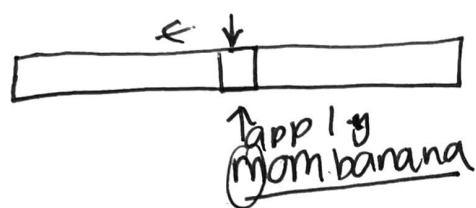


Lecture 9

FINISH linear calculation
of the suffix array.

$S = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \\ D & A & D & B & C & D & D & A & D & B & C & D & \$ & \$ & \$ \end{matrix}$

given suffix array searching pattern p in T ?



$p = \text{apple}$

run binary search on SA looking for p in T .

$$\log n \Rightarrow p \times \log |T|$$

SA creation in linear time single search

$$\underbrace{O(T)}_{\text{SA creation}} + \underbrace{p \log T}_{\text{time for search}}$$

suffix tree search time: $\underbrace{O(p)}_{\text{size of pattern}}$

SA(strings)

3-mer

indexes $i \% 3 = \{1, 2, 3\}$

RANK

①

Suff₁ = DAD

Suff₂ = ADB

Suff₄ = BCD

Suff₅ = CPD

Suff₇ = DAD

Suff₈ = ADB

Suff₁₀ = BCD

Suff₁₁ = CD\$

Suff₁₃ = \$\$\$

$3n$
 \Rightarrow
sort
via
radix
sort

13 \$\$\$ \rightarrow 1

2 ADB \rightarrow 2

8 ADB

4 BCD \rightarrow 3

10 BCD \rightarrow 1

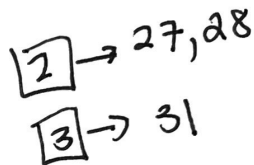
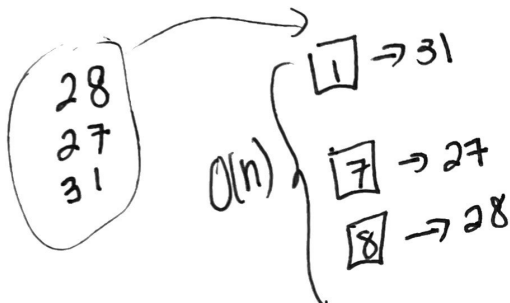
11 CD\$ \rightarrow 4

5 CDD \rightarrow 5

1 DAD \rightarrow 6

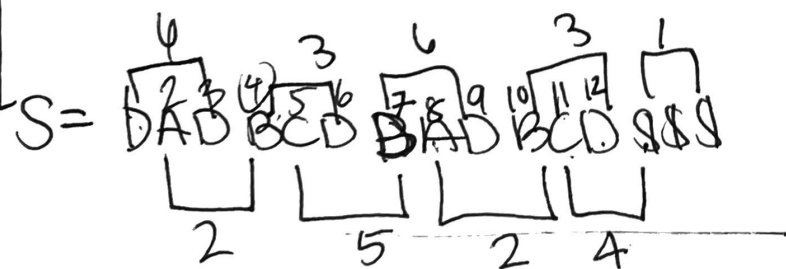
7 DAD \rightarrow 6

linear
work



m digits / buckets $n = \# \text{ of items}$

RADIX $m \cdot n$



DADBCDDADBCDDADBCDDADBCDDADBCDDAD

$S' = (6363) \mid 2524$

$S' = (6363) \mid 2524$
 $\underbrace{\hspace{1cm}}_{\frac{n}{3}} \quad \underbrace{\hspace{1cm}}_{\frac{n}{3}}$

② SA(8') ; $T(\frac{2n}{3})$

$S' \Rightarrow \textcircled{1} \textcircled{2} \textcircled{3} \textcircled{4} \textcircled{5} \textcircled{6} \textcircled{7} \textcircled{8} \textcircled{9}$
 $S' = 636312524$
 $\uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow \uparrow$
 $1 \ 4 \ 7 \ 10 \ 13 \ 2 \ 5 \ 8 \ 11$

