UNIVERSITY OF LIVERPOOL

Department of Computer Science
Sebastian Wild

Date: 2020-03-18
Version: 2020-03-09 23:16

# Tutorial 6 for COMP 526 – Applied Algorithmics, Winter 2020

**—including solutions—**

**It is highly recommended that you first try to solve the problems on your own before consulting the sample solutions provided below.**

## Problem 1 (Suffix trees and friends)

Consider the text $T =$ `abbabbaa$`.

What is $n$ here? (exactly follow the convention from the lecture!)

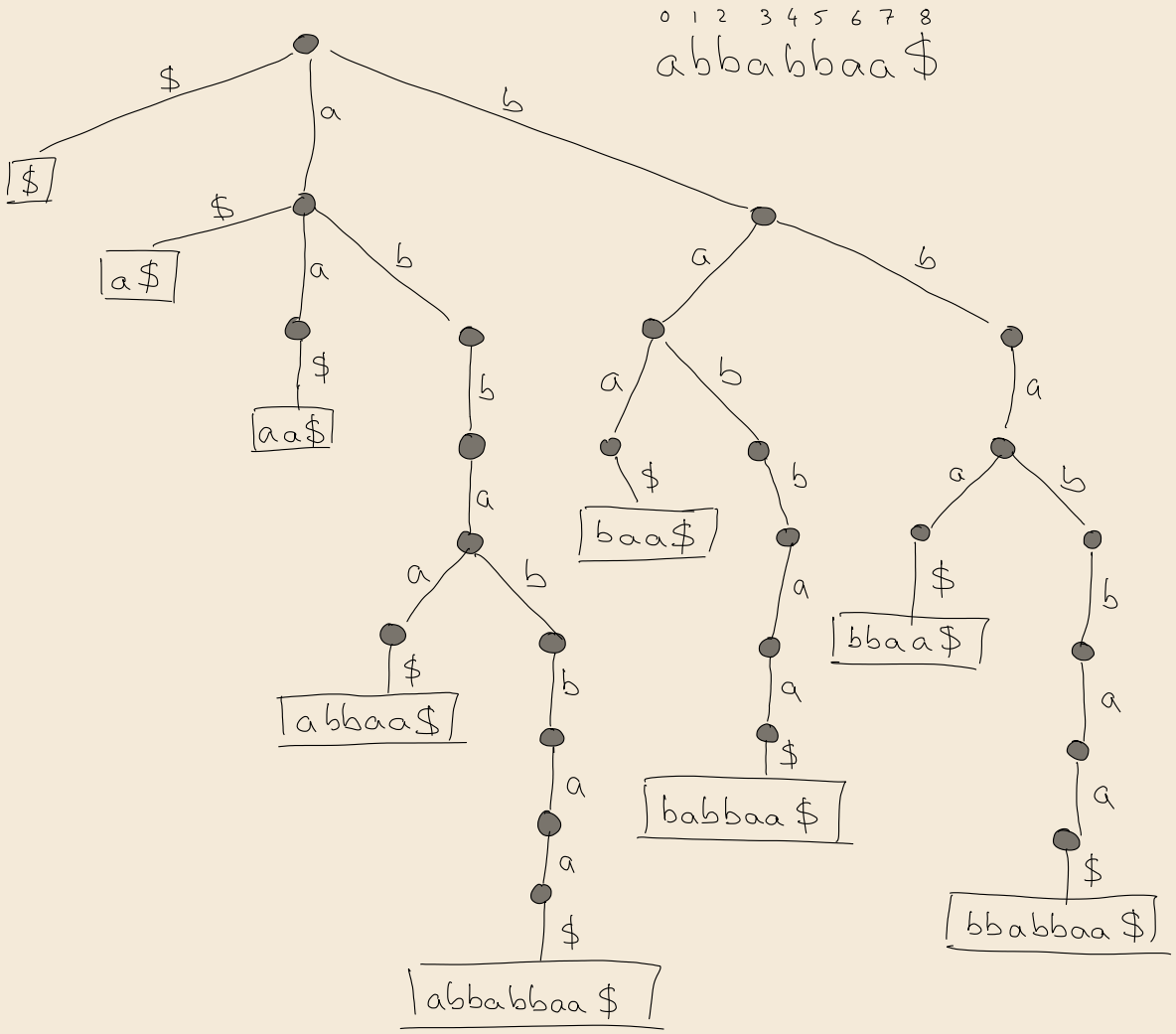
Construct/Draw the

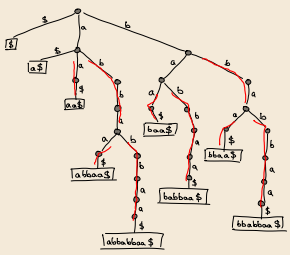
1. standard (not compacted) trie of all suffixes of $T$,
