

Comparative Study of Web Page Ranking Algorithms

Atul Kumar Srivastava¹, Mitali Srivastava², Rakhi Garg³, P. K. Mishra⁴
Department of Computer Science, Faculty of Science^{1,2,4}, Mahila Maha Vidyalaya³
Banaras Hindu University, Varanasi, (U.P.), INDIA^{1,2,3,4}

Abstract: With the exponential growth of information on web, getting relevant information regarding user query through search engines is a tedious job today. Several search engines use link analysis algorithms to rank the web pages according to the need. But these algorithms are still lacking with efficiency, scalability and relevancy issues. This paper put forward survey of various improved ranking algorithms and their pros and cons. Further, we have included comparative study of various ranking algorithms mainly PageRank and HITS based on computation environments like Sequential, Parallel. This will help scientist, researchers, and academicians working in this area to understand the existing algorithms and develop one which is need of today's environment.

Keywords: Web mining, Web structure mining, PageRank algorithm, HITS algorithm, Improved PageRank algorithms, Comparison of various page rank algorithms.

I. Introduction

World Wide Web (WWW) has become very popular and interactive medium to broadcast information. With the exponential growth of the Web, there is a huge amount of data and information available in web, so that accessing information from web with accuracy and speed is a big challenge for both people and software. Many problems with web exist like personalization of web pages, finding useful information, creating the knowledge based on extracted information and Learning about consumers and users. These problems arise due to multiplicity of data (i.e. structured, semi structured and unstructured) present in a web, containing many redundant information, due to dynamic nature of a web etc. [2, 11]

Web mining is emerged as a broad research area to solve these issues in last few decades. The web mining field is a converging field with Database, Information Retrieval and Artificial Intelligence [1]. Web mining is the application of data mining techniques to discover and extract the useful and previously unknown information from the web data. Web mining can be classified into three main categories according to the type of data to be mined [3, 5]:

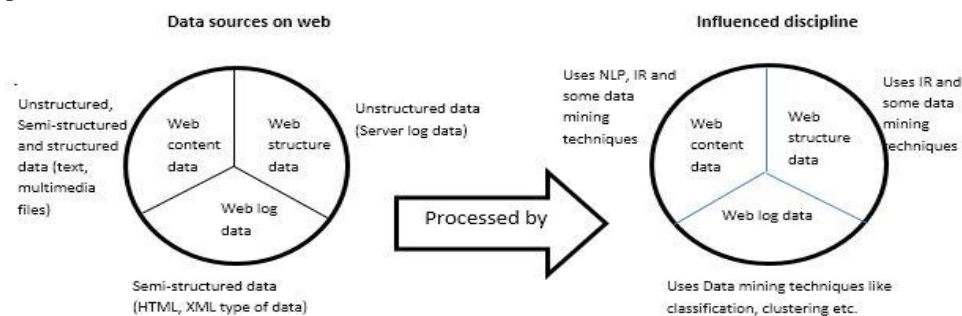


Figure 1: Categorization of web mining [3,5]

It is difficult to retrieve relevant and useful information from web because of existence of data in large amount and of its heterogeneous nature. As a result of huge and heterogeneous data available on WWW user get thousands or millions of web pages related to their query through search engine. Web users do not have much time and patience to go through all returned pages to find the relevant information that are of their interest and use [2, 3]. Web structure mining uses social network analysis and various link analysis algorithms to help web users to get information of their use on time.

Link analysis or web page ranking algorithms rank web pages based on authority, popularity and prestige of web pages. The role of link analysis algorithm is to select the web pages that are most likely be able to satisfy the user's need and provide them higher ranks [13]. There are two main Link analyses or ranking algorithms: PageRank (used by Google) and HITS. Page rank algorithms are used to rank web pages according to the relevancy. Link analysis algorithms try to keep the desired result within the top few pages, otherwise, the web search engine could be considered as unserviceable. There are following issues with the basic link analysis algorithms [3, 7, 14]:

- Uniformly sampling of web pages
- Discovering duplicate hosts
- To model in web graph

- d) To avoid web spamming
- e) To discover the content quality of web page

In this paper, we have reviewed page rank algorithms and focus on their pros and cons. Section II includes web structure mining in details. In section III, we reviewed two popular link analysis algorithms (sometimes referred as page ranking algorithm): PageRank and HITS algorithm, and their improvements over the decades. Section IV shows comparative study of ranking algorithms in environment like Sequential, Parallel and distributed. Finally, Section V concludes the Paper.

