Information Retrieval

16 Jan 2024

Q1

15 → 1, 3, 5, 6, 7, 8, 10, 16, 17, 22, 24, 44

16 → 2, 3, 5, 6, 7, 8, 9, 10, 16, 17, 20, 21, 22, 24

We rewrite 16 as follows

ref	copy list	extra nodes
	0 1 1 1 1 1 1 1 1 1 1 0	2, 9, 20, 21

# blocks	copy blocks
3	0, 0, 9, 0

0 can be deleted

Q2

A = {2, 5, 6, 9} B = {1, 2, 4} C = {1, 5, 6, 9}

$J(A,B) = \frac{1}{6}$

$J(A,C) = \frac{3}{5}$ ← The most similar pair

$J(B,C) = \frac{1}{6}$

	$\pi_1(x) = 3x \bmod 11$	$\pi_2(x) = x+5 \bmod 11$	$\pi_3(x) = 4x \bmod 11$
$A = \{2, 5, 6, 9\}$	6, (4), 7, 5	7, 10, (0), 3	8, 9, (2), 3
$B = \{1, 2, 4\}$	3, 6, (1)	(6), 7, 9	(4), 8, 5
$C = \{1, 5, 6, 9\}$	(3), 4, 7, 5	6, 10, (0), 3	4, 9, (2), 3

sketch (A) = [4, 0, 2]

sketch (B) = [1, 6, 4]

sketch (C) = [3, 0, 2]

$J(A, B) \simeq \frac{0}{3}$ $J(A, C) \simeq \frac{2}{3}$ $J(B, C) = \frac{0}{3}$

So the estimate obtained with min-hash would correctly find (A, C) as the most similar pair.

Q3

$A \to 1, 2, 3$

- $AA \to 1$
- $AB \to 1, 2$
- $BA \to 2$

$AC \to 3$

$CA \to 3$

$Q = BAAB$

↳ B, BA, AA, AB

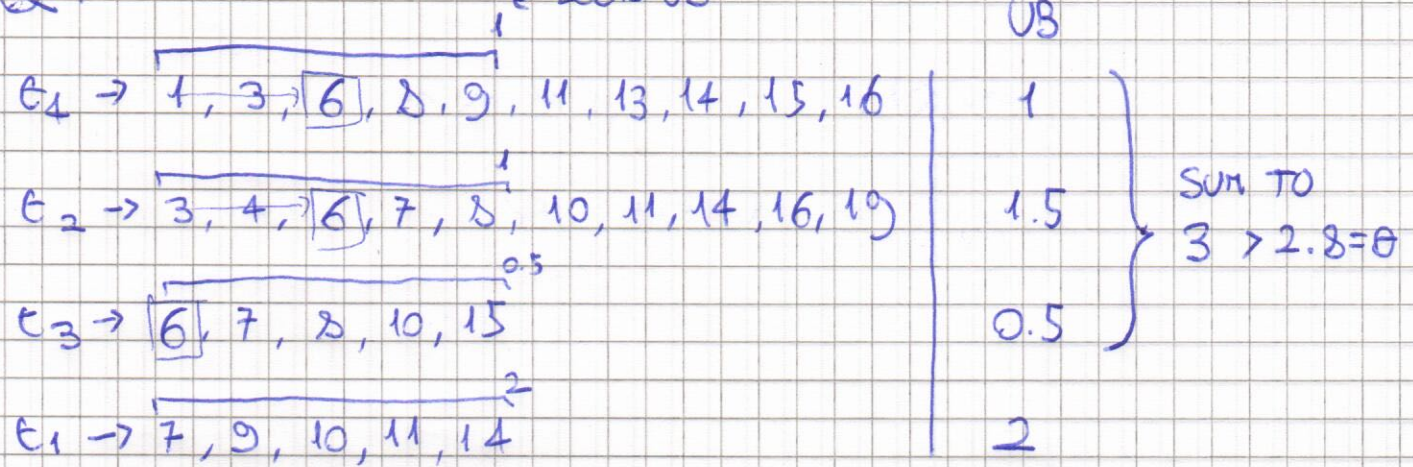
access the posting-lists of these bigrams and count #occurrences for each string

$\#s_1 = 2$, $\#s_2 = 2$, $\#s_3 = 0$ (implicitly derived)

the threshold is $= |Q| - K \cdot e = 4 - 2 \cdot 1 = 2$

therefore the candidate strings are s_1 and s_2.

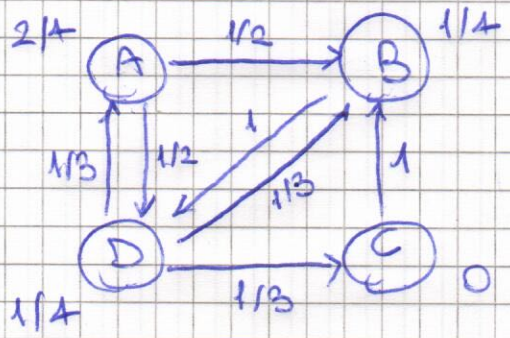
Q4



PIVOT IS 6 AND WE DO NOT COMPUTE ITS FULL SCORE BECAUSE THE LOCAL UBs OF BLOCKS CONTAINING IT SUM TO 2.5 < θ.

WE SKIP NO BLOCK (NOTE $t_1 \rightarrow 7$) AND MOVE ITERATORS POINTING TO 6 TO THE NEXT DOCID, SO THE STATE OF THE ITERATORS IS: $t_1 \rightarrow 7$, $t_2 \rightarrow 7$, $t_3 \rightarrow 7$, $t_4 \rightarrow 8$.

Q5



$$r(A) = \frac{1}{2}\left(\frac{1}{4}\cdot\frac{1}{3}\right) + \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{24} + \frac{1}{8} = \frac{4}{24}$$

$$r(B) = \frac{1}{2}\left(\frac{2}{4}\cdot\frac{1}{2} + 0\cdot 1 + \frac{1}{4}\cdot\frac{1}{3}\right) + \frac{1}{8} = \frac{1}{2}\left(\frac{3+1}{12}\right) + \frac{1}{8} = \frac{7}{24}$$

$$r(C) = \frac{1}{2}\left(\frac{1}{4}\cdot\frac{1}{3}\right) + \frac{1}{8} = r(A) = \frac{4}{24}$$

$$r(D) = \frac{1}{2}\left(\frac{2}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot 1\right) + \frac{1}{8} = \frac{1}{2}\,\frac{2}{4} + \frac{1}{8} = \frac{3}{8} = \frac{9}{24}$$