

# Basic String Algorithms

ynycoding

2025 年 7 月 31 日

# KMP 算法

有问题就讲讲

# EXKMP 算法

有问题就讲讲

# BZOJ4974 字符串大师

## 题面

有一个未知字符串  $s$ ，定义  $p[i]$  为前缀  $s[1 \cdots i]$  的 period，输入字符串长度  $n$  和数组  $p$ ，求出可能的  $s$  中字典序最小的一个。保证有解。

$n \leq 10^5$ 。

# BZOJ4974 字符串大师

## 题面

有一个未知字符串  $s$ ，定义  $p[i]$  为前缀  $s[1 \cdots i]$  的 period，输入字符串长度  $n$  和数组  $p$ ，求出可能的  $s$  中字典序最小的一个。保证有解。

$n \leq 10^5$ 。

模拟 kmp，每次会知道字符的相等或者不等关系。直接贪心即可。

## LOJ 10035

## 题面

给定若干个长度  $\leq 10^6$  的字符串，询问每个字符串最多是由多少个相同的子字符串重复连接而成的。如：ababab 则最多有 3 个 ab 连接而成。

$n \leq 10^6$ 。

## LOJ 10035

## 题面

给定若干个长度  $\leq 10^6$  的字符串，询问每个字符串最多是由多少个相同的子字符串重复连接而成的。如：ababab 则最多有 3 个 ab 连接而成。

$n \leq 10^6$ 。

过于简单

# [NOI2014] 动物园

## 题面

对于字符串  $S$  的每个前缀，计算出它有多少个 border，满足这些 border 的长度都不超过该前缀长度的一半。

$n \leq 10^6$ 。



# [NOI2014] 动物园

## 题面

对于字符串  $S$  的每个前缀，计算出它有多少个 border，满足这些 border 的长度都不超过该前缀长度的一半。

$n \leq 10^6$ 。

维护指针指向不超过一半长度的串。

# [NOIP2020] 字符串匹配

## 题面

小 C 学习完了字符串匹配的相关内容，现在他正在做一道习题。

对于一个字符串  $S$ ，题目要求他找到  $S$  的所有具有下列形式的拆分方案数：

$S = ABC$ ,  $S = ABABC$ ,  $S = ABAB \cdots ABC$ , 其中  $A, B, C$  均是非空字符串，且  $A$  中出现奇数次的字符数量不超过  $C$  中出现奇数次的字符数量。

更具体地，我们可以定义  $AB$  表示两个字符串  $A, B$  相连接，例如  $A = aab$ ,  $B = ab$ , 则  $AB = aabab$ 。

并递归地定义  $A^1 = A$ ,  $A^n = A^{n-1}A$  ( $n \geq 2$  且为正整数)。例如  $A = abb$ , 则  $A^3 = abbabbabb$ 。

则小 C 的习题是求  $S = (AB)^i C$  的方案数，其中  $F(A) \leq F(C)$ ,  $F(S)$  表示字符串  $S$  中出现奇数次的字符的数量。两种方案不同当且仅当拆分出的  $A, B, C$  中有至少一个字符串不同。

小 C 并不会做这道题，只好向你求助，请你帮帮他。

$n \leq 10^7$ 。

# [NOIP2020] 字符串匹配

枚举循环节  $AB$ ，利用 EXKMP 找到最大循环次数。利用桶和指针可以做到  $O(n)$ 。

# Trie

相信大家都会

## LOJ537

## 题面

现在给出一段 DNA 序列，请帮他求出这段 DNA 序列中所有连续  $k$  个碱基形成的碱基序列中，出现最多的一种的出现次数。

$n \leq 5 \times 10^6, k \leq 10$ 。

过于简单

# AC 自动机

有问题就讲讲

## LOJ10059

## 题面

有一个长度不超过  $10^5$  的字符串  $S$ 。Farmer John 希望在  $S$  中删掉  $n$  个屏蔽词（一个屏蔽词可能出现多次），这些词记为  $t_1 \sim t_n$ 。

