COMP6714 Proj_part2

Name: Zanning Wang

zID: z5224151

Part2:

U(A):4	U(B):4	U(C):3
(1, 4)	(1, 3)	(1, 1)
(3, 1)	(2, 4)	(2, 3)
(4, 4)	(4, 4)	(5, 3)
(5, 1)	(6, 2)	(6, 3)

In order to get a threshold(α), to reduce the number of full evaluations, we can make use of the upper bound of each term. Suppose we have the example, the top-k is 3 and the final result is [(8, 1), (7, 2), (8, 4)]. To get the α max, firstly we should calculate the sum of each document's upper bound which may 11, 7, 8. In current example, and we choose the minimum among the result which is 7 as the α max, if we set threshold smaller than α max, it will always return the correct result because the previous term total score will never exceed the minimum upper bound(sum of each term's upper bound)

If we set the threshold >= the α max, the smaller score may lost because their sum of upper bound will always no larger than α max, at current case, if we set the α max to 10, we rerun the same data, both (7, 2) and (8, 2) will lost. If the threshold smaller than 7, such as 5, even the program may run more full evaluations, it will still output the same result [(8, 1), (7, 2), (10, 4)].