II. Web Structure Mining

Web Structure mining is a process which discovers the structural summary about the hyperlinks and web pages. Web structure mining could be divided basically into two categories: Document structure analysis and Hyperlink analysis [9]. Document Structure analysis provides structural information of web pages i.e. how contents are organized in HTML and XML tags. On the other hand Hyperlinks analysis provides how the web pages are connected with each other. Link information is used to combine with content of web to measure quality of information. It is also useful to find community of web pages according to user's common interest. Many researchers have summarized the following some basic goals for determining the quality of individual web pages and group of web pages [5, 9]:

- a) To count the local links in Web tuples (rows in a web table) in web table (representation of web pages). Local links are those links which connects different web documents which are related to same server. This also specifies the completeness of a particular web site.
- b) To count the links in web tuples which are global in nature and the links spans different web sites. This describes visibility of web page and ability to relate same or other related documents across the web.
- c) To measure the frequency of identical web tuples that appears in the indexed web pages or among the Web tables.

Link analysis provides an effective tool to calculate the importance of web pages on any particular topics. To search any web page from the search engines involves two main steps: The first step extracts the relevant pages according to user's query while the second step rank pages based on their quality [8, 10]. There are many different methods proposed to rank the web pages using hyperlinks. Following section include the brief discussion about these methods.

III. Link Analysis Algorithms

Every web search engine uses its own ranking algorithm to put resulted web pages in decreasing order of relevance so that appropriate result would be on first page. Ranking algorithm uses hyperlink analysis of web graph to rank the web pages. There are many different method proposed to identify the relevancy of web pages, some users use random walk on web graph while others uses web graph structure. There are two main ranking algorithms: PageRank (uses by Google) and HITS. Both of them rank relevancy score of individual web page. Some common link analysis algorithms are discussed below:-

A. In-degree Algorithm:

This is the oldest algorithm which ranks the web pages according to the popularity of web pages [13]. Popularity of a particular web page is measured by the number of incoming links; if any page has many incoming links then the page is more popular than less number of incoming links. Mathematically, it can be shown by the following equation [13]:

$$R(p) = Id(p) \quad (1)$$

Where $Id(p)$ denotes the in-degree of that particular page and $R(p)$ denotes the rank of that page. This algorithm was previously used by AltaVista, HotBot and several other web search engines.

B. PageRank Algorithm

In-degree algorithm ranks the web pages according to the number of incoming links. Brin & Page [4], who is the founders of Google search engine, observed that the number of incoming links is not the only criteria to determine rank. Further, they extended this algorithm and observed that all incoming should not be given equal importance. A web page should assigned higher rank if it is referenced by many high ranked web pages. By using this concept they proposed the simplified algorithm called as PageRank algorithm which computes rank of web pages iteratively. Mathematical equation for this algorithm is stated as [4, 15]:

$$PR(u) = c \sum_{v \in B(u)} \frac{PR(v)}{Od(v)} \quad (2)$$

Where u and v are the web pages, $PR(u)$ and $PR(v)$ denotes PageRank values of web page u and v , $B(u)$ is the set of web pages point to page u , $Od(v)$ indicates the pages that link to page u , and c is normalization constant ($c < 1$). Table 1 summarizes few advantage and disadvantage of PageRank algorithm [4, 6, 27, 30]:

