International Journal of Scientific & Engineering Research, Volume 3, Issue 12, December-2012
ISSN 2229-5518

An Empirical Analysis of Page Ranking Algorithms

Dr Paras Nath Gupta[1], Pawan Singh[2], Punit Kr Singh[3], Sandeep Chaudhary[4], Pankaj P Singh[5], Deepak Sinha[6]

Abstract— The information containment in the World Wide Web has increased incredibly. With the high growth rate of increase of information, retrieving the relevant data has become cumbersome. To get rid of the complexity of information access the search engines are use which apply different strategy to collect filter and drag the required information. This paper describes the different page ranking algorithms and their empirical analysis by simulating them.These algorithims are PageRank, Weighted PageRank, and PRLV.The algorithms are compared on the basis of relevance values and number of relevant pages in resultant page set for different queries.

Index Terms— web structure, page ranking, web graph, hyperlink structure, weighted page ranking, PRLV, web searching.

1 Introduction

THE WWW is a collection of web pages which contain the data or information. One can move to another page from a page using hyperlink this way a complete structure is being made by the web pages and hyperlinks known as hyper-linked structure of web pages. It is a directed graph where web pages are nodes and hyperlinks are directed edges connecting pages. Search engines uses this structure to calculate the importance & relevance of pages with respect to user query and provide the user the web pages in sorted order of importance as page ranks. There are different strategies and methods based on different factors used to calculate the page rank.

2 Web Page Ranking Algorithms

Many Search Engines has developed the web page ranking strategy, taking into account the relevancy, importance and content scores to mine the data and organize them as per the user interest. A number of favoured web page ranking algorithms are mentioned within the following section.

2.1 PageRank Algorithm

S. Brin and Larry Page [1] developed a ranking formula utilized by Google, named Page Rank after Larry Page that uses the link structure of the World Wide Web to check the importance of pages of sites. This formula states that if a page has some very important incoming links there to then its outgoing links to various pages in addition becomes very important.

- Dr Paras Nath Gupta[1] , Associate Professor,A.M. College,Maghad University,Gaya, India. E-mail: drparasnathgupa.magadhuni@yahoo.com
- Pawan Singh[2] ,Assistant Professor, Department of Computer Science, IIMT Engineering College,Meerut,India.E-mail: pawansingh3@yahoo.com.
- Punit Kr Singh[3],Assistant Professor, Department of Computer Science, IIMT Engineering College,Meerut,India. E-mail: [3]punit20.singh@live.com
- Sandeep Chaudhary[4] Assistant Professor, Department of Computer Science, IIMT IET,Meerut,India. E-mail:sandeepmymail19@gmail.com.
- Pankaj P Singh[5] Assistant Professor, Department of Computer Science, IIMT Engineering College,Meerut,India.E-mail:creatorconcern@gmail.com
- Deepak Sinha[6] Assistant Professor, Department of Computer Application, IIMT Management College,Meerut,India.E-mail: dipu_**sinha**@yahoo.com

Therefore, it takes back links into thought and propagates the ranking through links. Thus, a page incorporates a high rank if the addition of the ranks of its back links is high. The formula is shown in eq. (1)

$$PR(u) = (1-d) + d \sum_{v \in B(u)} \frac{PR\ (v)}{N_v} \qquad (1)$$

Where u represents an internet web page, B(u) is that the set of pages that spot to u. PR(u) and PR(v) are the rank scores of page u and v, correspondingly. N_v denotes the quantity of outgoing links of page v. Where d is a dampening factor that is generally set to 0.85 , d may be thought of as the probability of users' following the links and could regard (1 − d) as the page rank allocation from non-directly linked pages. In Page Rank, the rank score of a page, p, is equally divided among its outgoing links. The values appointed to the outgoing links of page p are in tern used to successively calculate the ranks of the pages to that page p are pointing.

2.2 Weighted Page Rank Algorithm

Wenpu Xing and Ali Ghorbani [2] suggested an expansion to standard web Page Rank called Weighted Page Rank (WPR).This formula allocates larger rank prices to additional very important pages rather than distributing the rank value of an internet page equally among its outgoing connected websites. Each out-link web page gets a worth in line with its quality. The recognition of an web page is computed from the amount of in-links and out-links as Win (v,u) and Wout (v,u), consequently. Win (v,u) (as in eq. (2)) is that the weight of web link(v, u) computed on the premise of the count of in-links of website u and also the count of in-links of all reference web pages of web page v.

$$W^{in}_{(v,u)} = \frac{I_u}{\sum_{p \in R(v)} I_p} \qquad (2)$$

I_u and I_p represents the count of in-links of web page u and web page p, correspondingly. Wherever R(v) provides the reference web page list of web page v and Wout (v,u) (as in eq. (3)) represents the weight of link(v, u) computed based on the count of out-links of page u and the count of out-links of all

International Journal of Scientific & Engineering Research Volume 3, Issue 12, December-2012
ISSN 2229-5518

reference web pages of page v.

$$W^{out}_{(v,u)} = \frac{O_u}{\sum_{p \in R(v)} O_p} \qquad (3)$$

Where Ou and Op provides the count of out-links of web page u and web page p, consequently. Considering the importance of web pages, the particular Page Rank formula eq. (1) is changed as given in eq. (4).

$$PR(u) = (1 - d) + d \sum_{v \in B(u)} PR(v) W^{in}_{(v,u)} W^{out}_{(v,u)} \qquad (4)$$

The resultant order of pages availed by Page Rank and WPR is different. To compare the WPR with actual Page Rank, the authors categorized the resultant pages of a query into four groups based on their relevancy to the given query:

