Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 3
March 2024

Performance Evaluation of Page Ranking Algorithm Based on Counting of Link Visits (PRCLV) for Effective Information Retrieval on World Wide Web

Zaved Akhtar
Research Scholar (CSE),
Dr. APJ AKTU, Lucknow. U.P., India
Ravindra Kumar
(Ex) Director
NIET, Gr. Noida,U.P., India
Umesh Chandra Jaiswal
Professor (IT & Applications Deptt.)
MMMUT, Gorakhpur, U. P., India
Mahesh Kumar Singh
Associate Prof. (CS Deptt.)
ABES Engg.College, Ghaziabad, U.P., India

Abstract:- Due to an increasing number of web users and traffic, knowledge investigators rely greatly on engines for searching to extract useful information. The availability of innumerable textual, video, audio, and different kinds of content has raised the duty of search engines. The searching engine delivers relevant data about Internet users' queries based on interest, link order, and so on. It does not, however, ensure the accuracy of the data. The position module is a significant variable that decides the efficiently the search engine operates. Getting valuable information has demonstrated to be a debilitating undertaking. Site page ranker, a part that is professed to have been the vital game changer in Google's accomplishment, is quite possibly of the main component that solid the reception of page search administration. This article explains page ranking systems for internet mining that are based on content, structure, and usage mining. The suggested Page Ranking Algorithms is based on the number of link hits. Because each product has its own webpage dedicated to decryption, the top traffic links may be utilized to build a list of ideas for someone carrying out an online information search as well may use for develop web based recommendation system on the basis of users behaviors .

Keywords:- Information Retrieval, Data Mining, PageRank, Weighted PageRank, Web Graph, Web Mining.

1. Introduction

Web mining is the practice of employing tools for data mining to explore the internet for accessible data. This "hidden data," or "knowledge," can be found in the text of the websites, the WWW's link order, or the server logs [1]. The World Wide Web is a massive collection of hyperlinked and disparate material that includes text, photos, audio, video, and information. Because of the fast development of the resources accessible through the world's largest website and the rising expectations of users, managing content on internet sites and meeting the needs of customers is getting harder. We are unaware while being enveloped by facts. Users have begun to rely on finding data tools to seek, recover, filter, and analyze data.

A. Search engine [8] in response to user queries, the system examines its index finding appropriate papers (the paperwork which are presumably associated to the query and are expected to be fascinating), rates the relevant articles, and then shows them as results. The entire process may be broken down into a few different tasks:

B. Crawler [14, 15] The Crawler's job is to visit as many sites he can in order to obtain the pertinent details. The data provided is supposed to be remaining for future reference by the search tool.

C. ***Indexing*** The data that is produced by a crawler must be saved in order for the engine that searches to access it. Considering the user would be standing close to of his a computer system, time the interaction becomes an important aspect. To that end this material has been followed to save the time expected to investigate it.

D. ***Searching*** A web-based search tool is a graphical depiction of the individual interface necessary for finding the information at hand. It fills in as a channel for trades between the client and the information vault.

E. ***Ranking/Soring*** if a user find out regarding something generally (like a Data Science Course), there will be an incredible number of parts corresponding to his or her question, but just a fraction of this huge quantity of knowledge will prove truly fascinating for that individual. As result, engines that search employ ranking systems that organize the listings.

F. ***Recommended System***. Each its description has a distinctive set of links. As a consequence, each product and service offered by e-commerce enterprises has its own page. The ranking system assigns a score to each page, indicating its overall popularity [3, 4], which can be used to select the most popular products. The time of online organizations is here. A huge number of organizations offer web-based channels for buying and selling products, and PageRank contributes in deciding how pervasive the thing is.

There are seven sections in this paper. The initial segment characterizes web mining and their different varieties. The second section encompasses over the related study of web page ranking algorithms, the third portion explores problem identification, the fourth part come up with an implement a page ranked algorithm based on link hits, the fifth section is comparison assessment, the sixth chapter addresses how it is used in efficient data retrieval, online business, and so on, and the final part talks about the conclusion that following the paper.

