

# 字符串选讲

张尊喆

# 咒语

## 倒序考虑

- 如果在复制后位置的前面，则此次操作不起作用
- 如果在里面，则变换到对应位置
- 如果在后面，减去复制的长度即可

# 博得

考虑如何判断两个串相等？

- 记录每个字符和这个字符上次出现位置的距离，变成新的数组
- 两个字符串相等当且仅当新的数组相同
- 利用此性质做 KMP

# 能量水晶

考虑一段区间可以开  $k$  次根号当且仅当区间乘起来以后每个质因子的幂次都是  $k$  的倍数。

维护一个串，第  $i$  为表示第  $i$  个质数的幂次模  $k$  的余数，当这个串全 0 时即为可以开  $k$  次根号。

考虑哈希维护这个串，用 map 存一下左端点，在右端点处询问即可。

注意用自然溢出而不是普通哈希。

# 字符串谷

考虑题目所求为两者权值之差，考虑令串  $s_i$  的权值为  $v_i + \frac{1}{2} \sum_j LCP(s_i, s_j)$

那么如果两个串都被一个人选中，LCP 恰好被加了两遍

如果分别被两个人选中，一相减恰好消除。

所以直接对这个权值排序，从大到小以此选择即可。

## P2375 [NOI2014] 动物园

给出一个字符串  $S$ , 求  $\sum (num_i + 1)$ ,  
对于字符串  $S$  的前  $i$  个字符构成的子串, 既是其前缀又是其后缀, 同时满足该前缀和后缀不重叠, 这种字符串的数量记作  $num_i$

正常做一遍 KMP，再做一遍求答案，这一次维护指针  $j$ ，当  $j$  大于  $i$  的一半时额外跳一次 next。

## P3193 [HNOI2008] GT考试

阿申准备报名参加 GT 考试，准考证号为  $N$  位数  $X_1, X_2 \dots X_n$  ( $0 \leq X_i \leq 9$ )，他不希望准考证号上出现不吉利的数字。

他的不吉利数字  $A_1, A_2, \dots, A_m$  ( $0 \leq A_i \leq 9$ ) 有  $M$  位，不出现是指  $X_1, X_2 \dots X_n$  中没有恰好一段等于  $A_1, A_2, \dots, A_m$ ， $A_1$  和  $X_1$  可以为 0。

对于全部数据， $N \leq 10^9$ ， $M \leq 20$ ， $K \leq 1000$ 。



设  $dp[i][j]$  表示前  $i$  为数, 匹配了  $A$  前  $j$  位的方案数。

枚举下一位填什么,  $dp[i][ch[j][c]] + = dp[i - 1][j]$ ,  $ch[j][c]$  表示当前匹配  $A$  了前  $j$  位, 在加一个字符  $c$  会匹配多少位。

发现可以矩阵优化, 时间复杂度  $O(M^3 \log N)$

## P4696 [CEOI2011]Matching

对于整数序列  $(a_1, a_2, \dots, a_n)$  和  $1 \sim n$  的排列  $(p_1, p_2, \dots, p_n)$ , 称  $(a_1, a_2, \dots, a_n)$  符合  $(p_1, p_2, \dots, p_n)$ , 当且仅当:

- $\{a\}$  中任意两个数字互不相同;
- 将  $(a_1, a_2, \dots, a_n)$  从小到大排序后, 将会得到  $(a_{p_1}, a_{p_2}, \dots, a_{p_n})$ 。

现在给出  $1 \sim n$  的排列  $\{p\}$  和序列  $h_1, h_2, \dots, h_m$ , 请你求出哪些  $\{h\}$  的子串符合排列  $\{p\}$ 。

对于 100% 的数据, 有  $2 \leq n \leq m \leq 1\,000\,000$ ;  $1 \leq h_i \leq 10^9$ ;  $1 \leq p_i \leq n$ , 保证  $\{h\}$  中的元素互不相同, 且  $\{p\}$  是一个排列。

考虑 KMP 和 AC 自动机本质上需要满足什么条件？

如果  $s[1 \dots n] = t[1 \dots n]$  那么  $s[l \dots r] = t[l \dots r]$ 。

只要快速判断两个字符相等即可，但在这题中如果跳了 `next` 字符串竟是变化的！我们满足另一个条件也可以判断：直接判断两个字符串完全相等即可。

在这题当中，首先将  $a$  由记录第  $k$  大对应什么位置变成第  $k$  个位置是第几大。

题目条件等价于将  $h$  的子串  $s$  离散化之后和  $a$  完全相同。

例如  $s = \{21, 45, 23, 7, 8, 4, 43\} = \{4, 7, 5, 2, 3, 1, 6\}$  考虑匹配跳  $\text{nxt}$  的时候，这个排列是变了的，但是如果最后一个字符在之前的串中的排名相同，那么这个字符就是可以加入末尾的，即使其他数的排名变化了也是依然是相等的。

