CRITICAL JOURNAL ANALYSIS
OF THE ANATOMY OF A LARGE SCALE HYPERTEXTUAL WEB SEARCH ENGINE

INTRODUCTION

The following paper provides a critical analysis of journal article ‘The anatomy of a large scale hypertextual web search engine’ by the authors Sergey Brin and Lawrence Page. This journal article explores the World Wide Web in 1998, and focuses on search engines and their ability or inability to provide a quality search, in response to their users. With an emphasis on past technology that was used to sort results. The authors have researched the growing trend in user’s turning to search engines, such as Yahoo! and AltaVista for precise and relevant information. With this research they have come to believe that these search engines handle user queries with quantity rather than focusing on the quality of pages produced. They go on to argue that the greatest test of a search engine is the quality of results it produces, and technologies such as keyword matching and other related Information retrieval tools just aren’t cutting it on the web. This was being realized by other researches in the field. ‘By 1998, it was clear that too many results were being returned by the average search engine for the one or two keywords that were entered by the searcher.’

Not only do the young developers outline these issues they also claim to have a prototype of a large scale search engine that has shown better results to queries, compared to larger commercial engines that are used and relied upon by most users.

The name of the prototype, a little thing called Google, coming from the misspelling of googol. This creation which in 1998 was available on the Stanford university site was used for research purposes for both researchers and students alike. Google’s presence allowed research to be conducted on how search engines worked, on a very technical level which is the main basis of the journal article, providing descriptions of their advanced technologies including page rank and anchor text. As well as this they allow researches to take a look at the core functions of the search engine, including the web crawlers, indices, barrels etc, through diagrams and complex descriptions. This type of analysis was very scarce in 1998, as search engines deterred from

---

1 Page rank Google’s original sin
publishing their advances in search technology due to competitiveness. Leaving researches no where to turn, therefore the article and prototype attempted to provide a better understanding of this technology and with the quality of search results being produced as well as future developments, it was anticipated by the authors that Google could even make it to commercial standards.

**WHO ARE THEY AGAIN?**

Just looking at the title of the article one might think “*how am I ever going to understand this?*” Starting as most articles do the authors set the scene; the scene of course is searching the web of 1998. This was done by using their research conducted on various popular search engines, monitoring the number of user queries handled on a daily basis. They stated that user queries were growing, claiming that in March and April 1994 the World Wide Web worm handled an average of 1500 queries per day, talking about AltaVista statistics, etc. Yet users are only willing to sift through the first 10 results, they want relevant pages at their finger tips without having to look. However with all this research that represents a clear problem with search engines, they fail to provide adequate referencing to check these claims. From this the authors then continue to hypothesize the number of queries that search engines will handle by 2000. I didn’t know who these authors were which lead me to think, who are these people anyway? How can they make statements and base recommendations on some “*facts*” that have no reference, and why should the reader believe that search engines were so bad in 1998! One is forced to look up the authors and find out if they actually are who they say they are?

Okay so they are the inventors of the world renowned Google, the single most successful search engine of the world, it is safe to say they did know what they were talking about. But they weren’t renowned in 1998, not to say their credentials weren’t building because in the same year Google was in the top 100 search engines in PC magazine², with little experience and just a start up search engine it would have been hard to digest that little known Google was better then search giant Yahoo! So they have a prototype Google, their creation, a search engine they built within the Stanford University website, and this is the reason why they take such a biased side and represent the web to be absolutely hopeless. They make the reader believe that they have solved all these problems, Google is the savior.

² Strategies for e-business
Well I guess it is now, only because of the expertise and experience that these authors have now in 2005, only allows them to have authority to claim what they did in 1998. Which provided a problem, why would anyone believe them in 1998? Because they were graduates in computer engineering, maths and computer science, and they have a prototype.

A GIANT LEAP FOR SEARCH ENGINE TECHNOLOGY

Google’s prototype along with this article was one of the first in outlining what really makes a search engine tick! This analysis was done via a very long and detailed description with an aid of a diagram. As well as the formula for an innovated search tool called page rank. This description gives the reader great insight to the inner workings of a search engine, and Google technology. It is so in depth that those with the knowledge could replicate this page rank formula to other information retrieval systems.

