Unit-4

12/12/2021

String Matching

Find substring in mainstring
↓ ↓
Pattern – M. Text (length → n)

1) T = a b c d a b f a b e g
P = a b e

Sol:→

T = a b c d a b f a b e g
P = a b e

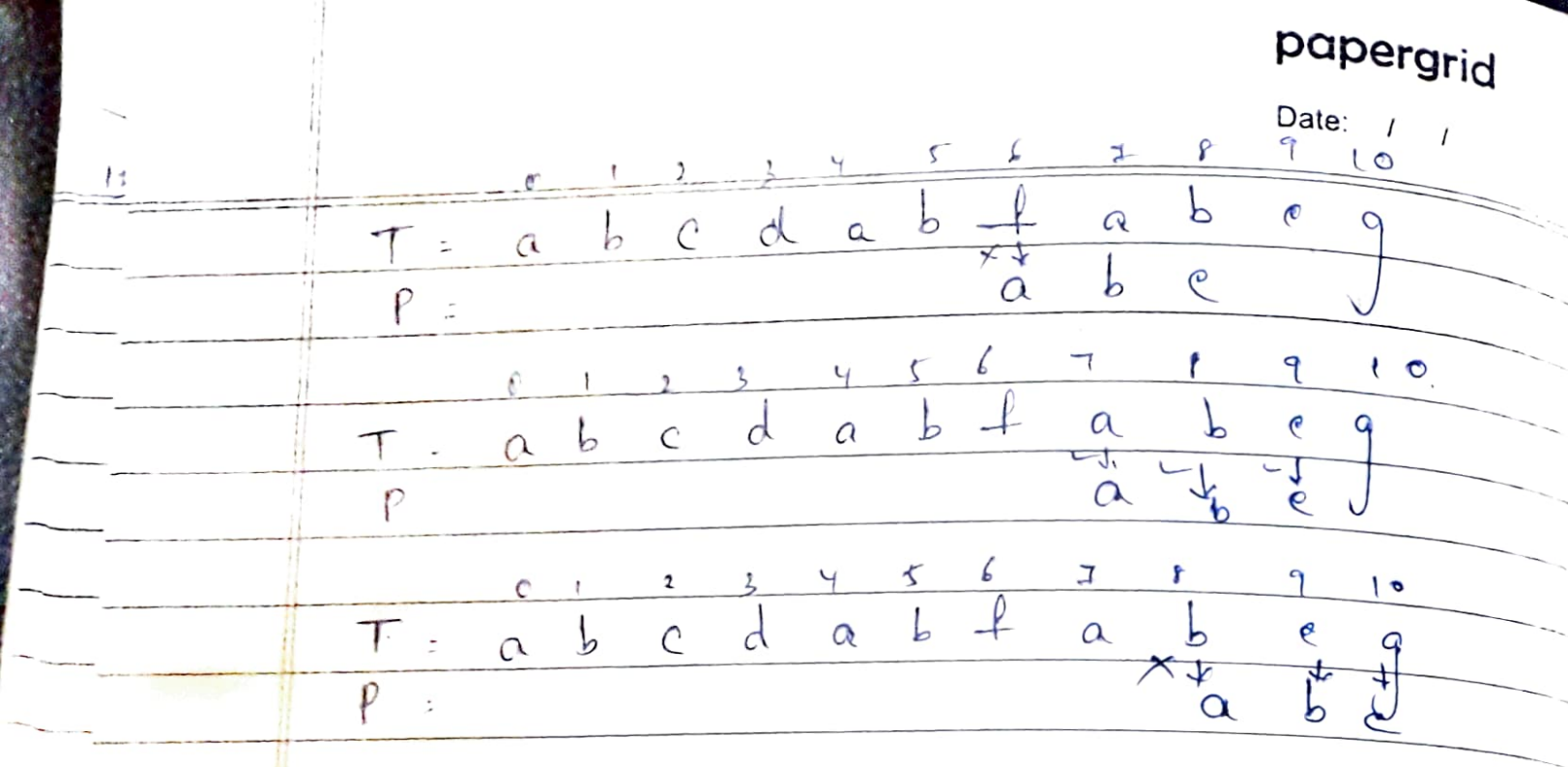
T = a b c d a b f a b e g
P: a b e

T = a b c d a b f a b e g
P: a b e

T = a b c d a b f a b e g
P: a b e

T = a b c d a b f a b e g
P: a b e

T = a b c d a b f a b e g
P: a b e



Naive-String-Matches (T, P)

1. n = T.length
2. m = P.length
3. for s = 0 to n-m
4. if P[1...m] == T[s+1...s+m]
5. print "Pattern occurs with shift s"

