

以 AGTCC 为例

步骤 1. 首先将字符串复制一遍生成长度为  $2n$  的字符串  $S = \text{AGTCCAGTCC}$

步骤 2. 建立一个长为  $n$  的 `int` 数组  $A$ , 使得  $A[x] = x$

步骤 3. 对  $A$  数组做快速排序

(排序的比较函数是对当前比较值  $x, y$ , 比较  $S.\text{substr}(x, n)$  和  $S.\text{substr}(y, n)$  的字典序)

这样做其实是对源字符串移位了  $x$  位以后生成的字符串与移位  $y$  位以后生成的字符串做了比较, 而且不需要存储矩阵来做, 只需要一个长度为  $2n$  的字符串即可。

最后  $A$  数组生成的排序就是我们 FM-index 需要的后缀数组。

步骤 4. 通过后缀数组可以生成最后一列的字符串。同时也可以生成  $C$  和  $OCC$  数组

最终结果导致每次索引能做更长的字符串, 所以总索引次数减小, 查询时间就会更快。