# Homework 3 solution

## 1

### a

Apply the Käräkkanen and Sanders' Algorithm:
$T =$ y a b b a d a b b a d p \$ . Augment $T$ as $T =$ y a b b a d a b b a d p \$ \$ \$ . Now let $s_i$ represent the three-alphabet substring starting at $i$ i.e $s_i = T[i : i+2]$.
$s_1 =$ y a b
$s_2 =$ a b b
.
.
$s_{12} =$ p \$ \$
$s_{13} =$ \$ \$ \$

$$T = \underset{1}{y}\ \underset{2}{a}\ \underset{3}{b}\ \underset{4}{b}\ \underset{5}{a}\ \underset{6}{d}\ \underset{7}{a}\ \underset{8}{b}\ \underset{9}{b}\ \underset{10}{a}\ \underset{11}{d}\ \underset{12}{p}\ \underset{13}{\$}\ \$\ \$$$

Step 1: Sort $S := (s_1, s_2, s_4, s_5, s_7, s_8, s_{10}, s_{11}, s_{13}) = (yab, abb, bad, ada, abb, bba, adp, dp\$, \$\$\$)$ using radix sort in $O(n)$ time.
The sorted sequence is: $(\$\$\$, abb, abb, ada, adp, bad, bba, dp\$, yab)$.
Excluding multiplicity the sorted sequence is $U = (\$\$\$, abb, ada, adp, bad, bba, dp\$, yab)$.
Compute ranks of $S = (s_1, s_2, s_4, s_5, s_7, s_8, s_{10}, s_{11}, s_{13})$ in $U$.

$$R := rank(s_1, s_2, s_4, s_5, s_7, s_8, s_{10}, s_{11}, s_{13}) = (8, 2, 5, 3, 2, 6, 4, 7, 1)$$

Step 2: Define $S' = (s_1, s_4, s_7, s_{10}, s_{13}, s_2, s_5, s_8, s_{11})$. Let $R'$ represent the ranks of $S'$ from Step 1.
Then $R' := rank(S') = (8, 5, 2, 4, 1, 2, 3, 6, 7)$.
Find suffix array of $R'$ (appended with \$) using recursion.
$SA(R') = [5, 6, 3, 7, 4, 2, 8, 9, 1]$.
$S'[SA(R')] = (s_{13}, s_2, s_7, s_5, s_{10}, s_4, s_8, s_{11}, s_1)$.
This implies that the suffix array of $T$ restricted to indices $1, 2 \pmod 3$ is

$$SA_{1,2} = [13, 2, 7, 5, 10, 4, 8, 11, 1]$$

Step 3 Define $t_i := T[i]SA^{-1}_{1,2}(i+1)$ , for $i \equiv 0 \pmod 3$, i.e. $t_3 = b6$, $t_6 = d3$, $t_9 = b5$, $t_{12} = p1$
`Note,` $SA^{-1}_{1,2}$ `is the inverse function of` $SA_{1,2}$`, i.e.` $SA^{-1}_{1,2}(x)$ `gives the index of value` $x$ `in` $SA_{1,2}$ `array.` $SA^{-1}_{1,2}$ `can be easily computed in O(n) time as an array that can be looked up in O(1) time.`
Find the suffix array of $T$ restricted to indices $0 \pmod 3$ as follows: Radix sort $(t_3, t_6, t_9, t_{12})$ in $O(n)$ time: $sort(t_3, t_6, t_9, t_{12}) = (t_9, t_3, t_6, t_{12}) \implies$

$$SA_0 = [9, 3, 6, 12]$$

Step 4 Merge $SA_{1,2}$ and $SA_0$ using *mergesort* type merging, where a comparison between $i' := SA_{1,2}[i]$ and $j' := SA_0[j]$ takes constant time by lexicographic comparison of

1. $T[i']SA_{1,2}^{-1}(i'+1)$ and $T[j']SA_{1,2}^{-1}(j'+1)$, if $i' \equiv 1 \pmod 3$
2. $T[i']T[i'+1]SA_{1,2}^{-1}(i'+2)$ and $T[j']T[j'+1]SA_{1,2}^{-1}(j'+2)$, if $i' \equiv 2 \pmod 3$

