

# A6-SequenceBio

Daniel Giesel, Barbara Mühlbauer

November 2022

## Excercise 1

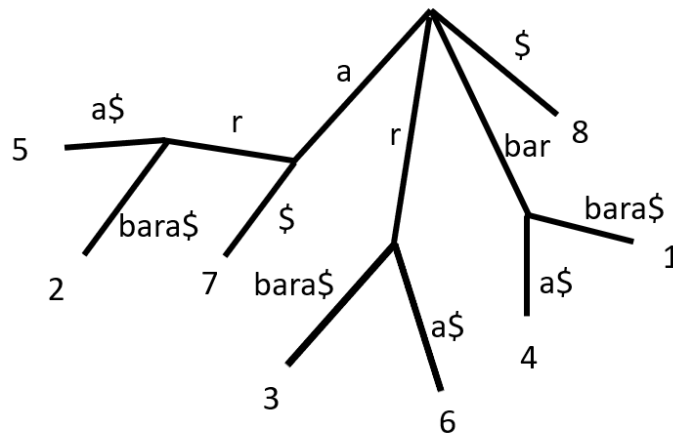
Right-minimal unique substrings are all substrings which are unique with non-unique proper prefixes. One method to find these substrings, is via a suffix tree.

In suffix trees, we know that the path to each leaf will result in a unique substring, as well as each internal node (i.e prefix) being non-unique as it appears in at least two substrings. This means that the label of any node with two children at least one of which is a leaf and the first letter of this leaf is a unique right-minimal substring, because while the prefix (label) occurs at least twice the prefix with the first letter is unique occuring only in the child. It is also right-minimal containing only the first letter of the leaf label, which is strictly necessary to make it unique. The condition of  $|u| \geq L$  can be fulfilled by either filtering out any found substrings with length lower than the desired length, or during the traversal of the tree checking whether the substring has reached a desired length before arriving at a leaf.

Using the tree below one gets the following right-minimal unique substrings:

barb,bara,ara,arb,rb,ra (a\$)

a\$ only appears in the tree and not in the actual sequence.



## Exercise 2

Implemented in the file `FindQueries<Names>.java`

## Exercise 3

Implemented in the file `FindQueries<Names>.java`

## Exercise 4

Implemented in the file `MUMS<Names>.java`