2. Web Mining

Information extraction is the method involved with extricating intriguing perceptions or patterns from enormous informational collections that is non-insignificant, certainly, beforehand obscure, and perhaps supportive. In this regard Web Mining plays a major role to extract effective and useful data and information from World Wide Web. Services of Web Mining can be classified into three distinct categories [6] as Mining of Web Content, Mining of Web Usage, and Mining of Web Structure:

A. Web Content Mining (WCM)

WCM defines a method of searching internet data as well as additional sources for data automatically. The priority is on the contents of the website rather than its links. It can be utilized on website or search engine results pages. WCM is distinguished by two points of view: the perspective of Information Gathering (IR) to be Views and the Database (DB) the view. A large portion of research in IR employs an "assemblage of phrases" to convey unorganized content, and is based on metrics regarding just one term in independence. With data that is semi-structured, all of the pieces make advantage of the HTML structure accessible within the texts. Web mining is a process that always tries to figure out the framework of a web page so as to turn it into a computerized representation.

B. *Web Structure Mining (WSM)*

Formal rundowns of pages and sites are made by WSM. In a normal Web diagram, hub are pages, and edges are associated across destinations that are connected to one another. While WCM works to a great extent on the format of within records, WSM tries to distinguish the connection design of associations at the between report layer.

Web design mining plans to track down the model that upholds the web's connected designs. The topology of hyperlinks, with or without a link descriptor, serves as the foundation for the model. This approach might be utilized to bunch sites and gives information, for example, equals and connections between Sites. Furthermore, critical derived information can be found in the web's connection structure, which can be utilized to sort or rank sites. It is feasible to peruse a connection from page A to page B and the creator of page B endorsing page A, for instance. In addition to term searching, novel methods that make use of this link structure can be used to create a Yahoo-like hierarchical or identify online groups. They frequently beat IR calculations as far as subjective exactness since they depend on additional

information than essentially the material of the page. While it is possible to change the connection design of the Internet on the neighborhood level, doing it on a worldwide scale is troublesome. Consequently, global link analysis algorithms have fairly robust anti-spam protections.

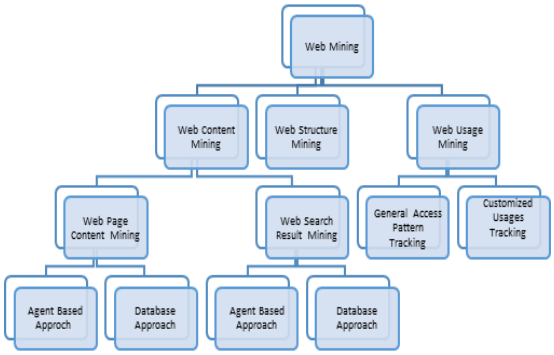

Fig 2.1 Web Mining Categories

Neelam Duhan, A. K. Sharma, Komal Kumar Bhatia [6]

C. *Web Usage Mining (WUM)*

worldwide Utilization Mapping (WUM) aims to identify customers navigation habits in web statistics as well as significant facts in additional data collected via interactions between users when the water the web. It addresses approaches for predicting how consumers will behave when they utilize the web. This sort of web digging accommodates the acquiring of Web information on access for Pages.

Website adequacy, insight for business, online customization; site modification, use description, webpage classification, and additional use cases are only a few of the use cases that are included under each of the three web mining categories stated above. Page Ranking [9] is frequently utilized by search engines in order for finding higher-importance sites. The PRCLV look at is executed by using the web's structure and web mining strategies for evaluating internet pages for its use in the online system of recommendations.

3. Related Works on Ranking Algorithms

The internet is vast and diverse, and a single query may be related to multiple websites. What reason a strategy/calculation is utilized for sorting out the total pages that are suitable to a purchaser's inquiry? All of the strategies treat internet pages as a connected graph, with page established as links and nodes designated as edge.

A. PageRank Algorithm (PR)

The PageRank algorithm, named after Larry Page, a co-benefactor of Google's web crawler, was developed by Sergey Brin and Larry Page, it investigates the world's connection construction to survey the worth of online locales [9]. It investigates remarks and spreads the position by means of connections. In this way, a site has a phenomenal position in the event that all of its associations' rankings is high. An easy overview of the site's position is provided below:

$$PR(p) = (1-c)\sum_{q\in I(p)} \frac{PR(q)}{O(q)} \qquad (3.1)$$

A normalizing factor c is employed in the algorithmic calculation of Google. It is crucial to keep in mind that 0c 1 since there are pages with no external links and their importance is removed. Later, Sitemap was modified as it came to light that not everyone can browse direct links on the the World Wide Web.

$$PR(p) = (1-d) + d\sum_{q\in I(p)} \frac{PR(q)}{O(q)} \qquad (3.2)$$

Where d is a hose factor that is much of the time set to 0.85 (any amount somewhere in the range of 0 and 1), d might be characterized as the likely number for people following the connections, and (1 d) may be considered the position of the page conveyed from non-straightforwardly related pages. Notice the arranged diagram under.

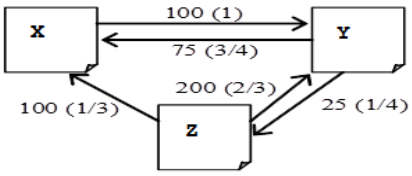

Fig 3.1 Example graph [a]

Rank of the pages X, Y and Z are calculated on damping factor =0.5 using (3.2). After calculation, ranking of the pages are as: PR (X) =1.2, PR (Y) =1.2, and PR (Z) =0.8.

B. Weighted Page Rank Algorithm (WPR)

Wenpu Xing and Ali Ghorbani [11] introduced Generalized PageRank (WPR) as a variant to normal Rank. It ranks pages based on their importance compared to the web graph's link structure [5].

Instead of splitting a website's rank score evenly in its outbound linked sites, this method assigns higher rank values to additional significant pages. Each outline page is assigned a value based on its popularity. The amount of in links and out links is used for assessing reputation.

$$PR(p) = (1-d) + d \sum_{q \in I(P)} PR(q) W^{in}_{(q,p)} W^{out}_{(q,p)}$$

(3.3)

For in links and out links, respectively, Win (q,p) and Wout (q,p) are provided.

$$W^{in}_{(q,p)} = \frac{I_p}{\sum_{v \in R(q)} I_v} \qquad 3.3.1$$

$$W^{out}_{(q,p)} = \frac{O_p}{\sum_{v \in R(q)} O_v} \qquad 3.3.2$$

Where Iv, Ip represent in-links whereas Ov, Op represent out-links for page v and page p, respectively. With damping factor d=0.5, page scores for pages X, Y, and Z are determined using (3.3); PR (X) =0.65, PR (Y) =0.93, and PR(Z) =0.60 are the final scores for pages X, Y, and Z, respectively.

C. Page Content Rank Algorithm (PCR)

Jaroslav Pokorny and Jozef Smizansky [10] created Page Content Rank (PCR) as a novel ranking method for page relevancy that makes use of WCM technology. It utilize different calculations that accept to be imperative to checking page content. The meaning of the words on the given page determines the significance of the page [7]. The import of an expression will be made sense of as it connects with a specific q. As its center order association, PCR uses an organization of neurons. In PCR, the total significance of all of the words on page P is used to determine its value. The second immediate is given as an example to promote the key word while reducing the others [4].

Page_importance(P)=sec_moment({importance(t): t ∈ P}) (3.4)

D. Hyperlinked Induced Topic Search Algorithm (HITS) [12, 13]: The strategy proposed expects for each inquiry subject an assortment of "authority" or "authoritative" pages or websites destinations that are relevant and normally visited zeroed in on the issue, as well as "hub" pages or websites links that incorporate valuable linking [2] with appropriate sites, containing links to many related authorities.

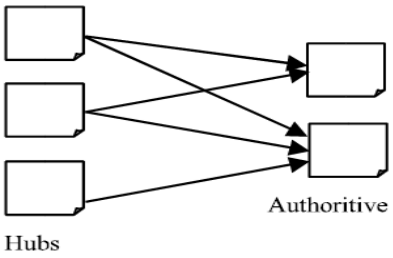

Fig 3.1 Authorities, Hubs [b]

Alan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal, and Panayiotis Tsaparas [13]

Working of HITS: The HITS has two stages of operation. In the Example stage, an assortment of the suitable pages connected with the set hunt is gathered, for example a sub-chart S of G containing significant position pages is recovered [3]. Using the result of the testing phase and the resulting calculation, the repetitive stage finds hubs and authority.

$$H_p = \sum_{q \in I(p)} A_q \qquad (3.4)$$

$$A_p = \sum_{q \in B(p)} H_q \qquad (3.5)$$

As per above, Hp represent weight of the hub and Ap represent weight of authorities. I (p) and B (p) are the reference and referrer destinations for page p, separately. What follows is an explanation of algorithmic contrasts.

