上机实验报告（9）

姓名：邱文静

班级：17应统(1)班

学号：117060400119 指导老师：林卫中老师

# 实验标题：组合数据类型续

# 实验目的：

* 运用列表管理采集的信息，构建数据结构
* 运用字典处理复杂的数据信息
* 运用组合数据类型进行文本词频统计

# 实验过程及结果：

1. 程序练习题6.4：文本字符分析

**源代码（1）：**

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)

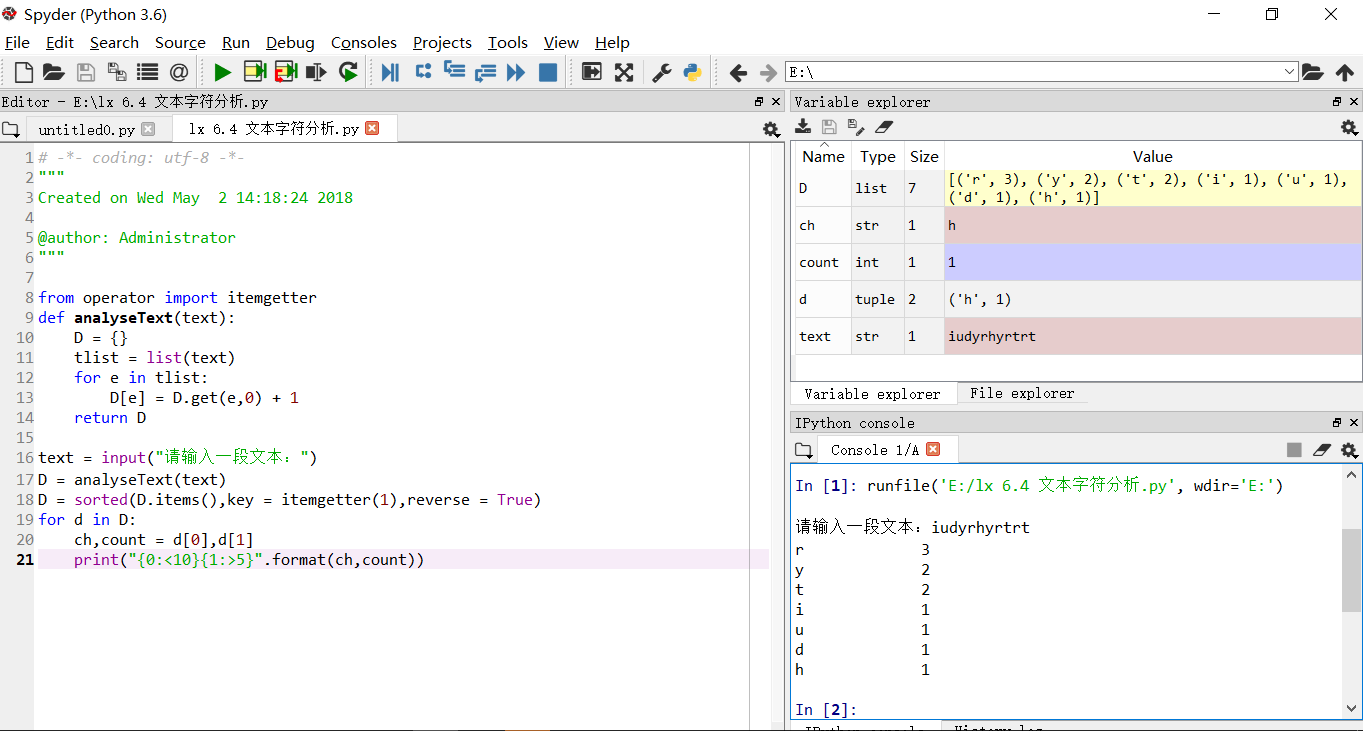
D = sorted(D.items(),key = itemgetter(1),reverse = True)

for d in D:

ch,count = d[0],d[1]

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

**执行结果：**



**源代码（2）：**

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())

