



4D INTERNATIONAL JOURNAL OF MANAGEMENT AND SCIENCE

ISSN-2250-0669

www.4dinternationaljournal.com

Volume-7, Issue-1-2016

Page Rank And Trust Rank Algorithms of Search Engine

*** Dr.Dora Norman**Dr. Prakash Millind**

Page rank is still the most revealing & critical metric that governs a domain's ability to rank. But since 2013 Google have chosen not to update & share this information any more in an attempt to make their search results harder to manipulate. PR & Algorithm updates always happen in real time. When Toolbar PR used to be updated, and a domain went from a PR2 to a PR3, the sites real PR didn't update overnight, only the Tool Bar Page rank did. The site was most probably already a PR3 two months before the Toolbar PR update. A domain's real PR increases within a 48hr period of Google crawling its latest inbound links. These incremental increases are in ratio to the quality and volume of new inbound links.

Key words: Page rank, Search result, Inbound links,

***/ **Asst Prof. & associate Prof.--Singapore Management College.**

Introduction

The most important reason why you want to get first page rankings is that you want to be visible to everyone that will need your product/service. Google, for example, is the #1 search engine and it provides for almost 12 billion searches a month. Approximately 1.17 billion unique searchers use it every month.

It's the number 1 search engine used in the US. It gets 67.5% of the US search market, and a whopping 87.1% of the mobile search market.

And if you are in business, you should also know that 93% of all buying decisions start with an online search. And the key to receiving traffic through Google is to gain first page rankings, because first page websites get 91.5% of Google traffic. Most people just don't really bother going beyond the first page. If your website's on the second page you only get to share 4.8% of the traffic along with all other websites on the page. With page 3, it shrinks lower to 1.1%

Since different search engines weigh different parts of their algorithms differently, they can all have similar relevancy while having significantly different search results. Some people are quick to ask how their sites are being penalized in one search engine because they rank well in others. While many sites do get penalized for aggressive spamming techniques, most sites that do not rank well usually just fail to satisfy the specific ranking algorithm for that specific engine in question.

- Google is primarily focused on site age and link based authority. They are much better than their competitors at determining which links are legitimate editorial citations.
- Yahoo! focuses a bit more on on-the-page content while still relying a decent amount on link popularity.
- MSN is like Yahoo!, but even easier to manipulate with low-quality links, aggressive anchor text, and keyword prominence.
- Teoma or Ask is focused on local communities. It is far easier to spend a good bit of time learning how to get links versus constantly tweaking pages over and over again. I do not worry much about things like keyword density. I know that Yahoo! usually likes content to be a bit more keyword rich and Google (which is spending millions of dollars to scan books) usually prefers more natural reading content. While the on-page criteria can vary from engine to engine, the ability of off-page factors to raise your rankings is limitless. Different search engines will evaluate different links in different ways. In general, they all like links pointing to your website, especially links from authoritative sites related to your field.

Thumbshots has a cool ranking tool that shows you how sites rank in different search engines, so you can compare how well different algorithms overlap one another. Due to heavy load, this tool is not available all the time.

MyriadSearch.com is another site that makes it easy to see the overlap between the major search engines.

ANALYSIS

Page Rank and Trust Rank Algorithms:-Google determines rankings of its search result listings using PageRank and TrustRank algorithms. It is important to understand these algorithms since the higher one's website ranks in search engine results, higher the potential to gain more targeted visitors. PageRank: The rank of a webpage in organic search results of Google is determined by PageRank.

$$PR(A)=(1-d) + d[PR(T1)/C(T1) + \dots + PR(Tn)/ C(Tn)]$$

where

PR(A) is Page Rank of web page A ,T1...Tn are web pages that point to page A

d is damping factor which can be set between 0 and 1. It is usually set to 0.85

C(A) are the number of links going out from web page A

PR(A) is based on the concept that a random surfer who is given a web page A keeps clicking on links at random until he gets bored. The surfer never hits the back button. On getting bored, the random surfer requests a random web page. The probability that a surfer visits a page A is PR(A). The damping factor d is the probability that at each page, the surfer gets bored and requests another random web page. A variation that is added to the PageRank calculation is that different damping factors may be assigned different pages T1...Tn which link to page A.

