第九周实验报告

实验课程：Python语言程序设计 姓名：韩天楠

学号：117060400112 班级：17应用统计学1班

指导老师：林卫中老师 时间：2018.5.4

1. 实验名称：第六周程序练习题。
2. 实验目的：
3. 理解列表概念并掌握Python中列表的使用；
4. 理解字典概念并掌握Python中字典的使用；
5. 运用组合数据类型进行文本词频统计。
6. 实验源代码：

6.4

from operator import itemgetter

def analyseText(text):

D = {}

tlist = list(text)

for e in tlist:

D[e] = D.get(e,0) + 1

return D

text=input("请输入一段文本:")

D = analyseText(text)

items = list(D.items())

items.sort(key=itemgetter(1),reverse=True)

for item in items:

ch, count = item

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

6.5

from datetime import datetime

from random import \*

def generateSamples1(n:int):

birthdays = []

days = [31,29,31,30,31,30,31,31,30,31,30,31]

for i in range(n):

month = randint(1,12)

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

someday = (month,day)

birthdays.append(someday)

return birthdays

def isRepetitive(ls):

set1 = set(ls)

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

return False

else:

return True

birthdays = generateSamples1(50)

c=0

for i in range(10000):

subset = sample(birthdays,23)

if isRepetitive(subset):

c = c+1

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

6.6

import jieba

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'}

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

words = jieba.lcut(txt)

counts = {}

for word in words:

if len(word) == 1:

continue

else:

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

for word in excludes:

del(counts[word])

items = list(counts.items())

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

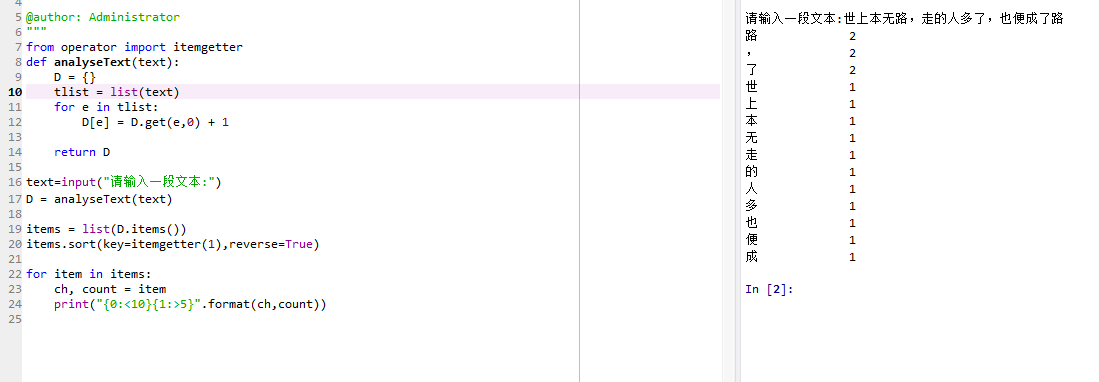
for i in range(15):

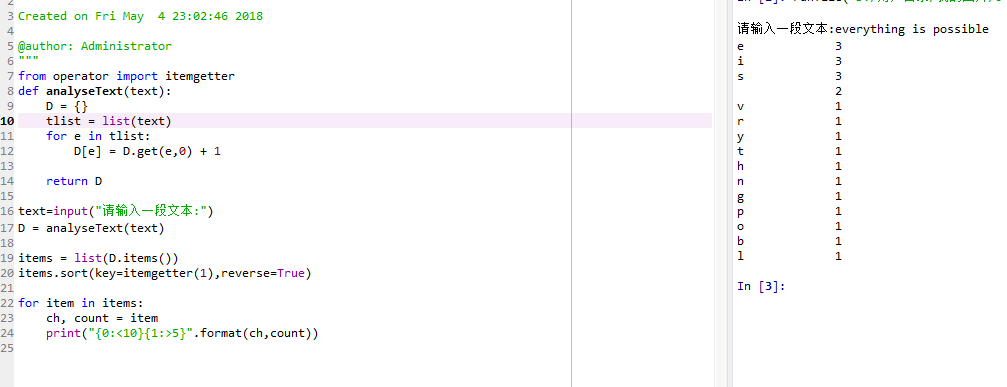
word, count = items[i]

print ("{0:<10}{1:>15}".format(word, count))