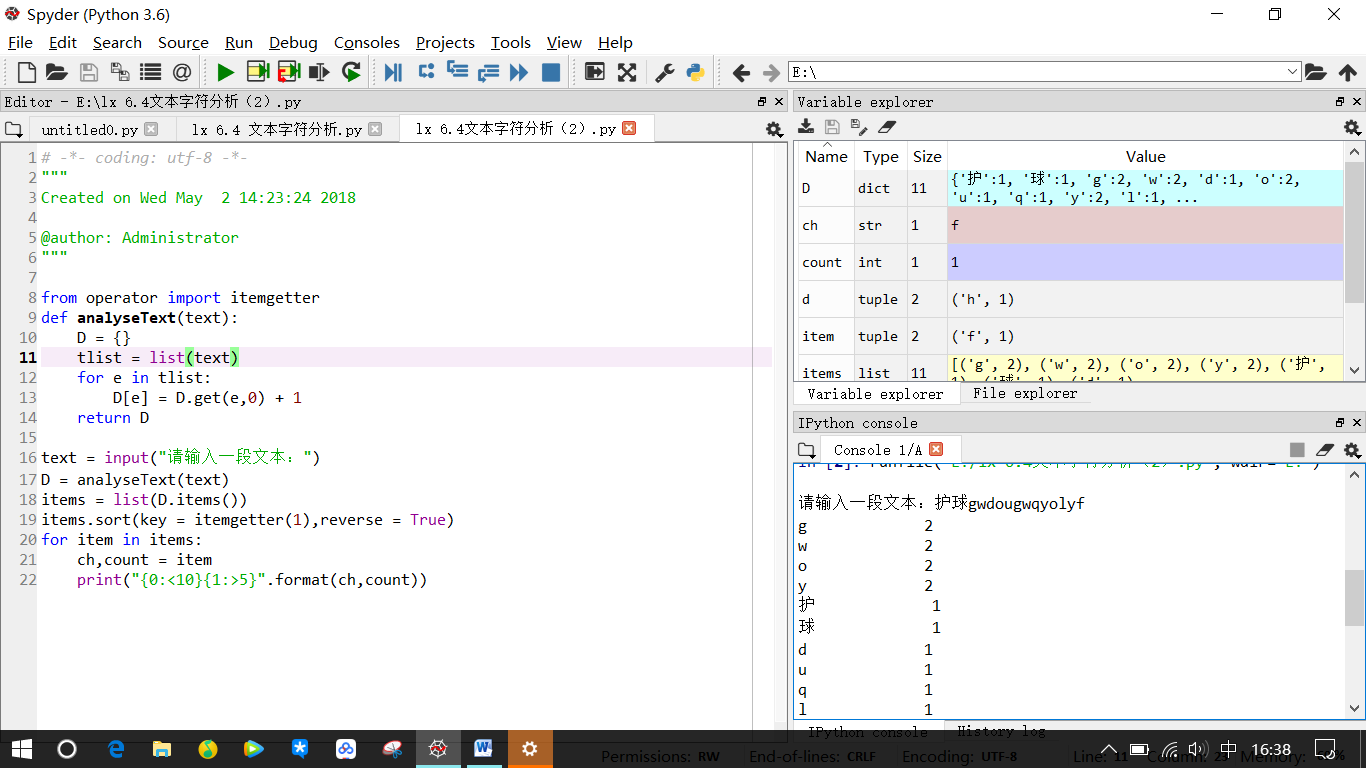
items.sort(key = itemgetter(1),reverse = True)

for item in items:

ch,count = item

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

**执行结果：**



**源代码（3）：**

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())

