Comparison of Page Ranking and Page Access Coefficient of Rank Values

Mrs. L.Rajeswari [#1] , Dr.S.S.Dhenakaran [#2]

[#1]*Computer Centre*
Alagappa University, Karaikudi
lr_eswari@yahoo.co.in

[#2]*Professor, Dept. of CSE*
Alagappa University, Karaikudi.
ssdarvind@yahoo.com

Abstract --The World Wide Web (WWW) is [1] rapidly growing in all aspects and is a massive, explosive, diverse, dynamic and mostly unstructured data repository. As on today WWW is the huge information repository for knowledge reference. Hence, information retrieval from the available web pages requires specific techniques with reduced complexity and increased speed. Online social networks involve a large quantity of information transmission that creates an increased impact on it. Ranking is an important component in most search engines to prioritize a search results and to offer the user an immediate list of the most relevant results to a query. Google Page Rank [2] (Google PR) is one of the methods for Google to determine a Pages' relevance or importance. Page Access Coefficient (PAC) [3] is used for ranking the web pages for information retrieval from the Social Network Websites. This paper compares the PAC with Page Rank in terms of Rank values.

Keywords – Page Rank, PAC, Precision, Search Query, Web Pages.

I INTRODUCTION

Ranking WebPages is an important mission as it assists the user to look for highly ranked pages that are relevant to the query [4]. From the social network web pages it becomes more complex to search for the relevant web pages since it is interconnected. Hence, page ranking algorithms are involved for the purpose. The most famous search engine Google uses the Hyperlink structure for ranking the web pages.

All are aware that the Google's Page Rank Based Searching algorithm to produce the search results. When user search or browse by simply typing their search query through search engines like Google. The search results are returned to users base on Page Rank of various Web pages. A Page Rank value is computed offline for each page and it does not depend on search queries. Page Rank can be calculated for collection of documents of any size. A page can have a high Page Rank if many pages point to it, or some pages having a high Page Rank point to it [5].

II BACKGROUND

Ranking retrieval systems and relevance feedback have been closely connected each other for the past 25 years of research. Relevance feedback was one of the fist features to be added to the basic SMART system, and is the foundation for the probabilistic indexing model. Ranking can be done without the use of relevance feedback, retrieval will be further improved by the addition of the query modification technique.

Google is based on back links. The more quality back links the higher Google Page Rank. If a website having Page Rank of more than 3 then it is said to be a good website. But if a website is having the Page Rank of more than 5 then the website is getting great traffic and the overall performance or the structure of the website is good enough. With high Page Rank, a website is usually seen on the top page of the Search Engine Results Page (SERP). With more than 6 Page Rank most of the advantages of the website is that the Webmaster will request link exchange with those website and users tend to interested in doing transactions with them. It acquires a reputation in the web world and it may further go beyond that Page Rank if the website is getting on a way like that before.

Drawbacks of Google Page Rank

- New pages have less Page rank and they take much time to get listed and gain high ranks.

609

- If someone is inaccurately quoting something on a web page then subsequent readers also quotes it on another web page, search engines index all of the inaccurate pages, and if end up with a mess where fiction is accepted as reality.
- Search results are based on the literal (keywords, tags, meta data) things but not on meaning. As a result, its precision of retrieval is often unsatisfactory [6].
- Page Rank is usually a numeric value between 0 and 10. A page with Page Rank 10 is considered to be the most important and a Page Rank 0 is considered to be the least important. The Page Rank score of many pages is alike which degrades the quality of ranking.
- Sometimes when a user searches on Google most of the websites are having low page Rank that is 2 or 3 or sometimes 0 and it is coming on top of the page of Google.
- It is a static algorithm that, because of its cumulative scheme, popular pages tend to stay popular generally.
- Popularity of a site does not guarantee the desired information to the searcher so relevance factor as also needs to be included.

- In Internet, available data is huge and the algorithm is not fast enough.

Nobody understand the logic behind the Google giving the rank to a website or what they considered to be the most according to them is best. To overcome the drawbacks of the Google Page Rank, a new algorithm Page Access Coefficient algorithm is developed for ranking the Web pages.