1. Very Relevant pages (VR)
2. Relevant pages (R)
3. Weak-Relevant pages (WR)
4. Irrelevant pages (IR)

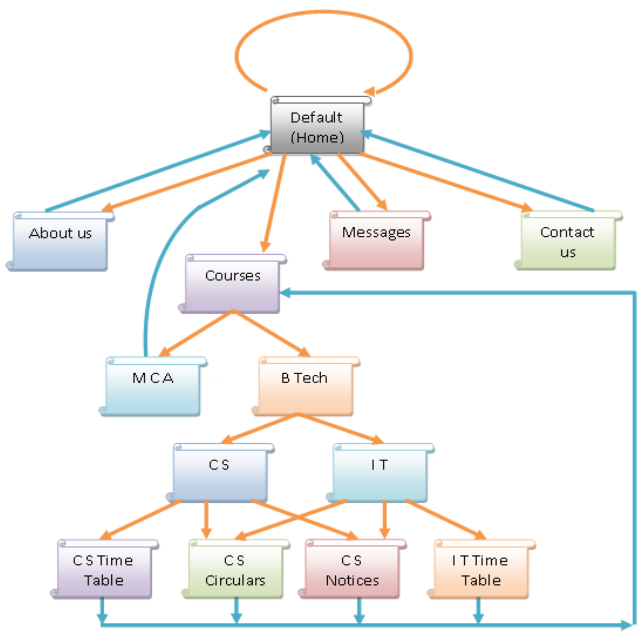

Figure 1: A Web Access Graph

Relevancy Rule: The relevancy of a page relies on its category and its position in the page-list. The higher the relevancy value is, the better is the result. The relevancy κ of a page-list is a function of its category and position as in eq. (5):

$$\kappa = \sum_{i \in R(p)} (n - i) \times W_i \qquad (5)$$

where i represents the ith web page in the end result web page-list R(p) and n represents the first n web pages selected from the end result web page list and Wi gives the weight of page i.

2.3 Page Ranking based on Links Visits (PRLV)

PRLV (Page Ranking based on Link Visits) [3] based on Web Structure Mining and Usage Mining is taking the user visits of pages/links to determine the importance and relevance score

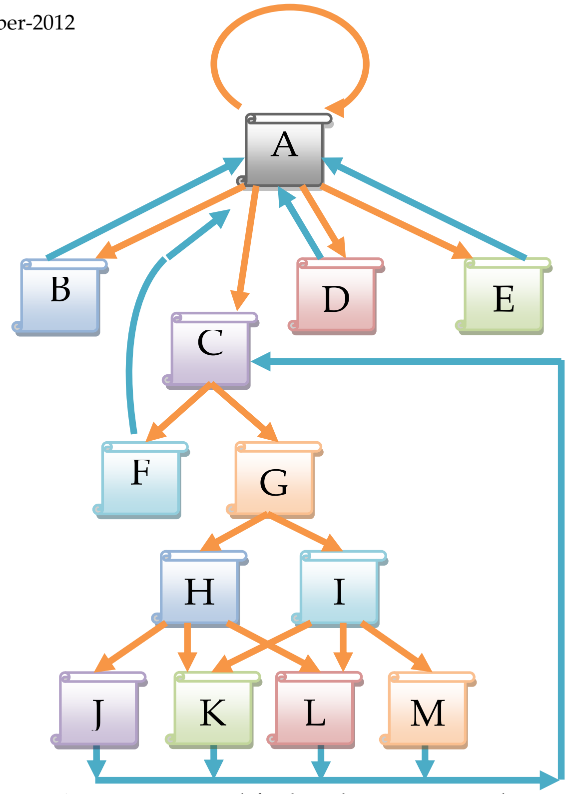

Figure 2: A Simplified Web Access Graph

of the web pages. To accomplish the task many subtasks are to be performed as:

1. Storage of user's access information (hits) on an outgoing link of a page in related server log files.
2. Fetching of pages and their access information by the targeted web crawler.
3. For each page link, computation of weights based on the probabilities of their being visited by the users.
4. Final rank computation of pages based on the weights of their incoming links.
5. Retrieval of ranked pages corresponding to user queries.