所以现在我们只需要快速判断一个数的排名，可以先预处理出  $p$  这个排列中每个数的前驱和后继，注意这里的前驱（后继）是位置在自己**前面**的前驱和后继。

那么在  $h$  这个串上匹配的时候，只用查看对应  $p$  上的前驱后继是不是真正的前驱后继即可。

## P5537 【XR-3】 系统设计

给定一棵  $n$  个点的有根树和一个长度为  $m$  的序列  $a$ ，接下来需要实现  $q$  个操作。

操作分两种：

1.  $1 \ x \ l \ r$  表示设定起点为有根树的节点  $x$ ，接下来依次遍历  $l \sim r$ 。当遍历到  $i$  时，从当前节点走向它的编号第  $a_i$  小的儿子。如果某一时刻当前节点的儿子个数小于  $a_i$ ，或者已经遍历完  $l \sim r$ ，则在这个点停住，并输出这个点的编号，同时停止遍历。
2.  $2 \ t \ k$  表示将序列中第  $t$  个数  $a_t$  修改为  $k$ 。
  - $1 \leq n, m, q \leq 5 \times 10^5$ 。
  - $1 \leq a_i \leq n$ 。

Hint: 如果每次起点为根节点怎么做

考虑到每个点从根节点到其的路径是唯一的, 设为  $p_x$ , 可以将操作 1 转化为从根节点开始, 在操作区间前加入一段序列。

考虑中间经过的某个节点  $x$ , 那么  $p_x$  一定为操作区间的一段前缀, 二分 + 哈希即可。

加速: 用线段树维护并在线段树上二分, 时间复杂度  $O(q \log n)$ .

# Manacher 算法

维护每个位置的回文半径



## [HDU5371] Hotaru's problem

给定一个长度为  $n$  的数列  $a[i]$ , 现在要求你找出最长的形似  $s + s^R + s$  的子串,  $s^R$  是  $s$  的反串。

$$N \leq 10^5, 1 \leq a[i] \leq 10^9$$

回文半径，考虑一个  $ss^R s$  有两个回文串构成，对应两个回文半径  $s_1, s_2$ ，保证  $s_1 + r_1 \geq s_2, s_2 - r_2 \leq s_1$  即可。

所以枚举  $s_1$ ，求  $[s_1 + 1, s_1 + r_1]$  这个区间  $s_2 - r_2$  的最小值即可。

## P6216 回文匹配

对于一对字符串  $(s_1, s_2)$ , 若  $s_1$  的长度为奇数的子串  $(l, r)$  满足  $(l, r)$  是回文的, 那么  $s_1$  的“分数”会增加  $s_2$  在  $(l, r)$  中出现的次数。

现在给出一对  $(s_1, s_2)$ , 请计算出  $s_1$  的“分数”。

- 对于 100% 的数据:  $1 \leq n, m \leq 3 \times 10^6$ , 字符串中的字符都是小写字母。

考虑求出回文半径，用 KMP 记录每个位置是否匹配。

统计答案发现是等差数列，利用记录前缀和  $\sum a_i \times i$  和  $\sum a_i$  即可。

## P5446 [THUPC2018] 绿绿和串串

定义一种字符串操作翻转，即将字符串  $S$  的前  $|S| - 1$  位复制一份，并翻转，然后放到字符串的末尾，如 `abc` 翻转后得到 `abcba`。

现在给定一个字符串  $S$ ，问有多少种长度的初始字符串  $R$ ，使得  $S$  是  $R$  翻转若干次后的前缀，显然，当  $|R| > |S|$  时一定存在答案，所以只要求出  $|R| \leq |S|$  的方案。

从只翻转一次入手，发现满足  $i + r[i] = n$  的位置满足条件。

那么更多次是类似的，当前仅当  $i + r[i]$  满足条件且  $r[i]$  能覆盖 1 到  $i$

Trie 树的基本操作：

插入，查询

01-Trie 的基本操作：

异或最大

## P3065 [USACO12DEC] First! G

给定  $n$  ( $1 \leq n \leq 3 \times 10^4$ ) 个总长不超过  $m$  ( $1 \leq m \leq 3 \times 10^5$ ) 的互不相同的字符串，现在你可以任意指定字符之间的大小关系。问有多少个串可能成为字典序最小的串，并输出这些串。



