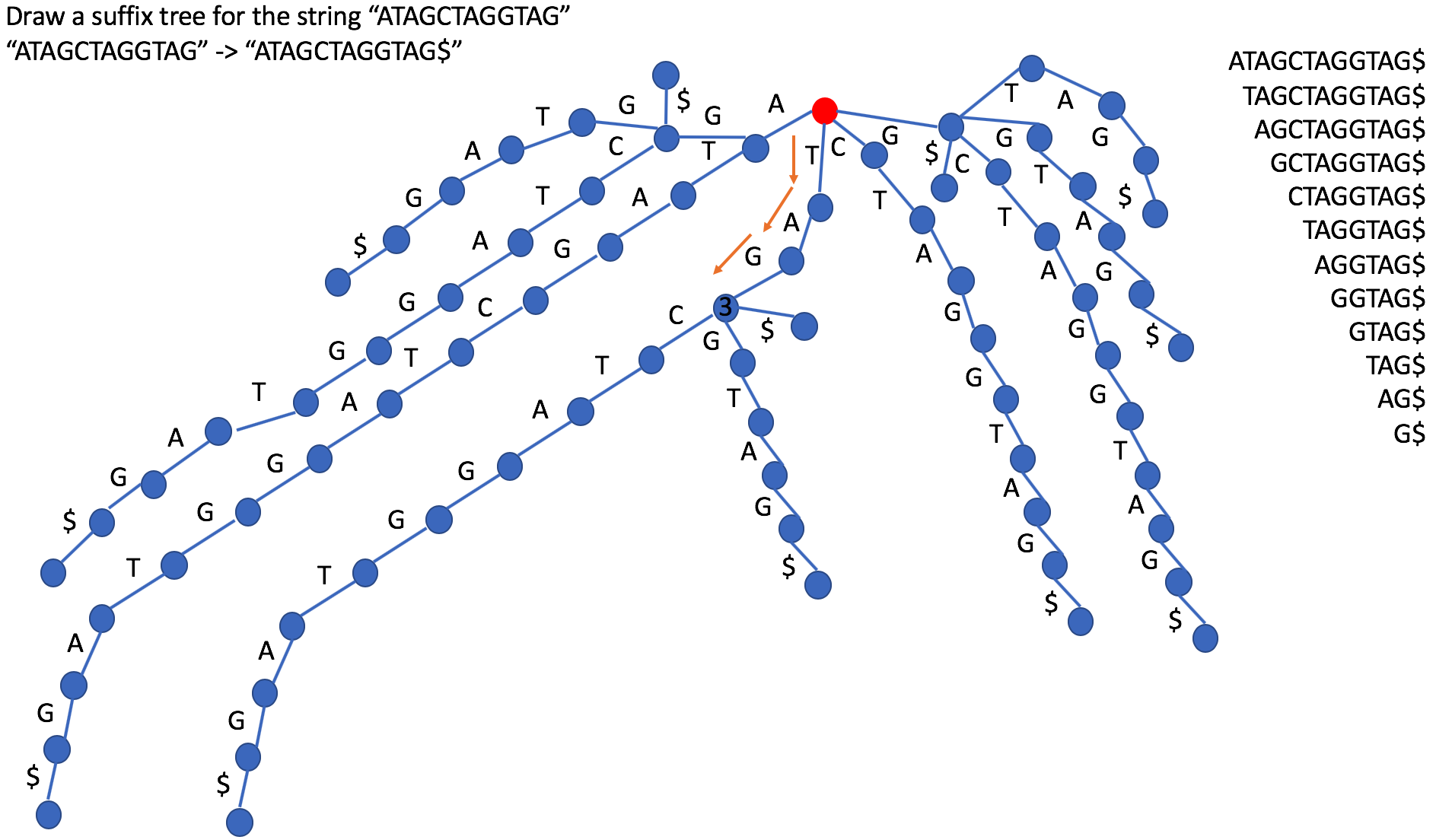
Homework 2 Answers:

Notes: there are many acceptable ways to accomplish the homework, these are just examples.

**Question1:**



1. See the figure above.
2. If you could find a path that goes from C -> T -> A starting from the root node, then CTA is a substring.
3. A suffix of a Trie by definition is a string of any length less than or equal to the longest pathway of the Trie, such that the last character of the suffix is the last character of the longest pathway in the Trie. If we could find a path that goes from G -> G -> T -> A and there is $ branching from A, then it is a suffix. The answer is No, thus GGTA is not a suffix.