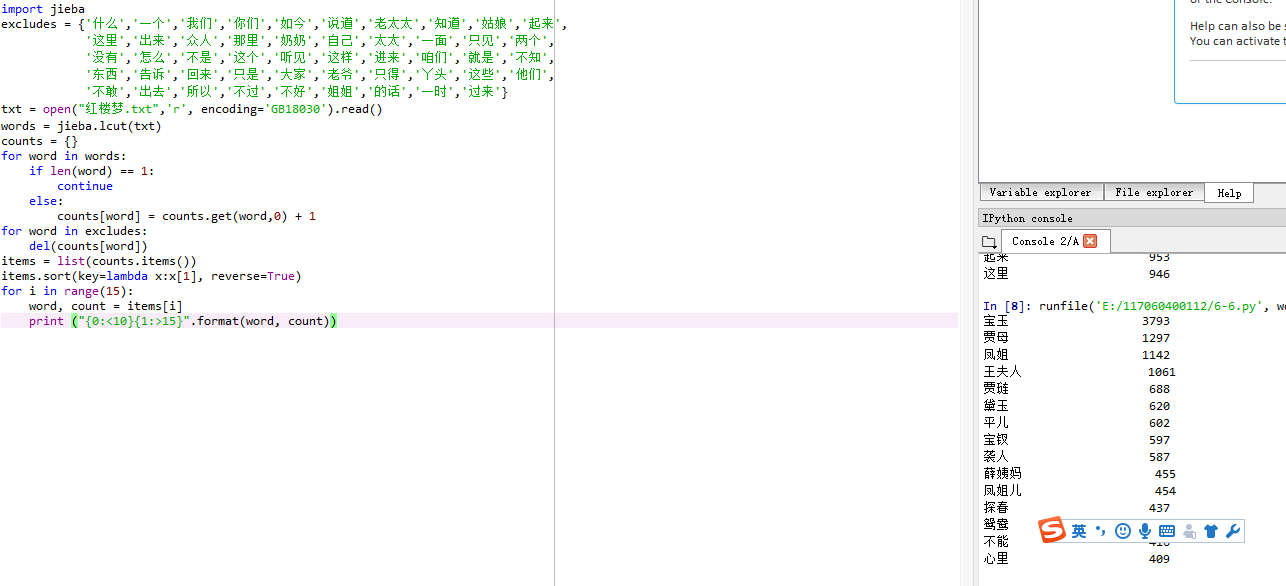
1. 实验运行情况：

6.4

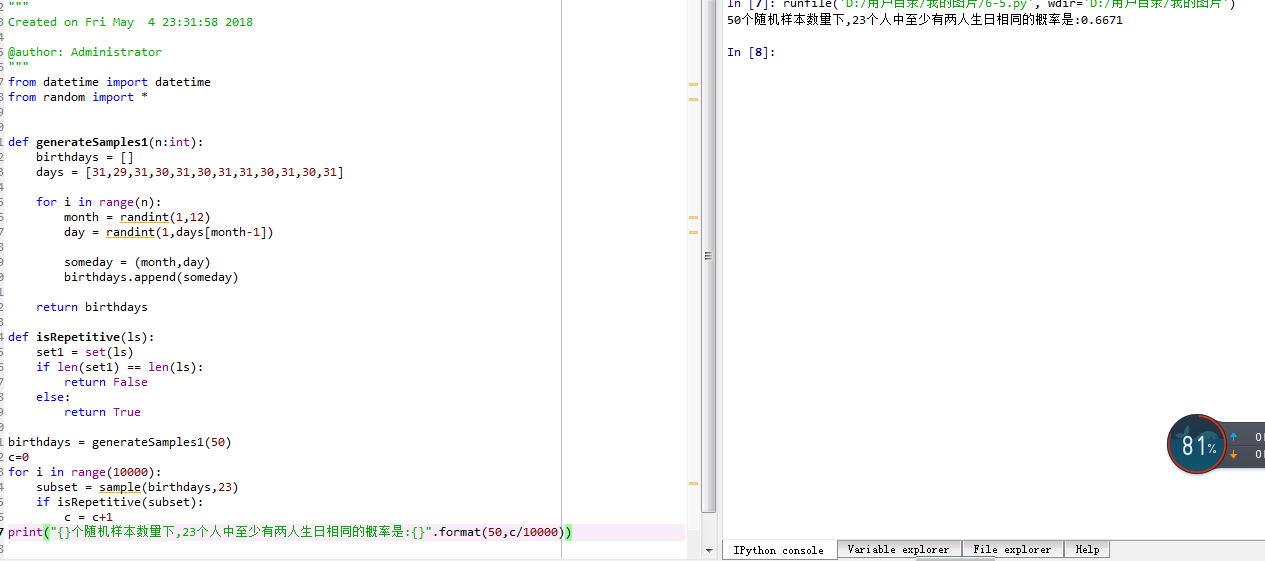




6.6



6.5



1. 实验总结：

本次实验主要了解了jieba库的安装及如何使用其对中文文档进行分词并进一步统计文档词频，并且再次介绍了元组、字典、列表的基本操作，通过本次学习学会了如何统计词频，并进一步分析，真正了解到编程在统计学方面的应用，使我对编程有了新的认识并提高了学习的兴趣，希望在大家的共同努力下，可以学好这门课程。