Table 1: Some observations of PageRank Algorithm

Advantage	Disadvantage
• Capable to fight with spam web pages: In PageRank algorithm a page assigned higher rank if the pages pointing to it have high rank. Due to this it is not easy task for a particular web page owner to add several incoming links to the pages which pointing to his/her page, i.e. it is not easy task to influence PageRank. • Global rank score computation: PageRank algorithm computes the rank score of web pages globally i.e. it computes of the rank web page of whole web graph unlike to HITS. • Query independent algorithm: It computes the rank of web pages of whole web graph globally and stored off-line rather than query time. At query time search engines just match the query keyword with some other strategies to rank the web pages and sorted them according to the rank value of web pages. • Efficient and fast computation: Due to global computation and query independent nature it is very fast and efficient algorithm. These two advantages play very important role in the success of “Google” search engines.	• Rank Sink: Let consider a situation where a sub graph of web graph in which two web pages point to each other and do not point to any other page, and other web pages point to any one of these two web pages. So during the iterative computation of rank this loop accumulates the rank score and do not distribute rank score to any other web page. • Dangling Node: Dangling nodes are the nodes in web graph that have no-out edges. So they are not able to distribute their ranks. It affect the model when the number of dangling nodes is too large. • Recency Search: The important factor in PageRank algorithm is the number of incoming link to any page. The old pages in web have more incoming link than newly crawled web pages so old pages get higher rank value than new ones. But relevant information may be in newer pages. • Topic drift: PageRank is query independent algorithm. So it is unable to predict whether a hyperlink is related to user’s query or subject of web pages. In result of this it may suffers with topic drift problem. • Link spamming (Page Cheating): Some commercials web sites use optimization of search engines try to get position of first page in search engines results. It is obvious that PageRank only consider number of incoming links not the content of pages.

C. HITS Algorithm

According to Brin and Page, PageRank algorithm is one level wait propagation scheme. Later Kleinberg proposed a different query dependent ranking algorithm which is based on two level weight propagation schemes [12]. In HITS algorithm rank of every web page is determined by two score rather than one in PageRank: First score is authoritative score a_i which measures the quality of page by containing relevant or valuable information regarding query, and Second score is hub score h_i which measures the quality of page by containing link of authority pages. Clearly, a good authority page is a page which pointed by many good hub pages and a good hub page is a page which points to many good authority pages. We can bipartite web graph according to hub and authority pages by making two parts of each page, where one set contains hub pages and other contains authority pages.

Computation of HITS: Computation of hits algorithm mainly done in two steps: In first steps collect the web pages on which hits algorithm applied and second steps perform hits computation in those web pages-

Step 1: HITS algorithm is applied to a subset of web graph. The subset of web pages is the collection of relevant pages depending on the query. The algorithm starts with a root set R_q of the top t pages from the result list of query q , by some content-depend rank algorithms (from a text based search engines ex- Alta-vista or HotBot). From the root set R_q a neighbored set N_q is obtained by [12, 13]:

$$B_G(i) = \text{Set of at most } d \text{ nodes which point to } i \in R_q \quad (3)$$

$$F_G(i) = \text{Set of all nodes which pointed by } i \in R_q \quad (4)$$

$$N_q = R_q + B_G(i) + F_G(i) \quad (5)$$

Where N_q denotes the neighbored set on which authority and hub score would be computed.

Step-2 Computing Authorities and Hubs: To calculate the hub and authority score for page Kleinberg defined that authority score of a page is the sum of all hub score of pages that point to it and hub score of a page is the sum of all authority score of pages that is pointed by it. There is mutual reinforcement relationship between hubs and authorities, by make use of this relationship updates the hub and authority score of each page iteratively [12].

$$a_p = \sum_{q:(q,p) \in E} h_q, \quad h_p = \sum_{q:(p,q) \in E} a_q \quad (6)$$

Where h, a represents n -dimensional vector of hub and authority score of n pages.

Issues with HITS algorithm: There are several issues with HITS algorithm [31, 33, 34]

- The main advantage of HITS algorithm is its query dependent nature i.e. it computes the rank of web pages based on user’s query. But due to query time evaluation this algorithm is not feasible with respect to time because search engines contain billion to trillion of query per day.
- This algorithm is not capable to fight with spam web pages as it computes the rank of web pages based on hub and authority score.
- This algorithm also suffers from topic drift problem. When it collect the web pages for root set it does not uses its own algorithm, so maybe it collects the web pages which are not relevant to query topic.