This provides a good basis for researchers to investigate search engine technology as other technologies focus on IRS’s’s and do not apply well to the information of the World Wide Web. They continue to go into the process that Google goes through to crawl the web and the documents it has also cached although because of the advanced technological language used by the graduates, the mere person would become lost in Information retrieval world, therefore the description would only be useful to those with a heavy IR background. But along with this the authors also try to dumb it down by using the example of a web surfer to explain page rank, which works quite effectively to explain a large concept.

As well as an analysis of their search engine the authors also provide a link to their prototype, which of course is no longer available, this invites readers and researchers alike to experience their creation and test out how Google actually works compared to other popular search engines.

Example of the formula they attempted to explain

\[
PR(A) = (1 - d) + d \left( \frac{PR(T_1)}{C(T_1)} + \ldots + \frac{PR(T_n)}{C(T_n)} \right)
\]
PROVING GOOGLE IS BETTER

Even though readers are endowed with the URL to test Google’s results and of course agree with everything Lawrence Page and Sergey Brin claim. They also provide an example of a search performed on the prototype querying Bill Clinton. In comparison to other searches performed on commercialized engines, Google came up with legitimate official white house web sites, where as one search engine, that is not named came up with Bill Clinton joke of the day, as the most relevant website. Showing all of the websites Google retrieves on Bill Clinton, which I might add is impressive for the young technology; it does not compare the search with another popular search engine. Even though in 1998 you would be able to test and make comparisons of your own, you can not do that now! And who can remember how searching was in 1998? This makes the information irrelevant, when it has potential to really prove Google, yet all we can do is rely on what the authors say as proof. Again the article takes a biased turn, and provides no evidence that their search is better, if only they had just taken a snap shot of another search engines results, just as they did of their own, we could really get a feel for how Google was ahead of the rest in 1998 and why it became so popular.

GOOGLE = PERFECTION

The benefits of Google are obvious and clearly outlined throughout the whole article, and in 1998 it would be hard to convince someone with this article that Google will not be a driving force in the search engine world, and that users will not adapt to it well. Of course it wouldn’t be hard because not one weakness of the prototype is outlined! There is no weakness paragraph or problems section, even the future developments section is short and does not identify if any of the technology will be enhanced or needs to be improved. Just as all the benefits and advantages of Google are interesting for researches I’m sure that the problems that they encountered and the weaknesses in the system would provide some benefit as well. Although this is only a short extract from the original longer version, I doubt that in the original the authors had gone into immense detail about the weaknesses of their prototype.
There are problems with Google as a search engine now, with Google bombs that take advantage of the very link structure\(^3\) the article talks about. Surely the authors knew about this issue and others. I can only imagine the problems that they faced as a novice search engine, yet why don’t we know about this? If the authors are so interested in providing details about the inner workings of a search engine, for research purposes, why don’t they give the full analysis the good and the bad? Maybe because it was going to be presented at a World Wide Web conference\(^4\) and these weaknesses could not be published.

Although they failed to spill the secrets about problems with mighty Google, the future section does outline projects and new technologies that they would like to implement before the engine is fully commercial. These search tools include the use of Boolean operators and the text surrounding the link as well as the anchor text used in keyword searching, in an attempt to provide users with a more relevant comprehensive search.

It seems that Google today has definitely met the author’s future expectations and more delivering Boolean operators like most search engines as well as thorough keyword matching.

**CONCLUSION**

This journal article gives us a great insight into the inner workings of both the prototype of Google and the creators Lawrence Page and Sergey Brin. Proving and detailed description of the system as well as the technology implemented that differed Google from search engines in 1998. The authors describe the issues that they believe Google addresses and the need to provide a much more detailed and relevant search to web users. The journal is strongly directed at researches and those with a knowledge background in information retrieval systems. Although it fails to explore any weaknesses which it is known to have. The authors are very opinionated on the state of search engines in 1998 and support this using a scarce amount of referencing that portraying most search engines as very sloppy. The article has a strong bias to the advantages of Google without providing solid evidence. Readers although are invited to test the prototype, the journal gets weaker as it gets older as search engines change information constantly. Therefore the claims that they made in 1998, e.g. about Bill Clinton search results from other search engines cannot be proven.

---

\(^3\) Google Time Bomb

\(^4\) Page rank Google’s original sin
Lastly the article outlines their future aspirations of Google and the search engine today has shown that it has met these goals, as well as other extensive Information retrieval tools, consequently becoming the world leader in search technology today.
