CHSTR - Editorial

PROBLEM LINK:

0
Practice
Contest

Author: Vasia Antoniuk
Tester: Mahbubul Hasan and Sunny Aggarwal
Editorialist: Balajiganapathi Senthilnathan
Russian Translator: Sergey Kulik
Mandarian Translator: Minako Kojima

DIFFICULTY:

Medium

PREREQUISITES:

Z function, combinatorics

PROBLEM:

Given a string and a query k, output the number of ways we can choose k equal substrings from the string.

SHORT EXPLANATION

Let $cnt[i]$ be the number of different substrings that occur in the string exactly i times. Then the answer for a particular $k(\leq n)$ will be : $\sum_{i=k}^{n} cnt[i] * \binom{i}{k}$ we can calculate cnt[i] using suffix array/Z function.

EXPLANATION:

Let $cnt[i]$ be the number of different substrings that occur in the string exactly i times.

For example, for the string "ababa", the array will be:
$cnt[1] = 4$ -> {"ababa", "abab", "baba", "bab"} occur exactly once
$cnt[2] = 4$ -> {"aba", "ab", "ba", "b"} occur exactly twice
$cnt[3] = 1$ -> {"a"} occur exactly thrice
$cnt[4] = 0$
$cnt[5] = 0$

Now suppose we want the answer for $k = 2$, i.e. number of ways to choose 2 equal strings. How do we calculate this using the cnt array?

$cnt[1]$ is of no use as we need 2 equal string.
$cnt[2]$ -> There are 4 different substrings that occur 2 times. So, we can add 4 to the answer.
$cnt[3]$ -> There is only 1 substring that occurs 3 times. But we need only 2 times, so we can choose any 2 of the 3. So, $\binom{3}{2} = 3$

Summing them, for $k = 2$ we get 7.

Now let us focus on calculating the cnt array. To calculate this let us loop through all suffixes of S.

For the suffix $P = S[i..N]$, let us calculate the array $Z[i..N]$ where $Z[i]$ is the maximal equal substring starting from i that matches a prefix of $S[i..N]$.

For example, for the above string consider i = 1, the whole string is the suffix "ababa". Here the Z array would be
$Z[1] = 5$ -> maximum prefix matching of "ababa" and "ababa"
$Z[2] = 0$ -> maximum prefix matching of "ababa" and "baba"
$Z[3] = 3$ -> maximum prefix matching of "ababa" and "aba"
$Z[4] = 0$ -> maximum prefix matching of "ababa" and "ba"
$Z[5] = 1$ -> maximum prefix matching of "ababa" and "a"

How is this array useful for calculating cnt?

From the above array we can conclude that we have 1 substring which can be choosen atleast 3 times. How? Because there are 3 entries in the above array that are greater than or equal to 1. Similarly we can deduce for 2 (2 times), 3 (2 times), 4 (1 time) and 5(1 time). So we increment cnt[3] once, cnt[2] twice and cnt[1] twice.

We do this for all suffix of S. The only thing left is to observe that cnt[i] now actually has how many different strings can be choosen *atleast* i times. We want it to be exactly i times. This is simple: we just subtract from $cnt[i]$ ($cnt[i+1] + cnt[i+2] \ldots + cnt[N]$).

We can compute the Z function in $O(n)$ time. See here for how to do that.

Time Complexity:

There are N suffixes and we take $O(N)$ time for each suffix so total time is: $O(N^2).

AUTHOR'S AND TESTER'S SOLUTIONS:

Author's solution
Tester's solution

june15 editorial

This question is marked "community wiki".

edited 41 mins ago
admin ♦♦
10.9k●346●472●486

asked yesterday
balajiganapath ♦♦
589●3●14●27
accept rate: 77%

6 Answers:

oldest newest most voted

1

This question was very similar to http://discuss.codechef.com/questions/43060/anusar-editorial.

link | award points

answered 3 mins ago
tapasjain01
51●2
accept rate: 0%

0

Can we solve it without z algorithm?

link | award points

answered 32 mins ago
rajatgoyal
99●4
accept rate: 0%

1 yes, i have solved with hashing using rabin karp algo see my solution http://www.codechef.com/viewsolution/7225121 the code is not quite readable but you can understand it.

raja44ever (12 mins ago)

@raja44ever need a little bit explanation of your approach. How did you decided the hash function. Please help ?

apptica (1 min ago)

0

I tried using a Trie data structure , and finding out cnt array in O(N^2). and the use DP to store the answers for 1 to n and the print appropriate answer or 0 if query is greater than N. But i am getting TLE for three cases in Subtask 3? is anymore optimizations required ???

link | award points

answered 26 mins ago
geek_geek
1
accept rate: 0%

0

I am using a Trie data structure and using that finding out how the cnt array in O(N^2) .and using DP to answer the queries in O(1) time .I am still getting TLE for SUbtask 3 except two cases .... i am using FAST IO is any further optimizations required??

link | award points

answered 24 mins ago
geek_geek
1
accept rate: 0%

0

Even I had initially used trie data structure to solve the problem and I am pretty sure my complexity was O(3*(n^2)). I got only one of the last subtask set accepted and rest all tle..The link to my solution is http://www.codechef.com/viewsolution/7156684 . Is it possible that the constant(3) multiplied to n^2 in the big Omega notation lead to TLE?? Does that happen often?? Plz someone answer.

link | award points

edited 5 mins ago

answered 6 mins ago
ohil17yo36
21●3
accept rate: 0%

0

Suffix Array + RMQ can also be used for solving the problem.

link | award points

answered 4 mins ago
dragonslayerx
11●1
accept rate: 0%

Your answer

[hide preview] community wiki

Post Your Answer