One can conclude from the Page Rank equation that:

1. The more inbound links a web page has, the higher the Page Rank
2. It is better to have inbound links from a web page that has high Page Rank and few out links over a webpage with high Page Rank and too many out links.

e.g. PR(X) = 4 and C(X) = 5 then $d[PR(X)/C(X)] = 0.85d$
 PR(Y) = 8 and C(Y) = 100 then $d[PR(Y)/C(Y)] = 0.085d$

PageRank forms a probability distribution over web pages, so the sum of all web pages' PageRanks will be 1. $PR(A)$ can be calculated using an iterative algorithm, and corresponds to principal eigenvector of the normalized link matrix of the web.

$$PR(A_1) + PR(A_2) + PR(A_3) + \dots + PR(A_n) = 1$$

$PR(A) = (1-d)$ if web page A has no inbound links.

There are hundreds of web pages added to the World Wide Web every moment. Since sum of PageRank of all web pages over the WWW is a constant i.e. 1, this means that as more pages are added to the WWW, PageRank of each web page gets constantly updated to accommodate the PageRank of new web pages'. Assume that, if a web page has no inbound links, $(1-d) \approx 0$. As inbound links increase the PageRank of a webpage, one can conclude that outbound links decrease the PageRank of a web page. This decrease in PageRank of a webpage due to outbound links is called PageRank Leak.

To ensure a high PageRank it is necessary that:

1. A web page should have high number of inbound links
2. A web page should have low number of outbound links

The Page Rank algorithm determines the importance of a web site by counting the number of inbound links. This concept can be manipulated by artificially inflating the number of inbound links to a web page. Page Rank also does not incorporate the quality of the web page in its calculations. Hence Google is developing the Trust Rank algorithm and has registered the trademark for Trust Rank on March 16, 2005.

Trust Rank: According to Gyongyi, Garcia-Molina and Pederson, the proposed algorithms for Trust Rank rely on the Page Rank algorithm. This algorithm takes into account, not only the inbound links to a web page but also the quality of the web page. To determine the quality of a web page, a panel of human experts will identify a set of reputable web pages that will act as the seed

for the spider. This algorithm is based on an empirical observation that: good pages seldom point to bad ones.

Thus it can be conclude that a web page can achieve higher Trust Rank if:

1. Reputable (good) web pages link to the web page
2. The web page does not link to any bad web pages
3. The web page does not mislead the search engine or employ search engine spam **Overlap analysis:-**A study conducted by Dogpile.com in collaboration with the University of Pittsburg and Pennsylvania State University in April 2005 and July 2005 reveals that only 1.1% of 485,460 first page search results were the same across Google, Yahoo!, MSN Search and Ask Jeeves. The study of search engine results for a given keyword over different search engines at the same time is termed as Overlap analysis and forms the basis of Meta search engines like Dogpile.com. Meta search engines send search queries to popular search engines and their results are displayed together on a single page. Since Google, Yahoo! Search and MSN Search are significant in terms of percent share of search queries answered, it is important to optimize the web pages to achieve top rankings in all three search engines.

Figure is a snapshot of overlap analysis performed on Google, Yahoo! Search and MSN Search conducted on August 25, 2005 at 18.50 EST for the keyword “free image library” and URL pattern “imageblowout.com”.

The screenshot displays the Googlerankings.com interface. At the top, there is a logo with a crown and the text 'GOOGLE RANKINGS'. Below the logo, it says 'Check the ranking of a web page in Google™ A free search engine optimization tool'. There are two columns of text: 'Search Engine Submission' and 'Top Search Engine Ranking'. Below this, it shows the keyword 'free image library' and the URL pattern 'imageblowout.com'. The search results are listed for Google, Yahoo!, and MSN. Google shows 'Position 556', Yahoo! shows 'Position: 8', and MSN shows 'Position: 13'. At the bottom, it says 'Time: 49.37 seconds' and 'URL PATTERN: IMAGEBLOWOUT.COM KEYWORD: "free image library" DATETIME: 18:50 EST SUNDAY, AUGUST 25, 2005'.

Figure Overlap analysis of imageblowout.com - Googlerankings.com snapshot

1. Yahoo Search displayed “imageblowout.com” on the page# 1 of search results
2. MSN Search displayed “imageblowout.com” on the page# 2 of search results
3. Google displayed “imageblowout.com” on page# 55 of search results

The reason for the substantial difference in the ranking between Google search results and Yahoo! Search and MSN Search results is due to proprietary relevancy algorithms used by these search engines. Yahoo! Search and MSN Search use content based relevancy algorithms. The title tag of imageblowout.com is “Image blow out – Free Image Library for Commercial Use”. This is an exact match to the keyword used to perform the search query, hence the higher rankings in Yahoo! Search and MSN Search.