D. Summary of Improved PageRank algorithms

Several researchers have proposed improved PageRank algorithms to overcome issues as stated in table 1 with basic PageRank algorithms. Categorizations of improved PageRank algorithm based on their issues are summarized in table-2:

Table 2 - Summarization of modified ranking algorithms based on handling issues

Algorithm designed by Authors	Handling issue	Ideology for the method	Mathematical formulation of the method
Sergey Brin and Lawrence Page [4]	Rank Sink Problem	Uses Random surfer model to get out of the loop by using damping factor	$PR(u) = \frac{1-d}{n} + d \sum_{v \in B(u)} \frac{PR(v)}{Od(v)}$
Sergey Brin and Lawrence Page [4]	Dangling Link	Remove the entire dangling link until all PageRank are calculated and after calculation simply add them.	Use stochastic matrix to find the dangling links
Sung Jin Kim and Sang Ho Lee [21]	Dangling Link	Define a new matrix A in which all dangling column have value $\left[\frac{1}{n}\right]_{n \times 1}$ and then compute the PageRank (A*) value by using this matrix	$A^* = (1-d)A + d[1/n]_{n \times n}$ Where A* denotes PageRank matrix
ShiguangJu, Zheng Wang and Xia Lv [28]	Recency Search	Proposed a new method which combines the last modification time of web pages with the in-link and out-link weight of concerned web pages.	$PR_t(p) = \alpha \frac{1}{n} + (1-\alpha) \sum_{q \in B(u)} PR(q) W^i W^o W^{decay}$ W^i and W^o is the In-link weight and out link weight of link (q, p) and W^{decay} is links decay weigh
Wenpu Xing and Ali Ghorbani [16]	Recency Search	Proposed a new method which considers the in-degree and out-degree of web pages and distributes the rank based on the importance of pages.	$PR(u) = \left(\frac{1-d}{n}\right) + d \sum_{v \in B(u)} PR(v) W_{(v,u)}^{in} W_{(v,u)}^{out}$ Where $W_{(v,u)}^{in}$ is the weight of incoming link and $W_{(v,u)}^{out}$ weight of outgoing link
Jing Wan and Si-XueBai [32]	Recency Search	Combines the time activity model (based on web site, user interest, content of web pages and web site developer) with the traditional PageRank algorithm.	$PR_{final} = TR \times PR$ Where TR denotes the time-activity importance vector of page. If TR value is more that means the web page importance is more at that time.
Zhou Cailan and Chen Kai [27]	Topic Drift Problem	Proposed a method which maintains user search log based on random query and click log file, which contains click time and information about URL's. This method add an attribute with PageRank algorithm i.e. Click weight.	$PR(i) = PR(i) \frac{S(i)}{T(i)}$ Where S(i) denotes click weight and T(i) denotes weight of click time
Xiaoyun Chen, BaojunGao and Ping Wen [22]	Topic Drift Problem	Proposed a method which distributed the rank of web page to its outgoing page is based on page similarity, and latent semantic model is used to determine the similarity between web pages.	$LPR(v) = d \sum_{u \in I(v)} LPR(u) \times \alpha + (1-d)$ Where $\alpha = \frac{Sim(u,v)}{\sum_{p \in O(u)} Sim(u,p)}$ denotes similarity between web pages.
Chia-Chen Yen and Jih-Shih Hsu [29]	Link Spam Problem	In the proposed method the PageRank score of web pages does not distributed equally by the number of out-degree, but only relevant web pages can share the score according to relevance degree.	$PR(u_i) = (1-d) + d \sum_{j \in IN(u_i)} \frac{R_{ji} \times PR(u_j)}{\sum_{k \in OUT(u_j)} R_{jk}}$ where R_{ji} is the relevancy score between web page j to i.
AndriMirzal and Masahi Furukawa [37]	Topic Drift problem	Proposed a link-viewer tool to observe the topic drift problem, Solve these issues in two steps: in first step find out the most relevant web pages for the root set by projection method. In next step extract out the web pages which do not have many out links from the root set.	Uses Projection and Base-set Downsizing method to filter out the irrelevant pages from root set.
BunditManask asembak and ArmonRungsa wang [18]	Computation time and Resource	This method speedup the PageRank computation by partitioning the URL index files by source URL into n files of same size called L_i : $0 \leq L_i \leq n$ and each files is executed on separate processor.	$\forall R_i[v] = \frac{(1-d)}{N} + (d \times R_i[v])$ Where $R_i[v]$ rank vector which computed parallel.
Yizhou Lu, Benyu Zhang, Wensi Xi et.al. [35]	Efficient Computation	Proposed method is based on two attributes of web graph i.e. "Power Law Distribution" and "Hierarchy Structure". First it finds the low scorer pages then compute the global PageRank by combining these low scorer pages.	Power law distribution is used to for finding the importance of web pages in web graph.
Atish, Anisur and Eli [36]	Scalability	Proposed a scalable distributed model which used Monte-Carlo method to compute PageRank of web graph.	Implement random surfer model based distributed algorithm to compute PageRank of web graph.

