字符串与数论筛法

黄洛天

THU，IIS

April 27, 2025

- Border 理论
- 字符串杂题
- 筛子

P6292

https://www.luogu.com.cn/problem/P6292

Border 理论

Lemma

p 是周期当且仅当 $|w| - p$ 是 *border*。

Lemma

p, q 是周期，且 $p + q \leq |w| + \gcd(p, q)$，则 $\gcd(p, q)$ 为周期。

Lemma

对于任意 $(x, 2x]$，长度在此区间内的 *border* 构成了一个等差数列。

Lemma

一个字符串的 *border* 集合可以划分为 $O(\log n)$ 个等差数列。

基本子串字典

解决的问题是，q 次询问区间 border 信息。可以做到求出 $O(\log n)$ 个等差数列。

基本子串字典

令 $N_k(i)$ 为 SA 第 k 轮后，表示 $s[i:i+2^k-1]$ 的字符串的 rank。注意这里需要保证相同字符串的 rannk 相同。

Lemma

对于字符串 u, v，若 $2|u| \geq |v|$，则 u 在 v 中出现位置构成了等差数列。

Lemma

对于字符串 u, v，若 $2|u| \geq |v|$，且 u 出现了至少 3 次，则公差为 u 的周期。

基本子串字典

考虑计算长度在 $(2^k, 2^{k+1}]$ 内的 border。令 $u = s[l : l + 2^{k+1} - 1]$ $u_0 = s[l : l + 2^k - 1], v = s[r - 2^{k+1} + 1 : r], v_0 = s[r - 2^k + 1 : r]$。令 S_1 表示 v_0 在 u 里的出现位置，S_2 表示 u_0 在 v_1 里的出现位置。注意到每个 border 都对应了 S_1 和 S_2 里的一个元素。

基本子串字典

因此我们需要快速对 S_1 和 S_2 求交。注意到若 $|S_1| \le 2$ 或 $|S_2| \le 2$ 是平凡的，否则我们有 S_1 的公差为 v_0 的周期，S_2 的公差为 u_0 的周期。

Lemma

若 $|S_1| > 2, |S_2| > 2$，则 u_0 的周期等于 v_0 的周期。

于是我们可以 $O(1)$ 计算 S_1 和 S_2 的交。

基本子串字典

求 S_1 和 S_2 只需要在 $N_k(l) \sim N_k(l+2^k)$ 中求出 $N_k(r-2^k+1)$ 第一次出现，最后一次出现，和出现次数即可。三个信息均可以离线扫描线，对于每个 N_k 的值维护一个 deque 做到。

求 N_k 采用后缀数组的做法，总时间复杂度 $O((n+q)\log n)$.

P8006

https://www.luogu.com.cn/problem/P8006

基本字串结构

定义 $occ(u)$ 表示 u 在 w 中出现次数。

定义 $ext(u)$ 表示 u 向两侧扩展，最长的一个字串使得其出现次数和 u 相等。

Lemma

把 $ext(u)$ 相等的字串称为一个等价类，则每个等价类画在二维平面上是阶梯状。

基本字串结构

Lemma

一块的一行为一个正串 *sam* 节点，一列为反串 *sam* 节点。

Lemma

所有本质不同的块的周长之和为 $O(n)$.

基本字串结构

构造方法：先求出正串 sam，若正串 sam 的一个节点只有一条出边，说明其不是某个块的最顶上一行，于是把他和他指向的节点合并。这样就能求出每块对应的正串 sam 节点编号集合，类似的可以求出反串 sam 节点编号集合。

loj 3723

https://loj.ac/p/3723

uoj 697

https://uoj.ac/problem/697

uoj 752

https://uoj.ac/problem/752

loj6222

求 $1 \sim n$ 内有多少个 powerful number，以及他们的和。
powerful number 是指每个质因子次幂至少为 2 的数字。
$n \leq 10^{25}, TL = 2S, ML = 1G$

loj6222 - solution

每个 powerful number 都可以表示成 a^2b^3 的形式，如果要求 b square-free，那么分解方式唯一。

因此可以枚举 b ，个数即为 $\sum_{i=1}^{n}\mu^2(i)\sqrt{\lfloor\frac{n}{i^3}\rfloor}$。因为 $\sqrt{\lfloor\frac{n}{i^3}\rfloor}$ 只有 $O(n^{1/5})$ 种，可以直接整除分块，然后转化为计算 $\sum_{i=1}^{n}\mu^2(i)$。

有经典公式 $\mu^2(x)=\sum_{i^2|x}\mu(i)$，上式可以变成 $\sum_{i=1}^{\sqrt{n}}\mu(i)\lfloor\frac{n}{i^2}\rfloor$，$\lfloor\frac{n}{i^2}\rfloor$ 只有 $O(n^{1/3})$ 种，继续整除分块，线性筛 $\mu(d)$ 前缀和。

loj6222 - solution

求和需要算

$$\sum_{i=1}^{n} \mu^2(i) i^3$$

$$\sum_{i=1}^{n} i^3 \sum_{j^2 | i} \mu(j)$$

$$\sum_{j=1}^{\sqrt{n}} \mu(j) j^6 \sum_{i=1}^{\frac{n}{j^2}} i^3$$

。

预处理出 $\mu(i) i^6$ 前缀和，每次 $O(1)$ 计算立方前缀和，积分一下时间复杂度就是 $O(n^{4/15})$。

uoj 188

https://uoj.ac/problem/188

loj6686

https://loj.ac/p/6686