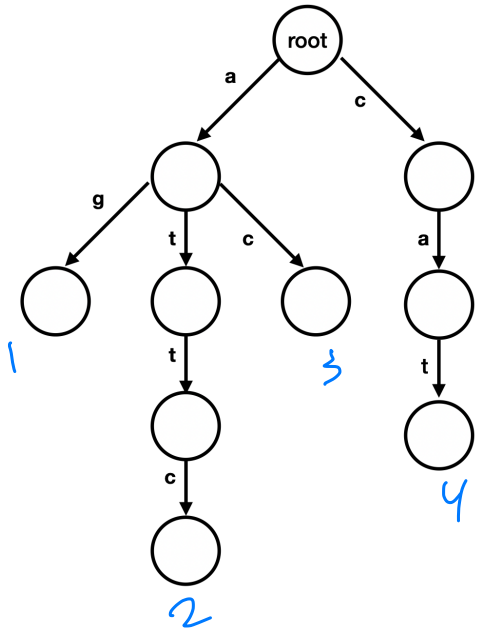
Problem 1

[10 points]

Below there is a Trie of some patterns.



(a) How many patterns are there in this Trie and what are they?

(b) Add the pattern *atcg*.

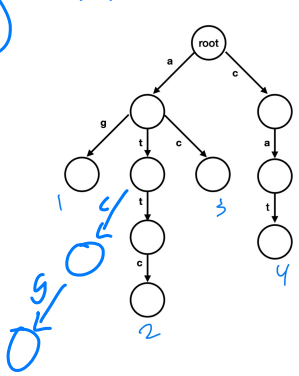
(c) Add the pattern *cag*.

(d) What happens if you try to add *att*? Can you think of anything to fix this?

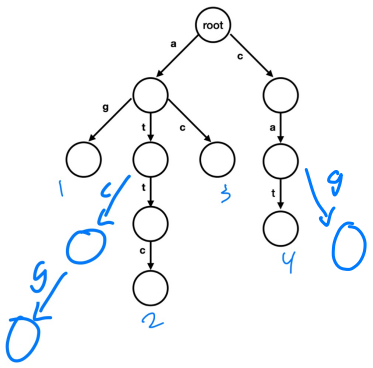
(A) ① AG
② ATTC
③ AC
④ CAT

4 Patterns

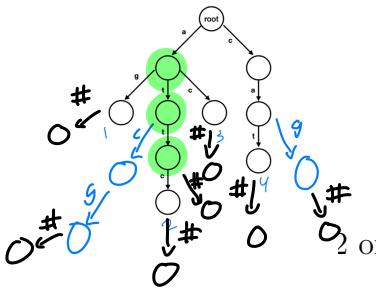
(B) ATCG



(C) CAG



(D)



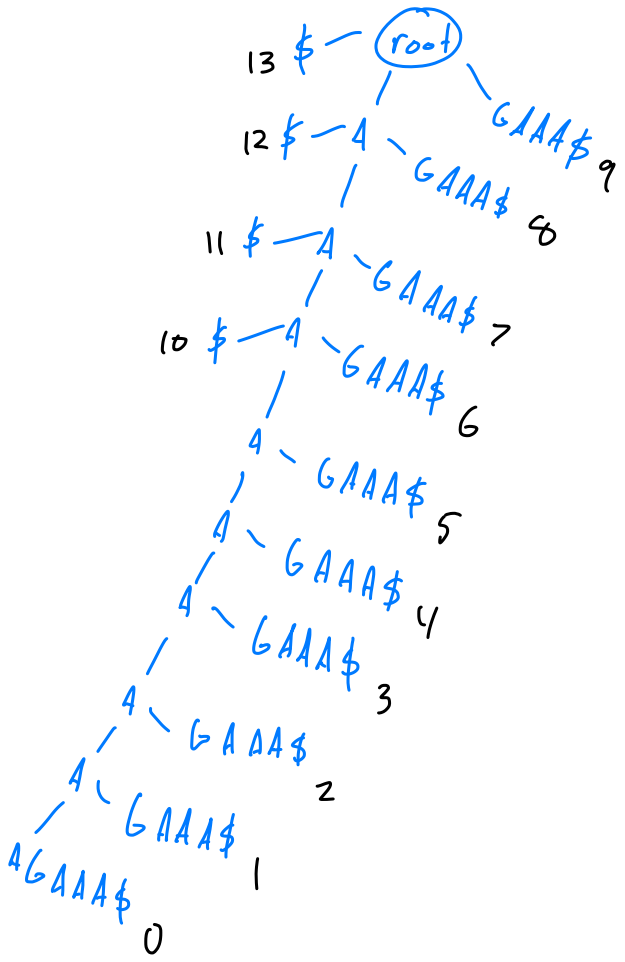
"ATT" is a substring of another pattern "ATTC". To solve this we could add End of pattern symbols such as # to our pattern Trie

