IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Page Ranking Algorithm for Ranking Web Pages

V. Banu Priya[1*], T. Meyyapan[2], SM Thamarai[3],

[1,2]Department of Computer Science, Alagappa University, Karaikudi, Tamil Nadu, India
[3]Alagappa Government Arts college, Karaikudi, Tamil Nadu, India

**Corresponding Author: banupriya0704@gmail.com Tel:7904387013*

Available online at: www.ijcseonline.org

Accepted: 19/Jul/2018, Published: 31/Jul/2018

Abstract- Billions of data related to the user queries is stored in several web Pages and it is growing each day. Many times, Query result is not satisfied to the user. Sometimes web pages display irrelevant data or insufficient data to the user. This type of problem is solved by using page ranking algorithm. Page ranking is based on the user queries. Page rank algorithm is massively used for ranking the web pages in order of most relevant in search engines World-Wide-Web. Page Rank work as main role in the process of web mining. Based on user query a rank list is associated with the listed web pages by any search engine. Therefore the web pages display higher Page ranks are listed in the top rank that helps the user to get a most relevant and useful information in minimum possible time. In this page ranking algorithm we can display both link and the content at a time. Many algorithms are used for page ranking such as Google page rank algorithm, Hyperlink-Induced topic search(HITS) algorithm etc., Using this algorithm we can easily eliminate the problem of older outdated web pages from our rank list. Page rank does not change only by the user click because most visited web pages are not useful or satisfied to the user. So we find the most visited web pages as well as total times spend by the user on the particular web pages. Based on current trend the particular web pages ranking is updated regularly.

Keywords- Data mining, Hyperlink-Induced Topic Search [HITS] algorithm, Page Rank Algorithm.

I. Introduction

In the world wide web an information system on the Internet is allows one documents to be connected to any other documents by hypertext, Multimedia and user to search for information by moving from one document to another. When we search any queries on the internet, there are many URL's has been opened. The large amount of information becomes very difficult to the users to find, filter the relevant information.web mining is one of the application of data mining technique to find relevant and useful information from web data. Using web mining technology we can access multiple data which is present at any location. In early stages of world-wide-web, search engines have developed various methods to rank web pages. Until today, the occurrence of a search phrase within a document is one major factor within ranking techniques of virtually any search engine. Page Rank is a vote for a particular web page on the web. Page rank gives the result of how the page is important .A link to a page counts as a vote of support. If there's no link on a web page there's no support for that particular web page. And also if there is more links and also irrelevant link on a web page there is also no support the webpage. In page ranking algorithm there are some tasks which will be performed.

1. Resources
2. Information selection and preprocessing
3. Generalization
4. Analysis

Web mining categories

Web mining categories include web content mining, web structure mining, web usage mining.

Web content mining

Web contain large amount of data, each and every website represent same information differently to the user. Our focus will be filtering, extracting of structured data from webpages.web content mining is used to find, filter extract mine useful and related information from webpage contents to the user.

Web structure mining

Web structure mining tries to discover the link structure of the hyperlinks at the inter document level. Based on the topology of the hyperlinks, web structure mining will categorizes the web pages and produce the information such as the similarity between sites or discovering web communication.

Web usage mining

Web usage mining is the process of filter or extract useful usage patterns from the web data. Our focus is based on finding patterns related to users query from webpages.web usage mining can provide useful information to improve a network's performance, to rebuild a website dynamically and also this technology include information about the pages, how long the user spend time with this website.

Existing Method

Panda Algorithm

How it works

Panda is called a 'standard score' for web pages: This score is used as a ranking factor. Initially, the panda was more filter than a portion of Google's rankings, but in January 2016, it was officially incorporated into Core Algorithm. The panda cycles have become too much, so both the fine and the review are now fast.

How to adjust

Regular content scenarios for content copy, thin content and key padding. To do that, you need a site crawler, such as SEO Power Suite's Internet Auditor. If you cannot be an e-commerce site and 100 percent of your personal content, you will try to use the original images there, and use the user reviews to make the product descriptions stand out from the crowd.

Penguin Algorithm

How it is works

The purpose of Google Penguin is that the below sites consider its connections to be handled. Since the end of 2016, Penguin is part of Google's core algorithm; But the panda, it works in real time.

How to adjust

How to adjust Keep track of the growth of your affiliate profile and enable regular audits with backup verification such as SEO Spiclus. In the abstract tool dashboard, you can see the progress map for the development of your link profile. Be careful for any unusual sharpness you need to look at the unnecessary back links.