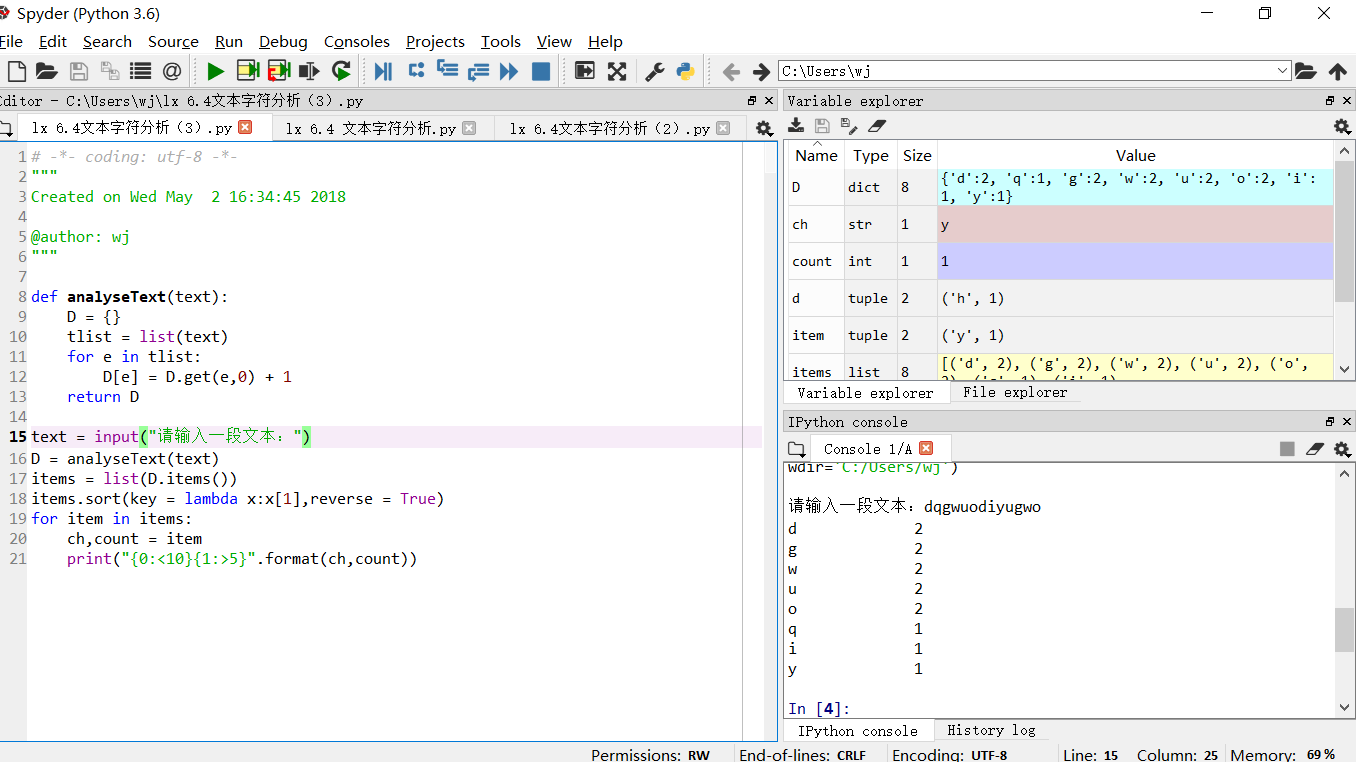
items.sort(key = lambda x:x[1],reverse = True)

for item in items:

ch,count = item

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

**执行结果：**



1. 程序练习题6.6：《红楼梦》人物统计

**源代码（1）：**

import jieba

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

items = list(counts.items())

