

**Advanced Sequence Analysis
Summer 2019**

Exercises

Number 10 (21.06.2019), Discussion: 28.06.2019

1. Construct the suffix tree of the string ABAAABBABBBBA\$ using the algorithm of Farach (1997). You may omit the details of the merging phases.
2. Describe a strategy, different from the one by Farach (1997) presented in class, to construct a suffix tree in linear time, independent of the alphabet size (for integer alphabets).
Hint: First construct a suffix array with LCP values, and then create the suffix tree from these in linear time. You may want to read the notes or watch the video by Eric Demaine that can be found here: <https://courses.csail.mit.edu/6.851/fall117/lectures/L16.html>.
3. Execute the search for the pattern BARB in the trivial NFA of the text RHABARBERBARBARA and in the corresponding BWT.
4. For a family of k sequences (S_1, \dots, S_k) of the same length n , the i th *positional Burrows Wheeler Transform* (pBWT) a_i ($1 \leq i \leq n$) is defined as follows: The array a_i is a permutation of the numbers $\{1, \dots, k\}$ with $S_{a_i[1]}[1..i] \leq \dots \leq S_{a_i[k]}[1..i]$ colexicographically, i.e. right-to-left lexicographically.
Give an algorithm to construct all pBWT arrays for increasing values of i in overall linear time $O(kn)$.