Backbone of Google:Testing Pagerank Algorithm

Roopali Salwan
Amity Institute Of Information Technology
Amity University,Rajasthan
Jaipur, Rajasthan

Deepali Salwan
TCS,Noida
Noida, Uttar Pradesh

ABSTRACT
Search Engine have changed our lives in every aspects and is now an important part of our lives covering each and every area ,it is now working as teacher,guide,entertainer,our own storehouse, information centre what not it's playing every role to support our living and it's our new friend which accompany us at every step. Whenever user have any query it's solution is just a click away type your query and just click on enter and Google solves your problem so it is concluded that it is 24*7 lifeline of this new era. So it's very much essential to know its backbone, logic which makes it working. Indeed it is our need to uncover the answer of some How's and Why's , just by one click Google yield result which is appropriate and so correct just in small time ,How it does that?.Though there are some flaws still Google is world best search engine, Technically speaking the basic reason behind the success of any project is the algorithm or the way how it is created; here also we would consider Google's searching algorithm "pagerank" which is used to rank pages which are accessed by users every now and then. This Paper through some light on rank given to the pages and how dangling node affects its efficiency and how just by small change you can make searching an easy and efficient job .

Keywords
Google,pagerank algorithm, dangling node,page ranking.

1. INTRODUCTION
In this section we will consider all the modes which are used till now to evaluate the web pages as a product is evaluated by the number of buyers who bought that product in the same manner web pages is examined or rather is ranked on the basis of how many users accessed that web pages.Firstly,We need to understand that How important is that to rank a paper we need to answer two important question "WHY PAGE IS RANKED?" and "WHAT IS THE IMPORTANCE OF THIS RANKING" or "WHAT IS THE PURPOSE OF THIS RANKING" .These question have simple answer not precisely restricted to technological front can be made understand to a layman as well like Why we need a computer moreover Microsoft office word for documentation though we have all possible resources required for any documentation paper ,pen ,ink and information, here answer to this is that there are certain evaluation criteria which calculate the efficiency and performance and when efficiency comes into existence there is another criteria that also considered i.e. time which walk hand in hand with performance so answer is computer is a machine that reduces our effort and t bime so is considered over manual documentation. Similarly work of search engine is to provide information required by the user as fast as possible and appropriate information searched should not be generalized it should be specialized with the time required for this should be less .So if a search engine has to mark its presence among it's multiple competitors it should make it's search criteria broaden as well as most appropriate and relevant information or web

pages should be shown with high preference and then so. Since 1998,Google had been working hard to standard which it has made for search engine. There are series of events which took place in these year and had given GOOGLE a place in web world such that GOOGLE is the new word for internet. What all is working at the back end What is the logic that is making Google efficient is the algorithm working behind this search engine. Considerably we need to study the algorithm

2. HISTORY OF PAGE RANKING ALGORITHMS
Ranking a page is a tough job it generally gives a scorecard of web pages that whether the webpage is worth visiting or accessing ,with this pay per click scenario it's difficult to find the real source where we can end up our search it's because some web pages are not full flourished to standalone and explain there existence and depend on other web pages to show their actual meaning and certainly navigation is whole and sole purpose of some web pages .So Question arise "How we can recognize the actual correct answer of our search" as search engine presents us with n number of solutions "How some links gain the higher or first position" [1].Now this problem is sorted with the help of algorithm which provides the step which in general search engine follows ,but as google is the world best search engine why not focuses on its page ranking algorithm.

3. HYPER TEXT INDUCED TOPIC SELECTION
This general algorithm states that web pages that are pointed to by many hyperlinks are called as authorities and these hyperlinks are called hubs ,here hubs and authorities are rated according to their score ,Now thus if a authority is to be rated high it should be pointed by large numbers of high scored hubs here a_p is the authority score for authorities and h_p is the hub score for hubs now here as we know that user type query to get correct or appropriate solution but here solution is not textual based which user expect that solution should be based on text entered in the query but the resultant is link based whichever authority have high score is placed high in the search though it doesn't make a relevant solution of its search .However in general scenario HITS was nota great success and wasn't widely used.

