

# Information Retrieval

8/2/2023

Q1 Compute intersection among the following 3 lists

$$T1 \rightarrow 3, 5, 6$$

$$T2 \rightarrow 2, 3, 6$$

$$T3 \rightarrow 2, 3, 4, 6$$

$$\langle 3, 2, 2 \rangle \quad \langle 3, 3, 3 \rangle \quad \langle 5, 6, 4 \rangle \quad \langle 5, 6, 6 \rangle \quad \langle 6, 6, 6 \rangle$$

Q2

$$D1 = \begin{matrix} \text{"a b c"} \\ 1 \ 2 \ 3 \end{matrix}$$

$$D2 = \begin{matrix} \text{"b d b"} \\ 2 \ 4 \ 2 \end{matrix}$$

$$D3 = \begin{matrix} \text{"a d a"} \\ 1 \ 4 \ 1 \end{matrix}$$

	D1	D2	D3
$h_1$ $3 \times \underline{m} \ 5$	3, ①, 4	①, 2, 1	3, ②, 3
$h_2$ $2 \times \underline{m} \ 5$	2, 4, ①	4, ③, 4	②, 3, 2

$$SK(D1) = [1, 1]$$

$$SK(D2) = [1, 3]$$

$$SK(D3) = [2, 2]$$

So the estimates according to the sketches of the Jaccard similarity between pairs of documents, is:

$$J(D1, D2) = 1/2$$

$$J(D2, D3) = 0 = J(D1, D3)$$

← the most similar pair! ▽

Q3 Web Graph comprehension

15 → 3, 5, 6, 7, 8, 10, 16, 17, 18, 22, 24, 26, 34

16 → 5, 6, 7, 8, 9, 16, 17, 20, 21, 22, 24, 29, 30

15 is copied "as is", so we detail only the list of 16

	outd	ref	copy list	extre nodes
16	13	1	0 1 1 1 1 0 1 1 0 1 1 0 0	9, 20, 21, 29, 30
			↓	
			#blocks	copy blocks
			7	0, 0, 3, 0, 1, 0, 1

↑  
1<sup>st</sup> bit

Q4

Sort the parking list according to their needs:

$T_3 \rightarrow \overset{0.4}{\text{1, 6, 9}}, 13, 15$

$ub_3 = 0.6$

$T_2 \rightarrow \overset{0.5}{\text{2, 6, 7}}, 8, 11$

$ub_2 = 1$

$\Theta = 2.2$

$T_1 \rightarrow \overset{0.3}{\text{5, 6, 12}}, 13$

$ub_1 = 0.4$

$T_4 \rightarrow \overset{0.5}{\text{6, 7, 8}}, 11$

$ub_4 = \frac{0.5}{2.5 \text{ Total}} > \Theta$

↑  
first delID

- According to the UBs, doc 6 should be evaluated for its score
- Given that blocks have size 3, and their overall sum is 1.7 which is smaller than  $\Theta$ , doc 6 must not be evaluated.
- The block to be discarded is the one with leftmost right extreme, hence the one of  $T_2$ .