2) T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0
P = 0 0 0 1

Sol:-

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

P = 0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

P = 0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

P = 0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

0 0 0 1 0

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

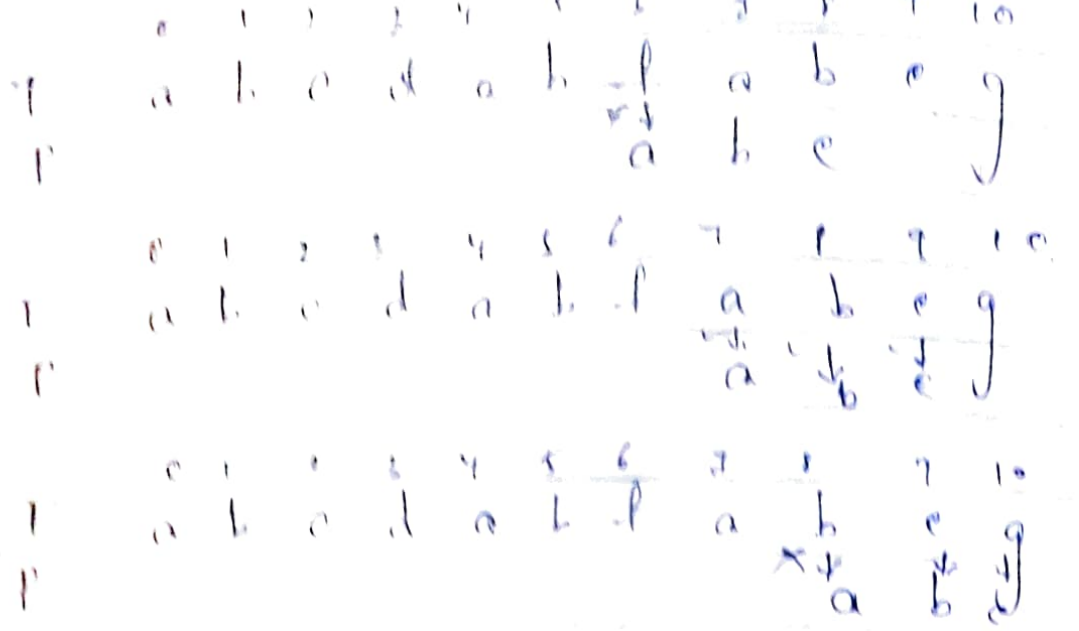
T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

P = 0 0 0 1

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

P - 0 0 0 1



Naive String Matcher (T, P)