4. PAGERANK EXPLANATION
In this technical era just by looking the beauty,beauty here referred to the graphics and the interface with which user is comfortable doesn't give web sites its higher rank but the users are more keen to know the logic behind and for knowing the logic behind Search Engines working .We should understand the "How these Search Engine Works" and "What is the work of SEO's". Now take an example of Google's page rank as we typed a query to find "Panther" i.e. the keyword we want to search we can see the search result here we haven't mentioned that whether we want to search panther animal or a software but

we have made pagerank to decide what kind of search result should be given to a user
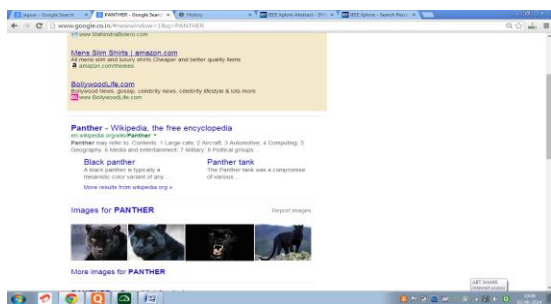
Source:www.google.com

We can see by a screenshot that first linked search by google is panther animal though there is a software name as "panther" .Now just to analyse the behavior of page rank algorithm and its searching criteria a new keyword is entered just to see the search result the word here which we typed was "apple" ,Screenshot shows the search result
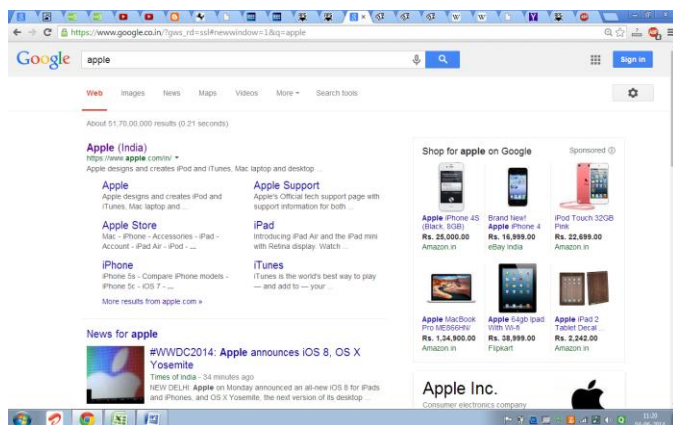
Source:www.google.com

You can see the difference in behavior for the keyword panther user need not to specify whether he wants to search an animal or a software but if you'll type keyword apple you'll easily see the difference in the behavior of page rank algorithm it will show the result as world most renowned apple company that design i-pods and i-ipads .So that was the general discussion about the behavior of page rank algorithm now we need to understand the logic behind if that's the case so why does it's behaved that way

5. MATHEMATICS BEHIND PAGE RANK ALGORITHM

From the time steam engine was invented we have seen people wandering and searching for logic behind the invention if wheel is moving "Why it is shaped circular not Rectangular" why just by rubbing two stones you get fire and as we all know need is mother of invention ,As when search engine was tried to be implemented some people must have had questioned about its feasibility "Why we need" so lets uncover these facts Search Engine as known used to search the keyword typed by user but what is the logic behind Google being number one search engine . "How some links are shown at first page of our search ?" and what is the need of page rank Suppose a naïve user say User A is surfing net without any intention of finding a particular question just a random user where he will be visiting random pages i.e. one page to another with less probability or null probability that he will ever move back or will navigate to the back page, but as he gets bored he start with surfing another random page. Thus Page rank is probability that any random surfer will visit this