SA: $12524 \Rightarrow \frac{S'}{5}$
 $24 \Rightarrow 8$
 $2524 \Rightarrow 6$
 $312524 \Rightarrow 4$
 $36312924 \Rightarrow 2$
 $4 \Rightarrow 9$
 $524 \Rightarrow 7$
 $6312524 \Rightarrow 3$
 $636312524 \Rightarrow 1$

original	RANK of suffix of all suffix $i \% 3 = \{1, 2, 3\}$
5	1
13	2
8	3
2	4
10	5
4	6
11	7
5	8
7	9
1	

$\frac{n}{3}$ $\left[\begin{array}{l} \text{Suff}_3 = \text{BCD DAD BCD} \$ \$ \$ \\ \text{Suff}_6 = \text{D DAD BCD} \$ \$ \$ \\ \text{Suff}_9 = \text{D BCD} \$ \$ \$ \\ \text{Suff}_{12} = \text{D} \$ \$ \$ \end{array} \right.$

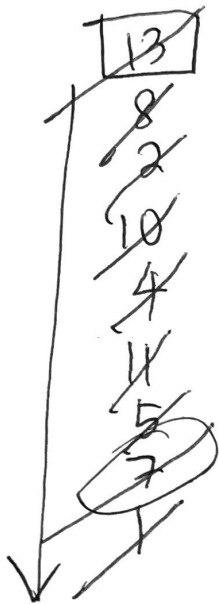
$\boxed{3} \ S[3].\text{suff}_4 \ D5$
 $\boxed{6} \ S[6].\text{suff}_7 \ D8$
 $9 \ S[9].\text{suff}_{10} \ D4$
 $13 \ S[12].\text{suff}_{13} \ D1$

RADIX \Rightarrow

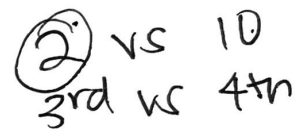
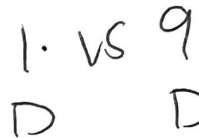
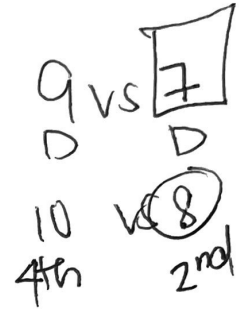
for ($i=1$ to n)
 $i = i \cdot 3$
 for ($j=1$ to i)
 print hello
 $O(n)$
 $O(n)$
 2
 3
 suffix array
 $\text{suff}_{i \% 3 = 0}$

13	D1
9	D4
3	D5
6	D8

SA₁₂



SA₀



RANK

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
D A D M C D D A D M C D S S S
9 3 5 7 8 2 4 6 1

S[7] suffix 8

S[12] suffix 13

SA[] = 13, 8, 2, 10, 4, 11, 5,
12, 7, 1, 9, 3, 6

comparing 2 suffix ^{i j} and

if (S[i] == S[j])

print suffix i

or suffix j based

on the rank of

suffix i+1 vs suffix j+1

i%3=1 j%3=0

i+1%3=2 j+1%3=1

in same list
O(1) rank comp.

int[] SA(char S[]) { }

// STEP 1

- $\frac{2n}{3} \rightarrow$ (a) create $\frac{2n}{3}$ kmers for suff $i \% 3 = \{1, 2\}$
 (b) sort kmers and rank via RADIX.
 (c) encode S into S' $O(\frac{2n}{3})$

// STEP 2

(2) ~~sort~~ SA_{1,2} = SA(S');
 decoding also

// STEP 3

$O(\frac{2n}{3}) \Rightarrow$ (a) RANK SA_{1,2}

$\frac{n}{3}$ (b) create a character string w/ letter and RANK of SA.

$O(n)$ (c) sort those. \Rightarrow SA₀ = [sorted 2 character strings]
 decode it.

STEP 4

merge SA_{1,2} and SA₀ by

if (S[i] != S[j]) print smallest

else

suffix ranking procedure. $O(1)$

$$\underbrace{O(n)}_{\text{step 1}} + \underbrace{T(\frac{2}{3}n)}_{\text{step 2}} + \underbrace{O(n)}_{\text{step 3}} + \underbrace{O(n)}_{\text{step 4}} =$$