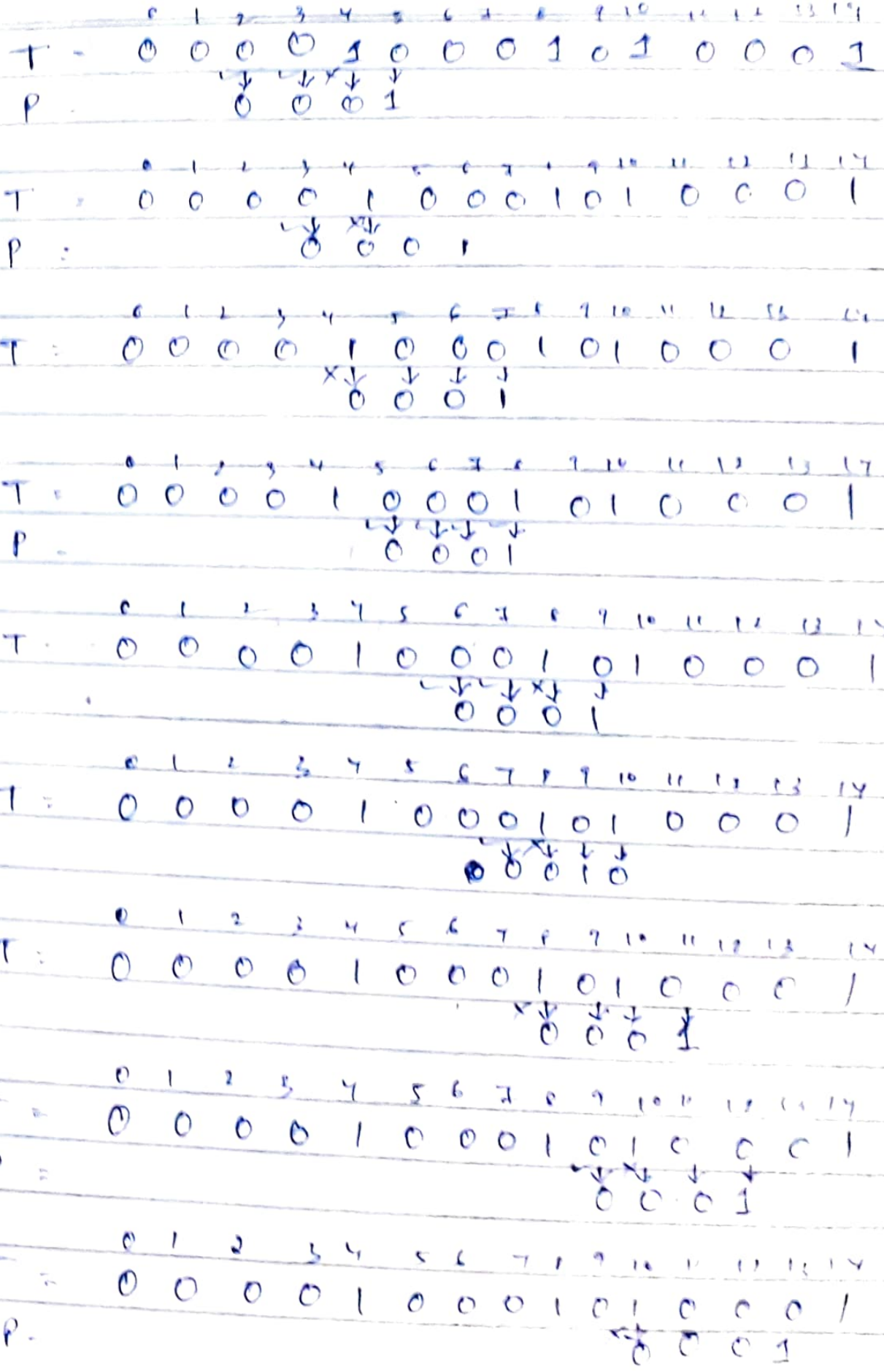
```
n = T.length
m = P.length
for s = 0 to n-m
    if P[1..m] = T[s+1..s+m]
        print "Pattern occurs with shift" s
```

2) T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0
P = 0 0 0 1

Sol: T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0
P = 0 0 0 1

T = 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1
P = 0 0 0 1

Date: / /



Date: / /

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

T: 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1

P: 0 0 0 0

S = 1

S = 5

S = 11.

Boyer Moore.

T: BESS_KNEW_ABOUT_BAOBABS.

