

AC自动机原理、优化、代码详解

前置知识

讲解044、讲解045 - 前缀树

讲解059 - 链式前向星建图、图的遍历，本节课题目1需要

讲解084、讲解085 - 数位dp，本节课题目2需要

讲解100、讲解101 - KMP算法

AC自动机原理讲解

AC自动机中对于fail指针的理解，涉及KMP算法

AC自动机中防止fail指针绕圈的优化，涉及三个场景

经过优化后

建立AC自动机 + 遍历文章，总的时间复杂度为 $O(\text{所有目标字符串的总字符数量} + \text{文章长度})$

这是讲解AC自动机最全面、最清晰的视频

AC自动机原理、优化、代码详解

自动机又叫确定有限状态自动机，是对信号序列进行判定的数学模型
比如，判定s1是否包含s2、判定s是否是回文，等等
自动机并不是具体的算法、数据结构，只是数学模型，更多是概念上的内容
每种自动机实现方式可能有多种

常见的自动机

前缀树，讲解044、讲解045

KMP自动机，理解KMP算法即可，讲解100、讲解101

回文自动机，理解Manacher算法即可，讲解103，下节课

后缀自动机(SAM)，理解后缀数组的用法即可，【挺难】课程里会安排讲述

AC自动机，本节课

AC自动机原理、优化、代码详解

AC自动机的功能

给你若干目标字符串，还有一篇文章，返回每个目标字符串在文章中出现了几次

AC自动机就是加上了fail指针的前缀树，经典的AC自动机，课上图解一下

fail指针含义：

AC自动机上的某个节点a，表示某个目标串的前缀串s的终止节点

所有目标字符串的前缀串，要求和s的某个后缀串完全一样，并且要求长度最大，并且不能是s的整体
满足这三点的前缀串假设为t，那么节点a的fail指针就指向t的终止节点

够绕的！意义何在？意义在于保留所有匹配成功的可能性！思想来自KMP算法，课上重点说明一下

一旦匹配失败，就通过fail指针绕圈找到能继续的节点，然后继续匹配

AC自动机原理、优化、代码详解

使用经典AC自动机，会出现fail指针绕圈的现象，具体有如下三个场景

1, 建立AC自动机时设置fail指针；遍历文章时，匹配失败去寻找支路。都需要fail指针绕圈

优化方式： 课上会重点图解，这是固定的优化，以后建立AC自动机都可以这么做

2, 遍历文章时，不知道是否命中了某个目标字符串，需要fail指针绕圈，进行词频传递

优化方式：只让当前节点收集词频，在遍历文章结束之后，再统一进行如下处理，防止绕圈

根据fail指针建立反图，然后利用图的遍历来汇总每个节点的词频，题目1的定制优化

3, 遍历文章时，不知道是否命中了某个目标字符串，需要fail指针绕圈，进行及时报警

优化方式：在设置fail指针时，把命中标记前移防止绕圈，题目2的定制优化

经过这样的优化，在遍历文章时，就不需要fail指针的跳转了，甚至可以忽略fail指针的存在
操作fail指针只需要发生在建立AC自动机时，或者文章遍历结束后的离线处理

AC自动机原理、优化、代码详解

题目1

AC自动机模版(威力加强版)

给你若干目标字符串，还有一篇文章

返回每个目标字符串在文章中出现了几次

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

经典AC自动机实现略过，因为fail的指针绕圈行为会让复杂度变高，经常无法通过测试
所以直接讲优化后的版本

本题需要理解

讲解059 - 链式前向星建图

本题防止fail指针绕圈的方式为：

遍历文章时，不进行fail指针绕圈，来到的节点收集词频即可

遍历文章后，根据fail指针建立反图(其实是一颗树)，用树的遍历来汇总每个节点的词频

AC自动机原理、优化、代码详解

题目2

数数(利用AC自动机检查命中)

我们称一个正整数 x 为幸运数字的条件为

x 的十进制中不包含数字串集合 s 中任意一个元素作为子串

例如 $s = \{ 22, 333, 0233 \}$

233是幸运数字, 2333、20233、3223不是幸运数字

给定 n 和 s , 计算不大于 n 的幸运数字的个数

答案对1000000007取模

测试链接 : <https://www.luogu.com.cn/problem/P3311>

本题需要理解

讲解084、讲解085 - 数位dp

防止fail指针绕圈的方式为: 设置fail指针时, 把命中单词的标识设置给fail指针上游的节点