2. suffix tree of $T$ (human version) with string labels on edges and leaves,
3. suffix tree of $T$ (computer version) as it is stored, i.e., offsets in nodes, starting index in leaves, first characters on edges,
4. suffix array $L[0..n]$ of $T$,
5. the inverse suffix array $R[0..n]$, and
6. the LCP array.
7. Annotate the internal nodes in the suffix tree with their string depth. Explain the connection between string depths and the LCP array.
8. Use the above structures to find the longest repeated substring in $T$.

## Solutions for Problem 1 (Suffix trees and friends)

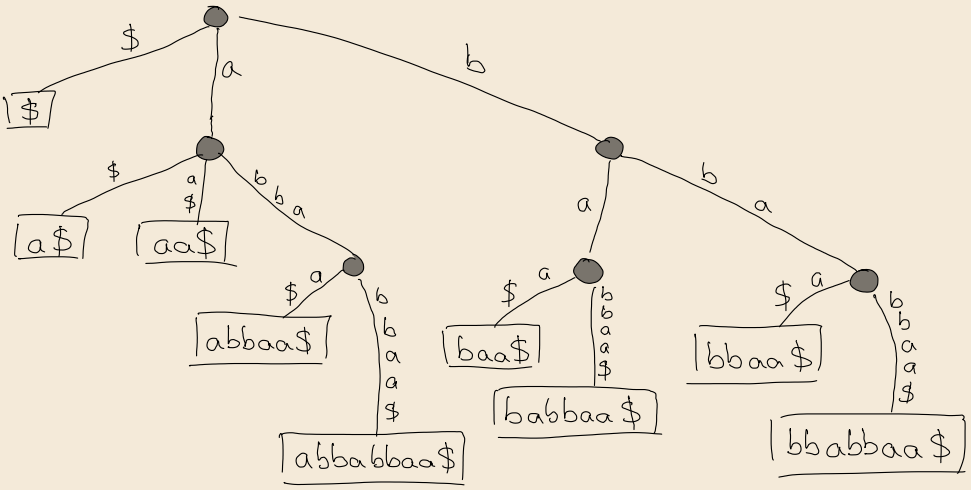
$n$ is the number of actual characters, not counting the end-of-word marker `$`, so $n = 8$.

1.



