

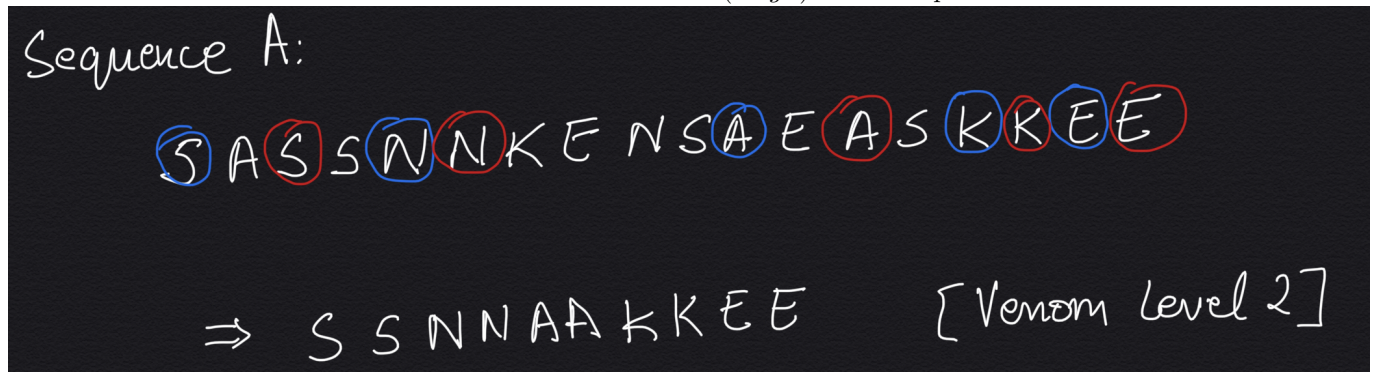
COMP3121 Homework Q1

Arth Sanskar Patel
z5228942

July 7, 2020

1 Answer

Here let sequence A be the DNA sequence of a given snake. Let B be the sequence SNAKE. Lets first find if B is a sub sequence of A. We can proceed in a similar manner to Q9 of Tutorial 3. We first find and mark the earliest occurrence of first letter of SNAKE (sequence B) in sequence A. Then for each subsequent letter of B, we find and mark the earliest occurrence of that letter in A which is after the last marked letter. If we reach the end of B before or at the same time we reach the end of A, we know that B is a sub sequence of A. This means given DNA of the snake is at least of venom level 1. Now before we do this process we initialise a hashmap and store the index of our found letters of B in sequence A. The whole process is done in $O(n \log n)$ time since going over letters of sequence of length n is done in $O(n)$ time and binary search inside the for loop will cost us a total of $O(n \log n)$. We again repeat the process described above but this time if we find a letter of B in A, we check if that index is already in our hashmap (this will take only $O(1)$ time), if it exists we ignore it and proceed forward. If it doesn't exist we take it into consideration and add its index to the hashmap. We repeat this until there are no sub sequences of B left in A. Now we go over A one more time and for each index check if its in the hashmap. If it is present in the hashmap, we move forward, otherwise we remove that letter. The total time taken by this algorithm is $O(n \log n) + O(n)$, there is a $+O(n)$ since we are going over the sequence in the end to remove letters. Hence the overall time taken will be $O(n \log n)$. An example is shown below of how this works.



As you can see, after the first process (index marked in blue), we will have our hashmap as:

$$HMap = [0, 4, 10, 16]$$

After the repeating the process (index marked in red), we will have our hashmap as:

$$HMap = [0, 4, 10, 16, 2, 5, 12, 15, 17]$$

Since repeating the process third time will not return any sequence as there are none left, we will go over the Sequence A and remove all the letter not in our hashmap (not colored blue or red) to give the venom level 2 as shown in the image.