CONCLUSION

One can conclude from the Page Rank equation that:

1. The more inbound links a web page has, the higher the Page Rank
2. It is better to have inbound links from a web page that has high Page Rank and few out links over a webpage with high Page Rank and too many out links.

Thus it can be conclude that a web page can achieve higher Trust Rank if:

3. Reputable (good) web pages link to the web page
4. The web page does not link to any bad web pages
5. The web page does not mislead the search engine or employ search engine spam

REFERENCES

- Brin S.(1998),’ The anatomy of a large scale hypertextual web search engine’ from www.seo2webdesign.com
- Poaolo Boldi,Massimo Santini & Sebatiano Vigna (1999),’ Pagerank as a function of damping factor’ from DSI University Study & www.mattcutts.com/blog/seeing-nofollow-links
- Gyongyi,Zoltan ,Hector Garcia-Molina, Jan Pedersen(1999),’ Combating webspam with trustrank’, from Stanford University & www.dmoz.org
- Robin nobles, susan o’neil (2000), ‘ Maximize web site traffic: build web site traffic fat and free by optimizing search engine placement’.From Wrox Publications, www.wrox.com/seopage

- Moses S. Charikar (2002), ‘ Similarity Estimation Techniques for rounding algorithms’ from ACM & www.dmoz.org
- Susan Sweeney(2002),’101 Ways to Promote Your Web Site : Filled with Proven Internet Marketing Tips ,Tools ,Techniques and Resources to Increase Your Web Site Traffic’.from www.google.com/support/webmasters
- Peter Kent (2004), ‘Search engine optimization for dummies’, from www.seo2webdesign.com
- David Viney(2005), ‘The SEO Expert Guide: How to get to the top of the Search Engine Rankings and stay there. From Tata MacHill Publications.
- Brian Moloney(Jan 2005),’Search engine optimization for the layperson’ from www.imagescape.com & publishers Imaginary Landscape, LLC.
- Gyongyi, Berkhin,Garcia-Molina & Pedersen(Oct 2005),’ Link spam detection based on mass estimation ‘ a technical report at Stanford University.
- Yi-Min Wang,Ming Ma,Yuan Niu,Hoo Chen (Oct 2006),’Spam Double-Funnel: connecting web spammers with advertisers’ from www.research.microsoft.com/saerchhanger
- Pat Bensky(2007), ‘5 simple steps to improve our website’s search engine ranking’ from www.seoresearchlabs.com
- Burns, Enid (2007),’Search engine results pages turn into destinations’ from www.clickz.com
- Joel Tachau(June 2007), ‘Analysis of three personalized search tools in relation to information search, publishers Information Architects.
- James Bake (2007),’ Search engine optimization : A road map to better listings’ from www.google.com/support/webmasters & www.sitemaps.org
- Bo Xing & Zhangxi Lin(2006),’The impact of SEO on online advertising market’, from ACM International Conference Proceedings Series,Vol. 156 & www.iamai.in
- Dan Thies(2007) ,’How to prosper with the new google, from www.seoresearchlabs.com[
- Arketi (March 2008),’Web watch survey, from center for media research & www.centerformediaresearch.com
- Eisenberg,Bryan (2008),’How to prioritize your destinations’ from www.clickz.com[19]
- Rainer Olbrich & Carsten D. Schultz(2008) ,’Search engine marketing & click fraud’ from www.fernuni-hegen.de
- Rajeev Misra (June 2009),’ White paper SEO’ from www.infoproworldwide.com & publisher is Infopro Worldwide.
- James Bake (2009),’Search engine optimization :Background knowledge & basic tips’, from Agriya Infoway Pvt. Ltd. & www.agriya.com
- Lee Underwood (1999) , ‘A Brief History of Search Engines’ from www.webreference.com.[23]
- Witten, Moffat and Bell, “ Managing Gigabytes”, 2nd edition, Morgan Kaufmann, 1999.
- Spink, Wolfram, Jansen and Saracevic, ”Searching the web: The Public and their Queries”.

- “Journal of the American Society for Information Science”, (2001) Queries are available from <http://www.mds.rmit.edu.au/hugh/queries>.
- Williams and Zobel, ” Compressing Integers for fast file access “, The Computer Journal (1999).