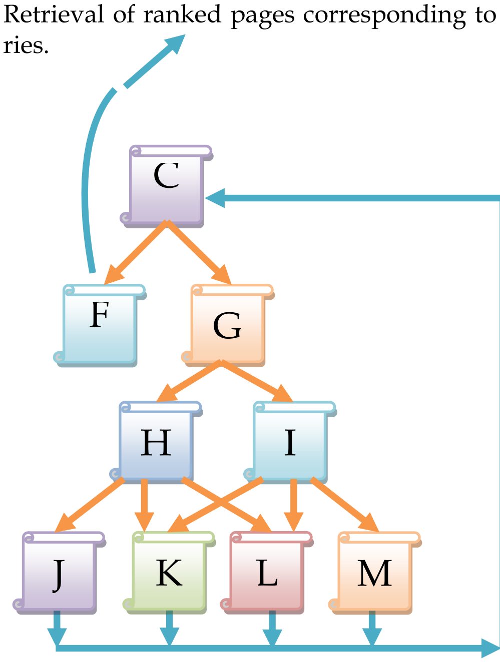

Figure 3: A Partial Simplified Web Access Graph

Table 1: Link Data Set

S No	C_URL	OUT_URL	Key 1	count	Rel	Key 2	count	Rel	Key 3	count	Rel	Key 4	count	Rel
1	http://localhost:1227/iimt_s/Default.aspx	http://localhost:1227/iimt_s/Default.aspx	Student	5	R	education	4	W R	IIMT	4	W R	India	1	I R
2	http://localhost:1227/iimt_s/Default.aspx	http://localhost:1227/iimt_s/courses.aspx	Student	5	R	education	4	W R	IIMT	4	W R	India	1	I R
3	http://localhost:1227/iimt_s/Default.aspx	http://localhost:1227/iimt_s/Messages.aspx	Student	5	R	education	4	W R	IIMT	4	W R	India	1	I R
4	http://localhost:1227/iimt_s/Default.aspx	http://localhost:1227/iimt_s/aboutus.aspx	Student	5	R	education	4	W R	IIMT	4	W R	India	1	I R
5	http://localhost:1227/iimt_s/Default.aspx	http://localhost:1227/iimt_s/contactus.aspx	Student	5	R	education	4	W R	IIMT	4	W R	India	1	I R
6	http://localhost:1227/iimt_s/courses.aspx	http://localhost:1227/iimt_s/BTECH.aspx	Student	2	W R	programme	2	W R	IIMT	1	I R	degree	6	R
7	http://localhost:1227/iimt_s/courses.aspx	http://localhost:1227/iimt_s/mca.aspx	Student	2	W R	programme	2	W R	IIMT	1	I R	degree	6	R
8	http://localhost:1227/iimt_s/Messages.aspx	http://localhost:1227/iimt_s/Default.aspx	Student	14	V R	quality	1	I R	IIMT	15	V R	technology	2	W R
9	http://localhost:1227/iimt_s/aboutus.aspx	http://localhost:1227/iimt_s/Default.aspx	Student	2	W R	quality	1	I R	IIMT	8	V R	college	2	W R
10	http://localhost:1227/iimt_s/contactus.aspx	http://localhost:1227/iimt_s/Default.aspx	Student	1	I R	meerut	5	R	IIMT	7	R	India	3	W R
11	http://localhost:1227/iimt_s/BTECH.aspx	http://localhost:1227/iimt_s/cs.aspx	Student	5	R	study	1	I R	IIMT	5	R	degree	2	W R
12	http://localhost:1227/iimt_s/BTECH.aspx	http://localhost:1227/iimt_s/it.aspx	Student	5	R	study	2	I R	IIMT	5	R	degree	3	W R
13	http://localhost:1227/iimt_s/mca.aspx	http://localhost:1227/iimt_s/Default.aspx	Student	1	I R	India	4	W R	IIMT	1	I R	college	3	W R
14	http://localhost:1227/iimt_s/cs.aspx	http://localhost:1227/iimt_s/cstimetable.aspx	Student	2	W R	degree	2	W R	IIMT	1	I R	technology	3	W R
15	http://localhost:1227/iimt_s/cs.aspx	http://localhost:1227/iimt_s/cscircular.aspx	Student	2	W R	degree	2	W R	IIMT	1	I R	technology	3	W R
16	http://localhost:1227/iimt_s/cs.aspx	http://localhost:1227/iimt_s/csnotice.aspx	Student	2	W R	degree	2	W R	IIMT	1	I R	technology	3	W R
17	http://localhost:1227/iimt_s/it.aspx	http://localhost:1227/iimt_s/ittimetable.aspx	Student	2	W R	degree	2	W R	IIMT	1	I R	Information	2	W R
18	http://localhost:1227/iimt_s/it.aspx	http://localhost:1227/iimt_s/cscircular.aspx	Student	2	W R	degree	2	W R	IIMT	1	I R	Information	2	W R
19	http://localhost:1227/iimt_s/it.aspx	http://localhost:1227/iimt_s/csnotice.aspx	Student	2	W R	degree	2	W R	IIMT	1	I R	Information	2	W R
20	http://localhost:1227/iimt_s/cstimetable.aspx	http://localhost:1227/iimt_s/courses.aspx	Student	5	R	Lecture	6	R	IIMT	5	R	Period	3	W R
21	http://localhost:1227/iimt_s/ittimetable.aspx	http://localhost:1227/iimt_s/courses.aspx	Student	1	I R	Lecture	4	W R	IIMT	1	I R	Period	2	W R
22	http://localhost:1227/iimt_s/cscircular.aspx	http://localhost:1227/iimt_s/courses.aspx	Student	1	I R	Faculty	5	R	IIMT	4	W R	Information	1	I R
23	http://localhost:1227/iimt_s/csnotice.aspx	http://localhost:1227/iimt_s/courses.aspx	Student	12	V R	Faculty	11	V R	IIMT	6	R	Information	3	W R

The weights are calculated for out-links of pages and ranks are computed by taking back-links into account.***Calculation of Visits (hits) of links:*** The hit count of every link is additionally hold on, that is set simply by count the distinct IP addresses visiting the page. The link_log contains knowledge all regarding the page URLs, their hyperlinks and total hit count of every address. The count of visits gathered in link_log is computed from the second log referred to as access_log, it's the format of the NCSA Combined Log Format. By process the access_log and count the distinct IP addresses or User_Ids visiting the address & stored in link_log.***Calculation of Rank Score:*** The search engines index most of the pages data as sporadically accessing from completely different internet servers databases by the online crawlers at the time of crawling. The operating of crawlers for PRLV is to fetch the pages in addition it access data as their hit counts hold on in link_logs and sent to the PRLV Calculator. Each link within the crawled web graph is assigned a weight, indicating its likelihood of being visited by the users. Thus, ranking is propagated iteratively through back coupled pages. Some terminologies used are:

Outbound link, a page p contains n hyperlinks in it and the set of outbound link is expressed by: O(p)= { o_1, o_2,…o_n },where every o_i is a web page which can be accessed from web page p.

