

信息学竞赛 字符串与哈希

主讲：李宁远



专注于拔尖创新人才培养

1 Hash

- 字符串哈希
- 树哈希
- 集合哈希
- Hash Table

2 线性结构字符串算法

- KMP
- Z Algorithm (exKMP)
- Manacher

3 树形结构字符串算法

- Trie
- AC 自动机
- End



字符串哈希

oi-wiki

卡固定模数哈希

改版好写的 AHU (?)

目标是判断两个有根树是否相同。尝试把每个出现的有根树编号成一个正整数。递归进行，当树是一个点时确定它的编号是 1，其孩子序列是空。

先计算出其每个子树的值，随后将其排序，这样就得到了一个它的孩子序列。使用一个 `map<vector, int>` 来映射孩子序列到编号。直接在其中查找，如果存在就得到了编号，不存在就令当前有 `cnt` 个不同子树，然后把它编号成 `cnt + 1`。

正确性是有保证的，如果小编号的有根树满足相同则编号相同，不同则编号不同，那么更大的有根树也会满足。相同的话排序后的孩子序列一定一样，不同的话就一定不一样。（从 uoj 树哈希模板代码看来的做法，感觉很厉害）

类字符串 hash

对集合 S , $h(S) = \sum_{i \in S} b^i \bmod p$.

异或哈希

整一个非负整数到非负整数的随机映射 f 。
对集合 S , $h(S) = \bigoplus_{i \in S} f(i)$ 。



例题

<https://www.luogu.com.cn/problem/P3792>

解法

不正确的哈希：维护区间 \min 、 \max 、 sum 、平方和。可以被卡掉：<https://www.luogu.com.cn/discuss/901587>

正确做法（换模数、扩大范围就无法轻易被卡）：异或哈希（保证值域于是离散化、维护区间最小值以确定值域区间，可以快速计算异或和）、多项式 hash：

<https://www.luogu.com.cn/article/c9en8yp9>

Hash Table

将要存的东西映射到一个小整数上，存在桶里。
会有冲突，需要解决。

OI-Wiki

当存入哈希表的数字是近似随机时，可以直接用二进制的后几位作为哈希值。当没有特别随机的时候，可以用前面的位异或上后面的位做哈希值。（例：买瓜）

Hash



KMP

线性结构字符串算法



树形结构字符串算法



KMP

OI-Wiki
也有哈希暴力。



Z Algorithm

OI-Wiki
也有哈希暴力。

Hash



线性结构字符串算法



树形结构字符串算法



Manacher

Manacher

OI-Wiki

例题

<https://codeforces.com/gym/105201/F>

做法

考虑 Manacher 的性质，可以发现只有暴力扩展时加边就是正确的。

Trie

OI-Wiki

例题 1

<https://www.luogu.com.cn/problem/P2922>

例题 2

<https://www.luogu.com.cn/problem/P11872>

AC 自动机

OI-Wiki

我们实际上希望构建一个结构，使得将一个串 s 扔到一个由一堆串组成的 Trie 上时，可以动态的得到 s 当前前缀中，作为该前缀的后缀出现过的，Trie 中的前缀的位置。这和 border 很类似。

构建 AC 自动机时，计算某前缀的“类 border”位置（即 *fail* 指针）时，不能暴力跳上一个的 *fail* 指针，需要额外计算一个对任意串，加某个字符时会 *fail* 到哪里停止的值。

kmp 实际上就是一个串上的 ACAM。

例题

<https://www.luogu.com.cn/problem/P5357>

解法

S 在 ACAM 上游走，对每个前缀，其相当于对每个 fail 树上到根的链有一次贡献。游走完后 dfs 一遍 fail 树即可。

Thank You

