

IR TERM PROJECT

Yash Jain 19CH10078

Group-13

My contributions in the Part-3 are as follows:

Task 3A(Relevance Feedback):

In this part, I was responsible for writing the code for the function implement the Rocchio's algorithm for modifying the query based on the relevant and non-relevant document set and initial query. In this code, I faced a problem while adding the vector space of the relevant and non relevant docs and initial query to form the modified query while adding two vector space I was directly updating the input object with making its copy and then return the result which was giving us wrong results. I solved this issue by declaring a new object and making the new a copy of the input object and then making changes in the new object to give the sum of two vector spaces.

I was also responsible for writing the code for calculating the MAP values and I also wrote the code for modifying the initial queries on the basis of the relevance feedback and ranking the document based on these modified queries. The issue while writing the MAP function was that there were some queries for which there were no relevant document in the ranking list thus, It was necessary to handle this edge cases for the correct implementation of the MAP function. Also, there were some documents which were not present in the gold standard document list (qrels.csv) thus I added a check condition in the code that if the document is not present in the gold standard document list then I took the relevance for that document as zero.

Documentation:

I took the charge of writing the documentation for part-3A as well as part-3B of the assignment. In which I documented, the explanation of the code and method to implement both the above tasks.