

Data Structures - Assignment no. 9

Remarks:

- Write both your name and your ID number very clearly on the top of the exercise. Write your exercises in pen, or in clearly visible pencil. Please write *very* clearly.
 - Give correctness and complexity proofs for every algorithm you write.
 - For every question where you are required to write pseudo-code, also explain your solution in words.
1. With path compression and union by rank, during the lifetime of a Union-Find data-structure, how many elements would have rank at most $\lfloor \log n - 5 \rfloor$, where there are n elements stored in the data-structure?
 2. Prove that in a set of n elements, a sequence of n consecutive FIND operations take $O(n)$ time in total.
 3. How would you use the edge-label compressed suffix tree given in Fig. 1 to find all occurrences of CGT in the search string? (Assume that you know which string gave rise to the tree.)
 4. Build the suffix tree that encodes the two strings CAGAGT and GAGAG. Show how to use your string to find the longest common substring to those two strings.
 5. In Fig. 1 there is a suffix tree for some search string with details graciously omitted. Suppose that each edge has a string of length exactly 2, except for the edges at the leaves, which are of varying lengths. What is the length of the longest string that occurs at least 8 times in the search string? How many such strings are there?
 6. In Fig. 2 there are four suffix trees, for four different strings, each of which begins with A. What are the four strings?

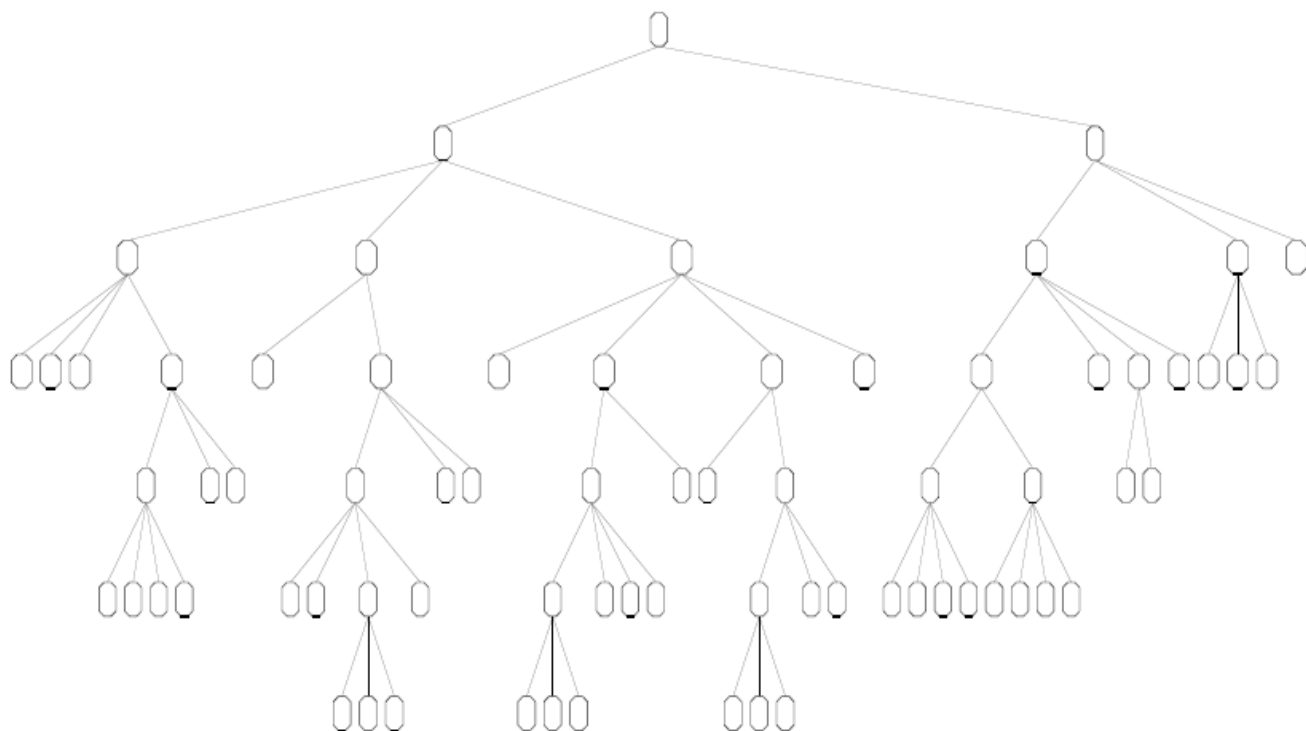


Figure 1:

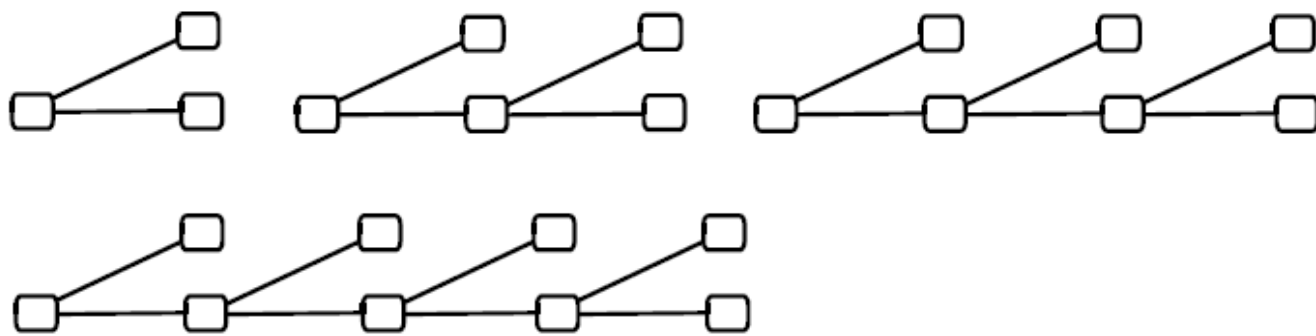


Figure 2: