

Advanced Sequence Analysis  
Summer 2019

Exercises

Number 5 (17.05.2019), Discussion: 24.05.2019

1. Which data structures could one use to store the branching vertices in a suffix tree?  
What is the resulting memory usage per branching vertex and what is the search time for a word of length  $m$ ?
2. Given the suffix tree  $T(t)$ , explain how all positions of the  $k$  occurrences of a substring  $s$  of  $t$  can be enumerated in  $O(|s| \cdot \log |\mathcal{A}| + k)$  time.
3. Given the suffix tree  $T(t)$ , explain how the number of different substrings of  $t$  can be computed in  $O(|t|)$  time.  
Explain why there is no factor  $\log |\mathcal{A}|$  in the asymptotic complexity, other than in Exercise 2 above.
4. The relationship between the suffix tree for a string  $t$  and for the reverse string  $t^{-1}$  is not obvious. However, there is a significant relationship between the two trees. Find it, state it, and prove it.  
*Hint:* Suffix links help.
5. Can Ukkonen's algorithm be implemented in linear time without using suffix links?