Table 3.1: Ranking Algorithms Comparison Chart

Algorithm	**Page Rank**	**WPR**	**PCR**	**HITS**	**SALSA, SimRank, Randomise HITS, etc**
Technique Used	Extracting Web Structures	Web Structure Mining	Web Content Exploration	Web Content Mining, Website	Web Page Mining, network Structure

				Structure Monitoring	Research
Description	Scores are determined at indexing time rather than on the fly. The pages have been arranged in order of meaning.	Scores are generated at indexing time. Unbalanced distribution of scores, pages and order by importance.	Scores are determined on the fly. Items gathered are relevant to the search query, i.e. pertinent paperwork have been returned.	On-the-fly computation of hub and power ratings. Articles that are simultaneously significant and pertinent show up.	On-the-fly assessment of hub and power scores. Pages that are equally relevant and important are presented.
I/P Parameters	Backlinks	Backlinks, forward Links	Content	Backlinks, forward Links	Backlinks, forward Links
Levels of work	n	1	Nil	n	n
Time complexity	O(log n)	< O(log n)	O(m*)	> O(log n)	> O(log n)
Relevancy	No	No	Yes	Yes	Yes
Importance	Yes	Yes	No	Yes	Yes
Rank of result	Low	High	Low	High	High
Stability in Results	High	High	Low	Low	High
Limitations	Ratings are computed at Indexing time, not on the fly. The pages have been structured in order of significance.	Relevance is neglected. The method determines values at a single layer.	The importance of pages has been completely neglected.	Difficulties with subject drifting and productivity.	Problems with subject drift, complexities. And velocity.

* N: the quantity of sites *m: how much expressions on each page.

4. Challenges on Page Ranking Algorithms

The key refers to and barriers with discussed algorithms can be summed up as follows:

A. Rank quality of PageRank

The talked about ranking methods were of great quality, as seen by the performance of Google (or their continued usage). On the other hand, multiple improvements may be added to it.

B. Data Mining Technique of PageRank

Page Rank philosophy just used Web Construction Learning and Web Material Mining procedures; it didn't utilize Web Use Mines are which could extraordinarily expand the nature of site pages positioned by the information that the client demands.

C. PageRank is Static in Nature

The relevance or rank value of every web page is unaltered in the calculation of PageRank. Only the web's link architecture impacts its ranking.

5. Algorithm for Page Ranking Based on the Count of Link Visits(PRCLV)

To determine efficiency of the web pages for user, it takes user behavior pattern search and accordingly assigned importance of the web pages. PRCLV (Page Rank Count of Link Visits) usage concept of Structure and Web Usage mining to complete the task from gathering use characterization using ultimate ranking assessment:

- Preservation of access details (hits) on a page's inbound link in pertinent network logs.
- The focused web crawler gets pages and access information.
- Weights are given to each page link based upon its probability of being clicked by visitors.
- The ultimate rank of webpages is calculated by the relative weights of their inbound relationships.
- Finding of selected pages suiting user requests.

A. Calculation of Visits (hits) of links

The following is the weighting of each friendly connection from page p to page o if p is a page with an outside interface set of O (p) and each outside interface is associated with a mathematical number addressing visit-count (VC):

$$Weight_{link}(p,q) = \frac{VC(p,q)}{\sum_{q' \in O(p)} VC(p,q')} \quad (5.1)$$

B. Page Rank based on Link Count Visits (PRCLV).

If p is a page in set B(p) with inbound-linked sections, and its ranking (PRCLV) is given by:

$$PRCLV(p) = (1-d) + d\left(\sum_{b \in B(p)} PRCLV(b) Weight_{link}(b,p)\right)$$

Where Weight link is the link weight derived by (5.1) and d is the PageRank damping effect. Page rank is acquired utilizing the cycle procedure. The picture beneath is an illustration of this. With d=0.5, these conditions might be effectively settled by utilizing the iterative travel, giving these outcomes using Figure [3.1 [a]] PRCLV: PRCLV(X) = 1.08, PRCLV(Y)= 1.26, PRCLV(Z)= 0.66.

