a refresh that significantly adjusted some nearby outcomes and changed how they now handle and decipher area prompts. Google guaranteed that Pigeon made nearer ties between the nearby calculation and core algorithm(s). It affected poor on and Off Page SEO. Prime focus in Pigeon update was on giving valuable, exact and significant local search results.

Google said this new SEO calculation improves their separation and area following parameters in more outcome situated way. The progressions made through Google Pigeon Update will likewise influence list items appeared in Google Maps as this refresh gives Google a chance to give query items in view of the client area and posting within reach in the local index. The primary reason behind acquainting Google Pigeon Update is with give inclinations to local list items in SERPs and that is the reason this is to a great degree gainful for nearby organizations.

8) HTTPS/SSL Update in August, 2014

In this update Google announced that it will give more preference to secure sites. Focus was on showing users a secure website, although encryption and security of a website didn't affect its ranking severely. But this was a bell for all websites to take securing measures for their websites to maintain their rank in future. They emphasized that this boost would begin out tiny however silent it'd increase if the modified tried to be positive.

9) Pirate in August,2012

In this update Google started penalizing the sites that were violating copyrights. In October, 2014 after first DMCA/ Pirate update Google launched one more update that combat digital media and software privacy. This targeted number of sites which caused a drastic drop in rankings of various sites.

10) RankBrain in October, 2015

It was a part of Hummingbird Algorithm but after various modifications in algorithm this update was rolled out. Using machine learning, Google tried to understand the meaning behind search queries and show the best matched results. So, along with artificial intelligence Google harmed the ranking of sites that had poor and shallow content, poor UX and lacked query specific relevant content.

According to Google it is the third most important ranking factor. So, create more content on the focused keywords and links them from the high authority pages is the best way to stand out in the competition.

11) Mobile Friendly Update in April, 2015

As the name mentions the update was made to boost mobile friendly pages. As 63% of world is using mobile phones today, while more than billion mobile subscribers are there in India only, so most of the users when in query will fire it via their phone on internet only. So, it was necessary to make websites mobile-friendly to improve user experience. Google's Mobile Update (otherwise known as Mobilegeddon) guarantees that versatile inviting pages rank at the highest point of portable hunt, while pages not upgraded for portable are sifted through from the SERPs or genuinely down-positioned.

12) Possum in September, 2016

It ensured that the results which a user gets are more specific to user's location and also varies more according to user's location. The simple concept was to show local business' address that are closer to user than the far one. At the same time, it also boosted business that were far from physical area of any city.

13) Fred in March, 2017

This latest update targeted the sites which were only ad-centred and focused only on making money through ads and having low content quality. No doubt earning through blog is legit in Google webmaster guidelines as long as the content is of high-quality and are relevant.

RESULTS & INTERPRETATION

A.   Impact of Backlinks Count

Since the basic algorithm and concept of Page Rank was based on links, so it is sure that it holds a major part in ranking of a website. But how much that is the question? So, based on the search queries executed, after 2 months the result was not so surprising. Website C scored on top in SEO, while Website B and Website A showed mixed result. From this result we can easily make out that the count and as well as quality i.e. DA of linked website matters equally in our Website Page Rank.

B.   Impact of Website's Age

After execution of search queries, we can observe that age played a little role in page ranking. Never minded few old websites scored on top while our website which was still young outperformed than various other websites as clearly our backlinks were of pretty good quality. Age will help a website if its content with time are improved and increased in relevancy of website's category.

C.   Impact of Content and Data Relevancy

After reading about various updates and executing search queries, it was easy to draw that content quality plays a major role on a page's ranking. As in almost every update Google has focused on content type and user experience. So, in order to score top smart stuffing of keywords along with good, relevant and high quality of content and data on website is required.

D.   Impact of Traffic

More number of visit and traffic on website will attract crawler to parse our websites more often and regular update in content will help our ranking. Traffic will also improve the reliability of a website which again will help in page ranking.

E.   Impact of Keyword Occurrence on Website

Keyword occurrence on websites is important especially in H1, H2, Meta data, etc. But obviously it's over usage and stuffing will attract penalty from Google. So, it's always better to user smart usage of keywords on our website as it is an important factor for page ranking as well as SEO.

F.   Impact of Domain Authority

As a result of higher domain authority, many pages scored up in search results. So, clearly pages do get a benefit from domain authority but one must never forget that its page quality that affects a domain authority so both are dependent on each other.

CONCLUSION AND OUTLOOK

A website's keyword count has significant impact on its own page score and ranking in Google search engine. That would simply mean that Google is more suitable when searching for particular keyword but in case of not getting exact match it also shows the relevant data. Thanks to RankBrain algorithm, which uses machine learning and knowledge graph helps in enhancing users experience in getting result. This is neither good nor bad, as users might get the exact match keyword result on later pages while in same case he might get a relevant data on top irrespective of exact match. So, experience will vary majorly from user to user. Along with keywords, there are many factors that Google takes care while calculating Page rank of a website, prominent in it is content of a website. So, data relevant to page category will surely boost its ranking while irrelevant

content might affect a website's rank negatively. At the same time, copyright issues and plagiarism are also important factor for ranking. A good and unique title is very important for a page anyway, however the additional Google is employed, and the additional an internet site owner ought to believe page's title well. Beside keyword counts, the title appears to be one amongst the foremost vital factors for the ranking rule. Therefore, it's even as vital that a title makes the reader curious and includes the foremost relevant keywords to realize high positions in Google. Since Google updates now uses machine learning so Google may consider broad match or relevant meanings, so website must be carefully made using the thinking of users what they search for. Otherwise, scoring up a page's rank will be difficult. In addition, web owners ought to consider carefully concerning the word they use in their web site as a result of this may decide whether or not their web site pages square measure found by Google users or not. Regularly updating blogs on website with appropriate usage of keywords is a good practice for improving rank as well as domain authority. Not to forget, there are many more factors which also affects a website's page ranking while some of the factors fall under the sub category of above mentioned categories, while some plays little role in ranking. Factors like submission of Sitemap, Site structure, Site up time, Server location, Users review are few of the factors not mentioned above but are also equally important in ranking.