FJ 在  $S$  中从头开始寻找屏蔽词，一旦找到一个屏蔽词，FJ 就删除它，然后又从头开始寻找（而不是接着往下找）。FJ 会重复这一过程，直到  $S$  中没有屏蔽词为止。注意删除一个单词后可能会导致  $S$  中出现另一个屏蔽词。这  $n$  个屏蔽词不会出现一个单词是另一个单词子串的情况，这意味着每个屏蔽词在  $S$  中出现的开始位置是互不相同的，请帮助 FJ 完成这些操作并输出最后的  $S$ 。

$n \leq 10^6$ 。

## LOJ10059

## 题面

有一个长度不超过  $10^5$  的字符串  $S$ 。Farmer John 希望在  $S$  中删掉  $n$  个屏蔽词（一个屏蔽词可能出现多次），这些词记为  $t_1 \sim t_n$ 。

FJ 在  $S$  中从头开始寻找屏蔽词，一旦找到一个屏蔽词，FJ 就删除它，然后又从头开始寻找（而不是接着往下找）。FJ 会重复这一过程，直到  $S$  中没有屏蔽词为止。注意删除一个单词后可能会导致  $S$  中出现另一个屏蔽词。这  $n$  个屏蔽词不会出现一个单词是另一个单词子串的情况，这意味着每个屏蔽词在  $S$  中出现的开始位置是互不相同的，请帮助 FJ 完成这些操作并输出最后的  $S$ 。

$n \leq 10^6$ 。

AC 自动机板子题



# Luogu2414

## 题面

输入一棵 Trie，其上有  $n$  个单词结点， $m$  组询问，每次询问第  $x$  个单词在第  $y$  个单词中出现了几次。

$n \leq 2 \times 10^5$ 。

# Luogu2414

## 题面

输入一棵 Trie，其上有  $n$  个单词结点， $m$  组询问，每次询问第  $x$  个单词在第  $y$  个单词中出现了几次。

$n \leq 2 \times 10^5$ 。

询问离线。每次枚举  $y$ ，对于  $x$  使用树状数组维护祖先的关键点个数。

# ARC141F

题面

# ARC141F

## 题面

首先假设是合法的，那么对于每个串可以利用 AC 自动机贪心地删除。如果无法删除完则非法。

如果一个串有两种删法，那么一定有字符串相交。只需要判断是否存在除了相交的部分其余部分不相等的串。

# Manacher

有问题就讲讲

# Luogu4555

## 题面

顺序和逆序读起来完全一样的串叫做回文串。比如 `acbca` 是回文串，而 `abc` 不是：  
`abc` 的顺序为 `abc`，逆序为 `cba`，不相同。

输入长度为  $n$  的串  $S$ ，求  $S$  的最长双回文子串  $T$ ，即可将  $T$  分为两部分  $X, Y$   
( $|X|, |Y| \geq 1$ ) 且  $X$  和  $Y$  都是回文串。

$n \leq 10^6$ 。

# Luogu4555

## 题面

顺序和逆序读起来完全一样的串叫做回文串。比如 `acbca` 是回文串，而 `abc` 不是：  
`abc` 的顺序为 `abc`，逆序为 `cba`，不相同。

输入长度为  $n$  的串  $S$ ，求  $S$  的最长双回文子串  $T$ ，即可将  $T$  分为两部分  $X, Y$   
( $|X|, |Y| \geq 1$ ) 且  $X$  和  $Y$  都是回文串。