字典序相关很容易想到 Trie 树

考虑一个字符串什么时候最小，在 Trie 树上每次都走的最左边！

可以得到当前字母是所有出边中最小的，当前字母和其他字母连有向边表示字典关系，最后拓扑排序即可。

时间复杂度  $O(26m)$

## CF178F3 Representative Sampling

有 $n$ 个字符串 $a_i$ , 从中选取 $k$ 个, 使它们两两之间的 $LCP$  (最长公共前缀) 之和最大。输出这个最大值。

$$k \leq n \leq 2000, |a_i| \leq 500$$

不难想到建 Trie 树, 然后考虑 DP

设  $f[x][y]$  表示  $x$  的子树中, 选了  $y$  个字符串的最大价值。

树上背包合并, 技巧:  $y$  这一维枚举到  $siz[x]$  大小, 时间复杂度  $O(n|a|k)$

注意到如果一个点只有一个儿子可以直接继承, 无需 DP, 所以只有  $O(n)$  个转移点, 时间复杂度  $O(nk)$

## P4551 最长异或路径

给定一棵  $n$  个点的带权树，结点下标从 1 开始到  $n$ 。寻找树中找两个结点，求最长的异或路径。

异或路径指的是指两个结点之间唯一路径上的所有边权的异或。

$1 \leq n \leq 100000; 0 < u, v \leq n; 0 \leq w < 2^{31}$ 。

求树上前缀异或和，一条异或路径为两前缀异或和的异或。

01-Trie 查询最大异或即可

## P5283 [十二省联考 2019] 异或粽子

给定一个长度为  $n$  的数组  $a_i$ ，你每次可以选择一个区间  $[l, r]$ ，其价值为  $val[l, r] = \oplus_{i=l}^r a_i$ 。求价值最大的  $k$  个不同区间的价值和。

对于所有的输入数据都满足： $1 \leq n \leq 5 \times 10^5, 1 \leq k \leq \min \left\{ \frac{n(n-1)}{2}, 2 \times 10^5 \right\}$ ， $0 \leq a_i \leq 4294967295$ 。

超级钢琴做法，查询一个区间异或最大值，Trie 树即可。

## P6072 『MdOI R1』 Path

给定一棵  $n$  个点的无根树，边有边权。你要选择两条简单路径，满足没有重合的点，且边权异或和之和最大。

对于 100% 的数据， $2 \leq n \leq 3 \times 10^4$ ， $1 \leq x, y \leq n$ ， $0 \leq w \leq 10^9$ 。



枚举一个分界点，那么就是在子树内找一条，子树外找一条路径。

子树内：启发式合并即可。

子树外：不好求，考虑整棵树最大的一条路径  $(u, v)$ ，如果分界点不在  $u, v$  到根的路径上，那么子树外就是这条路径的权值。否则暴力处理即可。

时间复杂度  $O(n \log n \log V)$

## P5410 【模板】扩展 KMP (Z 函数)

给定两个字符串  $a, b$ , 你要求出两个数组:

- $b$  的  $z$  函数数组  $z$ , 即  $b$  与  $b$  的每一个后缀的 LCP 长度。
- $b$  与  $a$  的每一个后缀的 LCP 长度数组  $p$ 。

对于 100% 的数据,  $1 \leq |a|, |b| \leq 2 \times 10^7$ , 所有字符均为小写字母。

## P7114 [NOIP2020] 字符串匹配

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。

对于所有测试点, 保证  $1 \leq T \leq 5$ ,  $1 \leq |S| \leq 2^{20}$ 。

容易想到枚举 AB 的长度  $L$ ，那么  $C$  就确定了，利用 border 判断 AB 是否是  $S[1 : iL]$  的循环节。

容易确定  $F(C)$ ，那么计算有多少  $k$  使得  $F(S[1 : k]) \leq F(C)$  即可。

时间复杂度  $O(n \log n + 26n)$

加速枚举，通过 exkmp 直接得到每种长度最多循环多少次  
剩下的部分为  $C$ ，那么  $F(C)$  只有两种可能取值，和循环次数有关。  
直接统计答案即可，时间复杂度  $O(26n)$  或  $O(n \log 26)$

加速统计答案，发现每次询问两种  $F(C)$ ，进一步发现  $F(C)$  要么是  $F(S)$ ，要么是  $F(S - (AB))$ 。

所以每次询问的 F 值最多和上次相差 1，拿个指针维护一下即可。

时间复杂度  $O(n)$

# 无题

给定两个字符串  $S(1 \leq |S| \leq 10^5)$  和  $T(1 \leq |T| \leq 10^5)$ , 每个字符串只包含 ' $A$ ', ' $B$ ', ' $C$ '.

有四种变换  $A \rightarrow BC, B \rightarrow AC, C \rightarrow AB, AAA \rightarrow \emptyset$

$10^5$  次询问, 每次询问  $S[l_i, r_i]$  能否变成  $T[L_i, R_i]$

性质:  $B \rightarrow AC \rightarrow AAB \rightarrow AAAC \rightarrow C, C \rightarrow B$

'B' 和 'C' 等价

性质:  $AB \rightarrow AAC \rightarrow AAAB \rightarrow B, B \rightarrow AC \rightarrow AB$

可以忽略 B 之前的 A

性质:  $B \rightarrow AC \rightarrow BCC \rightarrow BBB, A \rightarrow BC \rightarrow BB$

'B' 的个数只能增加偶数个, 且无法减少。



$S_1$  变成  $S_2$

1. 末尾 A 的数量
2. B 的奇偶性
3. 讨论：B 的个数相同，AAA 消除，不同则一定可行

# 作业

- CF808G Anthem of Berland
- P4824 [USACO15FEB] Censoring S
- P3546 [POI2012] PRE-Prefixuffix
- P3435 [POI2006] OKR-Periods of Words
- P3426 [POI2005] SZA-Template
- P1659 [国家集训队] 拉拉队排练
- P4555 [国家集训队] 最长双回文串
- P4287 [SHOI2011] 双倍回文
- P4302 [SCOI2003] 字符串折叠