

后缀数组复习

什么是后缀数组

将一个字符串的所有后缀按照字典序从小到大排序，我们记sa[i]表示排名第i的后缀在原串的开头位置。

后缀数组可以用来干什么

处理一系列关于子串/回文串/后缀子串/重复子串/公共子串/子串相似度的问题

后缀数组模版

build_sa函数

```
const int Len = 200000+5;
char s[Len];
int c[Len],sa[Len],val[Len],q[Len],newval[Len];
bool is_same(int a,int b,int hl,int len)
{
    return
    val[a]==val[b]&&((a+hl>len&&b+hl>len)||((a+hl<len&&b+hl<len&&val[a+hl]==val[b+hl]));
}
void build_sa(int len,int lim)
{
    int i,j,k;
    for(i = 0;i<lim;i++)c[i]=0;
    for(i = 0;i<len;i++)c[val[i]=s[i]]++;
    for(i = 1;i<lim;i++)c[i]+=c[i-1];
    for(i = len-1;i>=0;i--)sa[--c[val[i]]] = i;
    for(int d=1;;d++)
    {
        int hl = 1<<(d-1),id = 0;
        for(i = 0;i<len;i++)if(sa[i]+hl>=len)q[id++] = sa[i];
        for(i = 0;i<len;i++)if(sa[i]>=hl)q[id++] = sa[i]-hl;
        for(i = 0;i<lim;i++)c[i] = 0;
        for(i = 0;i<len;i++)c[val[q[i]]]++;
        for(i = 1;i<lim;i++)c[i]+=c[i-1];
        for(i = len-1;i>= 0;i--)sa[--c[val[q[i]]]] = q[i];
        lim = 0;
        for(i = 0;i<len;lim++)
        {
            for(j = i;j<len-1&&is_same(sa[j],sa[j+1],hl,len);j++);
            for(k = i,i = j+1;k<=j;k++)newval[sa[k]] = lim;
        }
        for(int i = 0;i<len;i++)val[i] = newval[i];
        if(lim==len)break;
    }
}
```

```
}  
}
```

求height(排名为i的后缀和排名为i-1的后缀的最长公共前缀)和rank(开头为i的后缀的排名)的函数

```
void build_rank(int len)  
{  
    for(int i=0;i<len;i++)  
        rnk[sa[i]]=i;  
}  
void build_height(int len)  
{  
    for(int i=0;i<len;i++)  
        if(rnk[i])  
        {  
            int j=0;  
            if(i)j=max(0,h[rnk[i-1]]-1);  
while(i+j<len&&sa[rnk[i]-1]+j<len&&s[i+j]==s[sa[rnk[i]-1]+j])j++;  
            h[rnk[i]]=j;  
        }  
}
```

From: <https://wiki.cvbbacm.com/> - CVBB ACM Team

Permanent link: https://wiki.cvbbacm.com/doku.php?id=2020-2021:teams:wangzai_milk:%E5%90%8E%E7%BC%80%E6%95%B0%E7%BB%84%E5%A4%8D%E4%B9%A0

Last update: 2020/05/09 22:59