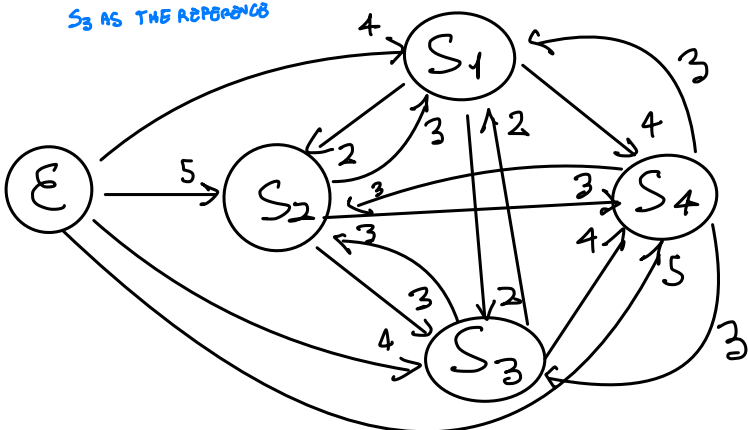


#1

$S = \{ \underset{S_1}{ABACO}, \underset{S_2}{BASCO}, \underset{S_3}{RACD}, \underset{S_4}{VASTO} \}$

INPUT	GZIP	# TUPLES
S_1	$\langle 0, 0, A \rangle \langle 0, 0, B \rangle \langle 2, 1, C \rangle \langle 0, 0, D \rangle$	4
S_2	$\langle 0, 0, B \rangle \langle 0, 0, A \rangle \langle 0, 0, S \rangle \langle 0, 0, C \rangle \langle 0, 0, D \rangle$	5
S_3	$\langle 0, 0, R \rangle \langle 0, 0, A \rangle \langle 0, 0, C \rangle \langle 0, 0, D \rangle$	4
S_4	$\langle 0, 0, V \rangle \langle 0, 0, A \rangle \langle 0, 0, S \rangle \langle 0, 0, T \rangle \langle 0, 0, D \rangle$	5
$S_1 S_2$	$\langle 4, 1, B \rangle \langle 2, 1, C \rangle \langle 5, 1, EOF \rangle$	3
$S_1 S_3$	$\langle 3, 1, B \rangle \langle 5, 3, EOF \rangle$	2
$S_1 S_4$	$\langle 4, 1, B \rangle \langle 2, 1, C \rangle \langle 5, 1, EOF \rangle$	3
$S_2 S_1$	$\langle 4, 2, S \rangle \langle 5, 2, EOF \rangle$	2
$S_2 S_3$	$\langle 0, 0, B \rangle \langle 4, 1, S \rangle \langle 5, 2, EOF \rangle$	3
$S_2 S_4$	$\langle 0, 0, B \rangle \langle 5, 2, C \rangle \langle 5, 1, EOF \rangle$	3
$S_3 S_1$	$\langle 0, 0, R \rangle \langle 4, 3, EOF \rangle$	2
$S_3 S_2$	$\langle 0, 0, R \rangle \langle 5, 1, C \rangle \langle 4, 1, EOF \rangle$	3
$S_3 S_4$	$\langle 0, 0, R \rangle \langle 5, 1, C \rangle \langle 4, 1, EOF \rangle$	3
$S_4 S_1$	$\langle 0, 0, V \rangle \langle 4, 1, S \rangle \langle 0, 0, T \rangle \langle 5, 1, EOF \rangle$	4
$S_4 S_2$	$\langle 0, 0, V \rangle \langle 5, 2, T \rangle \langle 5, 1, EOF \rangle$	3
$S_4 S_3$	$\langle 0, 0, V \rangle \langle 4, 1, S \rangle \langle 0, 0, T \rangle \langle 5, 1, EOF \rangle$	4

↑ COMPRESSION OF S_2 GIVEN S_3 AS THE REFERENCE



POSSIBLE SOLUTION:

$$E \xrightarrow{4} S_3 \xrightarrow{2} S_1 \xrightarrow{2} S_2 \xrightarrow{3} S_4$$

#2

$$D = \left\{ \underset{1}{ABBA}, \underset{2}{ABC}, \underset{3}{BABB} \right\}$$

$$\$A \rightarrow 1, 2$$

$$\$B \rightarrow 3$$

$$AB \rightarrow 1, 2, 3$$

$$BA \rightarrow 1, 3$$

$$BB \rightarrow 1, 3$$

$$BC \rightarrow 2$$

$$Q = ABCC$$

$$\$A \rightarrow 1, 2$$

$$AB \rightarrow 1, 2, 3$$

$$BC \rightarrow 2$$

$$CC \rightarrow \phi$$

$\geq |Q| - e \cdot k = 4 - 1 \cdot 2 = 2$ OF THE
K-GRAMS MUST MATCH, SO THE CANDIDATES

$$ARE \underset{1}{ABBA}, \underset{2}{ABC}$$

3

IDP

A → 1, 2, 3, 4

$$\log \frac{4}{4} = 0$$

AN → 4

$$\log \frac{4}{1} = 2$$

APPLE → 4

$$\log \frac{4}{1} = 2$$

HAVE → 1, 2

$$\log \frac{4}{2} = 1$$

I → 1, 2

$$\log \frac{4}{2} = 1$$

PEN → 1, 3, 4

$$\log \frac{4}{3}$$

PINEAPPLE → 2, 3

$$\log \frac{4}{2} = 1$$

	T_1	T_2	T_3	T_4	q
A	1×0	1×0	2×0	1×0	1×0
AN	0	0	0	1×2	0
APPLE	0	0	0	1×2	1×2
HAVE	1×1	1×1	0	0	1×1
I	1×1	1×1	0	0	0
PEN	$1 \times \log \frac{4}{3}$	0	$1 \times \log \frac{4}{3}$	$2 \times \log \frac{4}{3}$	0
PINEAPPLE	0	1×1	1×1	0	0

$$T_1 \cdot q = 1 \times 1 = 1$$

$$T_2 \cdot q = 1 \times 1 = 1$$

$$T_3 \cdot q = 0$$

$$T_4 \cdot q = 2 \times 2 = 4$$

So T_4 is the most similar text to q

#4

a) $u = 32$ $b = \lceil \log u \rceil = 5$
 $n = 9$ $w = \lceil \log \frac{u}{n} \rceil = 2$

3	0	0	0	1	1
4	0	0	1	0	0
12	0	1	1	0	0
20	1	0	1	0	0
24	1	1	0	0	0
27	1	1	0	1	1
28	1	1	1	0	0
29	1	1	1	0	1
31	1	1	1	1	1

$L = 11\ 00\ 00\ 00\ 00\ 00\ 11\ 00\ 01\ 11$

$M = \begin{matrix} 10 & 10 & 0 & 10 & 0 & 10 & 110 & 1110 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix}$

b) $GAP(S) = (3, 1, 8, 8, 4, 3, 1, 1, 2)$

$\gamma(GAP(S)) = 011\ 1\ 0001000\ 0001000\ 00100\ 011\ 1\ 1\ 010$

c) $GAP(S) - BASE = (2, 0, 7, 7, 3, 2, 0, 0, 1)$

PFOR DELTA 10 00 11 11 11 10 00 00 01

EXCEPTIONS 7, 7, 3

#5

$d(A) = h(c) = 2$

$d(B) = h(A) + h(c) + h(D) = 2 + 2 + 1 = 5$

$h(A) = d(B) + d(D) = 1 + 1 = 2$

$h(B) = d(D) = 1$