《Humongous String》解题报告

1 题目大意

对于字符集 $\Sigma=\{s_0,s_1,\ldots,s_{k-1}\}$，定义字符串序列 $T_i=\begin{cases} s_0 & i=0 \\ T_{i-1}s_{i \pmod k} & i>0 \end{cases}$，令 $S=T_0T_1T_2\ldots$，求 S 的长度为 n 的前缀有多少个本质不同的非空子串。T 组询问，每次询问给出 n,k。

2 数据范围

- $1 \le T \le 10^5$；
- $1 \le n_i \le 10^9, 1 \le k_i \le 10^9$。

3 解题过程

先特判掉 $k=1$。

考虑在 $S_i \not\equiv (S_{i-1}+1) \bmod k$ 的位置画上一个分隔符。那么 S 被分成若干块 $S=T_0|T_1|\cdots|T_{k-1}|T_kT_{k+1}|T_{k+2}\ldots$。用 B_i 表示第 i 个块。称形如 $T_{pk}T_{pk+1}$ 的块为大块，其余为小块。

有结论：对于 S 的子串 t，若其跨过了两个分隔符，那么串 t 只在该位置出现一次。证明考虑，假设块 u 是 t 的子串，可以根据分隔符的位置计算出 u 的末尾字符和 u 的长度，这唯一对应一个 B_i。

考虑计数，对于每个块 X 记 $cnt(X,i)$，表示有多少个以 X_i 结尾的子串在之前已经出现过。那么答案可以表示成 $\binom{n}{2}-\sum\limits_i\sum\limits_j cnt(B_i,j)$。

考虑如何计算 $cnt(X,i)$，记 X 中第一个 T_i 的长度为 b。那么之前已经出现过的子串是 i 的一个后缀。这里讨论 $k\ge 3$ 的一般情况，对于 $k=2$ 的情况由于 S 全部由大块构成，要特别讨论，但也是类似的。

- $b<k$

  当 $i<b$ 时，长度为 i 的后缀在上一个块完整出现过，答案为 i。当 $i=b$ 时，X_b 第一次出现，答案为 0。因此 $cnt(X,i)=\begin{cases} i & i<b \\ 0 & \text{else} \end{cases}$。

- $b=k$

  和 $b<k$ 类似，$cnt(X,i)=i$。

- $b \bmod k \neq 0$

  该块为小块。对于 $i \leq b-k$，$X_{1\sim i}$ 在 $b' = b-k$ 时出现过，同时可以再加上前一个块在上次出现过的后缀，即 $b' = b-k-1$ 的整个串。否则，当后缀不跨过分隔符时，子串在上一个大块完整出过。要讨论前一个是大块还是小块。因此：

$$cnt(X,i) = \begin{cases} i + (2b-3-2k) & i \leq b-k, b \bmod k = 2 \\ i + (b-1-k) & i \leq b-k, b \bmod k > 2 \\ i & i > b-k \end{cases}$$

- $b \bmod k = 0$

  该块为大块。和上面类似，但是当 $i > b+k-k-1$ 时，在上一个大块出现过的最长后缀为 $X_{k+1\sim i}$，因此 $cnt(X,i) = \begin{cases} i+b-k-1 & i \leq 2b+1-2k \\ i-k & \text{else} \end{cases}$。

考虑如何解决询问。n 可以分成前缀若干个块和一个不完整块的前缀两部分构成。一个完整块的 $\sum_i cnt(X,i)$ 可以用一个和 b 有关的二次多项式表示。那么分讨 $b \bmod k$ 的不同情况，经过一些处理后，前缀若干个完整块的贡献可以变成 $O(1)$ 个二次多项式 ax^2+bx+c 在 x 取一个前缀的取值的和。拆开每个项，正整数幂和可以求出。可以在 $O(1)$ 复杂度解决每次询问。

4 参考资料

1. 原题链接：`https://qoj.ac/contest/2103/problem/12329`。
2. 原题题解：`https://qoj.ac/download.php?type=attachments&id=2103&r=1`。