Overall, this review and study may raise more number of questions than it has answered. We highlighted the various factors, which now must be considered before and after the creation of websites. Implementation of Google AdWords and Google Analytics is also very much beneficial for a website's growth.

SUMMARY

Web search engines play a vital role in searching for anything or any data. To maximize their effectiveness, website owners and users should remember how search results are ranked on any portal. Most of the websites now regularly update themselves to maintain their rank. No doubt algorithms are hidden but still understanding the concept behind them is important. In each category, there are some concepts which are hidden and implemented in algorithms, understanding them can help a website's performance as well as boost its ranking in very positive manner.

Google Search Engine is one of the largest search engine used by people in world, though this paper was focused on Google search engine, but these factors are more or less same and affects in similar manner in search results of other search engines like Baidu, Bing, Duck Duck Go, etc. Not to forget that the algorithm is not known as a result, user don't know to what possibility Google can satisfy any users query.

We performed various steps to understand the Google's page ranking algorithm by reverse-engineering and came up with the following results:

1.   Overall, Google's page ranking algorithm depends heavily on a link inbound and outbound count. As a result, in order to improve a page's or a website's ranking it's important to create backlinks on relevant side just to avoid penalty from Google,

2.   Google's page ranking algorithm gives a lot of importance on words in the title i.e. the keywords used on and off page. Knowing this, website owners should think twice about any title they would give to their pages. All the relevant and less competitive keywords must be considered first and used in contents and it is sensible if you choose a long keyword title. Unnecessary stuffing of keywords should be avoided else it will attract a penalty.

3.	Google Scholar considers keywords more irrespective of their synonyms so wise use of words in content is benefitable. Search results are shown on basis of broad match, phrase match and exact match.

4.	It seems like Google does not put a weighting on the recurrence in which seek terms happen in the full content after panda refresh was take off. This implies a site won't be positioned higher for a specific pursuit in light of the fact that the hunt term happens all the more regularly in the full content. Along these lines, it could be helpful for web proprietors to refrain from a strict wording in their web substance and utilize more equivalent words. This would make their records less lucid yet more retrievable in Google. Not to overlook the proper measure of watchword event is as yet vital for any page's positioning.

5.	Google weighs the age of any page in very less amount, so if your page contents high quality of article, it can come up in ranking.

6.	Google crawls keywords from images caption, alt text and description, so keyword stuffing can be done their as it will help crawler to fetch keywords which we are targeting and at the same time will not harm the quality of website content.

7.	Google also considers the social commenting, no doubt its's no follow feature can prevent backlinks outbound but users interaction on a page in form of reviews and blogs can help with traffic interaction as well as updating content on a regular interval.

8.	Google's recent updates now consider mobile responsiveness also so while developing any website, developers must ensure that their websites are mobile responsive and also takes less load time. So, along with content quality, image optimization is also important factor to keep in mind while developing a website. These features can help in getting good results from SEO.

9.	Submission of site map is also important for SEO, as it helps anyone visiting site to go through whole structure of website and get traffic on all pages, over all affecting a site's domain authority. Also, it helps crawlers to fetch and retrieve data.

10.	Creation of contact page, and geo tagging of any company website will help in local SEO, as now Google updates also considers the location as an important factor while calculation a rank of any page. So, for that any company that wants to show its result based on location should use Google Business.

The study and attempt of this research carried out can only be seen as a first step. Many of the questions have stayed unanswered. Further research in future is required, with a lot more of sample data and more study and analysis in order to get a complete picture of Google's page ranking algorithm.Still, many answers and factors that affect page ranking were established. So, diversity in links, quality of links and content for both on page and off page and Domain factors must be considered first while performing SEO for any website.

REFERENCES

1) JöranBeel and Bela Gipp. "Google Scholar's Ranking Algorithm: An Introductory Overview" Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09).
2) Beel, J. (2009). Google scholar's ranking algorithm: An introductive overview (research in progress. In in Proceedings of 3rd International Conference on Research Challenges in Information Science (RCIS'09). IEEE.
3) Bhor, S. R., Dumbre, S. B., Bakare, S. A., &Raut, M. V. (2018). Intelligent Crawler for Deep Web Searching Using Page Rank Mechanism. International Journal of Current Trends in Science and Technology, 8(1), 23-28.
4) Muenchen, R. A. (2012). The popularity of data analysis software. UR L http://r4stats. com/popularity.
5) McGill, M. (1979). An Evaluation of Factors Affecting Document Ranking by Information Retrieval Systems.
6) Page, L., Brin, S., Motwani, R., &Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.

BIBLOGRAPHY

1) http://www.cs.sjsu.edu/faculty/pollett/masters/Semesters/Fall11/tanmayee/Deliverable3.pdf
2) https://www.geeksforgeeks.org/page-rank-algorithm-implementation/
3) http://pr.efactory.de/e-pagerank-algorithm.shtml
4) http://rapidwebseo.com/matt-cutts-does-domain-age-really-matter.php
5) https://backlinko.com/google-ranking-factors
6) https://unamo.com/blog/seo/30-important-google-ranking-factors-beginner-know
7) https://moz.com/google-algorithm-change
8) https://searchengineland.com/8-major-google-algorithm-updates-explained-282627
9) https://www.digitalvidya.com/blog/google-seo-updates/
10) https://www.edupristine.com/blog/google-algorithm-updates
11) https://karmasnack.com/how-bing-works/