****

**GOOGLE – HOW IT WORKS?**

Ever since, we first came across the World Wide Web, we have been wondering how a search engine works and how it displays the required search results at a lightning speed. First of all, we need to know, what a search engine is! It can be easily defined. It’s the world wide server which searches the entire web and displays out the required results. But, there are many mechanisms involved in it.

Many algorithms and programmes are designed to get the results instantly.

So, what are these phenomena? And what all algorithms are involved in the GOOGLE search engine?

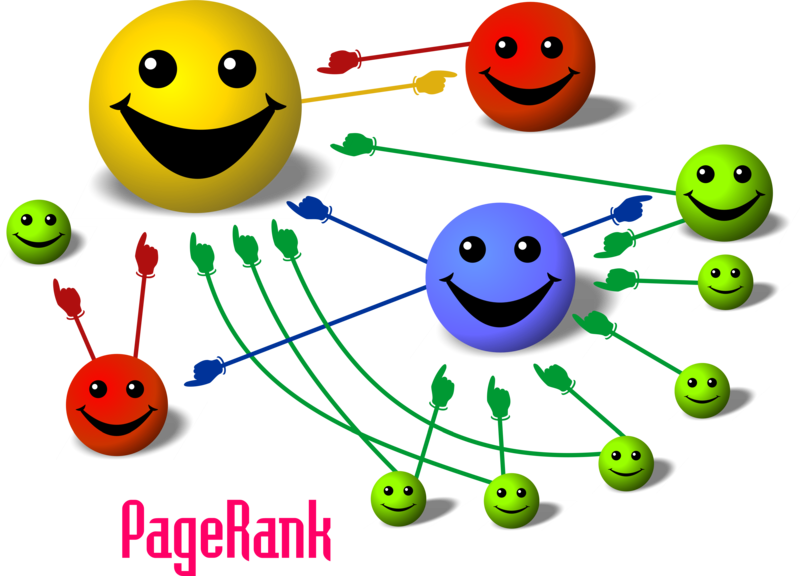
Let’s see what Larry Page and Sergey Brin explain about their simplified and efficient algorithm in this article.

But before that, let’s talk about the history of the company.

Google was first established as a search engine in 1996**.\*\*most of the sites say 1998\*\*** by **Larry page and Sergey Brin**. They both attended **Stanford University**. They want to step into the field of Web based search engines and make a stand in it. Their main aim was to create a web based search engine which should optimise and filter the results according to relevance with great speed.

All their main idea was of Larry page .who proposed an Algorithm Called **‘Page’ Rank. \*\*this is not given in any site that page refers to his name “page”… plz chk whr he got the info from. It should not b false naa\*\***

**PAGE RANK ALGORITHM:-**



**Page rank is the numeric value of 1-10 assigned by GOOGLE. It defines a page relevantly i.e. how important it is in the World Wide Web**.

When one page is linked to another page, it effectively casts a vote for that page .Google uses this to find the most voted page as their top search result. Page ranking is shortly called as PR.

**First of all let’s discuss some facts about page ranking.**

**\*Pages with higher PR appear top in the search engines.**

**\*PR is updated every week.**

**\*More the PageRank value, more relevant the page is.**

**\*According to Google, page rank is equally distributed among the links of the web page.**

**\*Page ranking is generally a probability distribution of a person randomly clicking on links will arrive at any page \*\*DOESN’T MAKE SENSE\*\***

Page rank requires many phases of iterations to find the exact page what we want.

First, let’s consider 4 Web pages A, B, C, D. These pages have many pages linked to it. Assume that there are no outgoing links from this page to any other pages.

First, the same page rank is initialised to all pages A, B, C, D. So the sum of the page ranks of all these 4 pages gives us the total number of webpages linked to these pages. Let each page in this example have a page rank of 1.

As we all know that is the person clicking the webpage (probability) will be between 0 and 1.

Next, these pages have some outbound links. It means that, these pages have some linkage among them like, from **A->B and A->C or A->D, D->B etc.**

If the only links in the system were from pages **B**, **C**, and **D** to **A**, each link would transfer 0.25 probability of the PageRank to **A** for a total of 0.75 in the 1st  iteration.

PR(A)= PR(B) + PR(C) + PR(D).\,

**PR (A) = 0.25 + 0.25 + 0.25 = 0.75**

**Outgoing linkages**

**0.25**

**0.25**

**0.25**

Suppose that page **B** has a link to pages **C** and **A**, while page **D** has links to all three pages. Thus, upon the next iteration, page **B** would transfer half of its existing value, or 0.125, to page **A** and the other half, or 0.125, to page **C**. Since **D** had three outbound links, it would transfer one third of its existing value, or approximately 0.083, to A.

**0.083**

**0.083**

**0.083**

**0.125**

**0.125**

This gives the formula as

PR(A)= \frac{PR(B)}{2}+ \frac{PR(C)}{1}+ \frac{PR(D)}{3}.\,

So, we can generalise this algorithm and formulate an equation which looks like this.

**PR(A)= \frac{PR(B)}{L(B)}+ \frac{PR(C)}{L(C)}+ \frac{PR(D)}{L(D)}. \,**

PR (A), PR (B), PR (C), PR (D) are the page ranks of the pages A, B, C, D.

L (A), L (B), L (C) are the number of out bounding links among the pages.

**\*\*grammar chking ke according it should be “IS” not “ARE”… but it doesn’t sound right.. so its up to u what to do \*\***

**In the general case, the PageRank value for any page µ can be expressed as:**

**PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}**

So this defines the most simplified general algorithm for page ranking.

Today, Google, by estimate runs over one million servers in data centres around the world. Using this simple Page Rank Algorithm, Google can process over one billion search requests and about twenty-four petabytes of user-generated data every day. Not sure how a big petabyte is? Well, here’s a picture to explain.

* **Vamshi Reddy( 2nd year Computer science)**