MATH-578A: Homework # 1

Due on Tuesday, March 10, 2015

Saket Choudhary
2170058637

Contents

Question # 1

Question # 2

Space required to store the nodes of suffix tree : O(n) Suffix tree has this important property that no internal node has less than 2 children, and there are n leaves in total. So the number of internal nodes is bound by $n-1$ and hence the total space is bound by $O(n)$.
For a GST, the space requirement is $O(\sum_{|S_i|}) + O()TODO$

Question # 3

At each node store the identity of the leaves below it whether they come from $S_1, S_2..., S_n$ etc. This can be done using depth first traversal occupying $O(m^2)$ space but only $O(m)$ in time. Then pattern P can be searched for. The identity of nodes where P search stops gives the occurence of P in all different strings.
TODO: Polish

Question # 4

We will assume N is a multiple of 2 and make a $O(n)$ query for a LSB of the n numbers. The number of 1 and 0 should ideally be same were all numbers present. Whichever occurence is smaller(0 or 1)($floor(\frac{N}{2})$), we do this for next $\frac{N}{2}$, then next $\frac{N}{4}$ and so on (total $O(2N)$) at each point storing which was lower $0 or 1$ and then recontruct the missing sequence.
TODO: Give Example with n=4.

Question # 5

Hash table sizes are often prime numbers to minimise probability of collisions of unlike quantities. Prime number modules ensures the number being divided go to maximum possible buckets(because the dividend and divisor are coprime)

Question # 6

Identifying longest common occuring substring, Travel to the deepest node that has a suffix link!

continued on next page...