Inbound Link to page p is a set B of m inbound links if: B= { b_1, b_2,…b_m}where each b_i is a web page from which web page p can be accessed .

Probability-Weight of Link - For a web page p O(p) is a outbound hyper link set where every outbound hyper link is associated a integer value Visit Count (VC).Then the weight for link between p and o web pages is given by eq.(6).

$$Weight_{link}\ (p,o) = \frac{VC(p,o)}{\sum_{o\ \in O(p)} VC\ (p,o)} \qquad (6)$$

Page Rank based on Link Visits (PRLV), If p is a page having inbound-linked pages in set B(p), then the rank (PRLV) is given by eq. (7):

$$PRLV(p) = (1-d) + d \sum_{b \in B(p)} PRLV(b).Weight_{link}\ (b,p) \qquad (7)$$

Where d is that the damping factor as is employed in PageRank, $Weight_{link}$ (.) is that the weight of the link. The PRLVcalculates the ranks of the pages with the assistance of knowledge gathered from internet crawler and passes these values to the net crawler.

2.4 Web Page Ranking using Link Attributes

Suppose that a user surfing the Web randomly and user is in a page, with certain probability. Ricardo [4] gives a variant WLRank assigns the ranking value R(i) to page I using the following eq. (8).

Table 2: Coding for nodes of figure 2

A	http://localhost:1227/iimt_s/Default.aspx
C	http://localhost:1227/iimt_s/courses.aspx
D	http://localhost:1227/iimt_s/Messages.aspx
B	http://localhost:1227/iimt_s/aboutus.aspx
E	http://localhost:1227/iimt_s/contactus.aspx
G	http://localhost:1227/iimt_s/BTECH.aspx
F	http://localhost:1227/iimt_s/mca.aspx
H	http://localhost:1227/iimt_s/cs.aspx
I	http://localhost:1227/iimt_s/it.aspx
J	http://localhost:1227/iimt_s/cstimetable.aspx
M	http://localhost:1227/iimt_s/ittimetable.aspx
K	http://localhost:1227/iimt_s/cscircular.aspx
L	http://localhost:1227/iimt_s/csnotice.aspx

Table 3: Iterative Calculation for PageRank

ITERATIONS	0	1	2	3	4	5	6	7	8	9	10	11	12	13
PR (C)	1	3.65	2.545	2.07538	1.87578	2.56744	2.27903	2.15646	2.10437	2.28489	2.20962	2.17762	2.16403	2.21115
PR (F)	1	0.675	1.80125	1.33163	1.13203	1.04721	1.34116	1.21859	1.1665	1.14436	1.22108	1.18909	1.17549	1.16971
PR (G)	1	0.675	1.80125	1.33163	1.13203	1.04721	1.34116	1.21859	1.1665	1.14436	1.22108	1.18909	1.17549	1.16971
PR (H)	1	0.675	0.53688	1.01553	0.81594	0.73111	0.69506	0.81999	0.7679	0.74576	0.73635	0.76896	0.75536	0.74958
PR (I)	1	0.675	0.53688	1.01553	0.81594	0.73111	0.69506	0.81999	0.7679	0.74576	0.73635	0.76896	0.75536	0.74958
PR (J)	1	0.53333	0.44125	0.40211	0.53773	0.48118	0.45715	0.44693	0.48233	0.46757	0.4613	0.45863	0.46787	0.46402
PR (K)	1	0.81667	0.6325	0.55423	0.82547	0.71237	0.6643	0.64387	0.71466	0.68514	0.6726	0.66727	0.68574	0.67804
PR (L)	1	0.81667	0.6325	0.55423	0.82547	0.71237	0.6643	0.64387	0.71466	0.68514	0.6726	0.66727	0.68574	0.67804
PR (M)	1	0.53333	0.44125	0.40211	0.53773	0.48118	0.45715	0.44693	0.48233	0.46757	0.4613	0.45863	0.46787	0.46402

ITERATIONS	14	15	16	17	18	19	20	21	22	23	24	25	26	
PR (C)	2.1915	2.18315	2.1796	2.1919	2.18677	2.18459	2.18366	2.18687	2.18553	2.18497	2.18472	2.18556	2.18521	
PR (F)	1.18974	1.18139	1.17784	1.17633	1.18156	1.17938	1.17845	1.17806	1.17942	1.17885	1.17861	1.17851	1.17886	
PR (G)	1.18974	1.18139	1.17784	1.17633	1.18156	1.17938	1.17845	1.17806	1.17942	1.17885	1.17861	1.17851	1.17886	
PR (H)	0.74713	0.75564	0.75209	0.75058	0.74994	0.75216	0.75124	0.75084	0.75067	0.75125	0.75101	0.75091	0.75087	
PR (I)	0.74713	0.75564	0.75209	0.75058	0.74994	0.75216	0.75124	0.75084	0.75067	0.75125	0.75101	0.75091	0.75087	
PR (J)	0.46238	0.46169	0.4641	0.46309	0.46266	0.46248	0.46311	0.46285	0.46274	0.46269	0.46286	0.46279	0.46276	
PR (K)	0.67476	0.67337	0.67819	0.67618	0.67533	0.67497	0.67622	0.6757	0.67548	0.67538	0.67571	0.67557	0.67552	
PR (L)	0.67476	0.67337	0.67819	0.67618	0.67533	0.67497	0.67622	0.6757	0.67548	0.67538	0.67571	0.67557	0.67552	
PR (M)	0.46238	0.46169	0.4641	0.46309	0.46266	0.46248	0.46311	0.46285	0.46274	0.46269	0.46286	0.46279	0.46276	

$$R(i) = \frac{q}{T} + (1-q)\sum_{j} \frac{W(j,k)R(j)}{\sum_k W(j,k)} \qquad (8)$$

where T is the total number of pages, q is the probability (damping factor) of leaving page p and

$$W(j, i) = L(j, i)\,(c + T(j, i) + AL(j, i) + RP(j, i)) \qquad (9)$$

where given a link from page j to page i we have:

