

# Tutorial 6

## CS 213: Data Structures and Algorithms Autumn 2021

1. This is concerning Prof. Garg's argument that the total number of unsuccessful comparisons, i.e., the number of shifts in KMP algorithm, is no more than  $|T|$ . Is this strictly true?
2. Review the KMP algorithm to see how it detects two overlapping occurrences of the pattern.
3. For a pattern  $P[1..n]$ ,  $h(i)$  is defined to be the smallest  $k > 0$  such that  $P[1] = P[k+1], \dots, P[i-k] = P[i]$ , but  $P[i-k+1] \neq P[i+1]$ . If there is no such match, define  $h(i)$  to be  $i$ . Fill in the table below.

$i$	1	2	3	4	5	6	7	8	9
$P[i]$	b	a	b	b	a	a	b	b	a
$h(i)$									

4. Suppose that there is a letter  $z$  in  $P$  of length  $n$  such that it occurs in only one place, say  $k$ , which is given in advance. Can you optimize on the computation of  $h$ ?
5. Compute the suffix tree for **abracadabra\$**. Compress degree 1 nodes. Use substrings as edge labels. Put a square around nodes where a suffix ends. Use it to locate the occurrences of **abr**.
6. Review the argument that, for a given text  $T$ , consisting of  $k$  words, the ordinary trie occupies space which is a constant multiple of  $|T|$ . How is it that the suffix tree for a text  $T$  is of size  $\mathcal{O}(|T|^2)$ ?