$n \leq 10^6$ 。

枚举划分点即可。

# 后缀数组

后缀数组存储后缀按字典序排序后的结果。  
常见的有  $n \log n$  的倍增，以及线性的 SA-IS。



# 倍增后缀数组

大家都会

## 后缀数组

## LOJ2377

给定一个长度为  $n$  的字符串  $S$ ，令  $T_i$  表示它从第  $i$  个字符开始的后缀，求：

$$\sum_{1 \leq i < j \leq n} \text{len}(T_i) + \text{len}(T_j) - 2 \text{lcp}(T_i, T_j)$$

$$2 \leq n \leq 500000$$

# 后缀数组

## LOJ2377

给定一个长度为  $n$  的字符串  $S$ ，令  $T_i$  表示它从第  $i$  个字符开始的后缀，求：

$$\sum_{1 \leq i < j \leq n} \text{len}(T_i) + \text{len}(T_j) - 2 \text{lcp}(T_i, T_j)$$

$$2 \leq n \leq 500000$$

简单题

# 后缀数组

## LOJ2083

如果一个字符串可以被拆分为 AABB 的形式，其中 A 和 B 是任意非空字符串，则我们称该字符串的这种拆分是优秀的。

现在有  $T$  组询问，每组询问给出一个长度为  $n$  的字符串  $S$ ，求出在它所有子串的所有拆分方式中，优秀拆分的总个数。

其中出现在不同位置的相同子串，我们认为是不一样的子串，它们的优秀拆分均会被记入答案。

$1 \leq T \leq 10, n \leq 30000$

# 后缀数组

## LOJ2083

如果一个字符串可以被拆分为 AABB 的形式，其中 A 和 B 是任意非空字符串，则我们称该字符串的这种拆分是优秀的。

现在有  $T$  组询问，每组询问给出一个长度为  $n$  的字符串  $S$ ，求出在它所有子串的所有拆分方式中，优秀拆分的总个数。

其中出现在不同位置的相同子串，我们认为是不一样的子串，它们的优秀拆分均会被记入答案。

$1 \leq T \leq 10, n \leq 30000$

先考虑求出以  $i$  结尾的 AA 的数量  $f_i$  以及以  $i$  开头的 BB 的数量  $g_i$ 。答案就是

$$\sum_i f_i g_{i+1}$$

可以枚举长度后求 lcp。由于有调和级数，复杂度为  $O(n \log n)$ 。

# 后缀数组

## LOJ2133

给定串  $S$ ，求  $\exists i \in [0, n)$  求有多少对后缀满足  $Len(lcp) \geq i$ ，以及满足条件的两个后缀的权值  $a_i$  乘积的最大值。

$n \leq 300000, |a_i| \leq 10^9$

# 后缀数组

## LOJ2133

给定串  $S$ ，求  $\exists i \in [0, n)$  求有多少对后缀满足  $Len(lcp) \geq i$ ，以及满足条件的两个后缀的权值  $a_i$  乘积的最大值。

$n \leq 300000, |a_i| \leq 10^9$

处理出后缀数组后排序维护一个单调栈即可。

## LOJ3089

给定一只由数字和 . 构成的字符串  $s$ 。给定  $m$  个特殊串  $t_1 \sim t_m$ ,  $t_i$  的权值为  $v_i$ 。需要在  $s$  中为 . 的位置上填入数字, 一种填入方案的价值定义为:

$$\sqrt[c]{\prod_{i=1}^c w_i}$$

其中  $W$  表示在该填入方案中, 出现过的特殊串的价值的可重集合, 其大小为  $c$ 。每个位置填入的数字任意, 最大化填入方案的价值, 并输出任意一个方案。

设  $s = \sum_{i=1}^m |S_i|$ 。

$1 \leq n, m, s \leq 1501, 1 \leq V_i \leq 10^9$ 。



## LOJ3089

给定一只由数字和 . 构成的字符串  $s$ 。给定  $m$  个特殊串  $t_1 \sim t_m$ ,  $t_i$  的权值为  $v_i$ 。需要在  $s$  中为 . 的位置上填入数字, 一种填入方案的价值定义为:

$$\sqrt[c]{\prod_{i=1}^c w_i}$$

其中  $W$  表示在该填入方案中, 出现过的特殊串的价值的可重集合, 其大小为  $c$ 。每个位置填入的数字任意, 最大化填入方案的价值, 并输出任意一个方案。

