**数据结构实验报告**九

**学号：** 117060400204 **姓名**： 李涛 **班级：应用统计2班**  **指导老师：** 林卫中

**实验名称**： 第九次上机实验

**实验要求：**文本词频的统计，jieba库的安装与使用

**实验题目：程序练习题**

**一**

import jieba

txt = open("红楼梦.txt","r",encoding='GB18030').read()

words = jieba.lcut(txt)

count = {}

for word in words:

if len(word) == 1:

continue

else:

count[word] = count.get(word,0) + 1

item = list(count.items())

item.sort(key=lambda x:x[1],reverse=True)

for i in range(20):

word,count = item[i]

print("{0:<10}{1:>5}".format(word,count))

excludes = {'什么','一个','我们','你们','如今','说道','老太太','知道','姑娘','起来',

'这里','出来','众人','那里','奶奶','自己','太太','一面','只见','两个',

'没有','怎么','不是','这个','听见','这样','进来','咱们','就是','不知',

'东西','告诉','回来','只是','大家','老爷','只得','丫头','这些','他们',

'不敢','出去','所以','不过','不好','姐姐','的话','一时','过来'

}

for word in excludes:

del(count[word])

二

import jieba

txt = open("红楼梦.txt","r",encoding='GB18030').read()

words = jieba.lcut(txt)

count = {}

for word in words:

if len(word) == 1:

continue

else:

count[word] = count.get(word,0) + 1

item = list(count.items())

item.sort(key=lambda x:x[1],reverse=True)

for i in range(20):

word,count = item[i]

print("{0:<10}{1:>5}".format(word,count))

excludes = {'什么','一个','我们','你们','如今','说道','老太太','知道','姑娘','起来',

'这里','出来','众人','那里','奶奶','自己','太太','一面','只见','两个',

'没有','怎么','不是','这个','听见','这样','进来','咱们','就是','不知',

'东西','告诉','回来','只是','大家','老爷','只得','丫头','这些','他们',

'不敢','出去','所以','不过','不好','姐姐','的话','一时','过来'

}

for word in excludes:

del(count[word])

includes = {'王夫人','凤姐','贾母','老太太'}

for word in words:

if word in includes:

if len(word) == 1:

continue

elif word == '凤姐' or word == '王夫人'：:

rword = '凤姐'

elif word == '贾母' or word == '老太太'：:

rword = '贾母'

count[rword] = count.get(rword,0) + 1

**三**

import random

def generateSamples1(n:int):

birthday = []

day = [31,29,31,30,31,30,31,31,30,31,30,31]

for i in range(n):

month = randint(1,12)

day = randint(1,day[month-1])

someday = (month,day)

birthday.append(someday)

return birthday

def isRepetitive(ls):

set1 = set(ls)

if len(set1) == len(ls):

return False

else:

return True

birthday = generateSamples1(50)

c = 0

for i in range(10000):

subset = random.sample.sample(birthday,23)

if isRepetitive(subset):

c = c+1

print("{}个随机样本数量下，23个人中至少有两人过生日的概率是：{}".format(50,c/10000))

一