**Question2:**

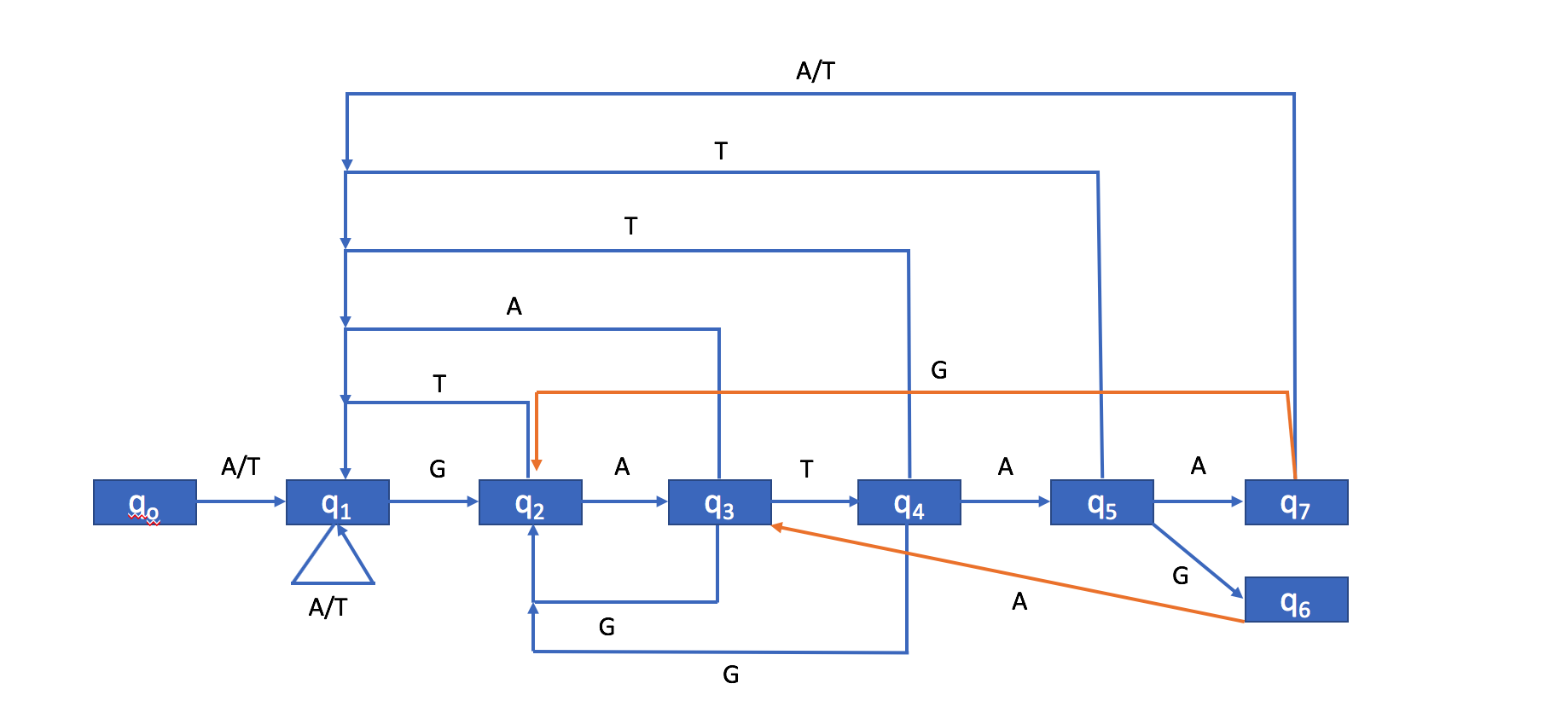
You can combine the suffix trees for both by adding a different symbol to the second string, for example “ATAGCTAG$” and “ACGCTAGG#”, and then draw the suffix tree for these two strings together. The longest common substring is the longest path that has both $ and # branching from it.

**Question3:**

A)

FSA Diagram:

If we consider that overlapping motifs are allowed, then the FSA diagram could be like this (orange lines are merely for contrast):



For test string

ATTTGATTTTTAGATAACACTAAAA, it will reach accepting state only once.

Transition Table:

