Information Discovery and Retrieval

The anatomy of a large-scale hypertextual Web search engine

The paper that i have studied, presents the Google Web search engine. Its presented by the founders and developers of Google Web search engine, Sergey Brin and Lawrence Page and discusses the following issues and themes:

- The creation of Google Web search engine
- Design goals
- System features: Page Rank, Anchor text, Indexes
- System Architecture
- Results and Performance

The paper is very hard to understand as its complicated and technical. It targets an audience which has an IT or Information Management background, possessing the qualifications and interests of the World Wide Web and in particular Web search engines. Other readings, such as Informer from BCS explains the way search engines retrieve information and websites and different classification models that can be used to make web searching as easy as possible for the user. It allows the reader to understand the content of the article, using less technical terms as possible and if technical terms are used, these are explained in such a manner that a person without the IT or Information Management background can understand the meaning of it.
The paper written by Sergey Brin and Lawrence Page was published in 1998. Since then, 7 years have gone by, the information contained may now be outdated, as technology and the Web has changed dramatically during that period time. The information stated might serve the purpose today of how Google started and what were the goals / objectives so the Google of 1998 could be compared to the Google and other search engines of today.

In the article, Brin and Page state that Google has the ability to improve its search quality limiting the amount of "junk results" that a web search engine might return back to the user. ‘Our main goal is to improve the quality of the Web Search engines. There is quite a bit of optimism that the use of more hypertextual information can help improve search and other applications’ (Brin & Page, 1998). This is compared to what Heather Hedden suggests in her article about the reliability of search engines. Her concerns are expressed that search engines often retrieve results or sites that have nothing in common with the searched term.

So the question is: Are Web search engines reliable?

In my opinion most of the times, web search engines are useful tools and have the ability to discover and retrieve information without retrieving “Junk results”. But when it does happen, people consider whether or not the methods used to retrieve these sites such as Page Ranks, the way a search engine ranks a URL according to a search criteria, may not work as well as people might think or suggest they do. It doesn't necessarily mean that because a web site has many "hits" (the number of times it has been accessed) that this web site might be the one that the user is looking for even though it may be highly ranked for that particular search
statement. There is still a long way to go for web search engines to perfect the way that information is discovered and retrieved, but at this point of time most users are content with the progress that has been made and the way that search engines go about finding quality information.

**Do anchor texts improve search quality?**

The method of Anchor texts is utilised by Google to enhance its search quality. Brin and Page state that it has several advantages, providing accurate descriptions of web pages and returns pages that have not been crawled. The report written by Glover, Tsioutsouliklis, Lawrence, Pennock and Flake have a slightly different opinion in that anchor text words are less powerful than that of full text alone, but combining both systems or different methods would have a much improved search engine that produces quality search results. Anchor texts do help as they make the links clearer for the users to determine the correct search results.

**Alternative methods for Information searching.**

Heather Hedden suggests in the Issue 15 of INFORMER that the A-Z classification model could work well in conjunction with a web search engine. This type of information searching allows the user to search for a particular topic with ease. But she also explains that this method should be only utilised on web pages that are generally classed as medium sized (30 to 600 pages) and changes to the pages are not made at frequent rate. To back this up, this classification and information retrieval system would be impossible to work on a Web based search engine such as Google and Yahoo as Google indexes around 8 billion pages and
would be very hard and close to insane to be able to use the A-Z classification model.

Conclusion

My general opinion of web search engines is that people can’t live without search engines these days. If people want to find something then they turn to the solution of Google, the quickest and easiest solution which lays in front of us. But even though these search engines are widely accepted by the community, they have drawbacks as to whether or not the information that is being searched, is being returned by the search engine. For example a search engine might return web sites that have the search word in the site but it may not be the information that the person is looking for, thus user’s time is wasted and does not help in the quest to find the results that the users have been seeking for and proves that there is still work to be done to these web search engines. But as previously stated web based search engines have come into our lives for good and with time they will be able to provide users with exceptional search capabilities and results.
References

