

字符串匹配算法（下）

字符串匹配算法（下）

多模匹配问题

AC 自动机的思想

预习资料：【数据结构】> 字符串统计

多模匹配问题

1. 有多个模式串的匹配问题，就是多模匹配问题
2. Step1: 多个模式串，建立成一棵字典树
3. Step2: 和文本串的每一位对齐匹配，模拟暴力匹配算法的过程

AC 自动机的思想

1. 当匹配成功文本串中的 she 时，也就意味着后续一定会匹配成功 he
2. she 对应了字典树中的节点 P，he 对应了字典树中的节点 Q
3. P 和 Q 就是等价匹配节点，如果从 P 引出一条边指向 Q，就可以加速匹配过程
4. 在 P 下面查找节点的操作，等价于在 Q 下面查找节点的操作
5. 这条等价关系边，通常在 AC 自动机上叫做【Fail 指针】
6. AC 自动机 = Trie + Fail 指针
7. 子节点的 Fail 指针是需要参照父节点的 Fail 指针信息的，最简单的建立方式，就是采用【层序遍历】
8. 没做优化的 AC 自动机，本质上是一个 NFA（非确定型有穷状态自动机）
9. 通俗理解：根据当前状态 p，以及输入字符 c，无法通过一步操作确定状态
10. 第二种理解：当前状态，并不代表唯一状态。

AC 自动机优化：使用路径压缩思想，使状态转移时可以一步跳转到目标状态。

优化以后的 AC 自动机，更像 DFA（确定性有穷状态自动机）。

预习资料：【数据结构】> 字符串统计

1. AC 自动机裸题
2. 解题的关键，在于如何维护每一个单词的计数量
3. 使用幼儿园必知必会的指针技巧维护的
4. **船长翻车**：题目数据中给出的单词会有重复