- L(j, i) is 1 if the link exists, or 0 otherwise, and c is a constant that gives a base weight to every link,
- T(j, i) is a value that depends on the tag where the link is inserted,
- AL(j, i) is the length of the anchor text of the link divided by a constant d that depends that estimates the average anchor text length in characters, and
- RP(j, i) is the relative position of the link in the page weighted by a constant b.

T(j, i) could also be a sequence of constants reckoning on the tag where the link is, if the link is inside a < h1> tag, a high T(j, i) worth is appointed and slightly less for < h2>, etc. The term AL(j, i) offers lots of value to links where the creator explained in further detail what net resource is being connected. RP(j, i) offers lots of weight to links that area unit at the beginning of the page rather that at the last of the page.

2.5 Time Rank: Based on visiting time

The Time Rank model [5] is generated by causing time primarily based visiting model. If the page content is related to keywords of user query then the user can remain page for a protracted amount, the opposite user can leave quickly giving short visiting time to the page. In Time Rank, every page has n-ranking scores that relies on n-number of topic, thus area unit TSPR(i) ,which is calculated offline. Next the user's keywords submitted in query area unit matched against the quantity of topics. The relative chance between questions alphabetic character and topic i is usually recommended on the premise of theorem theorem [6] as in eq. (12)

$$P_r\ (T(i) \mid q) \;=\; \frac{Pr(q, T(i))}{Pr(q)} \qquad (10)$$

$$= \frac{Pr(T(i)) * Pr(q \mid T(i))}{Pr(q)} \qquad (11)$$

$$\approx Pr(T(i)) * Pr(\,q \mid T(i)\,) \qquad (12)$$

Ti → is the topic i of each page.
Pr (T(i)) → means the proportion of pages related to topic i in the whole pages set.
Pr (Ti | q) → means the probability of the query q belonging to topic i.

The topic sensitive page rank used in the Time Rank is given by

$$TPSRt(j) = \alpha \sum_{i \in B} \frac{TSPR(i)}{\mid Fi \mid} + (1-\alpha).Et(i) \qquad (13)$$

Single jump probability 1/n is replaced by Et = {E(1), E(2)............ E(n)}, n is the no. of topics.

$$E(i) = \begin{cases} 1/n_t \\ 0 \end{cases}$$

Where n_t → number of pages related to topic.

There are unit n TSPR scores related to topics .It is calculated statically offline. When some period of search engines the time vector associated with topics for each page will be calculated

Table 4: Iterative Calculation for Weighted PageRank

ITERATIONS	0	1	2	3	4	5	6	7	8	9	10	11	12	
WPR (C)	1	3.65	1.581667	1.322771	1.275006	1.351856	1.291874	1.284366	1.282981	1.28521	1.28347	1.283252	1.283212	
WPR (F)	1	0.391667	0.767083	0.474069	0.437393	0.430626	0.441513	0.433015	0.431952	0.431756	0.432071	0.431825	0.431794	
WPR (G)	1	0.533333	1.284167	0.698139	0.624785	0.611252	0.633026	0.616031	0.613904	0.613511	0.614143	0.61365	0.613588	
WPR (H)	1	0.4625	0.363333	0.522885	0.398355	0.382767	0.379891	0.384518	0.380907	0.380455	0.380371	0.380505	0.380401	
WPR (I)	1	0.4625	0.363333	0.522885	0.398355	0.382767	0.379891	0.384518	0.380907	0.380455	0.380371	0.380505	0.380401	
WPR (J)	1	0.306667	0.276208	0.270589	0.27963	0.272573	0.27169	0.271527	0.271789	0.271585	0.271559	0.271554	0.271562	
WPR (K)	1	0.476667	0.354833	0.332356	0.368521	0.340294	0.33676	0.336109	0.337157	0.336339	0.336236	0.336217	0.336248	
WPR (L)	1	0.476667	0.354833	0.332356	0.368521	0.340294	0.33676	0.336109	0.337157	0.336339	0.336236	0.336217	0.336248	
WPR (M)	1	0.306667	0.276208	0.270589	0.27963	0.272573	0.27169	0.271527	0.271789	0.271585	0.271559	0.271554	0.271562	

and thence every page is allotted as page rank supported time visited as in eq (14).

$$TIMEPRt(j) = TSPRt(j) * T(t) \qquad (14)$$

Where time vector Tv = {T(1), T(2), ...T(n)}
T(i) - user's total visiting time of a page related to topic i. Time rank means that irrespective of similarity of similar link structure of two web pages, the page having longer visited time gets the high score.

