

<Week 1>

Title: The Anatomy of a Large-Scale Hypertextual Web Search Engine

Last three digits of GTID: 430

This seminal paper by Sergey Brin and Lawrence Page describes the development of Google, the search engine that changed the way the Internet is perceived and used by the billions that are connected to it today. The paper starts with their motivation for entering into search engine research and their design goals, which were to not only devise solutions for the problems that the search engines of the time faced such as scalability and performance quality but also to document research on the subject of search engines which had till then been mostly conducted at companies and the technical details of the algorithms and techniques in use were not open to the public. They describe how a typical search engine of the time functioned and how the architecture and functioning of Google differed. They describe in sufficient detail the major contributions of their work and the major logical components of the Google architecture which are the PageRank algorithm and the use of Anchor text. A constant comparison of their techniques, with respect to crawling, indexing and searching and a mention of what was standard at the time serve to emphasize the importance of their work along with the examples of searches and corresponding search results.

The paper is well written and without being too technically detailed, conveys the important concepts. However, considering the complexity of the architecture of Google Search and the large number of components, the text could have been supplemented with a better flow diagram. Not without its weaknesses, in practice the PageRank algorithm has been found to be open to manipulation or spamdexing with spammers manipulating ranks by creating huge number of web pages and linking to the web page whose rank needs to be elevated or by maintaining link farms. For example, a search for "PageRank" returns www.pagerank.net as the 6th most relevant result when it has little to do with the algorithm. To be fair however, when the paper was written Google was still operating as a research project and the authors did take such possibilities into account and also mentioned other cases where their system did not perform as expected.

The authors describe each component of the search engine and also provide justifications for their design choices. These serve to inform and guide any projects that plan to work in the area of Search Engine Optimization (SEO). By understanding why a design choice made sense in the technological environment when the paper was written, researchers may be able to make their own choices. The work described in the paper is highly innovative and has been the basis for a lot of research and subsequent work in the years following the publication of this paper. For example, the PageRank algorithm has been adapted for use in other areas as well such as a replacement for the Impact Factor used by the Institute for Scientific Information. Undoubtedly, the greatest contribution of the paper is the PageRank algorithm but also valuable are the observations made by the authors regarding the need for research in the area, of which the performance and quality of search results is relevant even today.