Lecture 11

LAST TIME
→ linear computation of LCP ⇒ linear creation of S.T.

PBT
→ query start BWT

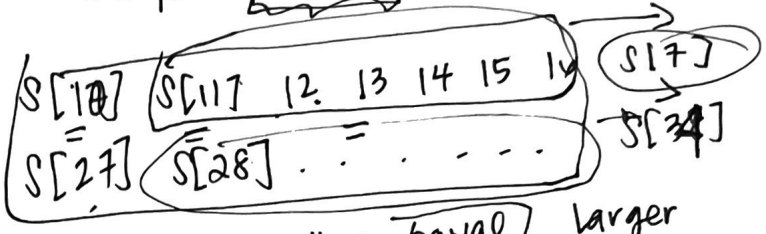
TODAY
→ SAMPLE QUIZ QUESTION
→ BWT FINISHED.

① Algo analysis as described on CANVAS

②

(A) S[17] ? S[34] smaller equal larger

Why



(B) S[11] ? S[28] ⇒ smaller equal larger

(C) match(suff28, suff11) ? = 17 - 11 = 6

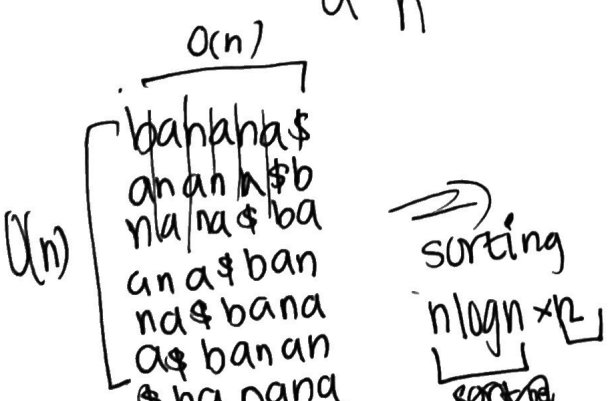
③ Run SA linear algo on a string S = ababab for example
Only top-level like we did in class.

BWT:

S = banana$

cat$
at$c
t$ca
$cat

O(n)



$banana (a)
a$banan (n)
ana$ban (n)
anana$b (a)
banana$ ($)
na$bana (a)
nana$ba (a)

BWT(s) = annb$aa

$n \times n \log n$

cost of comparing

comparisons

given n strings, length n^a w/ cost of comparing $O(1)$.

SORT TIME? $n \log n \times O(1) =$

RADIX items x digits

$n \times n = O(n^2)$

banana$
1 2 3 4 5 6 7

SORTED LIST of ROTATIONS

SUFF[7]		
	$banana$	S[6]
6	a$banan	S[5]
4	ana$ban	S[3]
2	anana$b	S[1]
1	banana$	S[7]
5	na$bana	S[4]
3	nana$ba	S[2]

createBWT(char *S[n];

int SA[n] = SA_linear(S); O(n)

O(n) for(int i=1 to n){

print(S[SA[i]-1]) if defined

else

print S[n];

}

S = cat$ (1 2 3 4)

SA = 4, 2, 1, 3

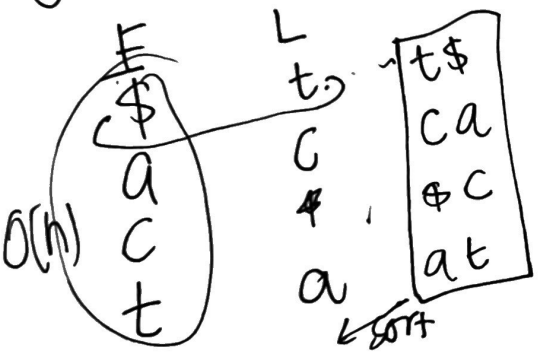
BWT = S[3], S[1], S[4], S[2]

