

PageRank

Content

1. What is it?
2. Applications
3. Facts about Page Rank algorithm
4. Application: Calculating PageRank for a Real Website Using the URL

What is it?

PageRank is an algorithm used primarily for rating the popularity of web pages. Although one may initially think 'Page' refers to the web page, it actually refers to the inventor, Larry Page (a founder of Google). This is why you'll find the capital 'P' wherever you find a reference to the algorithm.

The basic assumption is that the more inbound links a web page has across the web, the more valid the content the page contains. It's a crowd-sourcing algorithm of sorts; it relies on everybody online to create a network of links to different web pages, and classifies the validity of each page based on its overall popularity.

Applications

The most obvious application of PageRank is the Google search engine. In fact, much of the company's initial success may be attributed to the effectiveness of PageRank in organizing search results on Google.

It is important to note, however, that it is not only web pages which may use PageRank. Any data which can be modeled as a directional graph can be analyzed with PageRank.

Facts about Page Rank algorithm

Consider a directional graph, where web pages are modeled as nodes, and the links are modeled as vertices. Each vertex represents a link from one page to another.

While it may be tempting to simply count the total number of links pointing to each web page to discover which page is the most popular, PageRank is a bit more sophisticated.

We begin by assigning each node a score between zero and one. This score represents a probability distribution.

Application: Calculating PageRank for a Real Website Using the URL

The PageRank algorithm for the web calculates the PageRank based on the internal links within the input domain. It does not consider links from outside this source. Because this algorithm must complete a traversal of the site, the first run of the algorithm may take a few moments to complete.

You'll also see that the results are ranked starting with the URL with the highest PageRank, followed by the rest of the URLs in descending order.