

Music & Lyrics题解

雅礼中学 毛啸

2015 年 10 月 8 日

题意

给定 W 个字符串。

接着给定一个 N 个字符串，求出前面的 W 个字符串中的每个在所有 N 个字符串中的出现次数和。

W 不超过500， W 个字符串长度均不超过1000， N 不超过100， N 个字符串长度均不超过50000。

题解

一个简单的算法是对于每个句子的每个区间，看看是否有对应单词，如果有对应单词则更新对应单词的出现次数。可以用hash优化，我们枚举开始位置之后，对于每一个出现过的长度我们都用hash表算出他的hash值，看看是否等于某一个单词的hash值。hash函数如果定为字符串表示为 k 进制数模 p 的值，那么某一个区间的hash值就等于右端点hash值减去左端点hash值乘上 k 的区间长度次方的值，然而仍然难以通过本题。

此外我们也可以用KMP解决，KMP就是找出字符串每个前缀的border，也就是最长的字符串满足既是他的后缀又是他的前缀，这个东西考虑自己和自己匹配就很容易求得，匹配的时候如果失配就回溯到他的border即可再继续匹配。

我们注意到是border之间的关系类似父子关系，所以所有的前缀通过border关系可以形成一棵树，从这棵树可以引出border tree，虽然这个数据结构对解决本题没有太大作用，但是了解父子关系能形成树的事实对解决本题有很大作用。

此外我们还应该注意到，字符串匹配中任何时候只有当前匹配的长度是有用的，我们注意到，当前匹配长度确定且下一个字符确定的时候，接下来的匹配长度也是确定的，所以我们可以确定这一个转移，匹配长度和转移之间构成了一个确定性有限状态自动机。

然而上述算法很难通过本题，因为本题是多个字符串。

如果是多个字符串，那么border模型就不能用了，匹配长度也不能用了，因为我们必须存每一个字符串的border或匹配长度，这是不能忍受的。但是，上面介绍的确定性有限状态自动机却仍然可以用。

我们考虑将字符串全部插入一棵trie树。插入完毕后，我们BFS一遍这棵trie树，然后对于每个状态，求出一个fail值表示对于当前状态表示的字

符串，他的最长的一个在trie树中有对应节点的后缀所对应的节点，这个东西可以用类似KMP求border的方法解决。此外，我们可以对于所有trie树中不存在的边，利用fail值找到应有的转移。

此外，对于每一个状态，求一个last值，表示最长的一个满足他是这个状态所对应的字符串的后缀的字符串所对应的节点。

一个看似正确的方法是直接对于那N个字符串中的每一个在自动机上走一遍，每次访问到一个节点时通过他的last访问到所有匹配并更新对应的答案。

然而，这样做的复杂度是所有字符串的答案和，如果所有字符均为a，那么所有字符串的出现次数都接近那N个字符串的长度之和，这也是不能忍受的。

我们注意到，类似上面所述，fail指针实际上形成了一棵树，称为fail树。

我们每访问到一个节点，相当于将他到根的路径上所有有对应字符串的节点的访问次数都加一，也就是对于所有到根的路径的节点的访问次数加一之后，对每个字符串将他所对应的节点的访问次数输出。

所以我们现在的问题是，给定一棵树，支持将一个点到根的路径加一，询问一个点的权值。

用LCT，树链剖分显然是可以做的，但是我们可以做的更简单。

我们求出这棵树的dfs序之后，每一次加一相当于区间加，询问相当于单点询问。

用线段树或者差分之后树状数组均可解决。

另外我们注意到询问点权都是在最后。所以我们可以打两个标记，最后线性扫一遍。

此外我们可以直接在树上打标记，全部打完之后按逆dfs序将标记上传即可。

时间复杂度是读入复杂度，空间复杂度是读入字符串长度总和加上W个字符串的总和乘上字符集大小。

本题属于集训队作业中最简单的题之一。

官方题解

CC 上的题解