The statistics we take on the Penguin account are attached to the SEO Spikes' Penalty Risk Formula. To check the risks to penalty, go to the Linking Dashboard, click the Link Penalty Risks tab, select your links and click Update Risk. Once the check is complete, check the Penalty Risk Column and see each link over 50%.

Hamming Bird

How it works

Hamming Bird explains Google search queries better and gives results matching the search objective (as opposed to the individual words of the query). While the key words continue to exist, Hamming bird may be a question to a page, even if it does not have the right words in question. It is a natural language processing based on indirect semantic directory, co-occurring terms and synonyms.

How to adjust

Expand your key research and focus on feedback, not important. The bulk of such comments is Google's related search and Google Auto complete. You can find everything in the rank tracker's main research volumes. Use this intelligence to understand the language of your audience and to diversify your content. By creating comprehensive content that satisfies the purpose of the search, you will be successful on the basis of engagement and SEO. We'll see in detail later in this post when we discuss Rank Brain.

Possum Algorithm

How it works

Positive updating confirms that local records vary depending on the location of the search You can see it in local results because you are close to a business address. The results of the most similar questions, such as 'Dancing Tenver' and 'Dentist Denver Co', have also had a greater effect on the rankings. Interestingly, Bausum provided encouragement to businesses outside the city.

How to fix

Extend the main list and keep track of the specific location Ranking. Localized businesses must now target more important tags than they have used, since the Bass that was brought into the local SERPs was polluted. When you verify your rankings, make sure you are doing this from your destination (or better, a bunch of them). Priority list is the priority list for search engines.

Proposed method

Google use the page ranking algorithm for ranking web pages. But Google does not have a Patent Rights of page ranking algorithm. Page ranking algorithm patent rights belongs to Stanford University. The proposed method of this paper is rank list the web pages based on the quality of the web pages. Quality depends on the strong main content of the particular website. Content of the website is very clear, Accurate, Free of Errors, Authoritative and Trustworthy also the website look modern, be easy to use and it must be function correctly. Second point is to check visited and also most of time spend by users on the particular webpage. Third point how long the web page is loaded because users can not satisfied if the web page takes more time to load except network problem. So it is not a better website. It will reduce the mark from rank list. Fourth point we check whether the web page is regularly updated and proper maintenance. Proper Maintenance based on broken links, missing images, outdated information. And finally check the particular web page is not outdated. Based on the rank we give rank list to the web pages. Page rank rating is calculated by a page ranking algorithm and corresponds to the internet's normal link matrix. It is only calculate the value of a particular webpage. And it does not calculate the other pages that link to it. Using this Page Ranking Algorithm we can easily eliminate the problem of older outdated web pages

form our rank list. Page rank does not change only by the user click because most visited web pages are not always satisfied to the users. Links is an import factor in page ranking. Link or hyperlink is two different web pages that are linked together. Links are may be a Text, Images (or) a Buttons. Links is based on how relevant the link is to the particular web page. Page ranking have a two important links: that are Inbound Link and Outbound Links. Inbound link is normal link from our website to other websites. Outbound is a link from other websites to our websites. Links plays important role in rank list. In our proposed method also give a rank for inbound and outbound links.

Result and Discuss New pages have less page rank and they take much time to be get listed and gain high ranks. If someone inaccurately quote something on an webpage then subsequent readers also quotes it on another web page search engines index all of the inaccurate pages, and we end up with a mess where fiction is accepted as reality.

Search results are based on the literal (Keywords, tags, meta data)things but not on meaning.

PR(A) = (1-d) + d(PR(Ti) /C(Ti)) +……..+ PR(Tn)/C(Tn))

PR(A) is the Page Rank of Page **A**
PR(Ti) is the Page Rank of **Page Ti** Which link to Page A,
C(Ti) is the number of Outbound Links on **page Ti**
D is a Damping factor which can be set between 1 and 0
PR(Di) is the number of Dangling Links on **page Ti** which can be set 0 or 0.15

Damping factors:

It is considered to be equivalent to the normal amount (or average) of all web pages. Although the Page Rank formula does not have one page back links, it gets a small score of 0.15 (1 min D-factor).