ArmonRungsa wang and BunditManask amsak [19, 26]	Efficient Computation	Present the web graph in compact format so that after partitioning the graph into small parts it should be fit into the GPU's memory devices called cluster's. It takes advantage of GPU's and gives better result than parallel PageRank.	Compute the PageRank value Parallel on GPU system by using OpenMPI and GCC.
Yuan Wang and David J, DeWitt [24]	Scalability	This method proposed a framework which computes PageRank in distributed manner i.e. every web server computes PageRank locally over its own data and then result from every server merged to generate an indexed to determine global ranking.	Compute local Page Rank vector in individual web server $\gamma_{l(m)} = PR(G_{(m)}, V_{n_m}, V_{n_m})$ Where, $G_{(m)}$ represents the server m that contains n_m pages and V_{n_m} is the uniform column vector of dimension n_m . Compute the server rank vector $\gamma_s = PR(G_s, V_{n_s}, V_{n_s})$ Where G_s represent the server link graph.
Shumingshi, Jin Yu, and Guang Wen Yang [38]	Scalability	In this paper distributed page rank proposed by using open system PageRank method. Indirect transmission is used to reduce the communication overhead between servers.	$Y = BR$ Compute Y in distributed manner. Where Y be the rank vector B is square matrix with $B_{(u,v)} = \beta/d(u)$ and R ranks of all pages in the groups.

IV. Comparative Study

Table III, IV consecutively describes the comparative study of sequential algorithms and parallel Algorithms. These comparisons are based on parameter like method type, additional attribute, relevancy score, input, computation time, merit and issues.