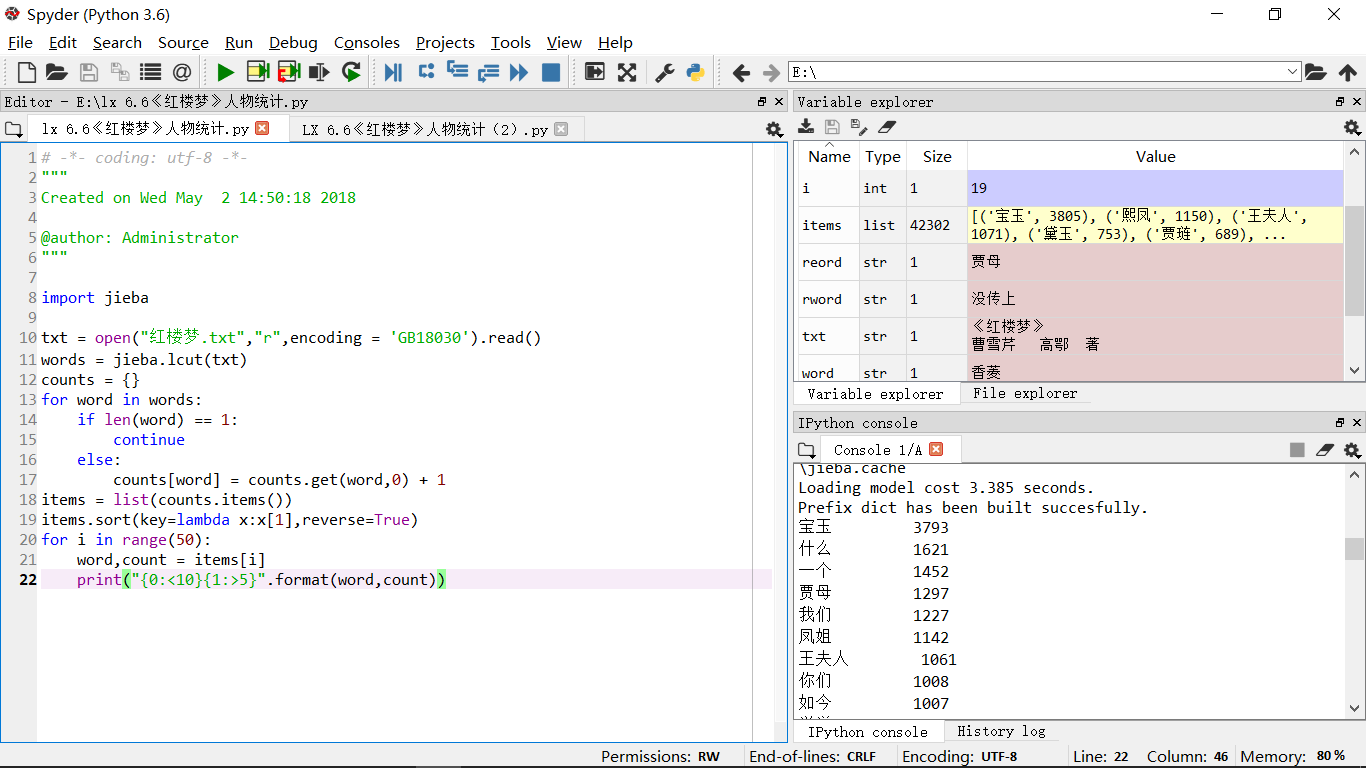
items.sort(key=lambda x:x[1],reverse=True)

for i in range(50):

word,count = items[i]

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

**执行结果：**



**源代码（2）：**

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','不能',

'她们','如此','银子','今日','答应','心里','二爷','二人','几个','这么',

'还有','只管','说话','一回','那边','这话','外头','自然','打发','哪里',

'今儿','罢了','那些','屋里','听说','问道','小丫头','如何','看见','人家','媳妇','不用'}

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

words = jieba.lcut(txt)

counts = {}

for word in words:

if len(word) == 1:

continue

elif word == '凤姐' or word == '熙凤':

rword ='熙凤'

elif word == '贾母' or word == '老母':

reord = '贾母'

elif word == '宝姑娘' or word == '宝钗':

rword = '宝钗'

elif word == '林姑娘' or word == '黛玉':

rword = '黛玉'

else:

rword = word

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

for word in excludes:

del(counts[word])

items = list(counts.items())

