

字符串问题 解题报告

清华大学 计算机科学与技术系

王聿中

清华大学 计算机类

吕时清

清华大学 交叉信息研究院

杨景钦

题目大意

- 给定一个字符串。从中取出若干个（可能相交的）字串，其中一些为 A 串，另一些为 B 串。
- 给定若干 AB 串之间的边。
- 规定，A 串 A_i 到 A_j 之间有单向边当且仅当存在一个 B 串，其为 A_j 的前缀，且与 A_i 有边相连。
- 求最长链。
- 所有值不超过 2×10^5

算法一

- 暴力建图跑拓扑排序
- 借助哈希可以加速

- 期望得分：10~40

算法二

- 针对测试点 4
- 由于 A 串唯一, 因此:
 - 要么答案为 -1 (有自环)
 - 要么答案为该 A 串的长度 (无自环)
- 借助哈希判断即可

- 时间复杂度: $O(n)$
- 期望得分: 10
- 结合其他可以获得更高的分数

算法三

- 对于测试点 1、4、5、6
- 我们发现所有 A 串前缀总数在一个可以接受的量级
- 将它们全部列举出来建点, 所有 B 串连向与其相等的前缀
- 每个 A 串拆出来的前缀再连向这个 A 串
- 跑拓扑排序即可

- 时间复杂度: $O(nm^2)$
- 期望得分: 40
- 结合算法二的期望得分: 60

算法四

- 考虑前 8 个测试点
- 考虑对大串建后缀数组
- 每个 A 串都是一个后缀的前缀
- 由于 B 串长度均小于 A 串长度, 因此一个 B 串能到达的一定是后缀数组上一段连续区间
 - 包含该 B 串所在位置的, height 最小值不小于该 B 串长度的极长区间
- 直接连边无法承受, 线段树优化建图即可
- 时间复杂度: $O(n \log n)$
- 期望得分: 80

算法五

- 对于一般数据，算法四的局限性在于，可能会有更长 B 串连到更短 A 串的情况（如果该 A 串在大串中后面紧接着的部分与 B 串后半段匹配的话）。
- 样例 1 的第 3 组数据即是一个例子。
- 我们只需求出每个 A 串在后缀数组中所在真实位置即可。
 - 换言之，就是把所有 A 串排序
- 按 height 数组从大到小合并，用并查集+链表维护即可。
- （可持久化线段树也是可以的）
- 时间复杂度： $O(n \log n)$
- 期望得分：100

算法六

- 事实上，除支配关系对应的边外，只需对于每个 A 串和 B 串（记为 s ），从是 s 前缀的最长 B 串连一条边到 s 。
- 对于大串求出后缀数组和 height 数组后，对于 $1 \leq i \leq n$ ：
 - 用单调栈维护所有仍然是 $S[SA[i] \dots n]$ 的前缀的 B 串序列；
 - 对于所有左端点为 $SA[i]$ 的 A 串和 B 串，在单调栈上二分找到长度不大于它的最长 B 串并连边。
- 存在相同的 B 串的情况需要细心处理，例如以 B 串编号作为第二关键字排序。
- 时间复杂度： $O(n \log n)$
- 期望得分：100

算法七

- 考虑将串翻转，则前缀限制转化为后缀限制。
- 即对于一个 B 串，它能够到达的 A 串为包含该 B 串作为后缀的所有 A 串。
- 考虑后缀自动机，我们在 SAM 上找到所有 A、B 串对应的节点，则每个 B 串能够走下去的节点就是它 parent 树上子树内的节点。
- 找点需要在 SAM 上倍增。
- 拆点建图即可。

- 时间复杂度： $O(n \log n)$
- 期望得分：100

花絮

写不动代码的
老年人

我本来以为是我的算法太麻烦了

Ctrl+Z

看了眼验题人也写了快 300 行

Ctrl+Z



你这题是d几t几啊...

d1t2

Ctrl+Z



惊了

蛮狠的

Ctrl+Z

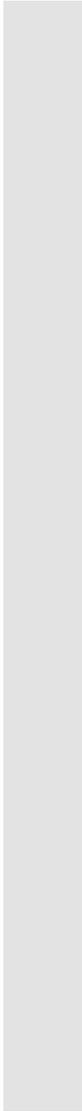
SAM 也有 200+ 行

Ctrl+Z



这好自闭啊

QA
&
Talks



FIN

Questions are welcome.