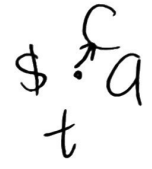
BWT = t c $ a

First LAST Algorithm

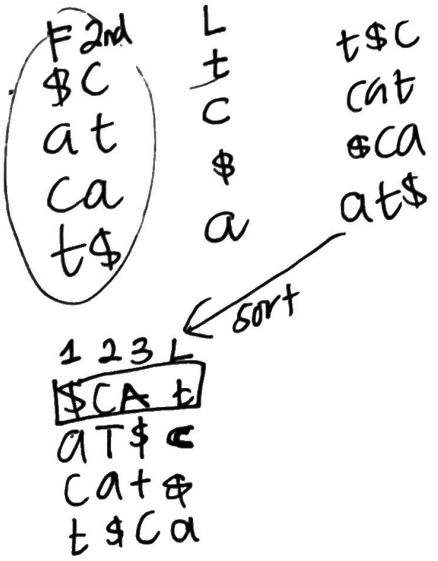
BWT(S) = S'

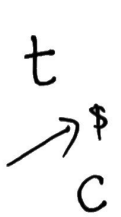
S' = t c $ a

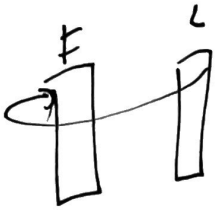




claims first row is a sort of the last row.

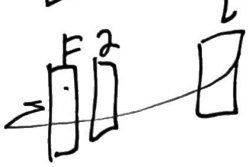






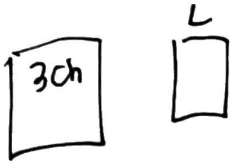
2 char ⇒ sort 2 char.

merge $n\log n \cdot O(2)$ $n \cdot 2$



3 char ⇒ sort 3

$n\log n \cdot O(3)$ $n \cdot 3$



4 char ⇒ sort 4

$n\log n \cdot 4$ $n \cdot 4$

⋮

$n-1$ sort $n-1$

$n\log n(n-1)$

$$\sum_{i=n-1}^{1} n\log n \cdot (n-i) \Rightarrow n\log n$$

$$= n^2 \log n$$

$$n\log n \sum_{i=2}^{n-1} i = n^2 \log n \cdot n^2$$

$$= n^3 \log n$$

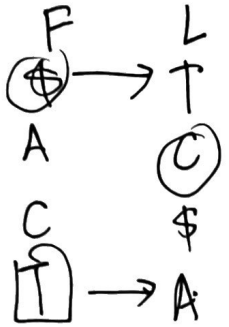
$$\sum_{i=2}^{n-1} n \cdot i = n \cdot n^2 = n^3$$

CAT$

BWT("CAT$")

$CAT
AT$C
CAT$
T$CA

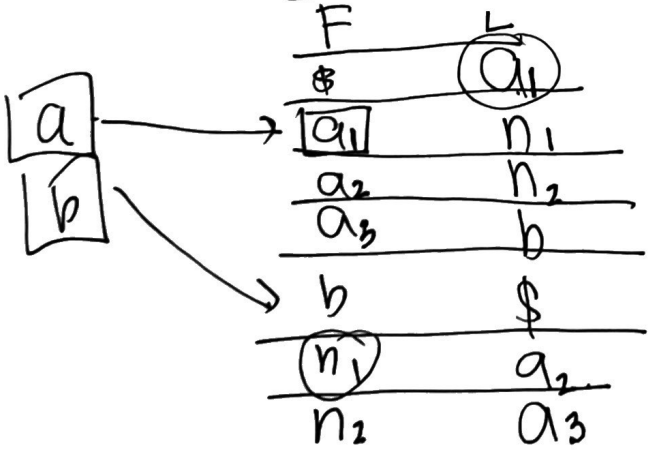
= TC$A



AT$

CAT$

BWT(T) = annb$aa



n1 a1 $
↑ ↗ b

~~$~~ b a3 n2 a2 n1 a1 $

BWT(S) = smnpbnnaaaaaa$a

smnpb N2 A5 $a

F	L
$	s
a_1	m
a_2	n_1
a_3	p
a_4 n_2	x b
a_5 n_3	y n_2
a_6	n_3
b	a_1
m	a_2
n_1	a_3
n_2	a_4
n_3	a_5
p	$
s	a_6

ana

pa

ma