Table 3 - Comparison of Sequential Ranking Algorithms

Parameter Methods	Method	Attribute used	Input	Relevancy	Complexity	Merits	Issues
HITS	Query dependent method which computes the importance of web page based on two score: Hub and Authority.	Hyperlink and content structure	Incoming link + Outgoing link + Content of web pages	< PR	<O(log N)	More relevance than PR based on content	Time complexity is more than PR to give output result, Topic Drift, link Spamming Problem etc.
PageRank (PR)	Query independent method which sorted the result based on the relevancy score of web pages	Random surfer model to avoid rank sink issue	Incoming link	> HITS	O(log N)	Less time to give output of search results than HITS.	Topic Drift, Recency Search and Dangling node Problem.
Weighted PageRank (WPR)	Modified basic PR by Computing the relevancy of web pages based on the weight of in-links and out-links.	Weight of incoming and outgoing links	Incoming link + Outgoing link	<PR but > HITS	<O(log N) but >HITS	More quality pages than PR	Lacking with Recency Search Problem
Latent Semantic Model(LSM)-PageRank	Modify PageRank by distributing the rank of web pages to outgoing link based on the having similar contents in pages	LSM is used to determine similarity between pages	Incoming links + similarity score between pages	< PR	> PR	More relevant pages than PR by resolving Topic Drift problem	More complex than PR due to finding the similarity score between pages.
Feedback of user click – PageRank [27]	Modification to PageRank by Combining PR method with the user feedback of search results.	HTTP log file for the feedback of the search result	PR + Click weight + Weight of click time	<PR and WPR	> PR	Resolve topic drift and Recency Search Problem	More computation time due to use of Web log Mining.
Time-Activity based PageRank [32]	Modify basic PR by Computing the rank value of web pages based on time activity curve.	Time-Activity important level model	PR + Time activity important degree	< PR	> PR	Solve Recency Search Problem	More complex than PR
Page Relevancy based PageRank [29]	Combined basic PR with the user feedback of result by calculating relevancy score between web pages.	Relevancy score to distribute rank of pages to other pages	Incoming links + relevancy score of web pages	< PR	> PR	Solve link spamming problem and convergence rate< PR	More complex to calculate rank value than PR.
Power-Rank Algorithm [17]	Computes the rank score in two step: In first step it filters out low score web pages	Power law distribution	Incoming links	= PR	<PR	Less computation time and filter out the	Possibility to filter out some new relevant pages

	and in next step it computes the global importance score of pages.					duplicate pages	
WTPR [28]	Associate basic PR with time attribute i.e. last modification time of web pages and hyperlinks	Log-file gives the modification time of page & link	In-links +Time-stamp of node & links	< PR	> PR	Solve Recency search issue and suitable for dynamic changes of web	More complex due to dynamic nature and takes more time than PR

Table 4 - Comparison of Parallel and distributed ranking algorithms

Parameter Algorithms	Method	Attribute used	Input	Relevancy	Time complexity	Merits	Issues
Yuan Wang David J.DeWitt [24]	Computes the rank score of pages on individual server then merge the overall rank of pages to get the result.	Distributed system to enable scalability	Intra server links, inter server links	Relevancy is same as PR	Less time than Sequential PR when dealing with large data.	Highly Scalable	Merging of rank value of web pages returned by individual server is too complex
BunditManas kasemsak and ArnonRungsa wang [18]	Divide the web graph in several parts then apply PageRank method on low cost Parallel system	MPICH v 1.2.5 for Parallel computation, and PC Cluster for storing Partitioned web graph	Equally partitioned web graph into each PC cluster	More quality pages are returned than sequential PR	Time complexity will decrease when no. of processors increases till the threshold time	Low cost parallel system to compute large web graph.	Takes more computation time when number of cluster increases
A. Rungsawang and B. Manaskasemsak [19]	Computes the PageRank value on Graphical Processing Unit (GPU) system.	GPU used for storage and to compute rank value on Meka cluster, and Open-MPI	Smaller size of web graph that fit in GPU cluster	Approximately Same as Sequential PR	Computation time is less than sequential PR	Fast computation with no constraint on the size of web graph	Takes more time on the copy of data between GPUs to CPU devices than other part of the task.
Sarma, Molla, GopalPanduranga, and EliUpfal [36]	Implement the fast random server based PageRank algorithm in distributed environment	Monte Carlo method with distributed environment	Same as PR	More quality pages than Sequential PR and HITS method	$O(\sqrt{\log n}/\epsilon)$ for undirected graph, $O(\sqrt{\log n}/\epsilon)$ for directed graph	Takes to less computation time than other distributed algorithm	Used in large-scale, distributed network & resource constrained system where computation time is important.

