

Information Retrieval

7/2/22

Q1.

$$S = (2, 4, 6, 10, 12)$$

$u = 12 + 1 = 13 \rightarrow$ numbers are encoded in 4 bits

2	0010
4	0100
6	0110
10	1010
12	1100

$$l = \lceil \lg_2 \frac{u}{n} \rceil = \lceil \lg_2 \frac{13}{5} \rceil = 2$$

$$\Rightarrow L = 1000101000$$

$$H = 101101010$$

base = 2 $\rightarrow S' = (0, 2, 2, 4, 2)$

0010101110 | 4 full representation

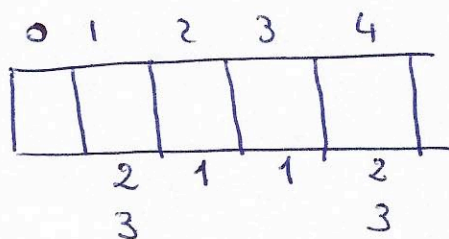
Q2

$$S = \{1, 2, 3\}$$

$$h_1(k) = 3 \times k \pmod{5}$$

$$h_2(k) = 2 \times k \pmod{5}$$

level 1

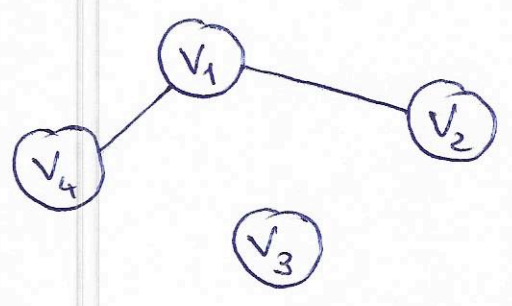


When we insert the key 3 we have a repeated minimum so the second level has not to be constructed even if there is a collision.

Q3. $V = \{00011, 00100, 01010, 11111\}$

Projections are $I_1 = \{1, 2\}$ $I_2 = \{2, 3\}$

	1 2 3 4 5	I_1	I_2
v_1	00011	00	00
v_2	00100	00	01
v_3	01010	01	10
v_4	10011	11	00



⇒ we have two connected components and two groups of similar sectors $\{v_1, v_2, v_4\}$ and $\{v_3\}$

Q4.

$D = \{ \underset{s_1}{aabb}, \underset{s_2}{abc}, \underset{s_3}{acac} \}$

- $\$a = 1, 2, 3$
- $aa = 1$
- $ab = 1, 2$
- $bb = 1$
- $bc = 2$
- $ac = 3$
- $ea = 3$

Since $e=1, k=2$, the #q-grams to be stored with a query string $Q = aabc$ is $L = 4 - 2 \cdot 1 = 2$

$\$Q \rightarrow \{\$a, aa, ab, bc\}$

These bi-grams are searched in the index above, and the portions below are scanned and counted the occurrences of every bi-gram in those strings

$\$a \rightarrow 1, 2, 3$	↳	s_1 includes 3 bi-grams
$aa \rightarrow 1$		s_2 " 3 "
$ab \rightarrow 1, 2$		" " 1 "
$bc \rightarrow 2$		s_3 " 1 "

Given the value of $L=2$, only the strings s_1, s_2 are taken as candidates. We thus compute DP matrix for the shortest one, namely $s_2 = a b e$

		a	a	b	e
0	0	1	2	3	4
a	1	0	1	2	3
b	2	1	1	1	2
e	3	2	2	2	1

← edit distance (s_2, Q)

Q5.

$$T_1 \rightarrow 2, 4, 5$$

$$T_2 \rightarrow 3, 4$$

$$T_3 \rightarrow 2, 3, 4$$

$$(2, 3, 2) \rightarrow \text{obj ID} = 2 \quad \text{counted } 2$$

$$(4, 3, 3) \rightarrow \text{obj ID} = 3 \quad \text{counted } 2$$

$$(4, 4, 4) \rightarrow \text{obj ID} = 4 \quad \text{counted } 3$$

$$(11, 11, 5) \rightarrow \text{obj ID} = 5 \quad \text{counted } 1$$