哈爾濱工業大學

实验报告

题　目：大数据高级数据结构设计与实践作业二

专　　业　大数据科学与技术

学　　号　1170300916

姓　　名　彭钰驯

课　　程　大数据高级数据结构设计与实践

日　　期　2020-4-25

一、实验要求

任务一：

根据输入的文件 wordlist1 构建 trie 树，实现统计一个任意输入的单词是否在文件中出现过。

任务二：

根据输入的文件 article1.txt 构建 trie 树，单词均以空格结束，实现统计单词词频的功能。

二、实验环境

系统环境：Windows10

IDE：Visual Studio

三、人员安排

一人完成

四、实验过程

4.1 Tire 树简介

英文字典树每一个内部结点都有 26 个子结点，树的高度为最长字符串长度一棵子树代表具有相同前缀的关键码的集合。例如“an”子树代表具有相同前缀 an-的关键码集合{and，ant}

4.2 Tire 树操作

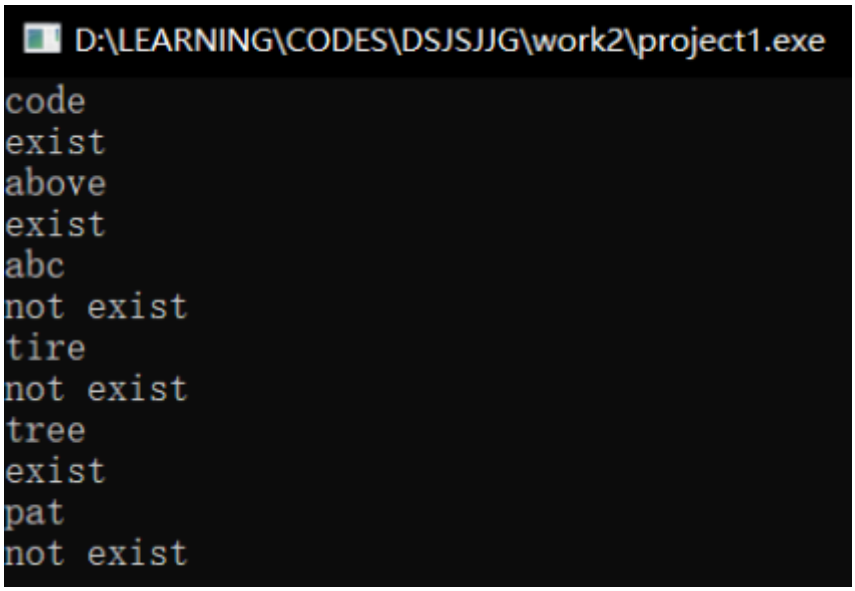
插入：首先根据插入纪录的关键码找到需要插入的结点位置。如果该结点是叶结点，那么就将为其分裂出两个子结点，分别存储这个纪录和以前的那个纪录。如果是内部结点，则在那个分支上应该是空的，所以直接为该分支建立一个新的叶结点即可。

查找：对待查询的的字符与树中的子节点进行比较，第一个字符与树根相比，如果不为空，则查询树根的所有子树并且和第二个字符相比。以此类推，如果进行到某一层，所有字符均已被匹配，且当前层存在标记，则查询成功。若在任意一层对应位置的字符的子树为空或者，最后一层不存在标记，则查询失败。

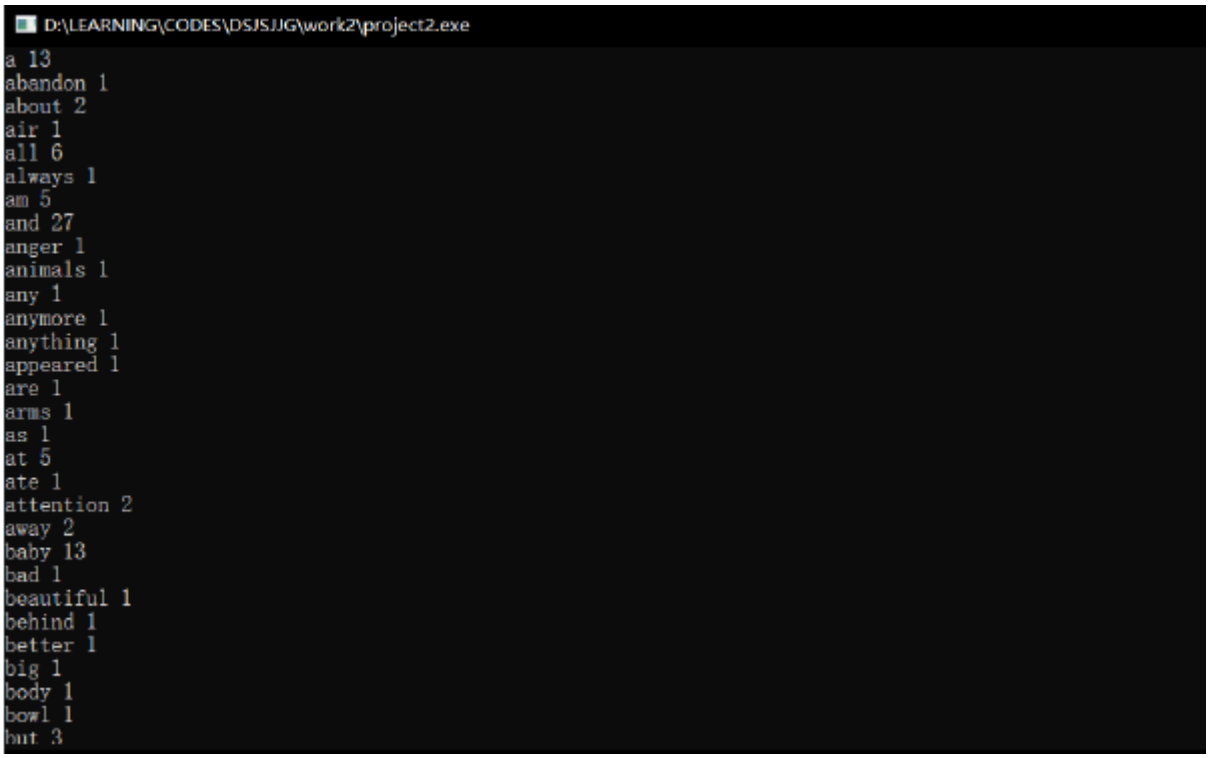
删除：根据插入纪录的关键码找到需要删除的结点位置。如果一个被删除结点的父结点没有其他的儿子，那么就需要合并。否则只需要将此分支设置为空即可。

五、实验分析

任务一：


D:\LEARNING\CODES\DSJSJJG\work2\project1.exe
code
exist
above
exist
abc
not exist
tire
not exist
tree
exist
pat
not exist

任务二：


D:\LEARNING\CODES\DSJSJJG\work2\project2.exe
a 13
abandon 1
about 2
air 1
all 6
always 1
am 5
and 27
anger 1
animals 1
any 1
anymore 1
anything 1
appeared 1
are 1
arms 1
as 1
at 5
ate 1
attention 2
away 2
baby 13
bad 1
beautiful 1
behind 1
better 1
big 1
body 1
bowl 1
but 3