



1.3.4 Search Engine Indexing & PageRank

Search Engine Indexing

A web crawler / spider traverses the internet by following a links.

When it gets to a site it records each word on the page along with the position of the word and adds an entry for the page under the word in the index.

PageRank Algorithm

Search engines need a way to order the pages that they display during a search.

There are many factors that contribute to a website's position in a search, the most common are thought to be:

- Using keywords in the <title> tag
- The age of your website
- How recently the website was last updated
- The frequency of updates
- The number and frequency of keywords appearing in the heading tags
- The relevancy of the domain name to the context
- Another factor is the PageRank of the website.

The PageRank algorithm can be used to generate a ranking for a website.

Effectively it is a popularity contest between websites defined by the number of votes or inbound links they receive. The weight of these votes depending on the PageRank of the inbound links.

The PageRank algorithm is defined as:

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

Where:

- $PR(A)$ is the PageRank of page A
- $C(Tn)$ is the total count of outbound links from web page n
- $PR(Tn)/C(Tn)$ is the share of the vote that page a gets from n
- d is the damping effect.
- A lower damping means less influence generated from inbound links.

