ASSIGNMENT3 QUESTION1

1. After the success of your latest research project in mythical DNA, you have gained the attention of a most diabolical creature: Medusa. Medusa has snakes instead of hair. Each of her snakes' DNA is represented by an upper-case string of letters. Each letter is one of S, N, A, K or E. Your extensive research shows that a snake's venom level depends on its DNA. A snake has venom level x if its DNA:

   - has exactly 5x letters
   - begins with x copies of the letter S
   - then has x copies of the letter N
   - then has x copies of the letter A
   - then has x copies of the letter K
   - ends with x copies of the letter E.

   For example, a snake with venom level 1 has DNA SNAKE, while a snake that has venom level 3 has DNA SSSNNNAAAKKKEEE. If a snake's DNA does not fit the format described above, it has a venom level of 0. Medusa would like your help making her snakes venomous, by deleting zero or more letters from their DNA. Given a snake's DNA, can you work out the maximum venom level this snake could have? Your algorithm should run in time $O(n \log n)$

Answer:

At first, we will go through the DNA of the snake one by one and count the number of letter S, N, A, K, E in DNA sequence. And determine the minimum number (M)of the 5 letters. It will take O(n). Clearly the maximum venom level (L) will less or equal the minimum number (M)of letters. Then use greedy strategy to see if it is possible to delete some of letter so make the rest sequences is exactly like $\underbrace{SS\ldots S}_{M}\underbrace{NN\ldots N}_{M}\underbrace{AA\ldots A}_{M}\underbrace{KK\ldots K}_{M}\underbrace{EE\ldots E}_{M}$.

We will make a loop to go through the DNA sequence in multiple times.

In every loop, We assume the current loop maximum venom level is X, we will count the letter 'S' num until it equal X, then count the letter 'N' in the rest sequence. Until all letter (snake) is count exactly equal to X(in order), it means this maximum venom level is bigger or equal to X and go to next loop(increase X), otherwise we cannot find it maximum venom level is less than X, try next loop(decrease X).

If we do not count follow the order 'S' 'N' 'A' 'K' 'E', Since only $\underbrace{SS\ldots S}_{M}\underbrace{NN\ldots N}_{M}\underbrace{AA\ldots A}_{M}\underbrace{KK\ldots K}_{M}\underbrace{EE\ldots E}_{M}$ can be venom level, that is invalid. If we do not count it equal X, if we count bigger than X it may decrease the venom level even make no venom level, which is not optimal and take more time to find. if we count less than X it will take more time to find the maximum venom level since we use binary search, that is not optimal.

If x = 0 mean the maximum venom level is 0. If x = M and fit the rules means the maximum venom level is M. Because there is no bigger value.

And we should start by X = M, and we know $\text{maximum venom level} \leq M$, then we can use binary search to determine the real correct value L by increase or decrease half current x. It will take $O(log\, n)$ time to find the correct value and every loop will take $0(n)$ time. The final will run in $O(n\, log\, n)$ time. i.e. if X = M not fit rule, try X = $\frac{M}{2}$, if pass, try X = $\frac{3M}{4}$, continue do that until X fit and x +1 not fit.

If we do not use binary search we have to go through all possible venom value and will take more time $O(n^2)$ to find the maximum venom level, which is not optimal answer.

To be noticed, the maximum venom level should be determined by no bigger value fit the rules. It means the real maximum venom level should be fit the rules but maximum venom level + 1 not fit. And in this approach, no matter where the maximum venom level is and the DNA sequence, we can always find optimal answer.