3 Experiment

To compare the different page ranking algorithms we simulate the working of these algorithms for the same we have carried out the following activities.

a) Generate a website with rich hyperlink structure so that different page rank could be calculated for this web site and then they may be compared on the basis of their working.
b) A web access graph is constructed which reflects he structure and connections of different web pages as shown in Figure 1.
c) For different links a data set is generated showing the URLs, Outgoing URLs, Keywords, frequency of keywords, relevancy as shown in Table 1.
d) Convert it in easy understanding simplified web graph as Figure 2 on basis of coding for nodes shown in table 2 selecting a partial graph for calculations as shown in Fig. 3.
e) Calculating different page rankings for it using iterative method.
f) Evaluating the results and give the empirical analysis of these algorithms

The PageRank values of pages in figure 3 calculated by eq. (1) Using d = .85.

$$PR(C) = .25 + .85 * \big(PR(J) + PR(K) + PR(L) \quad + PRM$$

$$PR(F) = .25 + .85 * (PR(C)/2)$$

$$PR(G) = .25 + .85 * (PR(C)/2)$$

$$PR(H) = .25 + .85 * (PR(G)/2)$$

$$PR(I) = .25 + .85 * (PR(G)/2)$$

$$PR(J) = .25 + .85 * (PR(H)/3)$$

$$PR(K) = .25 + .85 * \left(\tfrac{PR(H)}{3} + \tfrac{PR(I)}{3} \right)$$

$$PR(L) = .25 + .85 * (\frac{PR(H)}{3} + \frac{PR(I)}{3})$$

$$PR(M) = .25 + .85 * \left(\frac{PR(I)}{3}\right)$$

On the basis of the above equations the pagerank is calculated using iteration as shown in Table 3.
The calculation of Weighted PageRank algorithm for the web graph of Figure 3 calculated on the basis of eq. 2, 3, & 4 is as follows –

$$WPR(C) = .25 + .85 * (WPR(J) + WPR(K) + WPR(L) + WPR(M))$$

$$WPR(F) = .25 + .85 * (WPR(C) * \frac{1}{2} * \frac{1}{3})$$

Table 5: Iterative Calculation for PRLV

ITERATIONS	0	1	2	3	4	5	6	7	8	....	35	36	37	
PRLV (C)	1	3.65	2.545	2.075	2.012	3.063	2.624	2.438	2.413	....	2.6763	2.6763	2.6769	
PRLV (F)	1	0.454	0.995	0.769	0.673	0.66	0.875	0.785	0.747	....	0.796	0.796	0.796	
PRLV (G)	1	0.896	2.608	1.894	1.591	1.549	2.229	1.945	1.825	....	1.9791	1.9789	1.9789	
PRLV (H)	1	0.486	0.462	0.866	0.697	0.626	0.616	0.776	0.709	....	0.7174	0.7173	0.7172	
PRLV (I)	1	0.864	0.8	1.851	1.413	1.227	1.201	1.618	1.444	....	1.4652	1.4649	1.4648	
PRLV (J)	1	0.439	0.342	0.337	0.414	0.382	0.368	0.366	0.397	....	0.3855	0.3855	0.3855	
PRLV (K)	1	1.006	0.724	0.695	1.183	0.98	0.893	0.881	1.075	....	1.0036	1.0039	1.0038	
PRLV (L)	1	0.722	0.587	0.564	0.938	0.782	0.716	0.707	0.855	....	0.8004	0.8006	0.8006	
PRLV (M)	1	0.533	0.495	0.477	0.774	0.65	0.598	0.59	0.708	....	0.665	0.6651	0.6651	

$$WPR(G) = .25 + .85 * (\ WPR(C) * \frac{1}{2} * \frac{2}{3}\ )$$

$$WPR(H) = .25 + .85 * (\ WPR(G) * \frac{1}{2} * \frac{1}{2}\ )$$

$$WPR(I) = .25 + .85 * (\ WPR(G) * \frac{1}{2} * \frac{1}{2}\ )$$

$$WPR(J) = .25 + .85 * (\ WPR(H) * \frac{1}{3} * \frac{1}{5}\ )$$

$$WPR(K) = .25 + .85 * (\ \ WPR(H) * \frac{2}{5} * \frac{1}{3} + WPR(I) * \frac{2}{5} * \frac{1}{3}\ )$$

$$WPR(L) = .25 + .85 * (\ \ WPR(H) * \frac{2}{5} * \frac{1}{3} + WPR(I) * \frac{2}{5} * \frac{1}{3}\ )$$

$$WPR(M) = .25 + .85 * (\ WPR(I) * \frac{1}{3} * \frac{1}{5}\ )$$

On the basis of the above equations the Weighted PageRank is calculated using iteration as shown in Table 4.

The calculation of PRLV algorithm for the web graph of Figure 3 calculated on the basis of eq. 6 & 7 is as follows –

