eFactory

A Survey of Google's PageRank

Within the past few years, Google has become the far most utilized search engine worldwide. A decisive factor therefore was, besides high performance and ease of use, the superior quality of search results compared to other search engines. This quality of search results is substantially based on PageRank, a sophisticated method to rank web documents.

The aim of these pages is to provide a broad survey of all aspects of PageRank. The contents of these pages primarily rest upon papers by Google founders Lawrence Page and Sergey Brin from their time as graduate students at Stanford University.

It is often argued that, especially considering the dynamic of the internet, too much time has passed since the scientific work on PageRank, as that it still could be the basis for the ranking methods of the Google search engine. There is no doubt that within the past years most likely many changes, adjustments and modifications regarding the ranking methods of Google have taken place, but PageRank was absolutely crucial for Google's success, so that at least the fundamental concept behind PageRank should still be constitutive.

The PageRank Concept

Since the early stages of the world wide web, search engines have developed different methods to rank web pages. Until today, the occurence of a search phrase within a document is one major factor within ranking techniques of virtually any search engine. The occurence of a search phrase can thereby be weighted by the length of a document (ranking by keyword density) or by its accentuation within a document by HTML tags.

For the purpose of better search results and especially to make search engines resistant against automatically generated web pages based upon the analysis of content specific ranking criteria (doorway pages), the concept of link popularity was developed. Following this concept, the number of inbound links for a document measures its general importance. Hence, a web page is generally more important, if many other web pages link to it. The concept of link popularity often avoids good rankings for pages which are only created to deceive search engines and which don't have any significance within the web, but numerous webmasters elude it by creating masses of inbound links for doorway pages from just as insignificant other web pages.

Contrary to the concept of link popularity, PageRank is not simply based upon the total number of inbound links. The basic approach of PageRank is that a document is in fact considered the more important the more other documents link to it, but those inbound links do not count equally. First of all, a document ranks high in terms of PageRank, if other high ranking documents link to it.

So, within the PageRank concept, the rank of a document is given by the

rank of those documents which link to it. Their rank again is given by the rank of documents which link to them. Hence, the PageRank of a document is always determined recursively by the PageRank of other documents. Since - even if marginal and via many links - the rank of any document influences the rank of any other, PageRank is, in the end, based on the linking structure of the whole web. Although this approach seems to be very broad and complex, Page and Brin were able to put it into practice by a relatively trivial algorithm.

**eFactory
GmbH & Co. KG
sobek@eFactory.de**

Goethestraße 75
40237 Düsseldorf

Tel.: 0211 44 03 97-21
Fax: 0211 44 03 97-40