

Page Ranking

Nathaniel Everett
CS432





What exactly is page ranking?

- First off, consider what a search query is. When you make a search query, the engine tries to return results of the highest quality.
- PageRank was actually invented by Google's founders, Larry Page and Sergey Brin.
- PR evaluates both the quality and quantity of links to a webpage, determining relativity and score on a scale of 0 to 10.



TF rankings and IDF rankings

TF - term frequency, how often a term appears in a document.

IDF - inverse document frequency, how important a term is, weigh down frequency of terms while scaling up rare ones.

TF calculation: $TF(t) = (\text{\# of times a term } t \text{ is in a document}) / (\text{total number of terms in a document})$

IDF calculation: $IDF(t) = \log_e(\text{total number of documents}) / (\text{number of documents with term } t \text{ in it})$



Example problem

A document has 500 words with the word *faucet* appearing 12 times. The term frequency for *faucet* is $TF(\text{faucet}) = 12/500 = 0.024$

Say that there are 1000 documents and *faucet* appears in 150 of these. The inverse document frequency is $IDF(\text{faucet}) = \log(1000/150) \approx 0.824$



TF-IDF

This term is simply the multiplication of both the TF and the IDF!

From the previous example: $TF(\text{faucet}) * IDF(\text{faucet}) = 0.024 * 0.824 = 0.00198$

Intended to reflect how important a word is to a document, collection, or corpus, used for text mining, information retrieval, and user modeling. It can also be used for stop-word filtering for fields such as text summarization and classification.



Other page ranking methods

Google PageRank

Alexa



Google PageRank

Named after Google founder Larry Page.

Works by counting the number and quality of links to a page to determine the estimate of how important a website is.

Algorithm: $PR(A) = (1-d) + d(PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$

$PR(A)$ = PageRank of page A, $PR(T_i)$ = PageRank of pages T_i that link to page A

$C(T_i)$ = number of outbound links on page T_i , d = damping factor between 0 and 1



Google PageRank example

The more outbound links a page T has, the less page A will benefit from a link to it from page T

Weighted PageRanks are added up, then multiplied by the damping factor.

Check PAGE RANK of Web site pages Instantly

In order to [check pagerank](#) of a single web site, web page or domain name, please submit the URL of that web site, web page or domain name to the form below and click "Check PR" button.

Web Page URL: <http://myspace.com>

The Page Rank: 8/10

(the page rank value is 8 from 10 possible points)



Alexa

A more advanced page ranking tools, getting website traffic, statistics, and analysis.

Owned by Amazon

Measures popularity, visitor metrics, audience geography, upstream sites, linking sites, related sites, how fast a site loads, audience demographics



Alexa example



An
amazon.com
company

Features ▼

Resources ▼

Pricing

Log in

SIGN UP

myspace.com Traffic Statistics

Find similar sites to [myspace.com](#)

This site's metrics are estimated

Is this your site? [Certify your site's metrics.](#)

**GROW YOUR
BUSINESS USING
ALEXA**

- ✓ Keyword Difficulty Tool
- ✓ Competitor Keyword Matrix
- ✓ On-Page SEO Checker
- ✓ SEO Audit Tool
- ✓ Audience Overlap Tool
- ✓ Competitive Intelligence

How popular is [myspace.com](#)?



Alexa Traffic Ranks

How is this site ranked relative to other sites?



Global Rank

4,499 ▼ 460

Rank in [United States](#)

1,633



Conclusion

In summary, page ranking is a major ranking for Google and other engines to determine the relevancy and the popularity of websites.

With different tools using different algorithms, there are a number of ways to calculate a page's rank.

Term Frequency and Inverse Document Frequency are important to know for determining page rank.



References

Adams, Chelsea. "What Is Google PageRank, How Is It Earned & Does It Matter in 2016?" *Bruce Clay, Inc. Blog*, 9 June 2016, www.bruceclay.com/blog/what-is-pagerank/.

Tf-Idf :: A Single-Page Tutorial - Information Retrieval and Text Mining." *Tf-Idf :: A Single-Page Tutorial - Information Retrieval and Text Mining*, www.tfidf.com/.

Sobek, Markus. "The PageRank Algorithm." Google PageRank - Algorithm, EFactory GmbH & Co. KG Internet-Agentur, 2002, pr.efactory.de/e-pagerank-algorithm.shtml.