6714 PROJECT PART2

Name: Shengqian Zou Student ID: z5203970

If use the following 6 lines of code to replace Lines 28 to 34 of the ADCS2013 algorithm, it means that the threshold will not be updated, but set at the beginning. The initial size of the threshold has a direct relationship between the Top-k result and the algorithm time.

```
1 if |Ans| < k then
2 |   Ans ← insert(Ans, (c_pivot, s));
3 else
4 |   if s > minimum(Ans) then
5 |   |   Ans ← insert(Ans, (c_pivot, s));
6 |   |   Ans ← delete_smallest(Ans);
```

(Figure 1)

Task

1. According to the Assumptions, we already had top-k answer for Q and there is no two documents who achieve the same score for the query Q. So, if we want the algorithm (NB: the modified version of ADCS 2013 algorithm) will always return the correct top-k answers for Q. This needs to ensure that all results in top-k can enter the 14th line of the ADCS2013 algorithm(Figure 2).

if *tmp_s_lim* $> \theta$ **then**
  break, and continue from step 20
end if

(Figure 2)

In order to ensure that top-k (Q) can be entered in the IF statement in the above figure 2, it is necessary to satisfy that the minimum accumulated UB(tmp_s_lim) in Top-k is greater than the threshold(θ, now it is equal to α here). According the note, if we set α =min{score(D;Q) | D∈top-k(Q)}, the algorithm will always return the correct top-k answers. If we want to find a higher value for α here, we just need to satisfy that the minimum accumulated UB(tmp_s_lim) in Top-k is greater than the threshold. So, α=min{ the minimum accumulated UB(D;Q) | D∈top-k(Q)}. The value of the min{ the minimum accumulated UB(D;Q) | D∈top-k(Q)} should greater than min{score(D;Q) | D∈top-k(Q)}, because the score is accumulated the real score for each term and the score can not greater than the UB of the term.

A. When α_{max}=min{ the minimum accumulated UB(D;Q) | D∈top-k(Q)}, $\forall\alpha<\alpha_{max}$, the algorithm (NB: the modified version of ADCS 2013 algorithm) will always return the correct top-k answers for Q.

B. When α_{max}=min{ the minimum accumulated UB(D;Q) | D∈top-k(Q)}, $\forall\alpha\geqslant\alpha_{max}$, the algorithm (NB: the modified version of ADCS 2013 algorithm) will always return the correct top-k answers for Q.

2. If α_{max}=min{ the minimum accumulated UB(D;Q) | D∈top-k(Q)}, for $\forall\alpha<\alpha_{max}$, all the results in top-k can enter the 14th line(Figure 2), including the minimum accumulated UB(tmp_s_lim) in top-k(it will > α). So, it can make sure that it will not miss any results and return the correct top-k answers for Q. For $\forall\alpha\geqslant\alpha_{max}$, if the minimum accumulated UB(tmp_s_lim) in Top-k is equal to α

or small than α, some results of top-k can not enter the 14th line(Figure 2), it will miss some results. So the algorithm may return the wrong top-k answers for Q in some cases.