page (i.e. say page A) is called page rank of that particular web page .Similarly probability of a person to get bored and request another web page is called its damping factor d
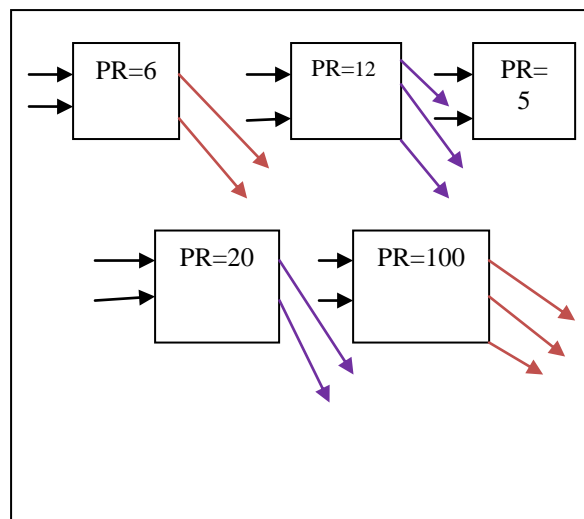
Assume there are four Web pages

Fig1:Collection of web pages with their respective page rank

Here we can see page rank given to initial page is 6 with two outgoing link now this page rank is like marketing high rated element is related to high rated element i.e. PR of first page that is 6 is mutually credited to the account of outgoing link we can see two outgoing link so each web page have in its account 6/2=3 i.e. 3 value now question arise why value 6 is divided by two 6 is the page rank of first web page and it has two outgoing link so value 3 is added to the account of outgoing links.

Page rank which was developed by Google and was name after Larry Page is used as an efficient algorithm for mining web pages . Where importance of web pages is done by analyzing the links associated with that pages and their respective rank and thus it is a voting scenario whichever web page gets more vote is most famous with high page rank

$$PR(U)= (1-d)+d*\sum_{V\in P(u)} PR(v)/N(v) \qquad (I)$$

where u and v represent web pages. P(u) is the set of pages that point to u. PR(u) are rank scores of page u. N(v) denotes the number of outgoing links of page v, d is a damping factor that is usually set to 0.85. d has been considered as the probability of users following links[4]

Now as we have seen the pages are given high pagerank on the basis of whether it is linked or pointed by web pages or links which have high page rank .

6. CLEARING DOUBTS RELATED TO PAGERANK

Also in a recent research it was said that don't get worried if one web sites is crowded by or is pointed by 25 links (or web sites) it doesn't add its page rank but if u have a website which is linked by just 6 links and that too with high popularity will yield you a high page rank It's just like judging one by his company, thus according to page rank algorithm a page has high page rank if its back links summed up to high value .One important issue that surrounds page rank is Dangling node problem or Dangling link problem those web pages that do not have any outgoing

links i.e. links that are linked to web pages like pdf file ,which has adverse effect on the rank of web pages.

7. WHY NOT RANK THE CONTENT INSTEAD OF A WEB PAGE

Why we are ranking a page just by looking whether it is associated with a pages with high page rank it is similar to judging a person by his company why not emphasizing on quality of it's content not always it's true that if searched page is crowded or pointed by web pages with high page rank it will give the relevant data as per user's requirement.

So let us see some pointed issue of Page Rank algorithm

I)Search Results are based on tags -metadata but not on meaning or content

Let us analyze this problem that if a search engine is for the use of people why we are running after pagerank it might be possible that page has high pagerank but it is not relevant for the user there should be a factor that will record or monitor user's behavior as taking an example

