

Question 1 :- how you calculate pagerank score

We first find the number of titles(N) in the collection. The formula has been used as is provided in the Lab 7 document. The damping factor has been taken as .85 and is subtracted from 1. The Result is divided by the number of titles(N). Therefore, we are deriving the page rank of a page from the page ranks of other pages. The damping value adjusts the derived value downwards. We have created 4 map-reduce pair where in first we derive the value of N. Then in the second procedure, we initialize the sum of pageranks for individual titles to 1. Then we run 8 iterations of the pagerank algorithm to get the final derived values. These derived values are then passed to the other procedure which sorts these values in decreasing order thereby giving us the appropriate result.

Question 2 :- What challenges or difficulties you have for this lab

Being new to map reduce programming I faced issues with understanding the API and getting things done. I had to look at a lot of sources on the internet to correctly understand how I need to implement this in the code. I also faced problem in the last map reduce where we had to sort the output and merge them into a single file.