Cheat Sheet Midterm 2

Pattern Searching Tools

1. Z-Algorithm
2. Suffix trees
3. Suffix Array
4. BWT

Creation time

1. Z-algorithm : $\boxed{O(P+T)}$
2. Suffix Tree: Brute force $\Rightarrow \boxed{O(T^2)}$

   linear $\Rightarrow$ a) create SA linearly

   b) create LCP linearly

   c) create suffix tree linearly

   $\Rightarrow \boxed{O(T)}$
3. Suffix Array: Brute force $\Rightarrow$ merge sort $\Rightarrow \boxed{T^2 \log T}$

   Divide & conquer $\Rightarrow \boxed{O(T)}$
4. BWT: Brute force $\Rightarrow \boxed{T^2 \log(T)}$ with merge sort

   or SA $\Rightarrow \boxed{O(T)}$

   inverse $\boxed{O(T)}$ by first. last method

Time to Search

① Z-algo → scan Z-array $O(P+T)$

② Suffix Trees :

$|P| \cdot \Sigma$ node edges

DFS $O(K) + O(K)$

→ $\boxed{O(P+K)}$ ← # of pattern in T

③ SA search : binary search

$\boxed{\log T(p) + (P) K}$ ← # of pattern in T

④ BWT Search: $\boxed{O(|P| + K)}$ note: DFS.. $(V+E)$

Steps to build SA in O(n) ← linear time

PS: follow lecture 9 from hand written notes in google drive

① 3 mers of % 3 : 1, 2 : list 1

② sort list 1

③ Create a ranking

④ Encode S in ranking

⑤ Create S'

⑥ $X = SA(S')$

⑦ Decode x to get sorted list 1 of suff i % 3 = 1, 2

⑧ list i % 3 = 0

⑨ encode list 2 in rank from Step 7

⑩ sort list 2 with radix sort

examples S, SA, LCP

given S, SA, LCP create suffix tree in O(n)

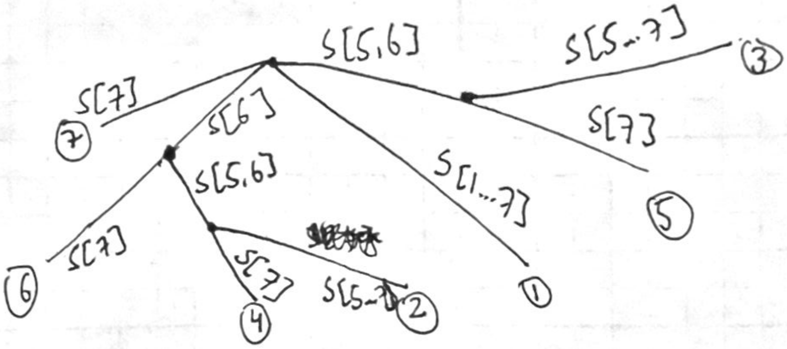
S= banana$
1234567

SA → follow 10 steps with 3-mers O(7)

to get SA: [7, 6, 4, 2, 1, 5, 3]

LCP = (Suff SA[i], suff SA[i-1])

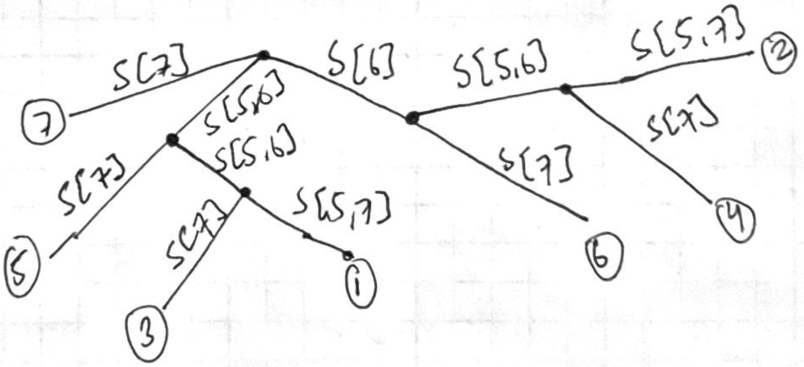
SA	7	6	4	2	1	5	3
LCP		0	1	3	0	0	2



d(v)
depth(v) = depth(v) # of characters of the path to v

ex: 2 S= ababab$

SA	7	5	3	1	6	4	2
LCP		0	2	4	0	1	3



ex 3: SA

	j	K	

suff$_K$ > suff$_j$

BWT Brute force:

① create all rotations of S

② Sort all rotations

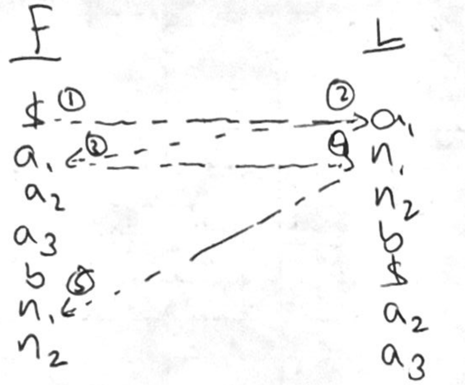
③ in BWT sorted matrix grad

$BWT(S) = A[1][n], A[2][n] \dots A[n][n]$

note: BWT = Last letter of each suffix array item in sequential order.

Example of linear time reverse BWT

$x = a\, n_1 n_2\, b\, \$\, a_2 a_3$



F	L
\$	a_1
a_1	n_1
a_2	n_2
a_3	b
b	\$
n_1	a_2
n_2	a_3

① \$ is last character in S

② comes before $\boxed{1}$ ~~\$~~

④ comes before $\boxed{2}$

$b\, a_3 n_2 a_2 n_1 a_1 \quad \$$