0.1 Citation

Brin, Sergey, and Lawrence Page. "Reprint of: The anatomy of a large-scale hypertextual web search engine." Computer networks 56.18 (2012): 3825-3833.

0.2 Summary

In this paper, Larry Page and Sergey Brin introduce their seminal project–the Google search engine. Despite Google's deep penetration into modern vernacular, it helps to first define what it actually is. In the words of Page and Brin, Google is "a large-scale search engine which makes use of the structure present in hypertext...designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems." The last part of this statement is probably the defining characteristic of Google: at the time, many search engines were already running in production, but most of them were highly ineffective. In fact, in November 1997, Page and Brin discovered that only one of the top four commercial search engines was able to find itself.

So what was going wrong? It turned out that many of the search engines already in production were not made to handle the Web of 1997. The number of documents on the Web (and therefore, search indices) was increasing at an astronomical rate, causing search engines to return lots of garbage results that would wash out relevant documents from a search. Diluting relevant results severely impacts a search engine?s effectiveness, since a user?s ability to read results is relatively static–most of the time a user only browses the first ten results of a search before giving up.

To resolve this issue, Page and Brin introduced PageRank: a citation (hyperlink) graph of the Web. Citations are a great way to measure the subjective importance a page–the more something is referenced online, the more likely it is to be important to a user. The reasoning behind this simple: citations are created by people.

Along with PageRank, Page and Brin introduced the idea of anchor text, which uses the text associated with a hyperlink to describe the page linked. This helps Google define pages effectively for two reasons. First, anchor text often defines a page better than the page itself. Second, anchor text helps give relevance to pages which cannot be defined otherwise. These are pages containing content such as videos or images.