Page	URL	Hits
http://localhost/webgraph/university.html	http://localhost/webgraph/jnu.html	45
http://localhost/webgraph/Research Areas IIT Delhi...	http://localhost/webgraph/Prof. K. K. Biswas.html	27
http://localhost/webgraph/university.html	http://localhost/webgraph/university of hydarabad....	44
http://localhost/webgraph/university.html	http://localhost/webgraph/IIT DELHI.html	47
http://localhost/webgraph/university.html	http://localhost/webgraph/university of delhi.html	27
http://localhost/webgraph/university.html	http://localhost/webgraph/university of pune.html	7
http://localhost/webgraph/university.html	http://localhost/webgraph/anna university.html	23
http://localhost/webgraph/university.html	http://localhost/webgraph/IIIt Hyderabad.html	12
http://localhost/webgraph/Research Areas IIT Delhi...	http://localhost/webgraph/Sanjiva Prasad IIT Delhi....	38
http://localhost/webgraph/Research Areas IIT Delhi...	http://localhost/webgraph/Sorav Bansal IIT Delhi.h...	32
http://localhost/webgraph/jnu.html	http://localhost/webgraph/IIT DELHI.html	13
http://localhost/webgraph/Du project.html	http://localhost/webgraph/Rakesh Kumar.html	23
http://localhost/webgraph/IIT DELHI.html	http://localhost/webgraph/Research Areas IIT Delhi....	18
http://localhost/webgraph/jnu.html	http://localhost/webgraph/Project in JNU.html	4
http://localhost/webgraph/Sanjiva Prasad IIT Delhi...	http://localhost/webgraph/Indian Council of Agricu...	10
http://localhost/webgraph/jnu.html	http://localhost/webgraph/university of delhi.html	30
http://localhost/webgraph/Research Areas IIT Delhi...	http://localhost/webgraph/Anshul Kumar IIT Delhi.h...	5
http://localhost/webgraph/Research Areas IIT Delhi...	http://localhost/webgraph/Naveen Garg IIT DELHI.ht...	21
http://localhost/webgraph/jnu.html	http://localhost/webgraph/Research Areas IIT Delhi....	23
http://localhost/webgraph/anna university.html	http://localhost/webgraph/anna university project....	15
http://localhost/webgraph/university of hydarabad....	http://localhost/webgraph/university of pune.html	21
http://localhost/webgraph/university of pune.html	http://localhost/webgraph/IIIt Hyderabad.html	60
http://localhost/webgraph/university of pune.html	http://localhost/webgraph/Vaibhav V. Kawade pune.h...	26
http://localhost/webgraph/university of delhi.html	http://localhost/webgraph/Du project.html	12
http://localhost/webgraph/IIIt Hyderabad.html	http://localhost/webgraph/R & D of IIIT Hyderabad....	22
http://localhost/webgraph/Project in JNU.html	http://localhost/webgraph/Arun K. Agni jnu.html	20
http://localhost/webgraph/Project in JNU.html	http://localhost/webgraph/V. Rajamani jnu.html	3
http://localhost/webgraph/Project in JNU.html	http://localhost/webgraph/Indian Council of Agricu...	60
http://localhost/webgraph/Project in JNU.html	http://localhost/webgraph/V. Subramanian jnu.html	12
http://localhost/webgraph/V. Rajamani jnu.html		0

Fig 5.1 the amount of hits on the websites

C. Experimental Performance Evaluation

A prototype is made to make sure that proposed algorithm is useful and works well. In this regard, a site is created using HTML, CSS, Bootstrap, JSP, JavaScript, and Apache Tomcat Server is utilized as the server to store data from clients. After data collection, We have performed practical part. Fig. 5.1 displays the web chart of the site. Customers may log in using their email address and access the various pages on the website, and data can be gathered by downloading the model on the local server. Figure 5.2 depicts the numerous URLs for each page as well as their respective visitor counts. Process the page position of every single association (page) using the condition (5.2), and subsequently demand the associations in a

lessening gathering of page rank, as displayed in Fig. 5.3.