III EXPERIMENTAL STUDY

In this method, the rank of Web page is calculated based on the number of incoming links of page, number of outgoing links of pages and the total number of pages. Another important advantage of the PAC is it makes simpler computations to retrieve information that are closer enough to the search query with no iteration.

Sample queries have been tabulated in Table 1 and Table 2 based on the existing and proposed methods as below.

The list of Web pages and their Page Rank received from PAC for the preprocessed query 'Object Oriented Programming' are given in Table 1.

Table 1
List of Web Pages Retrieved for the query 'Object Oriented Programming' based on Google PR.`

URL No.	URL	Page Rank of Google
1	https://en.wikipedia.org/wiki/**Object-oriented_programming**	7
2	searchsoa.techtarget.com/definition/**object-oriented-programming**	5
3	www.webopedia.com › TERM ›	0
4	www.tutorialspoint.com/cplusplus/cpp_**object_oriented**.htm	3
5	https://docs.oracle.com/javase/tutorial/java/concepts	6
6	www.codeproject.com › ... › Design and Architecture › General	6
7	www.javatpoint.com/java-**oop**s-concepts	0
8	https://en.wikibooks.org/wiki/**Object_Oriented_Programming**	4
9	https://www.coursera.org/learn/**object-oriented**-java	N/A
10	https://www.youtube.com/watch?v=lbXsrHGhBAU	5
11	www.businessdictionary.com/.../**object-oriented-programming-OOP**.html	N/A
12	https://www.youtube.com/watch?v=SS-9y0H3Si8	5
13	https://www.youtube.com/watch?v=SZqLEDH0x7A	5
14	https://www.quora.com/What-is-**object-oriented**-programming	3
15	www.mathworks.com/discovery/**object-oriented-programming**.html	5
16	c2.com/cgi/wiki?**ObjectOrientedProgramming**Language	4
17	https://help.sap.com/saphelp_nw70/helpdata/en/c3/.../content.htm	N/A
18	www.aonaware.com/**oop**1.htm	4
19	www.desy.de/gna/html/cc/Tutorial/tutorial.html	5
20	https://www3.ntu.edu.sg/home/ehchua/**programming**/cpp/cp3_**OOP**.html	2
21	**programming**.webcrawler.com	N/A

610

22	anandology.com/python-practice.../**object_oriented_programming**.html	N/A
23	www.purl.org/stefan_ram/pub/doc_kay_**oop**_en	4
24	www.freetechbooks.com/**object-oriented-programming**-f12.html	5
25	www.cs.utexas.edu/users/downing/cs371p	3
26	https://www.codecademy.com/courses/ruby-beginner-en-MFiQ6?...id..	3
27	harmful.cat-v.org/software/**OO_programming**	3
28	https://www.techopedia.com/definition/.../**object-oriented-programming**	N/A
29	https://msdn.microsoft.com/en-us/library/dd460654.aspx	4
30	www.cyclismo.org/tutorial/R/**objectOriented**.html	N/A
31	https://www.udacity.com/course/viewer#!/c-ud837/l-4619208555	N/A
32	bookboon.com/en/**object-oriented-programming**-using-java-ebook	4
33	www.cs.utexas.edu/users/downing/cs371p/	3
34	cs108.stanford.edu	5
35	eloquentjavascript.net/1st_edition/chapter8.html	N/A
36	www.python-course.eu/**object_oriented_programming**.php	3
37	https://www.edx.org/.../**object-oriented-programming**-iitbombayx-cs101-	N/A
38	www.lua.org/pil/16.html	5
39	www.desy.de/gna/html/cc/Tutorial/tutorial.html	5
40	avascript.info › Tutorial	6

Table 2
List of Web Pages Retrieved for the query 'Object Oriented Programming' based on PAC