A user is searching for keyword say Sky result would be as shown
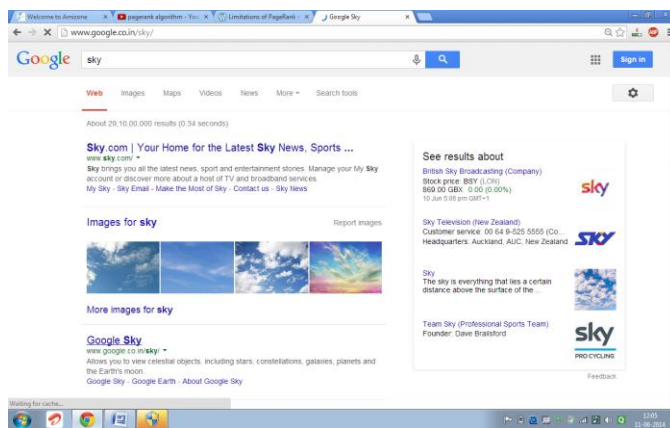
Source:www.google.com

According to the screen shot we can see that the link displayed

Must have high page rank that is why it was displayed at first position but here search engine is for the user's need there should be something to track user's behavior i.e. after searching for a particular keyword whichever link was visited by the user here just mouse click event is not the only source of information which can be used to judge a web page but here the main criteria is time for how much time user visited that web site i.e. though after searching a keyword sky we got the result shown above but we can see though Google sky is not the first link shown user click that link it means not always the search result is as per the need of user and just by clicking also i.e. by clicking a link we are not able to judge that the link is the best search result because there is a possibility that user clicked to a link and when he visited that

Webpage it was not up to the mark it means it was not giving the real data for which user was searching.
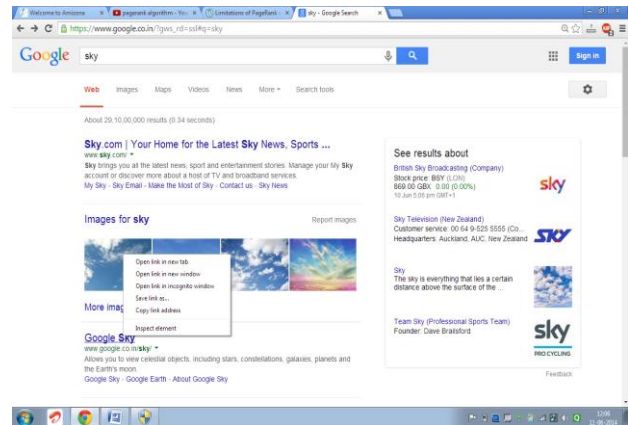
Source:www.google.com

So there should be a ground on which we can judge a web page if not a click because as we discussed a link can have many clicks but can be possible it is not providing appropriate data ,So there should be a factor which will add to value of a webpage that this page is relevant and that is time

Taking an example of a search where user typed some keywords

The link which should be placed first on the search should not based on it's related link rather it should be based on the content of that page and if the the content is appropriate user will stay at that page for enough time if there is immediate navigation means data was not relevant and if user stay say for 5-10 minutes on that page means it was relevant and this time will judge which link should be given high page rank .

8. CONCLUSION

Page Rank which is an efficient algorithm which makes Google number one search engine thought it still suffer from some of it's drawback like some pages that dosen't deserve high pagerank is allotted high pagerank just because it is surrounded by some web pages that have high page rank just because content is not seen or emphasized just by changing the approach and introducing time which can be the ground of selecting a best result for searched keyword

9. REFRENCES

[1] L. Page, S. Brin, R. Motwani, and T. Winograd, "The Page Rank citation ranking: bringing order to the web," Stanford Digital Libraries Working Paper, 1998.

[2] Page L, Brin S, Motwani R, et al. The pagerank citation ranking: Bringing order to the web[R]. Stanford Digital Libraries, 1999

[3] Wang Dong,Lei Jingsheng.An Improved Ranking Algorithms Based On PageRank[J].Microelectronics and Computer,2009,26(4):210-213.

[4] Ying Jin,Jing Zhang,Pengfei Ma,Weiping Hao,Shutong Luo and Zepeng Li "Applying PageRank Algorithm in Requirement Concern Impact Analysis". Computer Software and ApplicationsConference, 2009. compsac '09. 33rd Annual IEEE International (volume:1)