

Assignment 2.

ZHANG Chongzhi 1155077072.

1. (a). Suffix trie.

Location.

S = ^{1 2 3 4 5 6 7 8 9 0 1 2 3} G T A A C T G T A G T G \$ ✓

Suffix: TAACTGTA GTC ϕ ✓ 2

AACTGTA GTG \$ \checkmark \quad 3

ACTGTA GTG \$.\checkmark \quad 4

CTG TAG TG \$ \checkmark \quad 5

TG TAG TG \$ ✓ 6

G TAGTG \$ \checkmark \quad 7

TAGTG \$ \checkmark

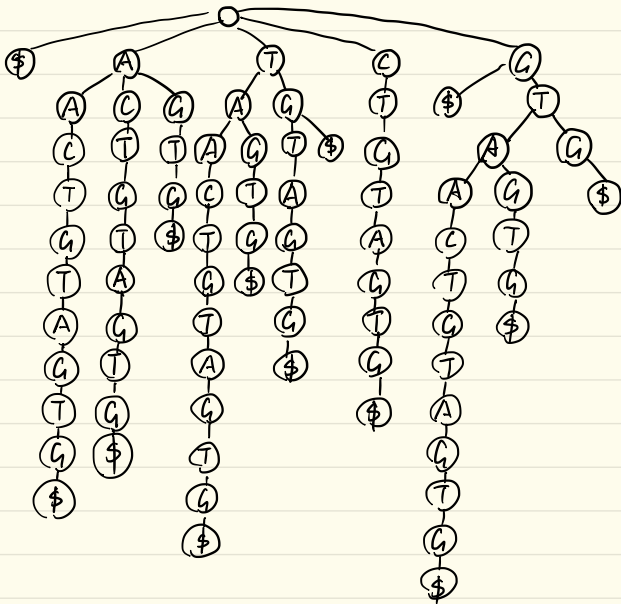
AGTG \$ ✓ 9

GTG \$ ✓ 10

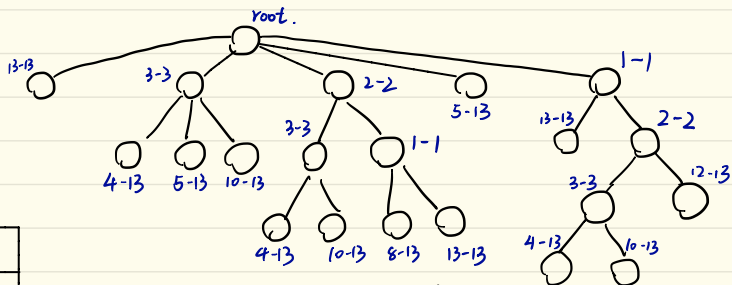
TG \$ \checkmark \quad \text{||}

G\$ ✓ 12

\$ ✓ 13.



(b). suffix tree.



(11). suffix array

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

Step: ①. Sort each level of the tree in lexicographic order.

②. Find the left most **Leaf** each time, calculate the **Location** of the suffix by using (the start location of the leaf) - (The depth of the leaf), and fill the **location** to the array one by one.

(d). ①. First gives all the suffixes. ② and then sort it in lexicographic order.

^{1 2 3 4 5 6 7 8 9 0 1 2 3}
 GTAAC TGTAGTG \$ 1
 TAACTGTAGTG \$ 2
 AACTGTAGTG \$ 3
 ACTGTAGTG \$ 4
 CTGTAGTG \$ 5
 TGTAGTG \$ 6
 GTAGTG \$ 7
 TAGTG \$ 8
 AGTG \$ 9
 GTG \$ 10
 TG \$ 11
 G \$ 12
 \$ 13.

\$
 AACTGTAGTG \$
 ACTGTAGTG \$
 AGTG \$
 CTGTAGTG \$
 G \$
 GTAACTGTAGTG \$
 GTAGTG \$
 GTG \$
 TAACTGTAGTG \$
 TAGTG \$
 TG \$
 TGTAGTG \$

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

(e). Because BTW is the last column of the sorted matrix.

Then each character of BTW is the one before [starting position of suffix array].

Thus for each position saved in Suffix array, Find that position in S and return the character before that position.

For instance, the first position is 13, and we put $13 - 1 = 12$ th character, which is G to the first position of b.

The following do the same procedure.

we have $b = \text{GTATAT\$TAGGCC}$

- (f). ①. Find all the rotation of s , and put them in a matrix.
 ②. Sort the rows of matrix by Lexicographic order.

ϕ GTAACTGATGTG
 AA CTGATGTG ϕ GT
 ACTGTAATG ϕ GTA
 AGTG ϕ GTAAC TGT
 CTGTAGTG ϕ GTA
 G ϕ GTAAC TGT
 GTAA CTGATGTG ϕ
 GTAGTG ϕ GTAAC T
 GTG ϕ GTAAC TGT
 TAA CTGATGTG ϕ G
 TAGTG ϕ GTAAC TGT
 TG ϕ GTAAC TGT
 GTAGTG ϕ GTAAC

- ③. Find the last column.
 we have $b = \text{GTATAT}\phi\text{TAGGGC}$

(g) f b

ϕ	G
A ₁	T ₁
A ₂	A ₁
A ₃	T ₂
C	A ₂
G ₁	T ₃
G ₂	ϕ
G ₃	T ₄
G ₄	A ₃
T ₁	G ₁
T ₂	G ₂
T ₃	G ₃
T ₄	C

- ①. Get the first column of Matrix by sorting b in Lexicographic order.

- ②. Mark the order of each character of both b and f .

In order to Find GTA,

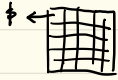
- ③. Find A₃ in f , before A, it is a T. Thus T₁ and T₂ has the chance.

- ④. Find T₁ and T₂ in f , before T, it is a G. Thus G₁ and G₂ meet the requirement.

- ⑤. Find the position of G₁ and G₂ in f , return the corresponding position in suffix array, which is 1 and 7.

(h). No.

Consider the first row of the sorted Matrix, the first character of first row must be a ϕ .



Thus, for all $n > 0$, the last character of first row have no chance to be a ϕ because there is only one ϕ in a row.

Since there is a position in b that ϕ can never appear, there's no doubt that the ϕ does not have a equal probability of all $n+1$ positions in b .

CSCI3220 2018-19 First Term Assignment 2

I declare that the assignment here submitted is original except for source material explicitly acknowledged, and that the same or closely related material has not been previously submitted for another course. I also acknowledge that I am aware of University policy and regulations on honesty in academic work, and of the disciplinary guidelines and procedures applicable to breaches of such policy and regulations, as contained in the following websites.

University Guideline on Academic Honesty: <http://www.cuhk.edu.hk/policy/academichonesty/>

Student Name: ZHANG Chongzhi

Student ID : 1155077072