URL No.	URL	Page Rank of PAC
1	https://en.wikipedia.org/wiki/**Object-oriented_programming**	8
2	https://docs.oracle.com/javase/tutorial/java/concepts	8
3	www.codeproject.com › ... › Design and Architecture › General	7
4	avascript.info › Tutorial	7
5	searchsoa.techtarget.com/definition/**object-oriented-programming**	7
6	https://www.youtube.com/watch?v=lbXsrHGhBAU	6
7	https://www.youtube.com/watch?v=SS-9y0H3Si8	6
8	https://www.youtube.com/watch?v=SZqLEDH0x7A	6
9	www.mathworks.com/discovery/**object-oriented-programming**.html	6
10	www.desy.de/gna/html/cc/Tutorial/tutorial.html	5
11	www.desy.de/gna/html/cc/Tutorial/tutorial.html	5
12	www.freetechbooks.com/**object-oriented-programming**-f12.html	5
13	cs108.stanford.edu	5
14	www.lua.org/pil/16.html	5
15	www.desy.de/gna/html/cc/Tutorial/tutorial.html	5
16	https://en.wikibooks.org/wiki/**Object_Oriented_Programming**	4
17	c2.com/cgi/wiki?**ObjectOrientedProgramming**Language	4
18	www.aonaware.com/**oop**1.htm	4
19	www.purl.org/stefan_ram/pub/doc_kay_**oop**_en	4
20	https://msdn.microsoft.com/en-us/library/dd460654.aspx	4
21	bookboon.com/en/**object-oriented-programming**-using-java-ebook	4
22	www.tutorialspoint.com/cplusplus/cpp_**object_oriented**.htm	3
23	https://www.quora.com/What-is-**object-oriented-programming**	3
24	www.cs.utexas.edu/users/downing/cs371p	3
25	https://www.codecademy.com/courses/ruby-beginner-en-MFiQ6?...id..	3
26	harmful.cat-v.org/software/**OO_programming**	3
27	www.cs.utexas.edu/users/downing/cs371p/	3
28	www.python-course.eu/**object_oriented_programming**.php	3
29	https://www3.ntu.edu.sg/home/ehchua/**programming**/cpp/cp3_**OOP**.html	2
30	www.webopedia.com › TERM ›	2
31	www.javatpoint.com/java-**oop**s-concepts	2

611

32	https://www.coursera.org/learn/**object-oriented**-java	2
33	www.businessdictionary.com/.../**object-oriented-programming-OOP**.html	2
34	https://help.sap.com/saphelp_nw70/helpdata/en/c3/.../content.htm	1
35	**programming**.webcrawler.com	1
36	anandology.com/python-practice.../**object_oriented_programming**.html	1
37	https://www.techopedia.com/definition/.../**object-oriented-programming**	0
38	www.cyclismo.org/tutorial/R/**objectOriented**.html	0
39	https://www.udacity.com/course/viewer#!/c-ud837/l-4619208555	0
40	eloquentjavascript.net/1st_edition/chapter8.html	0

IV RESULT AND DISCUSSION

From **Table 1,** Out of 40 pages, there are 2 pages have zero Page Rank.

10 pages have N/A (not available) Page Rank, since the web page is indexed by Google, but it is **not ranked** yet.

Relevant pages are computed by subtracting the pages having zero Page Rank and the pages having N/A.

Total number of pages retrieved = 40
Number of pages that are relevant = 28

Table 2 shows that by using PAC method, the web pages are ranked in ascending order. From Table 2, out of 40 pages, there are 4 pages have zero Page rank. Relevant pages are computed by subtracting the pages having zero Page Rank.

It is seen that the performance of PAC method of searching is better since it has brought more relevant pages.

Total number of pages retrieved = 40
Number of pages that are relevant = 36

Precision = Number of pages that are relevant / Total number of pages retrieved

The precision of Google for queries are shown in Table 3 as below. The PAC is also tested with the above queries. The precision of PAC for queries is shown in Table 4.

