Exercise sheet 6: BLAST

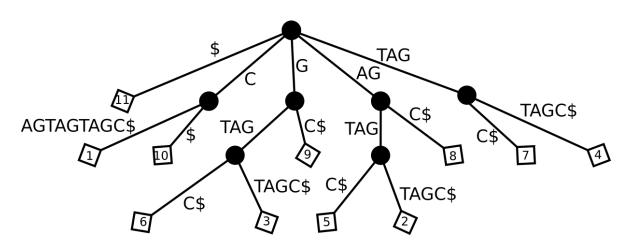## Exercise 1

You are given the text $T = CAGTAGTAGC$

Question 1A   Draw the corresponding suffix tree

Solution

CAGTAGTAGC

Question 1B   Describe the steps of a counting query for $P = TAG$

Solution

- start at root node
- locate outgoing edge that starts with T
- match subsequent characters of the pattern
- in the subtree rooted at $\overline{TAG}$ count the number of leaves $\Rightarrow 2$

Question 1C   Describe the steps of a reporting query for $P = AG$

Solution

- start at root node
- locate outgoing edge that start with A
- match subsequent characters of the pattern
- in the subtree rooted at $\overline{AG}$ report the labels of all leaves $\Rightarrow \{2, 5, 8\}$

Exercise 2

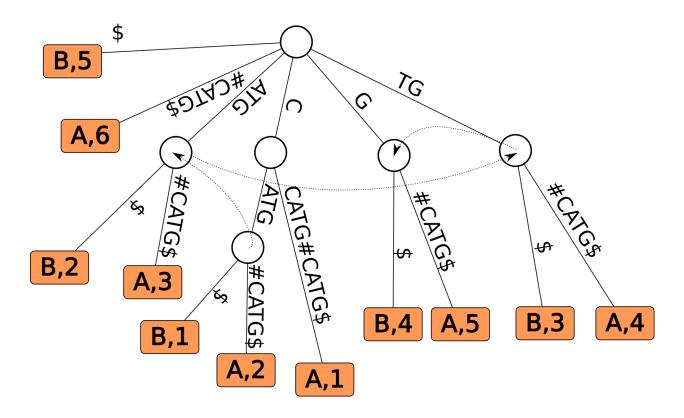**Question 2A**   Draw a generalized suffix tree for the sequences $A = CCATG$ and $B = CATG$.

Hint 1   Concatenate the two sequences using a unique character for splitting. e.g. $CCATG\#CATG\$$.
Dont forget to include suffix links

Formulae $sl(v) = w$

$v = \overline{cb}$

$w = \overline{b}$

$c : character, b : string$

remember: over lined strings are a representation for the node at that string

Solution

CCATG#CATG$

(Suffix tree with nodes: B,5 ($), A,6 (#CATG$), B,2 ($), A,3 (#CATG$), B,1 ($), A,2 (#CATG$), A,1 (CATG#CATG$), B,4 ($), A,5 (#CATG$), B,3 ($), A,4 (#CATG$), edges labeled C, ATG, #CATG$, G, TG)
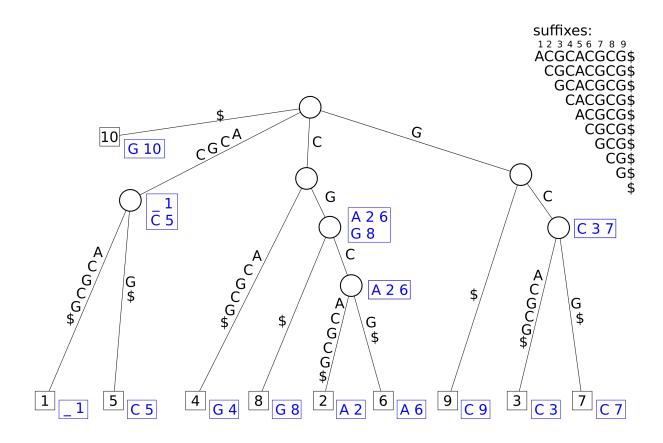
Question 2B Find the Maximal Unique Matches of the sequences $A = CCATG$ and $B = CATG$ using the tree from A)

Solution $CATG$ is the only MUM as $v = \overline{CATG}$ has no suffix links pointing to it

Exercise 3

Question 3A Draw a generalized suffix tree for the sequence $A = ACGCACGCG$.

Solution

suffixes:
```
 1 2 3 4 5 6 7 8 9
ACGCACGCG$
 CGCACGCG$
  GCACGCG$
   CACGCG$
    ACGCG$
     CGCG$
      GCG$
       CG$
        G$
         $
```

$
10   G 10

C G C A

C

G

10   C 5 1
     C̄

A C G C G $ G $

1   _ 1   5   C 5

G

A 2 6
G 8

A C G C G $ $ C

A 2 6

A C G C G $ G $

4   G 4   8   G 8   2   A 2   6   A 6

G

C

C 3 7

$

A C G C G $ G $

9   C 9   3   C 3   7   C 7

Question 3B   Find all maximal pairs of length at least 2

Solution $ACGC : (1, 5, 4)$

$CG : (2, 8, 2), (6, 8, 2)$

4