Project One

Submitted by: Ankita Muley

Date: 10/21/2013

Speculate why these documents have high PageRank values, i.e., why is it that these particular pages are linked to by (possibly) many other pages with (possibly) high PageRank values. Are all of these documents ones that users would likely want to see in response to an appropriate query? Which one are and which ones are not? For those that are not "interesting" documents, why might they have high PageRank values? How do the pages with high PageRank compare to the pages with many in-links? In short, give an analysis of the PageRank results you obtain.

Top 10 **pages** by PageRank

1. Page: WT21-B37-76 with Page Rank: 0.00267941005759
2. Page: WT21-B37-75 with Page Rank: 0.00152591699385
3. Page: WT25-B39-116 with Page Rank: 0.00146949352594
4. Page: WT23-B21-53 with Page Rank: 0.00137232204363
5. Page: WT24-B40-171 with Page Rank: 0.0012450717567
6. Page: WT23-B39-340 with Page Rank: 0.00124048970549
7. Page: WT23-B37-134 with Page Rank: 0.00120521440914
8. Page: WT08-B18-400 with Page Rank: 0.0011435408633
9. Page: WT13-B06-284 with Page Rank: 0.00112478163921
10. Page: WT24-B26-46 with Page Rank: 0.00108504628436

1st Rank WT21-B37-76, this page has highest page rank value. After looking in Lemur web interface it seems that the page is home page of the economist. As this is a home page every page in the sitemap points to the home page, hence this page has highest in link count, this we can verify from top 10 in link count. So it’s most likely whenever a user search for something like ‘news from economist’ or ‘economist’ he will end up on the home page. This will be an interesting document since it provides view for selected articles and thus justify highest page rank value.

2nd Rank WT21-B37-75, this page is copyright notice from the economist. It notifies public that the content has owner and it can’t be reproduced. This is an example of a page which has in link from most all pages but it’s least likely for any user to end up. As this page is important and it’s necessary to provide reference to this page it has high page rank value.

3rd Rank WT25-B39-116, this page is regarding security assurance requirement to guide people about the policies and procedure to use data without any security breach. Again this is one of the pages which are least likely for any user however; it’s a requirement to include it in every article in the paper. Because of the importance and high in link count it end up having high page rank.

4th Rank WT23-B21-53, this page provides information about web development team of sea exploring website. So it has content of the different domain which might be popular one. Also it incorporates keywords Web development which one of most used search term. Both the factors backlink from a popular website and keywords lead to high page rank. However, it is least interesting for any user.

5th Rank WT24-B40-171, this page has the archive of all the previously visited pages. It contains all the popular stories of the each day. As it contains featured stories when loading from cache, it has link from other documents. So it has high page rank. But it’s not interesting for any user.

6th Rank WT23-B39-340, this page provides financial information of companies sorted alphabetically. So it provides any user to view financial information of any company of his choice. It’s most likely to get most hits. Additionally it is useful and interesting to user. As we can see that this page contains backlink for each of the different company domain. Also it must be having link from each of company website. Also Backlink could be from popular company. Because of high backlink count and backlink from popular pages, high page hits it has high page rank.

7th Rank WT23-B37-134, this page contains disclaimer content. Again it is least interesting to any user. However it is required to provide disclaimer for any information or product. So it is high importance, hence high page rank.

8th Rank WT08-B18-400, this page contains kind of disclaimer, which is least likely for any user. However, importance, relevance leads to high page rank.

9th Rank WT13-B06-284, list of authorities from different univeristies.This page provides a user to see authorities of different universities in one go. It has backlink to high quality web sites. Hence high page rank.

10th Rank WT24-B26-46, this page is home page of Professor Milton from Michigan University. This page is interesting to user searching for Professor Information and his work. It has back link to popular university website. Hence high page rank.

Top 10 pages by In link Count

1. Page: WT21-B37-76 with In link Count: 2568
2. Page: WT18-B29-37 with In link Count: 2269
3. Page: WT01-B18-225 with In link Count: 2260
4. Page: WT23-B27-29 with In link Count: 1940
5. Page: WT21-B37-75 with In link Count: 1704
6. Page: WT27-B34-57 with In link Count: 1257
7. Page: WT27-B32-30 with In link Count: 1255
8. Page: WT08-B19-222 with In link Count: 1041
9. Page: WT08-B18-400 with In link Count: 1011
10. Page: WT10-B36-88 with In link Count: 946

1st Rank WT21-B37-76, this page has highest page rank value. After looking in Lemur web interface it seems that the page is home page of the economist. As this is a home page, every page in the sitemap points to the home page, hence this page has highest in link count. Also, it must have back link from numerous sites.

2nd Rank WT18-B29-37, this page is Environmental news state. It has not much high in link count however the page rank is not good. But might be possible it has one way of linking instead of diversity of in links.

3rd Rank WT01-B18-225, this page is online library for DRC. Though this page has high inlink count, it’s quite aged. It was last modified in year 1996, which depreciates its page rank.

4th Rank WT23-B27-29, this page is welcome page of sportsgate. Although, it may be interesting to a user interested in building links, but again, this page was last modified in year 1996 which reduces its page rank highly.

5th Rank WT21-B37-75, this page has the archive of all the previously visited pages. It contains all the popular stories of the each day. As it contains featured stories when loading from cache, it has link from other documents. So it has high page rank. But it’s not interesting for any user to see the archive unless he wants to dig into past. Additionally, it’s regularly updated since it requires adding popular stories of each day. Hence high page rank and high input link.

6th Rank WT27-B34-57, this page is message board of skeptic society. As we can see that this page has high link count however it not modified since year 1997.Hence the agedness of the page beats in link count, thus not high page rank. Also, no post has been posted since 1997, any user will be least interested in browsing this page.

7th Rank WT27-B32-30, this page contains frequently asked questions and answers in www board. This may be interesting to user if the www board hasn’t change in years. Also it has numerous in links however, lack of up gradation or can say changes leads to low page rank.

8th Rank WT08-B19-222, this page contains kind of information page includes terms and conditions applicable to td product and services. This page is not likely interest’s user.

9th Rank WT08-B18-400, this page contains internet security information. This page is not likely interest’s user. However it has lot of in links because we need to provide notice regarding security and backlink to popular site leads to page rank.

10th Rank WT10-B36-88, this page is site for gay, lesbian, bisexual community. This page may interest user. However, it might lack diversity of in link sources, which probable reason for low page rank.