24

$$T(n) = O(n) + T\left(\frac{2}{3}n\right)$$

$$a=1 \quad b=3/2$$

masters

$$n^1 \text{ vs } n^{\log_{3/2} 1}$$

$$\boxed{n^1} \text{ vs } n^0$$

$$\Rightarrow O(n)$$

Merge procedure

suff i

$$\textcircled{i+1}$$

suff j

$$\textcircled{j+1}$$

\Rightarrow

$$\boxed{SA_{12}}$$

$i \% 3 = 2$	$j \% 3 = 0$
$i+1 \% 3 = 0$	$j+1 \% 3 = 1$
$i+2 \% 3 = 1$	$j+2 \% 3 = 2$

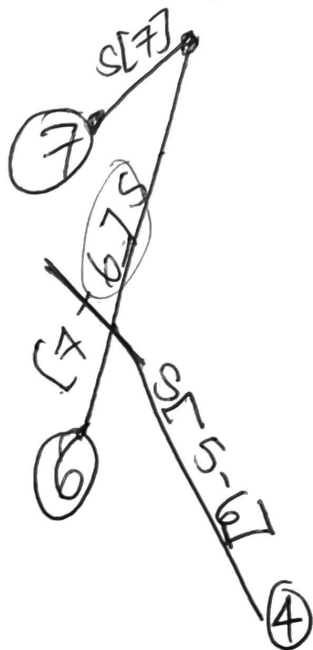
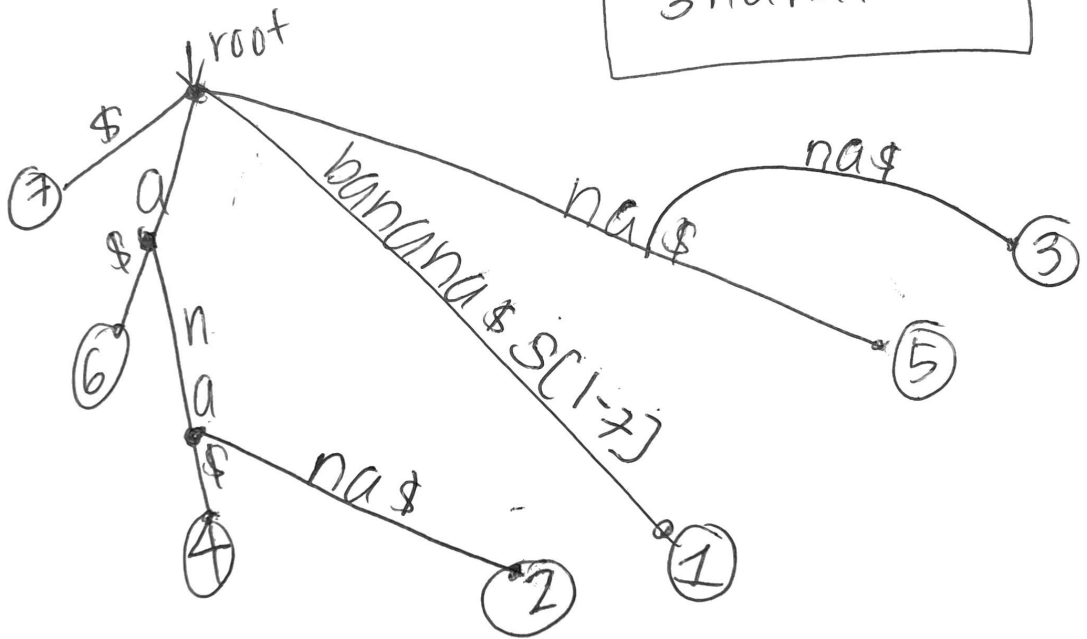
one more time

SAME LIST! SA_{12}

1 2 3 4 5 6 7
banana\$

SA 7 6 4 2 1 5 3
LCP

✓ 7\$
 ✓ 6 a\$
 ✓ 4 ~~ana~~ ana\$
 ✓ 2 ana na\$
 1 banana\$
 5 na\$
 3 nana\$



7 \$
 6 a\$
 4 ~~ana~~ ana\$
 2 ana na\$
 1 banana\$
 5 na\$
 3 nana\$

suff 6 : a\$
 suff 4 : ana\$

$S = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ a & b & a & b & a & b & \$ \end{matrix}$

$SA: \overset{\vee}{7}, \overset{\vee}{5}, \overset{\vee}{3}, 1, 6, 4, 2$
 $LCP: 0, 2, 4, 0, 1, 3$