$$PRLV(C) = .25 + .85 * (PRLV(J) + PRLV(K) + PRLV(L) + PRLV(M))$$

$$PRLV(F) = .25 + .85 * (\ PRLV(C) * \frac{12}{50}\ )$$

$$PRLV(G) = .25 + .85 * (\ PRLV(C) * \frac{38}{50}\ )$$

$$PRLV(H) = .25 + .85 * \left(\ PRLV(G) * \frac{5}{18}\right)$$

$$PRLV(I) = .25 + .85 * (\ PRLV(G) * \frac{13}{18}\ )$$

$$PRLV(J) = .25 + .85 * (\ PRLV(H) * \frac{2}{9}\ )$$

$$PRLV(K) = .25 + .85 * (\ \ PRLV(H) * \frac{5}{9} + PRLV(I) * \frac{4}{12}\ )$$

$$PRLV(L) = .25 + .85 * (\ \ PRLV(H) * \frac{2}{9} + PRLV(I) * \frac{4}{12}\ )$$

$$PRLV(M) = .25 + .85 * (\ PRLV(I) * \frac{4}{12}\ )$$

Table 6: Hit Count Data

OUT URL	CURRENT URL	HITS
/iimt_s/Default.aspx		5
/iimt_s/courses.aspx	/iimt_s/Default.aspx	18
/iimt_s/BTECH.aspx	/iimt_s/courses.aspx	38
/iimt_s/cs.aspx	/iimt_s/BTECH.aspx	5
/iimt_s/it.aspx	/iimt_s/BTECH.aspx	13
/iimt_s/cscircular.aspx	/iimt_s/it.aspx	4
/iimt_s/Default.aspx	/iimt_s/Default.aspx	9
/iimt_s/mca.aspx	/iimt_s/courses.aspx	12
/iimt_s/Default.aspx	/iimt_s/mca.aspx	12
/iimt_s/aboutus.aspx	/iimt_s/Default.aspx	5
/iimt_s/Default.aspx	/iimt_s/aboutus.aspx	6
/iimt_s/contactus.aspx	/iimt_s/Default.aspx	6
/iimt_s/Messages.aspx	/iimt_s/Default.aspx	6
/iimt_s/Default.aspx	/iimt_s/Messages.aspx	8
/iimt_s/cstimetable.aspx	/iimt_s/cs.aspx	2
/iimt_s/courses.aspx	/iimt_s/cscircular.aspx	9
/iimt_s/cscircular.aspx	/iimt_s/cs.aspx	5
/iimt_s/courses.aspx	/iimt_s/csnotice.aspx	13
/iimt_s/csnotice.aspx	/iimt_s/cs.aspx	2
/iimt_s/courses.aspx	/iimt_s/cstimetable.aspx	8
/iimt_s/default.aspx	/iimt_s/contactus.aspx	7
/iimt_s/ittimetable.aspx	/iimt_s/it.aspx	4
/iimt_s/courses.aspx	/iimt_s/ittimetable.aspx	4
/iimt_s/csnotice.aspx	/iimt_s/it.aspx	4
/iimt_s/aboutus.aspx	/iimt_s/default.aspx	5
/iimt_s/courses.aspx	/iimt_s/default.aspx	18

Table 7: The Number of Relevant Pages and Relevance Value for Query IIMT

	Number of Relevant Pages			Relevany Value			
Size of page set	PageRank	Weighted PageRank	PRLV	PageRank	Weighted PageRank	PRLV	
3	1	1	1	0	0.5	0.5	
6	1	2	2	1.5	2	2.9	
9	3	3	3	4.8	6.1	6	

Table 8: The Number of Relevant Pages and Relevance Value for Query Student

	Number of Relevant Pages			Relevany Value			
Size of page set	PageRank	Weighted PageRank	PRLV	PageRank	Weighted PageRank	PRLV	
3	1	1	1	0.2	0.7	0.7	
6	1	2	2	2.3	2.8	3.8	
9	3	3	3	7.2	9.2	9.1	

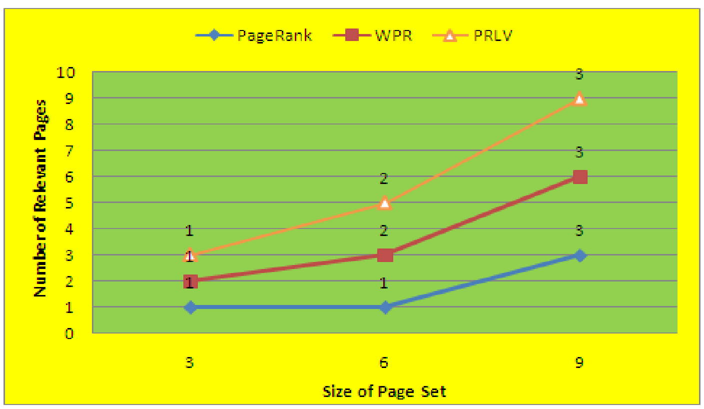

Figure 4: Number of Relevant Pages Verses Page Set for Query IIMT

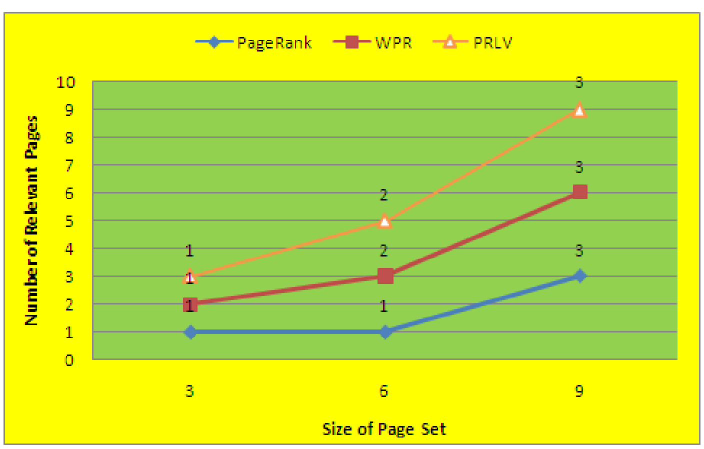

Figure 5: Number of Relevant Pages Verses Page Set for Query Student

On the basis of the above equations the PRLV is calculated using iteration as shown in Table 5 and the hit count of a web site IIMT made for simulation is calculated as shown in Table 6.