Conclusion

Page rank algorithm computes the page ranks of at the time of user query. But this modification algorithm that give results at the time of indexing as well as at the time of user query. This algorithm is produce a better satisfied result the user query. Page ranking Algorithm gives the user to find those pages that users really look for, show users which pages are the best Load fast and safest.

REFERENCE

[1] Combined Approach for " Page Ranking In Information Retrieval System Using Context and TF-IDF Weight" s.Gupta,V.Jain,P.Bhadana Research Paper | Journal Paper Vol.2,Issue.6,pp 39-42,Proceedings of the IEEE International Journal of Computer Sciences and Engineering, June-2014.

[2] Diefenbach D., Thalhammer A. (2018) PageRank and Generic Entity Summarization for RDF Knowledge Bases. In: Gangemi A. et al. (eds) The Semantic Web. ESWC 2018. Lecture Notes in Computer Science, vol 10843. Springer, Cham

[3] Thalhammer A., Rettinger A. (2016) PageRank on Wikipedia: Towards General Importance Scores for Entities. In: Sack H., Rizzo G., Steinmetz N., Mladenić D., Auer S., Lange C. (eds) The Semantic Web. ESWC 2016. Lecture Notes in Computer Science, vol 9989. Springer, Cham

[4] N. Duhan, A. K. Sharma and K. K. Bhatia, "Page Ranking Algorithms: A Survey", Proceedings of the IEEE International Conference on Advance Computing, 2009.

[5] Kohlschütter C., Chirita PA., Nejdl W. (2006) Efficient Parallel Computation of PageRank. In: Lalmas M., MacFarlane A., Rüger S., Tombros A., Tsikrika T., Yavlinsky A. (eds) Advances in Information Retrieval. ECIR 2006. Lecture Notes in Computer Science, vol 3936. Springer, Berlin, Heidelberg

[6] R. Kosala, H. Blockeel, "Web M ining Research: A Survey", SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining Vol. 2, No. 1 pp 1-15, 2000.

[7] P Ravi Kumar, and Singh Ashutosh kumar, "Web Structure Mining Exploring Hyperlinks and Algorithms for Information Retrieval", American Journal of applied sciences, 7 (6) 840-845 2010.

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

[9] J. Kleinberg, "Authoritative Sources in a Hyper-Linked Environment", Journal of the ACM 46(5), pp. 604-632,1999.

[10] Lages, J., Patt, A. & Shepelyansky, D. Eur. Phys. J. B (2016) 89: 69. https://doi.org/10.1140/epjb/e2016-60922-0, Springer Berlin Heidelberg.

[11] D. Cohn and H. Chang, "Learning to probabilistically identify Authoritative Documents". In Proceedings of 17th International Conf. on Machine Learning, pages 167-174. Morgan Kaufmann, San Francisco, CA, 2000.

Authors Profile

Ms.V.Banu Priya B.Com(CA)., MCA., from Madurai Kamaraj Univrsity, Madurai and M.phil(CS)., from Alagappa Univrsity, Karaikudi, Tamilnadu in 2011, 2014 & 2018. She has attended workshops at National levels. She has presented 2 numbers of research papers in National conferences.

Dr. T Meyyappan M.Sc, M.Tech., M.B.A., M.Phil, Ph.D. currently, Professor, Department of Computer Science, Alagappa University, Karaikudi, TamilNadu. He has organized conferences, workshops at national and international levels. He has published 90 numbers of research papers in National and International journals and conferences. He has developed Software packages for Examination, Admission Processing and official Website of Alagappa University. As a Co-

Investigator, he has completed Rs.1 crore project on smart and secure environment funded by NTRO, New Delhi. As principal Investigator, he has completed Rs. 4 lakhs project on Privacy Preserving Data Mining funded by U.G.C. New Delhi. He has been honoured with Best Citizens of India Award 2012 research areas include Operational Research, Digital Image Processing, Fault Tolerant computing, Network security and Data Mining.

SM. Thamarai currently, guest lecturer, Alagappa Government Arts College, Karaikudi, received her Diploma in Electronics and Communication Engineering, Department of Technical Education, Tamilnadu in 1989 and her B.C.A. M.Sc. (University First Rank holder and Gold medalist), M.Phil. (First Rank holder) degrees in Computer Science(1998-2005) from Alagappa University. She has published 27 research papers in International, National Journals and conferences. She received her Ph.D. degree in Computer Science in 2014. Her current research interests include Operational Research and Fault Tolerant Computing.