V. Conclusion

Web structure mining is one of the categorization of web mining which mines the hyperlink structure of web graph. Now days search engines are producing trillions of web pages based on user's search query. Ranking algorithms plays a vital role to find the quality of web pages. So efficiency, relevancy and scalability of ranking algorithms are important issues. Various ranking algorithms like HITS, PageRank and their modifications were proposed. This paper has included study of PageRank algorithm and their modification based on disadvantages of basic PageRank algorithm. It also explores modified algorithms in various environments used by researchers like parallel distributed and their pros and cons. Sequential modified PageRank algorithms almost resolved issues (Rank Sink, Dangling node, topic drift etc.) with basic PageRank algorithm but they are still lacking with some other prospects like scalability, relevancy etc. In Future Parallel, Distributed, Grid environment could be used to solve Topic Drift, Recency search and scalability issues by applying on various modified PageRank algorithms.

References

- [1] R. Kosala, HendrikBlockeel (2000), "Web Mining Research: A Survey", ACM SIGKDD Explorations, Vol.2 Issue 1, Page(s):1-15.
- [2] S. Chakrabarti, B. E. Dom, D. Gibson, J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins (1999), "Mining the Link Structure of the World Wide Web" IEEE Computer, Vol.32, Page(s):60-67.
- [3] A. Mülle(2011), "Web Mining in Social Media: Use Cases, Business Value and Algorithmic" by Social Media-Verlag Publication.
- [4] S. Brin, L. Page(1998), "The Anatomy of a Large-scale Hyper textual Web Search Engine" Proceedings of the Seventh International World Wide Web Conference, Page(s):107-117.
- [5] Wang Jicheng, Huang Yuan, Wu Gangshan, Zhang Fuyan (1999), "Web mining: knowledge discovery on the Web", IEEE International Conference, Vol.2 Page(s):137 – 141.