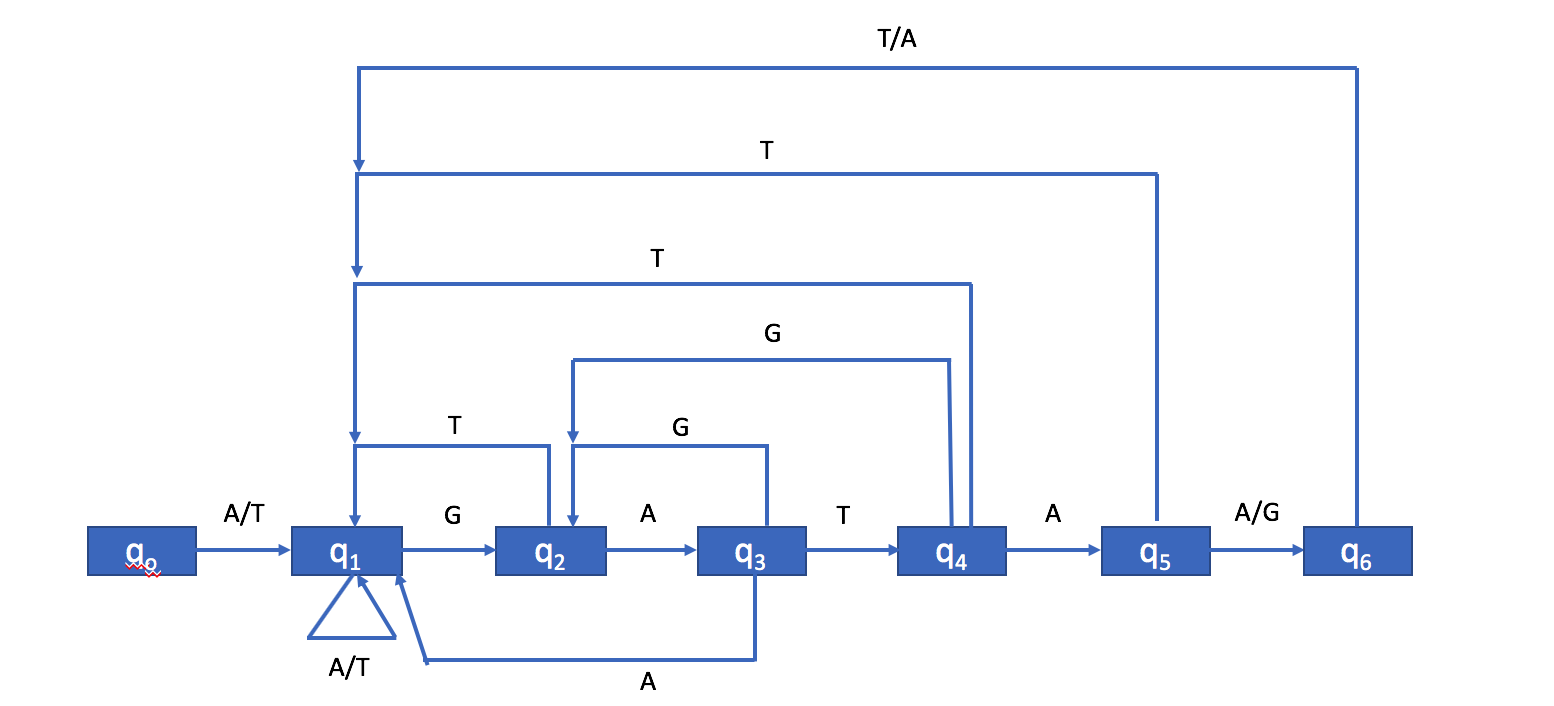
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | T | G | C |
| q0 | q1 | q1 | q0 | q0 |
| q1 | q1 | q1 | q2 | q0 |
| q2 | q3 | q1 | q0 | q0 |
| q3 | q1 | q4 | q2 | q0 |
| q4 | q5 | q1 | q2 | q0 |
| q5 | q7 | q1 | q6 | q0 |
| q6 | q3 | q1 | q0 | q0 |
| q7 | q1 | q1 | q2 | q0 |

Transition States:

From q0

|  |  |
| --- | --- |
| A | q1 |
| T | q1 |
| T | q1 |
| T | q1 |
| G | q2 |
| A | q3 |
| T | q4 |
| T | q1 |
| T | q1 |
| T | q1 |
| T | q1 |
| A | q1 |
| G | q2 |
| A | q3 |
| T | q4 |
| A | q5 |
| A | q7 |
| C | q0 |
| A | q1 |
| C | q0 |
| T | q1 |
| A | q1 |
| A | q1 |
| A | q1 |
| A | q1 |

If we consider that overlapping motif is not allowed, then the FSA diagram could be like this:



Transtion Table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A | C | G | T |
| q0 | q1 | q0 | q0 | q1 |
| q1 | q1 | q0 | q2 | q1 |
| q2 | q3 | q0 | q0 | q1 |
| q3 | q1 | q0 | q2 | q4 |
| q4 | q5 | q0 | q2 | q1 |
| q5 | q6 | q0 | q6 | q1 |
| q6 | q1 | q0 | q0 | q1 |

Transition States:

From q0

|  |  |
| --- | --- |
| A | q1 |
| T | q1 |
| T | q1 |
| T | q1 |
| G | q2 |
| A | q3 |
| T | q4 |
| T | q1 |
| T | q1 |
| T | q1 |
| T | q1 |
| A | q1 |
| G | q2 |
| A | q3 |
| T | q4 |
| A | q5 |
| A | q6 |
| C | q0 |
| A | q1 |
| C | q0 |
| T | q1 |
| A | q1 |
| A | q1 |
| A | q1 |
| A | q1 |

B)

Please see the hw2 question 3b.py file

C)

Please see the hw2 question 3c.py file