Web Page URL	PR	PRCLV
localhost:8080/webgraph/introduction.jsp	0.672	0.8739
localhost:8080/webgraph/laptop.html	0.4391	0.1236
localhost:8080/webgraph/mobile.jsp	0.3721	0.64
localhost:8080/webgraph/laptop/dell.jsp	0.5612	0.1265
localhost:8080/webgraph/electronics/lg.jsp	0.423	0.675
localhost:8080/webgraph/mobile/nokia.jsp	0.421	0.238
localhost:8080/webgraph/electronics/tv/panasonic.jsp	0.867	0.6821
localhost:8080/webgraph/electronics/lg.jsp	0.423	0.8734
localhost:8080/webgraph/mobile/apple.jsp	0.426	0.012
localhost:8080/webgraph/introduction/tv/samsung.jsp	0.869	0.142
localhost:8080/webgraph/electronics/tv/panasonic.jsp	0.867	0.69
localhost:8080/webgraph/introduction/cd/pan.jsp	0.698	0.046
localhost:8080/webgraph/mobile/xiumi.jsp	0.427	0.9348
localhost:8080/webgraph/laptop/lenovo.jsp	0.5613	0.1265
localhost:8080/webgraph/mobile/apple.jsp	0.426	0.082
localhost:8080/webgraph/electronics/samsung.jsp	0.869	0.642
localhost:8080/webgraph/mobile/nokia.jsp	0.421	0.638
localhost:8080/webgraph/electronics/tv/panasonic.jsp	0.867	0.724

Fig 5.2 Page Ranking (both PR and PRCLV)

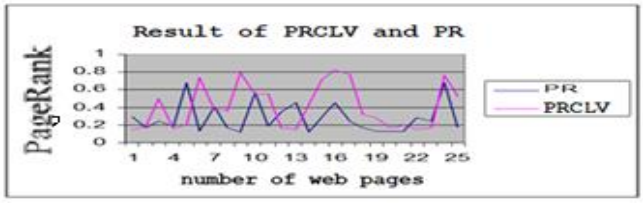

Fig 5.3 PRCLV variant with PR

6. The Comparison of PRCLV,PR, and WPR

Table 6.1 Page Rank Algorithms Compared

Table 6.1 clearly demonstrates that PRCLV is superior to PR and WPR.

Algorithm Parameter	*PageRank (PR)*	*Weighted Page Rank (WPR)*	*Page Rank Counts on Link Visits (PRCLV)*
Description	Scores are determined while ordering. The pages are arranged in a significant order.	Scores are computed during indexing. The pages are organized in order of significance.	Scores are computed during indexing. Pages are categorized by significance and relevancy.
Mining Technique Used	Web Structure Mining	Web Structure Mining	Web Usage and Structural Mining
Rank Distribution	Outgoing links have equal rank distribution.	The ranks of incoming links are distributed equally.	Outgoing links are assigned ranks based on their likelihood of being visited.
I/P Parameters	Page inbound links	Pages' inbound and outbound associations.	International sources, outward links, and link visit counts.

Working levels	n*	n*	n
Complexity	O(log n)	O(log n)	> O(log n)
The Characteristics of Rank	Less dynamic (rank fluctuates with link structure)	Less dynamic (rank fluctuates depending on link structure)	More fluid (rank fluctuates based on the number of visits and link arrangement)
Page relevance	No	No	Yes
The significance of page	Yes	Yes	Yes
The standard of the outcome	Low	High	High
Advantages	Rank computations with minimal effort and bother.	Rank calculation that requires the least amount of effort and complexity.	Because user input is taken into consideration, the pages provided are of excellent quality and relevance. As pages are filtered according to users' information demands, search space may be greatly reduced.