Table 3

Sl.No.	QUERY	Number of Web Pages Retrieved	Number of Web Pages Relevant	Precision of Google
1	Object Oriented Programming	40	28	0.70
2	Image Processing	40	21	0.53
3	Software Engineering	40	26	0.65
4	Numerical Method	40	27	0.68
5	Environmental Science	40	25	0.63
6	Mahatma Gandhi	40	21	0.53
7	Jawaharlal Nehru	40	22	0.55
8	Abdul Kalam	40	24	0.60

Precision of Google for Queries

612

Table 4

Precision of PAC for Queries

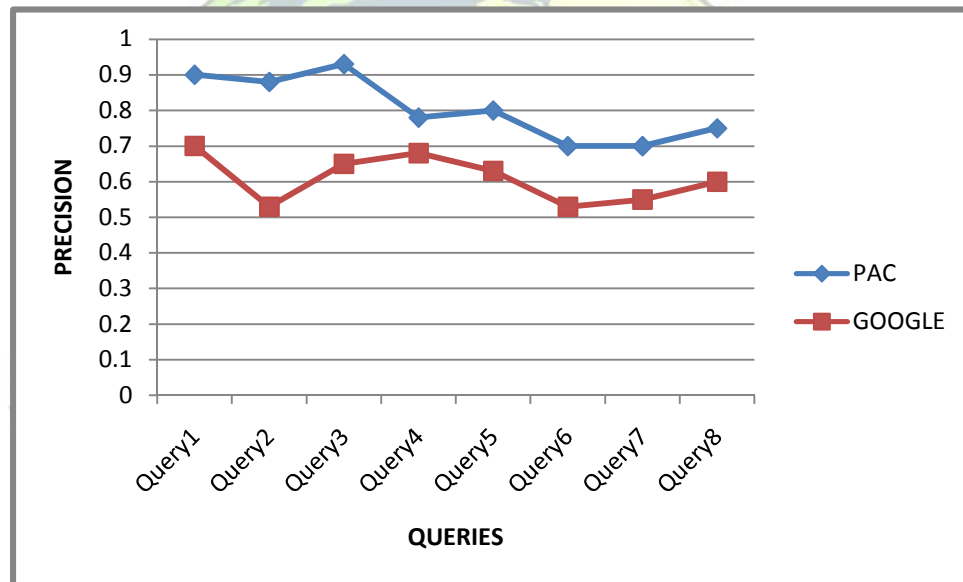| Sl.No.	QUERY	Number of Web Pages Retrieved	Number of Web Pages Relevant	Precision of PAC
1	Object Oriented Programming	40	36	0.90
2	Image Processing	40	35	0.88
3	Software Engineering	40	37	0.93
4	Numerical Method	40	31	0.78
5	Environmental Science	40	32	0.80
6	Mahatma Gandhi	40	28	0.70
7	Jawaharlal Nehru	40	28	0.70
8	Abdul Kalam	40	30	0.75

Figure 1. Precision of Queries

The figure above displays the precision of PAC and Google for the same set of synthetically generated queries.

V CONCLUSION

With the rapid growth of www and the user's demand on knowledge, it is becoming more difficult to manage the information on www and satisfy the user needs. Therefore, the users looking for better search query techniques to locate and filter the necessary information. This paper also studies the performance of Page Rank and PAC algorithms. The PAC algorithm is compared with existing Page Rank algorithm. The comparative analysis shows that the PAC efficiently retrieves relevant web pages than Page Rank and precision of PAC is also better than Page Rank.

References

[1]   Laxmi Choudhary and Bhawani Shankar Burdak, " Role of

613

Ranking Algorithms for Information Retrieval", pp: 1-17, Banasthali University, Jaipur, Rajasthan.

[2] Page L., Brin S., Motwani R. and Winograd T., "The PageRank Citation Ranking:Brining Order to the Web", Standford University, Draft, 2008.

[3] Rajeswari L., Prof. Dhenakaran S.S., "Page Access Coefficient Algorithm for Information Filtering in Social Network", IJCET@IAEME, Volume 4, Issue 3, May-June (2013), pp.60-69.

[4] NeelamTyagi, Simple Sharma, "Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Page", in International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-2, Issue-3, July 2012, pp:441 – 446.

[5] Arun K.Pujari, 'Data mining Technique", Universities Press, 2001.

[6] Vandana Dhingra and Komal Bhatia, "Compratve Analysi of Ontology Ranking Algorithm", International Journal of Information Technology and Web Engineering, Vol.7,No.3, pp.55-66,2012.

[7] www.prchecker.info/check_page_rank.php