4 Outcome Of Comparison Of Various Web PageRanking Based Algorthms

The query of topics "IIMT" and "Students" are used for evaluation of the PageRank, Weighted PageRank and PRLV. The comparison is done on the basis of number of relevance pages for both query and Relevance values calculated indicating the importance of order of web pages generated through different algorithms for this the column charts and line charts are used.In the result of query for IIMT and Student the number of relevant pages are been counted from the set of pages. The sets are constructed from the query results as first three pages, six pages and nine pages. The relevancy values are calculated using relevancy equation [2], and values along with the position of page in the result list which is sorted on the basis of different page rank.

As per the result shown in Table 7, 8 the Weighted PageRank and PRLV both gives similar number of relevant pages in each set which is greater than number of relevant pages generated in some set through PageRank for both the query IIMT and Student as shown in Figure 4, 5. The relevancy which is shown by relevance value [2], of pages in result set of Weighted PageRank is better than PageRank and relevancy of pages in result set of PRLV is better than Weighted PageRank in some cases as shown in Figure 6, 7. If the size of page set increased, the resultant relevancy value for Weighted PageRank and PRLV is almost similar.

5 Conclusion and Future Scope

This paper gives the working of different page ranking algorithms used to retrieve the relevant pages through search engines. The centre of attraction is the comparative analysis based on empirical study which shows importance of different ranking algorithms for the web pages. As well as the attention is drawn towards the relevance calculation of different page ranking algorithms through the simulation.

The result shows that when the size of page set to be searched is small the PRLV is best out of PageRank, Weighted PageRank and PRLV but if the size of page set to be searched is large the performance is similar.These results can be considered to generate a new algorithm in future to eliminate the problems and enhance the working of ranking algorithm to get the better relevant results.

Acknowledgment

We wish to thank Dr. Sohan Garg, Dr. A. K. Solanki, Dr. S. B. Gupta and Dr. Deepa Sharma for their valueable guidance and support during the work.

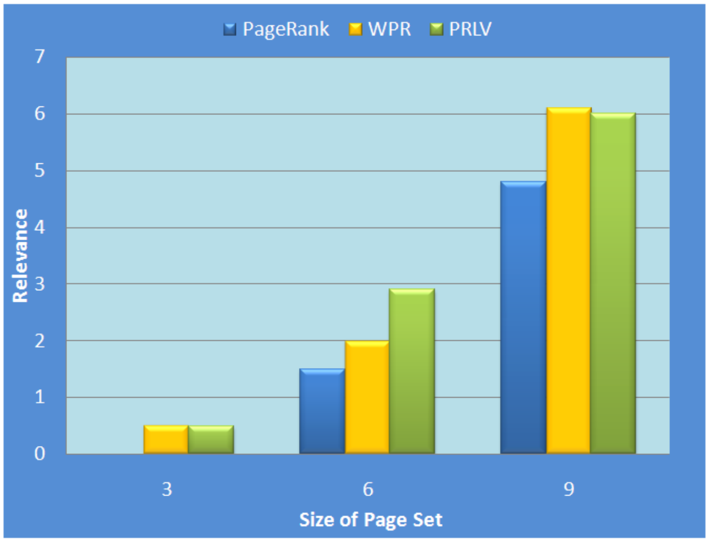

Figure 6: Relevance Verses Page Set for Query IIMT

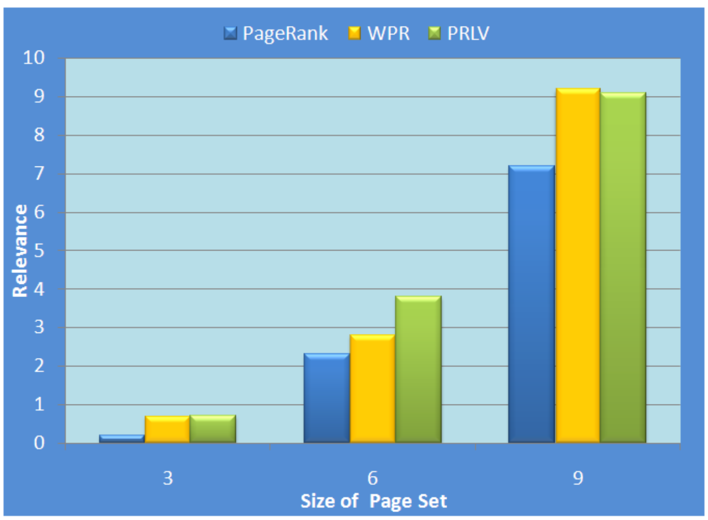

Figure 7: Relevance Verses Page Set for Query Student

REFERENCES

[1] S. Brin and L. Page (1998). The Anatomy of a large-scale hypertextual Web search engine. In seventh international World Wide Web Conference, Brishbane, Australia, 1998.

[2] Wenpu Xing, Ali Ghorbani. Weighted PageRank Algorithm[C], Proceedings of the Second Annual Conference on Communication Networks and Services R -esearch, IEEE, 2004.

[3] Sharma, A.K., Duhan, N. and Kumar, G. (2010) A Novel Page Ranking Method based on Link- Visits of Web Pages. International Journal of Recent Trends in Engineering and Technology, Vol. 4, No. 1, pp 58-63.

[4] Ricardo Baeza-Yates and Emilio Davis ,"Web page ranking using link attributes" , In proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, PP.328-329,2004

[5] H Jiang et al., "TIMERANK: A Method of Improving Ranking Scores by Visited Time", In proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008.

[6] Dimitri, P. Betsekas and John N. Tsitsiklis, Introduction to Probability. Athena Scientific,2002.