IR TERM PROJECT

Yash Jain 19CH10078
Group-13

My contributions in the Part-2 are as follows:

Task 2A(tf-idf vectorization):
In this part, I was responsible for calculating the tf-idf using the **anc.apc** scheme.Thus, the I used the tf and df values the for word in each query to calculate the score of the query document pair according the anc.apc scoring scheme and ranked the documents accordingly.
The issue which I faced during the task was related to the terms which were not there in any of the documents. Since, while calculating the tf-idf using the anc.apc we need the max value of the term frequency but if the max value of the term frequency is zero then it will lead to erroneous values for the score thus it was important to handle this edge case otherwise it result would have contained wrong results.Thus, corrected the error by making the td-idf equal to 0 whenever the max value of the term frequency was zero.

Task 2B(evaluation):
In this part of the task, I made some modification to the compute_value function which was calculating the MAP and NDCG for each query. The changes which I made helped us successfully store the map and ndcg values for all the queries into the output text file.I passed a empty diction in function as an argument which stored the values of the map and ndcg for each query and then I iterated over the dictionary and write the value of map and ndcg for each query to the output text file.

Documentation:
I took the charge of writing the documentation for part-2A as well as part-2B of the assignment.In which I documented,the explanation of the code and method to implement both the above tasks.