

# 字符串专题

吴戈

2024 年 2 月 14 日

但是 kmp 算法是基于均摊而实现的。有的时候我们希望避免均摊，该如何改进？

对于  $\Sigma$  较小的情况，可以维护每一个字符的 fail 节点，做到  $\mathcal{O}(n \Sigma)$ 。

但是 kmp 算法是基于均摊而实现的。有的时候我们希望避免均摊，该如何改进？

对于  $\Sigma$  较小的情况，可以维护每一个字符的 fail 节点，做到  $\mathcal{O}(n \Sigma)$ 。

对于  $\Sigma$  较大的情况，可以使用可持久化数组，来做到  $\mathcal{O}(n \log \Sigma)$ 。

## 定义

对于一个长度为  $n$  的字符串  $s$ , 定义函数  $z[i]$  表示  $s$  和  $s[i, n]$  的最长公共前缀的长度。 $z$  被称为  $s$  的 Z 函数。

## Problem

给定  $S, T$ , 对于每个  $i$  求  $S_i$  换为  $T_i$  后  $S$  的最长 Border。  
要求线性。

## Problem

给出  $n$  个小写字符串  $s_1 \sim s_n$ ，从中选出一个以字符串作为元素的子序列按顺序拼接，要求拼接出来的串长度恰为  $k$ ，求这些串里面字典序最小的，保证有解。

$n \leq 2000$ ,  $k \leq 10^4$ ,  $\sum |s_i| \leq 10^6$ 。

## 定义

AC 自动机是以 Trie 为基础，以 KMP 为思想的自动机。一般用于解决多模式匹配。

## Problem

维护一个字符串集合  $S$ ，支持三种操作：

- 向  $S$  中插入字符串  $s$ 。
- 从  $S$  中删除字符串  $s$ 。
- 给定字符串  $t$ ，求  $S$  中所有串一共在  $t$  中出现多少次。

输入字符串总长度  $L$  不超过  $3 \times 10^5$ 。



## Problem

给定一个字符串  $S$  和一个序列  $W$ ，初始时它们都为空。在线完成  $n$  次操作。每次操作在  $S$  后面添加一个字符  $c_i$ ，在序列  $W$  后面添加一个数字  $w_i$ 。二者均加密。

对于子区间  $[L, R]$ ，若  $S[L : R] = S[: R - L + 1]$ ，则其权值为  $\min_{i=L}^R \{W_i\}$ ，否则其权值为 0。

每次操作后，你都要求出当前的串的所有子区间的权值和。

$1 \leq n \leq 6 \times 10^5$ ， $0 \leq w_i < 2^{30}$ ， $c_i$  为小写字母。

## Problem

给定  $n$  个串  $s_1 \dots s_n$ ,  $q$  次询问  $l, r, k$  表示  $s_{l \sim r}$  在  $s_k$  中的出现次数之和。  
 $n, q, \sum |s_i| \leq 10^5$ 。

## Problem

给定字符串  $t$  和  $n$  个字符串  $s_i$ 。

给定  $m$  个询问，第  $i$  个询问给你  $l_i, r_i$ ，询问  $t[l_i, r_i]$  中  $s_1, s_2, \dots, s_n$  中的字符串出现了多少次。

形式化地，统计满足  $t[a, b]$  在  $s_1, s_2, \dots, s_n$  中出现过且  $l_i \leq a \leq b \leq r_i$  的  $(a, b)$  的个数。

$|t| \leq 5 \times 10^6, 1 \leq n, m \leq 5 \times 10^5, \sum |s_i| \leq 10^6$ 。

## Problem

给定长为  $n$  的串  $s$  和  $m$  个串  $t_1 \dots t_m$ , 定义  
 $h(s, l, r) = \sum_{i=1}^m \text{occur}(s[l, r], t_i) \times w_i$ , 其中  $\text{occur}(s, t)$  表示  $t$  在  $s$  中的出现次数。

求  $\sum_{i=1}^n \sum_{j=i}^n h(s, i, j)^k$ , 其中  $k = 5$ 。

$1 \leq T \leq 5$ ,  $1 \leq n, \sum |t_i|, m \leq 5 \times 10^5$ ,  $|\Sigma| = 26$ 。

前者能以  $\mathcal{O}(n)$  复杂度求出一个字符串的所有本质不同的回文子串的算法。

后者是能存储一个串中所有回文子串的高效数据结构!!

## Problem

求有多少字符集大小为  $|\Sigma|$ , 长为  $n$  的字符串满足其不存在长度为  $k$  的回文子串, 答案对  $10^9 + 7$  取模。

$1 \leq |\Sigma| < 10^9 + 7, 1 \leq n \leq 1000, 2 \leq k \leq 25$ 。

部分分:  $k \leq 7$ 。

## Problem

将给定字符串分为偶数段。假设分成了  $s_1, s_2, \dots, s_k$ , 求  $s_1 = s_k, s_2 = s_{k-1} \dots$  的方案数。  
 $|S| \leq 10^6$ 。

## Problem

给你两个长度为  $n$  的字符串  $s$  和  $t$ 。你可以翻转  $t$  的若干个不相交的区间，要求最终  $s$  和  $t$  相同。问你最少翻转几个区间，并输出方案。

$|s| = |t| \leq 5 \times 10^5$ 。