

后缀自动机的基础应用

ZYQN

成都石室中学

zyqn1227@gmail.com

2017 年 3 月 24 日

1 基础知识

- 字符串匹配
- 子串定位
- 广义后缀自动机的复杂度

2 经典例题

- APIO2014 - 回文串
- AHOI2013 - 差异
- 日常考试 - 加密
- Feyat cup1.5 - Str
- 日常考试 - String

3 参考资料

字符串匹配

有一些问题中，我们需要进行模式串匹配
我们把用文本串在 SAM 上走的时候，需要维护当前匹配的长度
为了明确问题我们来看一道例题

题目大意：

给定不超过 10 个字符串，每个字符串长度不超过 10^5
请你求出这些串的最长公共子串

解决方案是，我们随便拿一个串 A 出来建 SAM
然后拿剩余的串到这个 SAM 上去匹配，这是时候我们就需要时刻维护匹配的长度 L

显然走转移边会使 $L + 1$ ，但没有转移边跳 Fail 链的时候呢？

我们给定一个结论，如果跳 Fail 跳到结点 w 的时候第一次有了转移边
那么当前匹配的长度为 $\text{Max}(w)+1$

现在我们来证明这个结论：

对于串 A 来讲，这个长度为 $\text{Max}(w)+1$ 的子串肯定存在，且是当前最大的子串

对于文本串 B 来讲，因为是从结点 p 跳 Fail 链跳到 w 的，所以 w 代表的串是 p 代表的串的后缀

那么 p 代表的串在 B 中都出现了， w 代表的串也一定存在

又因为这个跳 Fail 链本身是一个贪心的过程，所以 $\text{Max}(w)+1$ 已经是最优解

子串定位

在有一些题目中，我们需要快速找到一个子串在后缀自动机上对应的结点

这个是可以做到 $\log n$ 的

具体做法是在 Fail 树上倍增

广义后缀自动机的复杂度

广义后缀自动机相信大家都会了
但建立的复杂度是一个玄学

这里给出结论：

在线的 DFS 建立的复杂度是 $O(\sum \text{叶子的深度})$

离线的 BFS 建立的复杂度是 $O(\text{最大的叶子深度} \cdot \text{字符集大小})$

详细的证明可以参考 2015 年的国家集训队论文

给定长度为 n ($n \leq 300000$) 的字符串 s

定义 s 中一个子串 t 的出现值为：出现次数 $\cdot t$ 的长度

问 s 的所有回文子串中，出现值最大是多少

给定长度为 n ($n \leq 300000$) 的字符串 s

定义 s 中一个子串 t 的出现值为：出现次数 $\cdot t$ 的长度

问 s 的所有回文子串中，出现值最大是多少

我们可以用回文树来做

但用 SAM 加上 manacher 应该是更常见的写法吧？

给定长度为 n ($n \leq 500000$) 的字符串 s , 定义 T_i 为从 i 开始的后缀
 $\text{len}(x)$ 表示字符串 x 的长度, $\text{lcp}(x,y)$ 表示字符串 x, y 的最长公共前缀
的长度

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

给定长度为 n ($n \leq 500000$) 的字符串 s , 定义 T_i 为从 i 开始的后缀
 $\text{len}(x)$ 表示字符串 x 的长度, $\text{lcp}(x,y)$ 表示字符串 x, y 的最长公共前缀
的长度

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

我们枚举 $\text{lcp}(x,y)$, 于是这题变成了 Fail 树上的树上 DP 问题

现在有一个包含 $n(n \leq 10^5)$ 个小写字母的字符串 S ，其下标从 0 开始。我们用下列方式得到它对应的密码：

1. 光标初始位于下标 0 处
2. 当光标位于 i 时，找到一个最长的子串 $S[j, j+len-1]$ ，满足 $j < i$ ，使得 $S[i, i+len-1] = S[j, j+len-1]$ 。然后在密文中追加 len 、 j 两个整数，并将光标往后移动 len 个位置。若对于最大的 len ，有多个合法的 j 与之对应，我们会选择最小者
3. 若不存在合法的 len 、 j ，则在密文中依次追加 -1 和 $S[i]$ 对应的 ASCII 码，并将光标往后移动一位
4. 当光标位于下标 n 时，加密完成，最终得到的密文会是一行偶数个用空格隔开的整数。

现给定某个字符串 S ，请你输出对应的密文。

这个东西我们可以在 Fail 树上 $O(n)$ 染色
具体的方法可以参考我博客的题解

给定两个长度分别为 $n, m (n, m \leq 10^5)$ 的小写字母字符串
定义两个字符串匹配为它们至多有一位不同
求这两个字符串的最长公共子串

给定两个长度分别为 $n, m (n, m \leq 10^5)$ 的小写字母字符串
定义两个字符串匹配为它们至多有一位不同
求这两个字符串的最长公共子串

这是国家集训队论文里的例题
复杂度可以做到 $O(n \log n)$

先给定一个长度为 $n(n \leq 50000)$ 的字符串 S

之后给定 $m(m \leq 50000)$ 个操作 (强制在线), 操作分两种

1. 在字符串 S 末尾加入一个字符
2. 将 S 的子串 $[l, r]$ 取出来, 设为字符串 T , 询问 T 中出现至少两次的字符串的最长长度

先给定一个长度为 n ($n \leq 50000$) 的字符串 S

之后给定 m ($m \leq 50000$) 个操作 (强制在线), 操作分两种

1. 在字符串 S 末尾加入一个字符
2. 将 S 的子串 $[l, r]$ 取出来, 设为字符串 T , 询问 T 中出现至少两次的字符串的最长长度

我们需要配合 LCT、函数式线段树来解决该问题

具体可以参考我博客上的解题报告

- [1] 刘研绎《后缀自动机在字典树上的拓展》
- [2] 张天扬《后缀自动机及其应用》
- [3] 陈立杰《后缀自动机》