

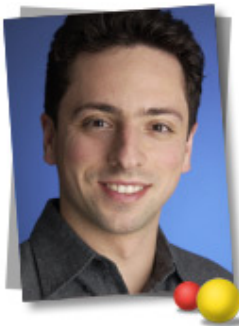
The importance of PageRank to Search Engine Optimisation

PageRank is a score out of 10. It is an indication of the relative importance of a webpage based on the number of other 'important' web pages that link to it. PageRank is one of the two factors identified by Google founders Brin and Page in 1998 as essential in improving the quality of search engines. It is one of the cornerstones that originally made Google such an effective search engine. Google has since evolved although PageRank still forms part of the complex algorithm used by Google today.

The problem with search engines

In their seminal paper on search¹, Brin and Page identified the following problems with search engines:

1. High quality human maintained engines such as Yahoo. These were subjective based on the reviewers own opinion. They were expensive to maintain due to the amount of resource needed and therefore slow to add new material. Finally they couldn't cover esoteric subjective well.
2. The automated search engines also had problems in that they utilized simple keyword matching techniques which returned too many low quality matches. In addition there were an increasing number of advertisers trying to mislead the search engines by stuffing keywords into pages.



Vision for Google

Brin and Page wanted to improve search quality, to make it easy to find anything on the web. Whilst they identified the fast growing use of search engines, very little academic research had been done in this field. They applied academic principles to address the issues highlighted above.

The importance of PageRank

They identified two key features that would enable the delivery of very high precision results in a search engine, one of these is PageRank. PageRank provides an indication of quality of any page by utilising the link structure of the web. Brin and Page mapped the number of incoming links of any given page. But they didn't just count the number of links, they gave each one a quality indicator as well, this means that a few high quality links can be more important than many low quality links.

The rationale for this is they felt the more people that were linking to a page the more important it is. Links from high quality sites are more important than low quality sites.

Understanding the PageRank formula

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

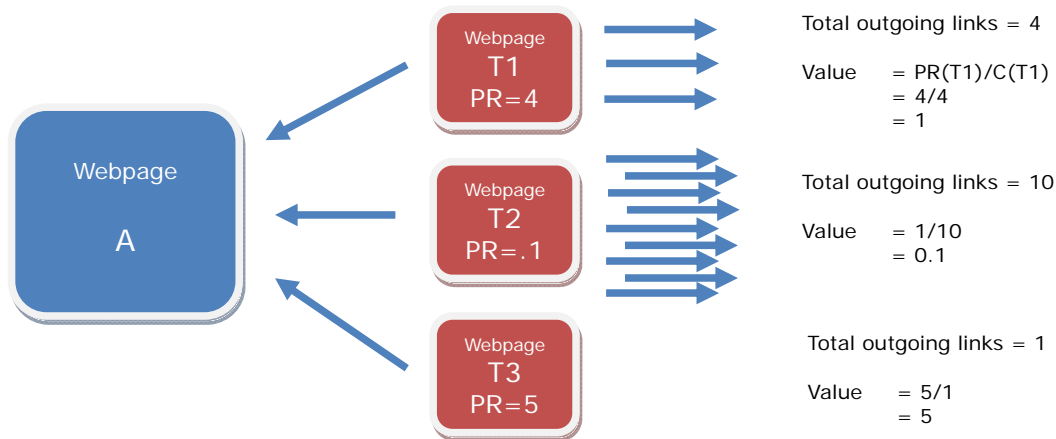
PR = PageRank

A = Web page in question

d = Boredom factor. An indication of how long a browser stays on the page for. This is given a default starting value of 0.85 and then adjusted from there.

T1... Tn = The pages that link to page A.

C = The number of links coming out of the T page.



Therefore:

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

$$PR(A) = (0.15) + 0.85(1 + 0.1 + 5)$$

$$PR(A) = 5.3$$

What is most interesting about this is the importance of quality links. One link from a page like T3 is worth 50 links from pages like T2.

Is PageRank still valid?

Considering this was created in 1998, a long time in search engine development, how important is PageRank today? Well, an important paper² published in 1999 by Bharat* and Mihailaspoke described the shortcomings of PageRank. They wrote that PageRank is a 'query independent' indicator so that it cannot distinguish between a page that is authoritative in general or one that is authoritative on the query (or search phrase) in question.

So for example a page might cover several topics and have a high PageRank, yet of all the topics it covers it is only really renowned for one of them. PageRank will not distinguish between the topics, it will simply indicate this is an authoritative page.

The authors outlined an alternative approach to PageRank, they called it Hilltop. Hilltop identifies expert pages for specific queries and keeps a record of these 'authority' pages on the web. A page is classed an authority if many other authority pages link to it, of course the links must be related to the particular topic. In tests they found Hilltop delivered high precision results than Google and other popular commercial search engines. It is now commonly considered that the Hilltop approach is used by Google in addition to PageRank (as well as other factors).

Notes:

- The expert pages linking to the target page must be non-affiliated so they should not come from the same company. A set of rules exists to identify affiliate pages.
- Expert pages are selected using a calculation based on incoming links from other pages where the text around the link is related to the query key phrase.



* Just after writing this paper Krishna Bharat became a principle scientist at Google working in the area of Web Information Retrieval.

What Google Say

Google still talk about PageRank as being a key part of their technology. *"PageRank relies on the uniquely democratic nature of the web by using its vast link structure as an indicator of an individual page's value. In essence, Google interprets a link from page A to page B as a vote, by page A, for page B. But, Google looks at considerably more than the sheer volume of votes, or links a page receives; for example, it also analyzes the page that casts the vote. Votes cast by pages that are themselves "important" weigh more heavily and help to make other pages "important." Using these and other factors, Google provides its views on pages' relative importance. Of course, important pages mean nothing to you if they don't match your query. So, Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines dozens of aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query."*

Summary

Google PageRank revolutionized search engines. However it does have a major flaw, it delivers an absolute measure of importance, not one related to the search phrase. PageRank has been refined to assess the authority of the linking page for a given search phrase, an approach taken from the Hilltop paper. PageRank is still important but it has to be considered alongside other very important factors. Understanding this is fundamental to high quality SEO.

When searching on Google, many websites with a low PageRank are listed above pages with a higher PageRank, this demonstrates that Google rely on other factors as well. The best advice is to focus on building a well structured website with good content to attract and retain visitors. If this is achieved a reasonable PageRank is likely to follow.

References

¹ *The Anatomy of a Large-Scale Hypertextual Web Search Engine* by Sergey Brin and Lawrence Page Computer Science Department, Stanford University, Stanford, CA 94305, USA.

² *Hilltop: A Search Engine based on Expert Documents* By Bharat and Mihaila