

Information Retrieval

17/1/23

Q1

$$S = (4, 6, 10, 12, 17, 24)$$

$$\text{gaps} = 4, 2, 4, 2, 5, 7$$

$$\text{gamma} = 00100, 010, 00100, 010, 00101, 00111$$

$$w = 5, \quad l = \lceil \log_2 \frac{u}{n} \rceil = \lceil \log_2 \frac{25}{6} \rceil = 3$$

$$n = 6$$

$$u = 25$$

$$h = w - l = 5 - 3 = 2$$

4	00	100
6	00	110
10	01	010
12	01	100
17	10	001
24	11	000

\longleftrightarrow \longleftrightarrow
 h l

$$L = 100 \ 110 \ 010 \ 100 \ 001 \ 000$$

$$H = 110 \ 110 \ 10 \ 10$$

$$\text{base} = 4, \quad b = 2 \rightarrow \text{esc} = 11$$

$$S' = 0, 2, 6, 8, 13, 20$$

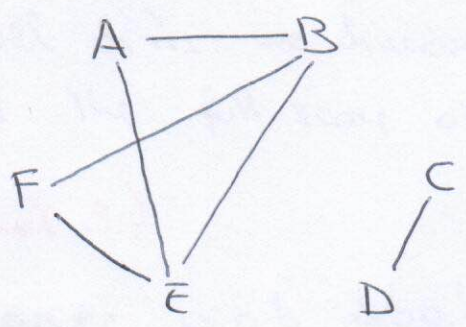
$$\text{gap}(S') = 0, 2, 4, 2, 5, 7$$

$$00, 10, 11, 10, 11, 11 \mid 4, 5, 7$$

2

A = 00000
 B = 00100
 C = 01001
 D = 01101
 E = 10011
 F = 10111
 1 2 3 4 5

	I ₁	I ₂	I ₃
A	00	00	00
B	00	01	00
C	01	10	01
D	01	11	01
E	10	00	00
F	10	01	11



Most similar vectors
 {A, B, E, F}
 {C, D}

Q3

D = {aa₁cc, ac₂b, ab₃ab} k=2 → pre-append \$

- \$a → 1, 2, 3
- aa → 1
- ac → 1, 2
- cc → 1
- eb → 2
- ab → 3
- ba → 3

Q = aacb ⇒ \$aacb

\$a → 1, 2, 3	} s ₁ ⇒ 3 s ₂ ⇒ 3 s ₃ ⇒ 1
aa → 1	
ac → 1, 2	
cb → 2	

↳ Q - k * e = 4 - 2 * 1 = 2
 so candidates are
 s₁ = aacc and s₂ = acb

The shortest candidate is acb

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

← if is a correct match for e=1

24

we sort by list's head:

$t_3 \rightarrow 1, 4, \textcircled{6}, 7, 13, 15$

$t_2 \rightarrow 2, 3, 5, 7, 8, 11$

$t_1 \rightarrow 5, \textcircled{6}, 7, 8, 11$

$t_4 \rightarrow \textcircled{6}, 7, 8, 11$
pivot

UB	
0.6	
1	1.5
0.4	2
0.5	2.5

$\Theta = 2.2$

The pivot does not occur in t_2 , so that after locating it in all lists we discover that $\sum UB_i = 1.5 < \Theta$ so that the full score of 6 is not computed, indeed.

Q3 (point 3)

$$D_1 = \{aaeb, acb, abeb\}$$

$$D_2 = \{aeb[s_1], aac[s_1],$$

$$eb[s_2], ab[s_2], ae[s_2],$$

$$bab[s_3], aab[s_3], abb[s_3], abba[s_3]\}$$

$$Q = aaeb \rightarrow Q^- = \{acb, aab, aac\}$$

$$Q \text{ in } D_1 \Rightarrow \emptyset$$

$$Q \text{ in } D_2 \Rightarrow \emptyset$$

$$Q^- \text{ in } D_1 \Rightarrow s_2;$$

$$Q^- \text{ in } D_2 \Rightarrow s_3, s_1$$

} so we return all strings as candidates.