Recall:

$$T = \underset{1}{y}\ \underset{2}{a}\ \underset{3}{b}\ \underset{4}{b}\ \underset{5}{a}\ \underset{6}{d}\ \underset{7}{a}\ \underset{8}{b}\ \underset{9}{b}\ \underset{10}{a}\ \underset{11}{d}\ \underset{12}{p}\ \underset{13}{\$}\ \$\ \$$$

$$SA_{1,2} = [13, 2, 7, 5, 10, 4, 8, 11, 1]$$

$$SA_0 = [9, 3, 6, 12]$$

Using the above technique our merging iterations are as follows: (we can skip $i = 1$ as we know that $SA_{1,2}[1]$ correspond to the trivial suffix). ("$\sim$" represents compare symbol)

$SA_{1,2}[2] \sim SA_0[1] \iff ab6 \sim ba8 \implies SA_{1,2}[2] < SA_0[1]$, $SA = [13, 2, \_]$
$SA_{1,2}[3] \sim SA_0[1] \iff a7 \sim b5 \implies SA_{1,2}[3] < SA_0[1]$, $SA = [13, 2, 7, \_]$
$SA_{1,2}[4] \sim SA_0[1] \iff ad3 \sim ba8 \implies SA_{1,2}[4] < SA_0[1]$, $SA = [13, 2, 7, 5, \_]$
$SA_{1,2}[5] \sim SA_0[1] \iff a8 \sim b5 \implies SA_{1,2}[5] < SA_0[1]$, $SA = [13, 2, 7, 5, 10, \_]$
$SA_{1,2}[6] \sim SA_0[1] \iff b4 \sim b5 \implies SA_{1,2}[6] < SA_0[1]$, $SA = [13, 2, 7, 5, 10, 4, \_]$
$SA_{1,2}[7] \sim SA_0[1] \iff bb5 \sim ba8 \implies SA_{1,2}[7] > SA_0[1]$, $SA = [13, 2, 7, 5, 10, 4, 9, \_]$
$SA_{1,2}[7] \sim SA_0[2] \iff bb5 \sim bb4 \implies SA_{1,2}[7] > SA_0[2]$, $SA = [13, 2, 7, 5, 10, 4, 9, 3, \_]$
$SA_{1,2}[7] \sim SA_0[3] \iff bb5 \sim da7 \implies SA_{1,2}[7] < SA_0[3]$, $SA = [13, 2, 7, 5, 10, 4, 9, 3, 8, \_]$
$SA_{1,2}[8] \sim SA_0[3] \iff dp1 \sim da7 \implies SA_{1,2}[8] > SA_0[3]$, $SA = [13, 2, 7, 5, 10, 4, 9, 3, 8, 6, \_]$
$SA_{1,2}[8] \sim SA_0[4] \iff dp1 \sim p\$0 \implies SA_{1,2}[8] < SA_0[4]$, $SA = [13, 2, 7, 5, 10, 4, 9, 3, 8, 6, 11, \_]$
$SA_{1,2}[9] \sim SA_0[4] \iff y2 \sim p1 \implies SA_{1,2}[9] > SA_0[4]$, $SA = [13, 2, 7, 5, 10, 4, 9, 3, 8, 6, 11, 12, 1]$

Therefore, $SA(T) = [13, 2, 7, 5, 10, 4, 9, 3, 8, 6, 11, 12, 1] \equiv [2, 7, 5, 10, 4, 9, 3, 8, 6, 11, 12, 1]$, omitting the trivial suffix.

## b

$SA = [2, 7, 5, 10, 4, 9, 3, 8, 6, 11, 12, 1]$. Binary search: start with index $i = |T|/2 = 6$. ("$\sim$" represents compare symbol)

1. Compare suffix $T_{SA[6]} \sim P \iff bad \sim abb \implies T_{SA[6]} > P$. Set $i \leftarrow i/2 = 3$
2. Compare suffix $T_{SA[3]} \sim P \iff ada \sim abb \implies T_{SA[3]} > P$. Set $i \leftarrow \lceil i/2 \rceil = 2$
3. Compare suffix $T_{SA[2]} \sim P \iff abb \sim abb \implies T_{SA[2]} = P$. Found $P$ at $SA[2] = 7$. Set $k \leftarrow i$. Set $i \leftarrow i/2 = 1$.
4. Compare suffix $T_{SA[1]} \sim P \iff abb \sim abb \implies T_{SA[1]} = P$. Found $P$ at $SA[1] = 2$. Since $SA[k+1]$ i.e. $SA[3]$ has already been checked, so we stop here (otherwise we had to continue on the right of $k$ as well to possibly find more instances of $P$).