Emily Zhang

CS24 Seminar

30 April 2015

Research Paper Summary

For the second research summary I read *The Anatomy of a Large-Scale Hypertextual Web Search Engine* by Sergey Brin and Lawrence Page, which is really interesting because it basically describes Google. It is apparently very difficult to design a search engine that will handle millions of queries and have to crawl through millions of web pages. Storage space must be used efficiently, and fast crawling technology is needed to index all of the web documents. They paid special attention to scale well to extremely large data sets, which is practical because it's very important to think ahead to what will happen in the future in regards to technology. They designed Google with several goals in mind, including improved search quality (because apparently some search engines at the time couldn't even find itself if its own name was entered into its search engine), academic search engine research such as building systems that a reasonable number of people could actually use at the same time, as well as building architecture that could support novel research activities on large-scale web data. Google has many interesting features such as PageRank, which calculates an objective measure of the page's importance that corresponds well with people's subjective idea of importance. It is very interesting how they used an intuitive model of user behavior to help them conceptualize PageRank. I do not think they have any oversights as the web crawler now known as Google became enormously successful and is flourishing in the present economy.

Source: http://infolab.stanford.edu/pub/papers/google.pdf