设  $s = \sum_{i=1}^m |S_i|$ 。

$1 \leq n, m, s \leq 1501, 1 \leq V_i \leq 10^9$ 。

取对数后可以变成分数规划, AC 自动机上 dp 即可。

## CF1313E

给定三个串  $a, b, s$ ，其中  $a, b$  长度为  $n$ ， $s$  长度为  $m$ ，求出四元组  $(l1, r1, l2, r2)$  的个数，满足：

- $[l1, r1]$  和  $[l2, r2]$  的交集非空。
- $a$  中位置  $l1$  到  $r1$  组成的子串与  $b$  中位置  $l2$  到  $r2$  组成的子串拼起来恰好是  $s$ 。

$n \leq 10^5$ 。

## CF1313E

给定三个串  $a, b, s$ ，其中  $a, b$  长度为  $n$ ， $s$  长度为  $m$ ，求出四元组  $(l1, r1, l2, r2)$  的个数，满足：

- $[l1, r1]$  和  $[l2, r2]$  的交集非空。
- $a$  中位置  $l1$  到  $r1$  组成的子串与  $b$  中位置  $l2$  到  $r2$  组成的子串拼起来恰好是  $s$ 。

$n \leq 10^5$ 。

先 exkmp 求出  $a$  每个位置向后和  $b$  每个位置向前最长延伸长度。剩下的可以树状数组解决。

## UOJ172

<https://uoj.ac/problem/172>

## UOJ172

<https://uoj.ac/problem/172>

由于 border 构成  $\log$  个等差数列，我们可以对于每个等差数列跑一个同余最短路。  
切换模数的时候也可以用一个同余最短路。

## UOJ476

贾樟柯在《山河故人》里说,「生活就是重复」。在生活中,人们总是喜欢重复自己做过的事情。语言就是一个很经典的例子。

比如,我们在表示疑问时,总不满足于使用一个问号「?」;使用一连串的问号「????」总是显得比较有力。

在表示抱歉时,一句「对不起」总显得不够情愿;连着表示「对不起对不起」才足够表达自己的真诚。

A 国就是一个喜欢重复的国家。在这个国家中,一个基本句子可以用一个长度恰好为  $m$  的小写字母字符串表示。为了表达自己对重复的喜爱,A 国的人们总喜欢把自己想要表达的句子重复无限多次。

有时,这样的重复是充满意义的。A 国的人们把一个字典序小于给定的字符串  $s$ , 且长度和  $s$  相同的小写字母字符串称为一个有意义的语义片段。他们想知道,有多少个不同的基本句子(即长度恰好为  $m$  的小写字母字符串)在经过无限重复后,可以从中找出至少一个有意义的语义片段?

$n, m \leq 2000$ 。

## UOJ476

首先补集转化。考虑如何判断一个字符串是否合法。对于与  $s$  的 lcp, 可以使用 kmp 来维护, 一旦比 fail 链上任何字符更小就 break。

考虑使用 dp 来做这个过程。由于是循环, 我们考虑在充分长之后沿着字符串走在  $s$  的 kmp 上的节点会有循环节。令  $f_{i,j,k,0/1}$  表示从  $i$  号节点出发, 当前走了  $j$  步到达  $k$  号节点, 中间是否已经合法的方案数。最后统计  $f_{i,|s|,i}$  的答案即可。复杂度  $O(n^3)$ 。发现环可以分为经过根节点和不经过根节点的两种。不经过根节点可以直接判断 (转移唯一), 经过根节点可以枚举起始节点和第一次转移到根节点前走的距离, 预处理出  $f_{i,x}$  表示从根出发走  $i$  步到达  $x$  的方案数。最后快速算出第一次转移到根节点前的方案数, 因为自动机一直不经过根节点前走的路径是唯一的。使用多项式求逆好像可以做到  $O(m \log m + n)$ 。