P: BAOBAB

Bad symbol table

c	A	B	S	O	_	*
t(c)	1	2	6	3	6	6

m - 1 - j.

m → no. of letter

j → index no

→ We have to consider rightmost letter

Good shift table

K	Pattern	d_2
1	BAOBAB	2
2	BAOBAB	5
3	BAOBAB	5
4	BAOBAB	5
5	BAOBAB	5
~~6~~	~~BAOBAB~~	~~5~~

T: BESS_KNEW_ABOUT_BAOBABS

P: ~~BAOBAB~~

P: BAOBAB

$d_1 = t(c) - k$

$= 6 - 0$

$= 6$

T: BESS_KNEW_ABOUT_BAOBABS

P: BAOBAB

$d_1 = 6 - 2$

$= 4$

$d_2 = 5$

$= \text{Max}\{d_1, d_2\}$

$= \text{Max}\{4, 5\}$

$= 5.$

T: BESS_KNEW_ABOUT_BAOBABS.

P: BAOBAB

$d_1 = 6 - 1$

$= 5$

$d_2 = 2$

$= \text{max}\{d_1, d_2\}$

$= \text{max}\{5, 2\}$

$= 5$

1:

T: BESS_KNEW_ABOUT_BAOBABS

P: BAOBAB$

String is matched.

H.W

T: JIM_SAW_ME_IN_A_BARBERSHOP

P: BARBER

Bad symbol table

C	A	B	E	R	*
H(c)	4	2	1	3	6

Good shift table

K	Pattern	d_2
1	BARBE R	3
2	BAR BER	3
3	BAR BER	3
4	BAR BER	3
5	BAR BER	3

T: JIM_SAW_ME_IN_A_BARBERSH

P: BARBER

$d_1 = \max(H(c), -k)$

$= \max(4, 0)$

$= 4.$

T: JIM_SAW_ME_IN_A_BARBERSHOP

P: ~~BARBBARBER~~

$d_1 = \max(t(c), k)$

$\max(1, 0)$

$= 1$

T: JIM_SAW_ME_IN_A_BARBERSHOP

P: BARBER

$= \max(t(k) - k, 1)$

$\max(6 - 0, 5)$

$= 6.$

T: JIM_SAW_ME_IN_A_BARBERSHOP

P: BARBER

$= \max(t(c) - k, d_2)$

$= \max(2, 6)$

$= 2.$

T: JIM_SAW_ME_IN_A_BARBERSHOP

P: BARBER

$= \max(t(c) - k, d_2)$

$= \max(4 - 0, 0)$

$= 4.$

Date: / /

T: JIM - SAW - ME - IN - A - BARBERSHOP

P: BARBER

String is matched.

Q WOWWOW

Good shift table..

K	Pattern	d_2
1	W O W W O W	2
2	W O W W O W	5
3	W O W W O W	5
4	W O W W O W	5
5	W O W W O W	5

#. when there is first match don't select last one

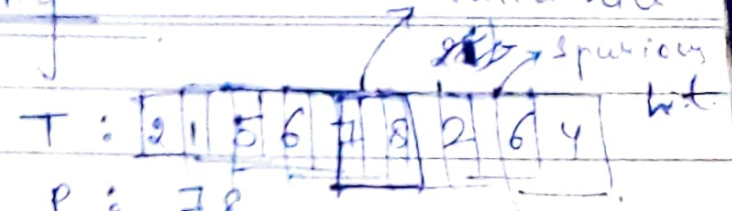
Q KUYULAYU

Good shift table

K	Pattern	d_2
1	KUYULAYU	6
2	KUYULAYU	4
3	KUYULAYU	4
4	KUYULAYU	4
5	KUYULAYU	4
6	KUYULAYU	4

Date:

Hashing



21 = 8 x

15 = 2 x

56 = 4 x

67 = 2 x

78 = 0 ✓

82 = 4

26 = 0