- [6] Taher H. Haveliwala (2003), "Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search" In IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No4, Page(s):784-796.
- [7] Monika R.Henzinger, R. Motwani and C. Silverstein (2002), "Challenges in web search engines" SIGIR Forum, 36(2): Page(s):11-22.
- [8] Bing Liu(2011), "Web Data Mining, Exploring Hyperlinks, Contents, and Usage Data" Second Edition, Springer.
- [9] P Desikan, J Srivastava, Vipin Kumar, and Pang-Ning Tan(2002), "Hyperlink Analysis: Techniques and Applications" Page(s):1-42.
- [10] A.Broder (2002), "Web Searching Technology Overview", advanced school and Workshop on Models and Algorithms for the World Wide Web.
- [11] Qingyu Zhang, Richard S. Segall (2008)," Web Mining: A Survey of Current Research, Techniques, and Software" International Journal of Information Technology & Decision Making, Vol.7 Issue 4, Page(s):683-720.
- [12] J.Kleinberg (1999), "Authoritative sources in a hyperlinked environment", Journal of ACM, Vol.46 Issue 5, Page(s):604-632.
- [13] A.Borodin, G.O.Roberts, J.S.Rosenthal, P.Tsaparas (2005), "Link analysis ranking: algorithms, theory, and experiments", ACM Transactions on Internet Technology, Vol.5 Issue 1, Page(s):231-297.
- [14] F.McSherry (2005), "A Uniform Approach to Accelerate PageRank Computation", Proceedings of the 14th World Wide Web Conference, Page(s):575-582.
- [15] PavelBerkhin (2005), "A survey on PageRank computing", Internet Mathematics 2, Vol.1, Page(s):73-120.
- [16] Wenpu Xing and Ali Ghorbani (2004), "Weighted PageRank Algorithm", Proceedings of the Second Annual Conference on Communication Networks and Services Research (CNSR'04), IEEE, Page(s):305-314.
- [17] Lu, Y., Zhang, B., Xi, W., Chen, Z., Liu, Y., Lyu, M.R., Ma, W.Y (2004), "The PowerRank Web Link Analysis Algorithm" In: 13th WWW conference, Page(s):254-255.
- [18] BunditManaskasemsak and ArnonRungsawang (2005), "An Efficient Partition-Based Parallel PageRank Algorithm" Proceedings of 11th International Conference on Parallel and Distributed Systems, IEEE, Vol.1, Page(s):257-263.
- [19] A. Rungsawang and B. Manaskasemsak (2003), "Fast PageRank Computation using PC Cluster", In Proceedings of the 10th European PVM/MPI User's Group Meeting, Vol.2840, Page(s):152-159.
- [20] K. Sankaralingam, S. Sethumadhavan, and J.C. Browne (2003)," Distributed PageRank for P2P Systems", In Proceedings of the 12th IEEE International Symposium on High Performance Distributed Computing, Page:58.
- [21] Sung Jin Kim and Sang Ho Lee (2002), "An Improved Computation of the PageRank Algorithm". In Proceeding of the European Conference on Information Retrieval (ECIR), Page(s): 73-85.
- [22] Xiaoyun Chen; BaojunGao; Ping Wen(2009), "An Improved PageRank Algorithm Based on Latent Semantic Model" In proceeding of: Information Engineering and Computer Science, Page(s):1-4.
- [23] R Khare, D Cutting, K Sitaker, Rifkin (2004), "Nutch: A flexible and scalable open-source web search engine". Oregon State University
- [24] Yuan Wang David J. DeWitt (2004), "Computing PageRank in a Distributed Internet Search System" Proceedings of the 30th VLDB Conference, Toronto, Canada, Vol.30, Page(s):420-431..
- [25] Ali Mohammad ZarehBidoki, Nasser Yazdani (2007), "Distance Rank: An intelligent ranking algorithm for web pages", Information Processing and Management in Elsevier, Vol. 44 Issue 2, Page(s):877-892.
- [26] B. Manaskasemsak and A. Rungsawang (2004), "Parallel PageRank computation on a Gigabit PC cluster" In Proceedings of the 18th International Conference on Advanced Information Networking and Applications, Vol.1, Page(s):273-277.
- [27] Z. Cailan, C. Kai,LiShasha (2011), "Improved PageRank Algorithm Based on Feedback of User Clicks", in IEEE, Page(s):3949-3952.
- [28] ShiguangJu, Zheng Wang, Xia Lv (2008), "Improvement of Page Ranking Algorithm Based on Timestamp and Link", International Symposiums on Information Processing, Page(s):36-40.
- [29] C. Yen, Jih Hsu (2009), "PageRank Algorithm Improvement by Page Relevance Measurement", In FUZZ-IEEE, Page(s):502-506.
- [30] D. Gleich, L. Zhukov, P. Berkhin (2004), "Fast parallel PageRank: A linear system approach Technical report" Yahoo! Research Labs.
- [31] Xinyue Liu and Hongfei Lin1 and Cong Zhang(2012), "An Improved HITS Algorithm Based on Page query Similarity and Page Popularity" in Journal of Computer, Page(s):130-134.
- [32] J. Wan, XueBai, "An Improvement of PageRank Algorithm Based on the Time-Activity-Curve" In IEEE conference, Page(s):549-552.
- [33] D. Fetterly, M. Manasse, M. Najork and A. Ntoulas (2006), "Detecting spam Web pages through content analysis," Proc. 15th WWW Conference, Page(s):83-92.
- [34] Y Asano, Y Tezuka, T. Nishizeki(2007), "Improvements of HITS Algorithms for Spam Links" In Proceeding APWeb, Page(s):479-490.
- [35] Yizhou Lu, Benyu Zhang, Wensi Xi, Zheng Chen, Yi Liu, Michael R. Lyu, and Wei-Ying Ma "The PowerRank Web Link Analysis Algorithm" In 13th International Conference on World Wide Web, Page(s):254-255.
- [36] A. D. Sarma, A. R. Molla, G. Pandurangan, E. Upfal(2013),"Fast Distributed PageRank Computation" In 13th WorldWide conference, Page(s):11-26.
- [37] AndriMirzal, Masashi Furukawa (2010)," A Method for Accelerating the HITS Algorithm" on JACIII 14(1): Page(s):89-98.
- [38] S. Shi, J. Yu, G. Yang, D. Wang (2003), "Distributed Page Ranking in Structured P2P Networks" on ICPP, Page(s):179-186.