

---

---

---

---

---



## Assignment 03

Q.03.1 (Shaique + Sandhya)

Given:  $T$  +  $S$ ,  $t$  strings with  $|S| \leq |t|$

we have suffix tree of  $S$  (shorter string)  
with suffix links.

we have Matching Statistics  $M$

(a)

To Prove:- computation of  $M$  in  $O(n)$

Proof:-

1. Initialise Empty suffix tree for string  $S$
2. Traverse the suffix tree of  $S$  and compute the following values for each node:-

- Depth: The depth of the node in the tree  
(The length of substring it represents)
- Start Position:- The position in  $S$  where the substring represented by node starts

3. Initialise the array  $M$  with zeroes.

4. Traverse string  $t$  with each position  $i$ .

- Use the suffix trees of  $S$  to find

LCP b/w the suffix of  $t$  starting at  $i$  and any suffix of  $S$ .

- Update  $M[i]$  with length of his longest common substring.

S. Return  $M$

Since each character in 'S' is scanned only once,

computation time  $\propto$  length of S

$\therefore$  Time Complexity  
is  $O(n)$

Example:-

finding M through slide Example-

T = TAATA\$ ; S = TTA\$

---

T	A	A	T	A	\$
	A	A	T	A	\$
		A	T	A	\$
			T	A	\$
				A	\$
					\$



(b)

The longest substring is given by me  
 $P[i:]$  if at the  $i^{\text{th}}$  position, we have  
a maxima. As shown in table below.

T =	T	A	A	T	A
P =	T	T	A		
	1	2	3		
	↑	↑			
	$i=1$	$i=2$			

$P[i:]$

∴ To make this table, Each character  
of T & P is read once.

we have time complexity =  $O(|S| + |T|)$

Q.03.3

(Shaique + Sandhya)

Given POS & LCP Array  $\rightarrow$  reconstruct  
Suffix + rep.

Step 1: Creating root node

Use a stack to track the  
distance b/w root node and another  
variable called active-node

Step 2 For  $i \in [1, \text{len}(\text{pos})]$

$\rightarrow \text{lcp\_value} = \text{lcp}[i]$

$\rightarrow \text{while } [\text{active\_node} - \text{root\_node}]$   
 $>$   
 $\text{lcp\_value}$

pop nodes from stack till

$$[\text{active\_node}] \leq \text{lep\_value}$$

Update active\_node  $\rightarrow$  top of stack:-

pos = [10, 9, 4, 7, 2, 5, 0, 8, 3, 6, 1]

lcp = [-1, 0, 1, 1, 3, 3, 5, 0, 2, 2, 4, -1]

$\rightarrow \text{pos}[0] = 10 \rightarrow \$$

$\rightarrow \text{pos}[1] = 9$  and  $\text{lcp} = 0$ , which means

no common character in blue

$\therefore$  we can add 'a' to the text

\_\_\_\_\_a\$



$$\rightarrow pos[2] = 4$$

$$lca[2] = 1$$

(deum = 1

-	-	-	-	<del>a</del>	<del>b</del>	a	b	<u>a</u>	<u>a</u>	<u>\$</u>
0	1	2	3	4	5	6	7	8	9	10

possible fix = sabaaababa\$

~~~~~