

42913

Social and Information Network Analysis – Assignment 3

Wandy Halim Vedant

13177011 [Student ID]

# Table of Contents

[Table of Contents 1](#_Toc42988645)

[Abstract 2](#_Toc42988646)

[Keyword 2](#_Toc42988647)

[Introduction 2](#_Toc42988648)

[Methods 2](#_Toc42988649)

[Related Work 2](#_Toc42988650)

[Dataset 2](#_Toc42988651)

[Methods 2](#_Toc42988652)

[Result 2](#_Toc42988653)

[Conclusion 2](#_Toc42988654)

[Reference 2](#_Toc42988655)

## Abstract

In this paper, we analyse the search querries and associated URL of the website from searchers coming from web search engines. The queries of the website are calculated based on the page rank . Thus, the findings in this report are based on Webspam Corpus data consisting of 200k web pages as well as their link structures and the html content. The implications of this paper are to create a better understanding of how the search engine works and to create the search engine algorithm to prototype the searching features and as relevant as it could to the user’s query request. The search engine will apply a page rank metric and we will investigate other features that could beneficially improve the ranking.

## Keyword

Search Engine, Google, Page Rank, Prototype

## Introduction

In this era, the amount of information flow across the internet is snowballing every day. Search engines like google, allow users to find the information in ease, just as simple as input and simple click. Google search engine has been a big impact on every aspect, and it highly depends on how the search query will return. One of these examples is the paper research by Bank. (Bank, 2011) shows that the increase in search queries is associated with a rise in trading activity and stock liquidity. This example shows that the search engine result might create a massive impact in the economic property of those related company.

The importance of the website is inherently a subjective matter (Langville, 2004). He believes that the subjective matter depends on the reader's interest, knowledge, and attitudes, which can be made objective by describing those pages in page rank. This allows a more methodological approach to rate the web pages. Having this in hand will effectively measure the person’s interest and returning the query based on his relevance. In this paper, we will create an algorithm of search engine prototype, which are able to return a set of pages based on the user search query. The mapping will be done based on the dataset from 200k webpages from Webspam Corpus.

## Related Work

Several related works on this paper consist of research on several approaches of page rank and search engine.Research by (Tsoi 2006) implemented a quadratic programming method to produces a set of parameters that can be used for ranking all the pages in the web. Tsoi shows that the method can be applied in building a customised version of page rank, which will easily be adapted to the search engine query request from that single user.

Another related work paper regarding algorithm of the search engine is by (Cone, 2006). They manage to implement a search query without using page rank , however they use session-based search algorithm. The paper presents a method of updating an internet search engine databased based on the database results of the user selection of specific web page listings from the general web page provided to the user. The database will then be updated and prioritize the web listings selected with the most relevance with the keywords. Furthermore, it will first present the most popular web page listings in a succeeding search using the same keyword for each search entry.

A study by (Fagroud, 2019 ) focuses on the issues of information related to IoT and search engine . With the development of IoT and increases in the number of IoT devices . The new type of search engine will become possible. The paper defines it search engines and the steps for the development, issues and advantage, as well as the impact of the search engine in terms of IoT. They divide the approach of IoT search engines into two approaches. The first approach is the search engines have search tools which allow the research of the connected device such as webcam, powerplant, refrigerator, etc. in the world. The device will then be listed and tracked based on a set of information like city, country, OS, and open port. The second approach is to think of the IoT itself if a search engine in which will enable its users to search the information with any topic related to IoT. The search tool contains all of the IoT based information which is accessible across the internet.

Another finding is by (Faqeeh, 2014 ). They proposed a new search engine entitled Topical search engine.This search engine collects all websites accessible on the internet, which content related to a specific topic of IoT, and their database contains 210,000 URLs . The various queries were submitted in order to test their solution and manually evaluate the returned page to determine which ones are relevant and which ones are not. The result of their papers proved that the proposed search engine is fairly quick and more accurate compared to the big search engine giant such as Google, Yahoo and Bing.

## Dataset

## Methods

**Algorithms and network metric go here**

## Result

**experimental report, analysis and visualization**

## Conclusion

## Reference

Bank, M., Larch, M. & Peter, G. 2011, 'Google search volume and its influence on liquidity and returns of German stocks', *Financial markets and portfolio management*, vol. 25, no. 3,p. 239.

Brin, S. & Page, L. 1998, 'The anatomy of a large-scale hypertextual web search engine'.

Cone, J., Franklin, G., Ryan, G. & Stalker, W. 2006, 'Adaptive search engine', Series Adaptive search engine Google Patents.

Fagroud, F.Z., Lahmar, E.H.B., Amine, M., Toumi, H. & Filali, S.E. 2019, 'What does mean search engine for IOT or IOT search engine', paper presented to the *Proceedings of the 4th International Conference on Big Data and Internet of Things*, Rabat, Morocco, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3372938.3372958>.

Faqeeh, M., Al-Ayyoub, M., Wardat, M., Hmeidi, I. & Jararweh, Y. 2014, 'Topical search engine for Internet of Things', IEEE, pp. 829-35.

Langville, A.N. & Meyer, C.D. 2004, 'Deeper inside pagerank', *Internet Mathematics*, vol. 1, no. 3,pp. 335-80.

Page, L., Brin, S., Motwani, R. & Winograd, T. 1999, *The pagerank citation ranking: Bringing order to the web*, Stanford InfoLab.

Tsoi, A.C., Hagenbuchner, M. & Scarselli, F. 2006, 'Computing customized page ranks', *ACM Trans. Internet Technol.*, vol. 6, no. 4,pp. 381–414.