Information Retrieval 15/12/2020

Q1: First of all the sequence is re-coded via gaps → 1, 1, 1, 7, 2, 1

$\gamma(1) = 1$, $\gamma(7) = 00111$, $\gamma(2) = 010$

Ford Delta uses b = 1 ⇒ 0, 0, 0, 6, 1, 0

b = 2 ⇒ 00, 00, 00, 11, 01, 00 ‖ 6

Q2: We first construct the Inverted List

a → 3
black → 3
book → 1, 2, 3
really → 2
white → 1, 2, 3

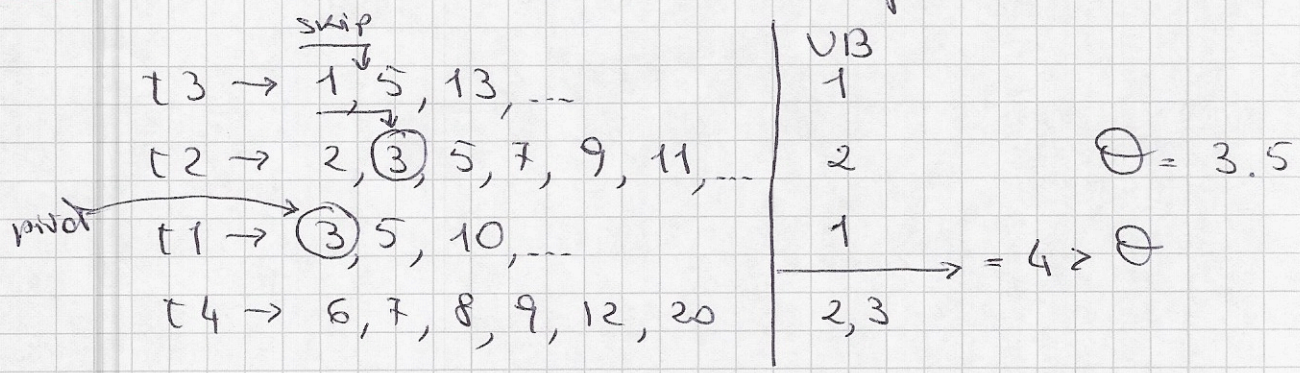
idf = $\log_2 3/1 \doteq L$
" = $\log_2 3/1 \doteq L$
" = $\log_2 3/3 = 0$
" = $\log_2 3/1 = L$
" = $\log_2 3/2 = L-1$

	T_1	T_2	T_3
a	0	0	$1 \cdot L$
black	0	0	$1 \cdot L$
book	0	0	0
really	0	$1 \cdot L$	0
white	$1 \cdot (L-1)$	$2 \cdot (L-1)$	0

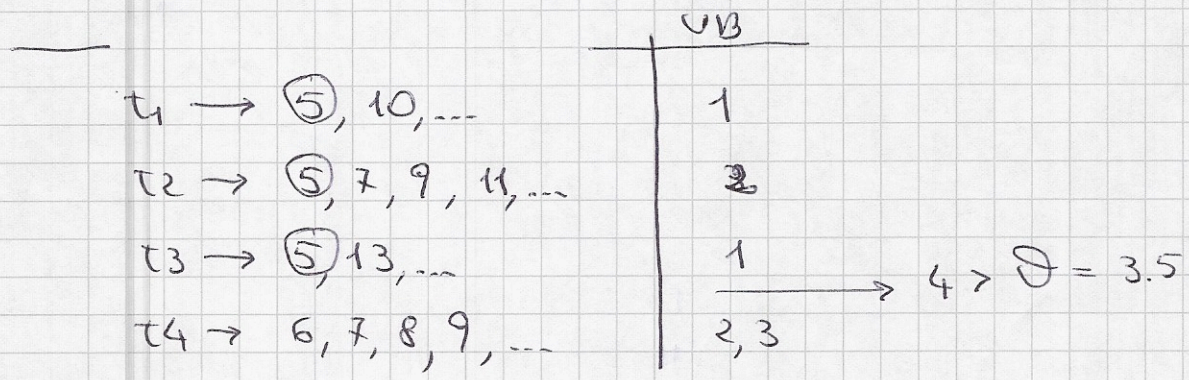
q = "black book" ⇒ $[0, 1 \cdot L, 1 \cdot 0, 0, 0]$

The most similar text is T_3

Q3. We sort the lists for increasing iterators:



* Hence the pivot is the docID 3 in t1's posting list
* Then we skip to 3 in T2 and T3, but 3 is found only in T2, but not in T3.
* $UB_2 + UB_1 = 3 < \Theta$ so it is wasteful to compute the score of 3 $\Rightarrow$ 3 is discarded



they are already sorted, the min-heap has been not changed because 3 was discarded, hence Θ is the same as before. The pivot is 5, the sum of the $UB_1 + UB_2 + UB_3 > \Theta$ and all they contain 5 so 5 is evaluated in T's full score.

Q3

Referring to the previous case, the lists are

$t1 \rightarrow$ [5, 10, 12] $UB^L = 1$

$t2 \rightarrow$ [5, 7, 9, 11, 12, 13] $UB^L = 1$

$t3 \rightarrow$ [5, 13, 20] $UB^L = 1$

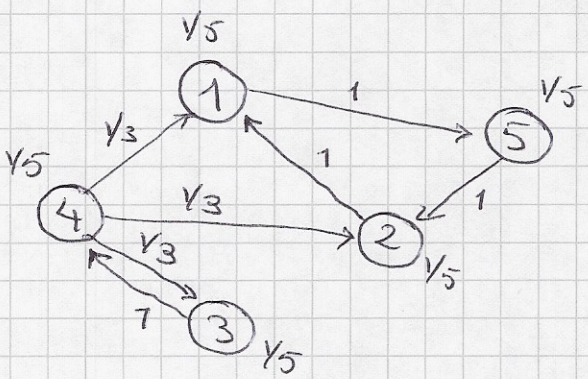
$t4 \rightarrow$ [6, 7, 8, 9, 12, 20] $UB^L = 3,3$

We have selected again 5 because of the global UBs by the local UBs of the blocks including 5 have now smaller than 9 and now the block 5 is not fully evaluated.

The block with the right extreme which is the leftmost one is t_2, and thus t2 moves to (11), and 11 moves to (10) and t3 to (13).

Q4.

$\alpha = 1/2$



$$Pr(1) = \frac{1}{2}\left(\frac{1}{5}\times\frac{1}{3} + \frac{1}{5}\times 1\right) + \frac{1}{2}\times\frac{1}{5}$$

$$Pr(2) = \frac{1}{2}\left(\frac{1}{5}\times\frac{1}{3} + \frac{1}{5}\times 1\right) + \frac{1}{2}\times\frac{1}{5}$$

$$Pr(4) = \frac{1}{2}\left(\frac{1}{5}\times 1\right) + \frac{1}{2}\times\frac{1}{5} = Pr(5)$$

$$Pr(3) = \frac{1}{2}\left(\frac{1}{5}\times\frac{1}{3}\right) + \frac{1}{2}\times\frac{1}{5}$$