Task 3: Qualitative analysis:

Top 5 pages

|  |  |
| --- | --- |
| **WT21-B37-76**  **WT21-B37-75**  **WT25-B39-116**  **WT23-B21-53**  **WT24-B40-171** | **0.00267941005759**  **0.00152591699385**  **0.00146949352594**  **0.00137232204363**  **0.0012450717567** |

|  |  |  |  |
| --- | --- | --- | --- |
| **Pages** | **Total No. of inlinks** | **Ranking** | **Description** |
| WT21-B37-76 | 2568 | 1 | The Economist Homepage |
| WT21-B37-75 | 1704 | 5 | The Economist copyright notice |
| WT25-B39-116 | 169 | 280 | Security Assurance Requirement. |
| WT23-B21-53 | 198 | 178 | Infor. about the web development team |
| WT24-B40-171 | 270 | 106 | Evening news on every week night. |
| WT23-B39-340 | 274 | 102 | Financial Information of the companies |

**Brief Analysis:**

**WT21-B37-76:** has the highest page rank, as indicated by the table above, that is **0.00267941005759** and it is ‘The Economist Home Page’. It can be seen that pages with higher inlinks are ranked lower than the ones having a lower inlinks.

**WT21-B37-76:** has the highest page rank, as indicated by the table above, that is **0.00267941005759** and it is ‘The Economist Home Page’. It has the highest number of inlinks. Every inlink to this page in turn also have inlinks. So basically, there is no page connecting to this page which is a sink node. It is particularly interesting as the search for recent events would hopefully lead to this page being in the list of returned queries. In addition, users would like to know the changes in economic policies and how that affects their daily lives. Hence this could be the reason for its popularity.

**WT21-B37-75, WT25-B39-116:** Its very unlikely that these pages would turn up in the list of returned queries as these are copyright notice, the security assurance requirement documents.

**WT23-B21-53**: provides information regarding the web development team and could be possibly linked by various documents. Users could possibly require that information for contact/complaints/comments purpose. Hence, though not very much useful, it still has a higher PageRank. It is interesting to notice that all the inlinks are from the documents with document IDs commencing with WT23. Each and every document would have copyright information which might contain the link to the web development team’s page.

**WT24-B40-171:** is the evening news on every weeknight. News pages normally have various links in-order to keep people updated with the most recent happenings. They are not only utilized to keep track of the most recent happenings but also to supplement present information. Hence, there would be many articles/blogs/web pages which link to this page. People require news and hence news is a very common query which makes this information interesting and important.

As suggested by Wikipedia, PageRank is a probability distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page.

The second document (WT21-B37-75) has a high number of inlinks (1704) and ranks fifth when sorted according to the number of inlinks. Although, for the third document (WT25-B39-116) the number of inlinks is not that high (169, ranked at 280 when sorted according to the number of inlinks). This data (inlinks) can be seen in two ways, firstly by the number of inlinks within these 5 documents i.e which document has an inlink from any other document amongst these 5 documents.

Therefore inlinks actually tell us the importance of a page,which can be easily seen in the case of the first and second documents:

**WT21-B37-76: WT21-B37-75,**

As can be seen, the above answers why the second document is ranked higher. It has an inlink

from the first document which has the highest number of inlinks. This increases the likelihood

of a person arriving at this document.

To analyze why the third document has a higher PageRank, we can look at what links are linking to the third document. WT25-B39-116 has two links that have high counts compared to other links. **WT25-B38-117:5 and WT25-B39-116:169.** There is one self link and there is only one document which has five inlinks (WT25-B38-117)

This link is a beginning of a book, more precisely, the first part titled ‘Introduction and General Model’. This seems an interesting page to appear in the results of a person’s search. Although the number of inlinks are not as high enough, which means the five documents linking to this document should have more inlinks.