

Tasks Details

[Check out Codility training tasks](#)

Medium	1. GenomicRangeQuery	Task Score	Correctness	Performance
	Find the minimal nucleotide from a range of sequence DNA.	100%	100%	100%

Task description

A DNA sequence can be represented as a string consisting of the letters A, C, G and T, which correspond to the types of successive nucleotides in the sequence. Each nucleotide has an *impact factor*, which is an integer. Nucleotides of types A, C, G and T have impact factors of 1, 2, 3 and 4, respectively. You are going to answer several queries of the form: What is the minimal impact factor of nucleotides contained in a particular part of the given DNA sequence?

The DNA sequence is given as a non-empty string $S = S[0]S[1] \dots S[N-1]$ consisting of N characters. There are M queries, which are given in non-empty arrays P and Q , each consisting of M integers. The K -th query ($0 \leq K < M$) requires you to find the minimal impact factor of nucleotides contained in the DNA sequence between positions $P[K]$ and $Q[K]$ (inclusive).

For example, consider string $S = \text{CAGCCTA}$ and arrays P, Q such that:

$P[0] = 2 \quad Q[0] = 4$
 $P[1] = 5 \quad Q[1] = 5$
 $P[2] = 0 \quad Q[2] = 6$

The answers to these $M = 3$ queries are as follows:

- The part of the DNA between positions 2 and 4 contains nucleotides G and C (twice), whose impact factors are 3 and 2 respectively, so the answer is 2.
- The part between positions 5 and 5 contains a single nucleotide T, whose impact factor is 4, so the answer is 4.
- The part between positions 0 and 6 (the whole string) contains all nucleotides, in particular nucleotide A whose impact factor is 1, so the answer is 1.

Write a function:

```
def solution(S, P, Q)
```

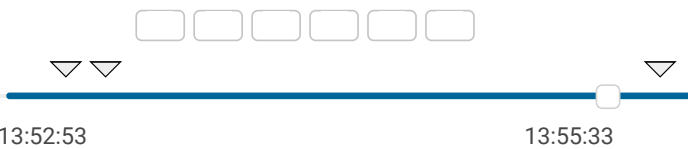
that, given a non-empty string S consisting of N characters and two non-empty arrays P and Q consisting of M integers, returns an array consisting of M integers specifying the consecutive answers to all queries.

Result array should be returned as an array of integers.

Solution

Programming language used:	Python
Total time used:	3 minutes ?
Effective time used:	3 minutes ?
Notes:	not defined yet

Task timeline ?



Code: 13:55:33 UTC, py, final, [show code in pop-up](#)
score: 100

```
1 # you can write to stdout for debugging purposes, e.g.
2 # print("this is a debug message")
3
4 def solution(S, P, Q):
5     M, letters = [], {'A':1, 'C':2, 'G':3, 'T':4}
6     for idx in range(len(P)):
7         min_coeff = 4
8         for key in list(letters.keys()):
9             if(key in S[P[idx]:Q[idx]+1]):
10                 min_coeff = letters[key]
11                 break
12         M.append(min_coeff)
13     return M
```

Analysis summary

The solution obtained perfect score.

Analysis

For example, given the string $S = \text{CAGCCTA}$ and arrays P, Q such that:

```
P[0] = 2    Q[0] = 4
P[1] = 5    Q[1] = 5
P[2] = 0    Q[2] = 6
```

the function should return the values $[2, 4, 1]$, as explained above.

Write an **efficient** algorithm for the following assumptions:

- N is an integer within the range $[1..100,000]$;
- M is an integer within the range $[1..50,000]$;
- each element of arrays P, Q is an integer within the range $[0..N - 1]$;
- $P[K] \leq Q[K]$, where $0 \leq K < M$;
- string S consists only of upper-case English letters A, C, G, T.

Copyright 2009–2021 by Codility Limited. All Rights Reserved. Unauthorized copying, publication or disclosure prohibited.

Detected time complexity: **$O(N + M)$**

expand all	Example tests
▶ example example test	✓ OK
expand all	Correctness tests
▶ extreme_sinlge single character string	✓ OK
▶ extreme_double double character string	✓ OK
▶ simple simple tests	✓ OK
▶ small_length_string small length simple string	✓ OK
▶ small_random small random string, length = ~300	✓ OK
expand all	Performance tests
▶ almost_all_same_letters GGGGGG..??..GGGGGG..??..GGGGGG	✓ OK
▶ large_random large random string, length	✓ OK
▶ extreme_large all max ranges	✓ OK

The PDF version of this report that may be downloaded on top of this site may contain sensitive data including personal information. For security purposes, we recommend you remove it from your system once reviewed.