Limitations	The relevance of pages is not taken into account in the rank computation.	The relevance of pages is not taken into consideration in the rank determination.	Crawlers must expend more effort to retrieve page visit numbers from web servers.
	All connections are seen as equally valuable.	All links are considered as equally helpful.	Extra computations to determine link weights.

the top most few pages, thus search space can be reduced to a large scale during. It's given a

7. Application of PRCLV for Effective Information Retrieval

Every online store that sells commodities has product pages on its own website that describe the product's specs. Because each page has its own URL, the page rank may be used to assess how popular something is depending on just how successful the hyperlinks were.

A recommendation system aims to gauge the user's interest in a certain area of information based on previous interactions. When a consumer interacts with the website of an online store, they contribute implicit or explicit data about what they like, such as clicks, ratings, responses, and so on. PageRank is critical in connecting an email ID with the rank of a page in order to generate a suggested list for a specific customer.

8. Conclusion and Future Work

Mining intelligent facts from an immense amount of data is a tough execution, and the World Wide Web plays an important role in information acquiring and propagation. Rank methods are used to efficiently search for relevant information. Different web ranking methods have been used in various approaches. The PRCLV analyzes the user's surfing information rather than link structure for establishing the rank of content. Because it takes surfing information into account, the PRCLV method is far more flexible than other ranking techniques for efficient and effective online data search because its provides a required content in direction that we can use Web Mining technique in many applications to make an effective system to fulfill the need of the web users. Further it gives an idea to predict user's activity on web to make an system to provide them contents according to users interest.

References

[1] T. B.-Lee, R. Cailliau, A. Luotonen, H.F. Nielsen, A. Secret. "The World-Wide Web. Communications of the ACM", 37(8):76–82, 1994.

[2] Giannoulakis, S., Tsapatsoulis, N. "Filtering instagram hashtags through crowd tagging and the HITS algorithm". IEEE Trans. Comput. Soc. Syst. 6(3), 592–603, 2019.

[3] Xiaocui Zhang and Huilin Wu, "PageRank Algorithm and HITS Algorithm in Web Page Ranking", Springer Nature Switzerland AG 2021 MMIA 2021, AISC 1384, pp. 389–395, 2021.

[4] Sharma Prem Sagar, Yadav , Dumka Ankur. "Improve the Response Time with Content Mining based Web Page Ranking", 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), IEEE Access, May 2022.

[5] Hemangini, S., Apurva, A.,"An improvement of link analysis algorithm to mine pertinent links: weighted HITS algorithm based on

additive fusion of graphs by query similarity". Int. J. Comput. Appl. 176(24), 21–27, 2020.

[6] Neelam Duhan, A. K. Sharma, Komal Kumar Bhatia, "Page Ranking Algorithms: A Survey", 2009 IEEE International Advance Computing Conference (IACC 2009) Patiala, India, 6-7 March 2009.

[7] Jin L, Feng L, Liu G, Wang C., "Personal Web Revisitation by Context and Content Keywords with Relevance Feedback", IEEE Trans Knowl Data Eng 29(7):1508–1521, 2017.

[8] Andrei Broder. "A taxonomy of web search", Technical report, IBM Research, 2002.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd. "The pagerank citation ranking", Bringing order to the web. Technical report, Stanford Digital Libraries SIDL-WP- 1999-0120, 1999.

[10] Jaroslav Pokorny, Jozef Smizansky, "Page Content Rank: An Approach to the Web Content Mining".

[11] Wenpu Xing and Ali Ghorbani, "Weighted PageRank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR 04), IEEE 2004.

[12] C. Ding, X. He, P. Husbands, H. Zha, and H. Simon. "Link analysis: Hubs and authorities on the world", Technical report:47847, 2001.

[13] Alan Borodin, Gareth O. Roberts, Jeffrey S. Rosenthal, and Panayiotis Tsaparas, "Finding authorities and hubs from link structures on the world wide web", In World Wide Web, pages 415–429, 2001.

[14] B. Pinkerton." Finding what people want: Experiences with the web crawler", In The second Internation WWW Conference Chicago, 1994.

[15] J. Cho, H. Garc´ýa-Molina, L. Page. "Efficient crawling through URL ordering", Computer Networks and ISDN Systems, 30(1–7):161–172, 1998.