

ITCS 414 – Information Retrieval and Storage

Report Project 2

PageRank: Towards Ranking the WWW Documents

1. Summarize and explain the paper within 1 paragraph (not more than 1 page). You can assume that the readers will have some IT background, but not an expert in IR. So the language that you use must not be too technical, and avoid too many jargons.

Answer: As you know, PageRank is an algorithm that Google invented and use in nowadays. However, there is always a room for improvement the algorithm. So, we observe a research of the in-depth nature of PageRank algorithm. Researchers discovered that, sometimes, while converging, most of high Pageranks tend to be slow-converging pages. On the other hand, fast-converging pages are mostly low PageRank. These discover determined that PageRank algorithm could be improved in the view of efficiency as high PageRank are likely the result that user are query to, but it is the one which takes long computation time. In order to solve this situation, the researcher came up with a new way of algorithm called as Adaptive PageRank. Adaptive PageRank is an improvement of PageRank by eliminating redundant computation because PageRank usually computes rank of pages that have already converged. By eliminating redundant computation, it speeds up the computational time of PageRank. Furthermore, computational time of Adaptive PageRank could be speeded up faster by adding some pruning criteria. This improvement is called as Modified Adaptive PageRank. The implementation is that it does not recompute the components of the PageRanks of the pages that have not yet converged due to links from the pages that have converged to a given tolerance. Therefore, these two PageRank improvements improve the computational time of PageRank around 20% and 30% respectively..

2. Answer the following question. How could you apply the technique(s) you read in the paper to a novel problem that has not been solved before. Make sure that you first describe what the problem is and why this problem is worth solving, before suggesting how such a problem could be solved using the methods presented in the paper.

Answer: Our opinion of novel problem is that the retrieval time of search engine is going to be longer in everyday because new web pages are created every day and it stacks to the current web pages which are many. Hence, our observation of improving PageRank algorithm would be worthy for solving these situation as it speeds up the retrieval time around 20 ~ 30%. Hence, the way to solve this situation is that implementing this algorithms, Adaptive PageRank or Modified Adaptive PageRank, to search engine which use normal PageRank algorithm.