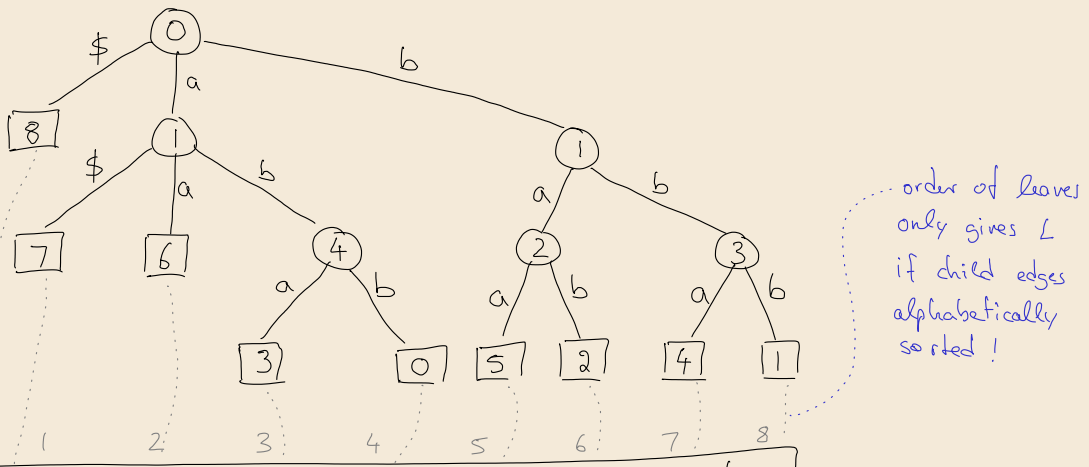



2.

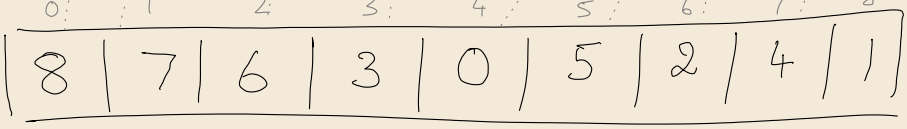



3.

0 1 2 3 4 5 6 7 8
abbabbaa$




4. L



**5.** R

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 4 | 8 | 6 | 3 | 7 | 5 | 2 | 1 | 0 |

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| a | b | b | a | b | b | a | a | $ |

**6.** L

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 8 | 7 | 6 | 3 | 0 | 5 | 2 | 4 | 1 |

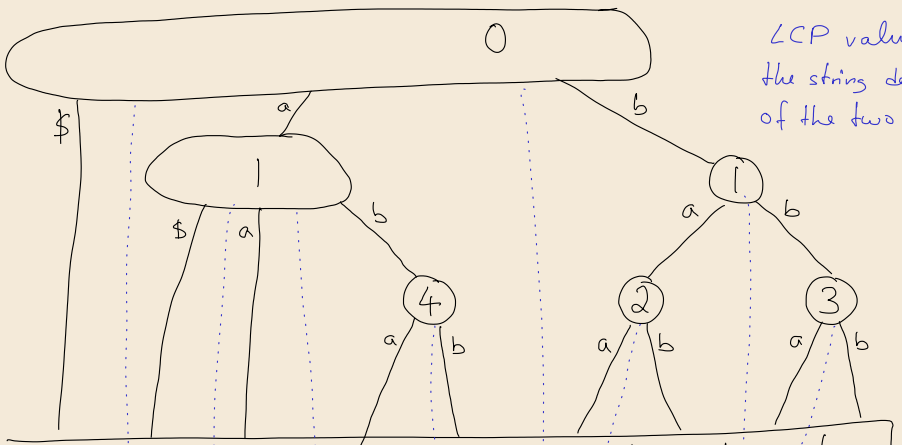
LCP

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 0 | 2 | 1 | 3 |

| $ | a | a | a | a | b | b | b | b |
|---|---|---|---|---|---|---|---|---|
| | $ | a | b | b | a | a | b | b |
| | | $ | b | b | a | b | a | a |
| | | | a | a | $ | b | a | b |
| | | | a | b | | a | $ | b |
| | | | $ | b | | a | | a |
| | | | | a | | $ | | a |
| | | | | a | | | | $ |
| | | | | $ | | | | |

**7.**

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| a | b | b | a | b | b | a | a | $ |




LCP values are exactly the string depths of the LCA of the two adjacent leaves

L

| 8 | 7 | 6 | 3 | 0 | 5 | 2 | 4 | 1 |
|---|---|---|---|---|---|---|---|---|

LCP

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 0 | 2 | 1 | 3 |

**8.** The largest LCP value is LCP[4] = 4.

This corresponds to the LCP of leaves [3] and [0],

so the longest repeated string is their first 4 characters: abba