Problem 2

[5 points]

How many leaves in SuffixTree(*AAAAAAAAAGAAA*)? Make sure to include the character '$' and this time we count it as its own suffix.

```
suffix array of AAAAAAAAAGAAA$
13 $
12 A$
11 AA$
10 AAA$
0 AAAAAAAAAGAAA$
1 AAAAAAAAGAAA$
2 AAAAAAAGAAA$
3 AAAAAAGAAA$
4 AAAAAGAAA$
5 AAAAGAAA$
6 AAAGAAA$
7 AAGAAA$
8 AGAAA$
9 GAAA$
```



14 leaves

Problem 3

[15 points]

Below is a Genome:

$$ABRACA\#AA\$$$

Yes, what the hell is '#' doing here? Well, why not? Let's adhere to some ordering that I just made up:

$$\$,\ A,\ B,\ C,\ \#,\ R$$

For the following problems, be sure to keep track of which occurrence the character is by using a subscript like we did in class.

(a) Compute the cyclic rotations of the Genome.

(b) Compute the $M()$ matrix.

(c) Write down the BWT.

A)

cyclic rotation matrix of: ABRACA#AA$

ABRACA#AA$	4
BRACA#AA$A	7
RACA#AA$AB	10
ACA#AA$ABR	5
CA#AA$ABRA	8
A#AA$ABRAC	6
#AA$ABRACA	9
AA$ABRACA#	3
A$ABRACA#A	2
$ABRACA#AA	1

B)

$	$ABRACA#AA	1
A $	A$ABRACA#A	2
A A	AA$ABRACA#	3
A B	ABRACA#AA$	4
A C	ACA#AA$ABR	5
A #	A#AA$ABRAC	6
B	BRACA#AA$A	7
C	CA#AA$ABRA	8
#	#AA$ABRACA	9
R	RACA#AA$AB	10

C) A A # $ R C A A A B

1 2 3 4 5 6 7 8 9 10

Problem 4

[15 points]

Alright, here is one of these BWT things:

$$\text{BWT(Text)} = \text{cfgggg\$catdctc}$$

Someone is kindly giving you the following table:

i	FirstColumn	LastColumn	LastToFirst(i)
0	$\$_1$	c_1	2
1	a_1	f_1	7
2	c_1	g_1	8
3	c_2	g_2	9
4	c_3	g_3	10
5	c_4	g_4	11
6	d_1	$\$_1$	0
7	f_1	c_2	3
8	g_1	a_1	1
9	g_2	t_1	12
10	g_3	d_1	6
11	g_4	c_3	4
12	t_1	t_2	13
13	t_2	c_4	5

Show how you reconstruct each letter using the First-Last property in each iteration as we learned in class (refer to Figure 9.12 in textbook). In other words, show the partial M(Text) matrix and the two arrows as in Figure 9.12 to indicate each letter you are reconstructing. To save space, you may reconstruct up to three letters for each partial M(Text) matrix. Be sure to label the edges according to the order of how they should be traversed.

sort BWT BWT

i	FirstColumn	LastColumn	LastToFirst(i)
0	$\$_1$	c_1	2
1	a_1	f_1	7
2	c_1	g_1	8
3	c_2	g_2	9
4	c_3	g_3	10
5	c_4	g_4	11
6	d_1	$\$_1$	0
7	f_1	c_2	3
8	g_1	a_1	1
9	g_2	t_1	12
10	g_3	d_1	6
11	g_4	c_3	4
12	t_1	t_2	13
13	t_2	c_4	5

$D_1 G_3 C_3 G_4 C_4 T_2 T_1 G_2 C_2 F_1 A_1 G_1 C_1 \$_1$