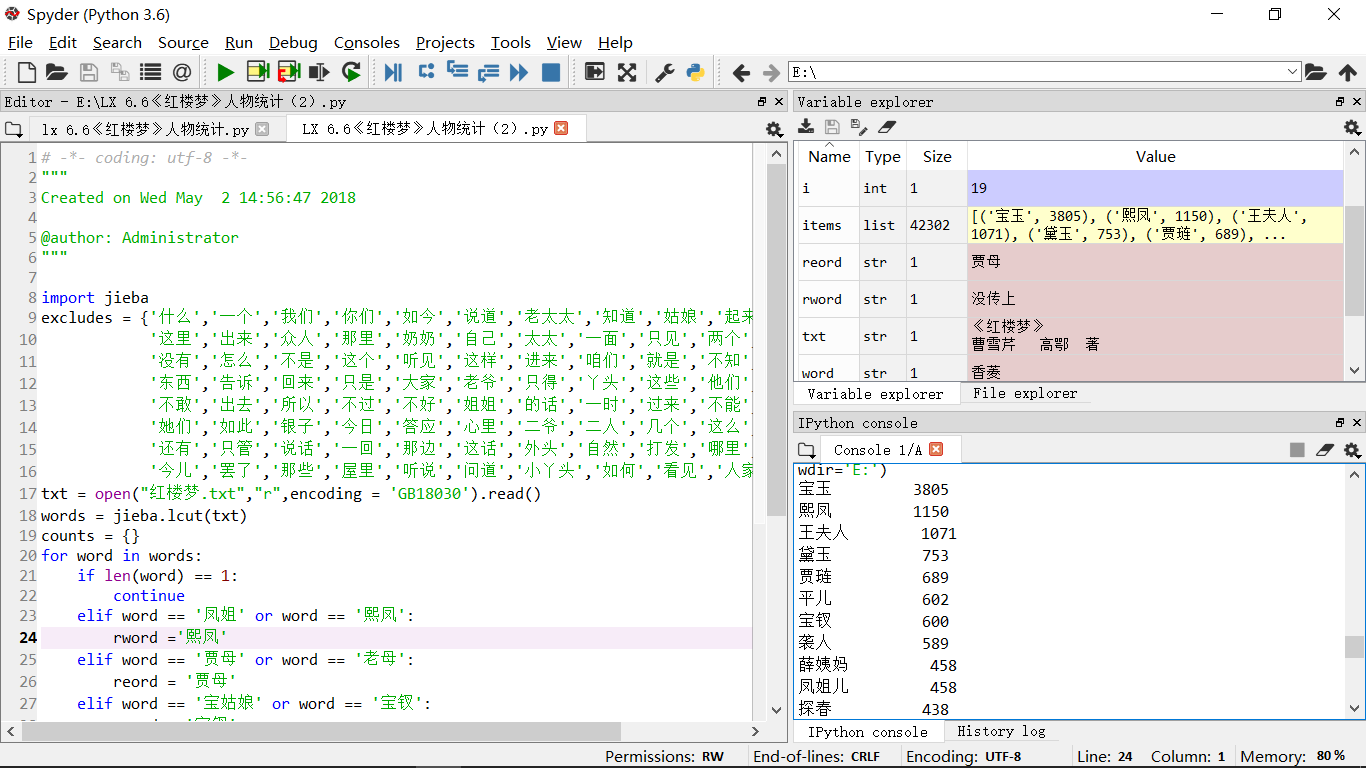
items.sort(key=lambda x:x[1],reverse = True)

for i in range(20):

word,count = items[i]

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

**执行结果：**



1. 程序练习题6.5:生日悖论分析

**源代码：**

from random import\*

def generateSample1(n:int):

birthday = []

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)

birthday.append(someday)

return birthday

def calSameBirthdayProb(birthday:list,n:int):

num = 0

for i in range(n):

people = sample(birthday,23)

pset = set(people)

if len(pset) != len(people):

num += 1

return num/n

def main():

while True:

n = int(input("请输入一个整数:"))

if n < 23:

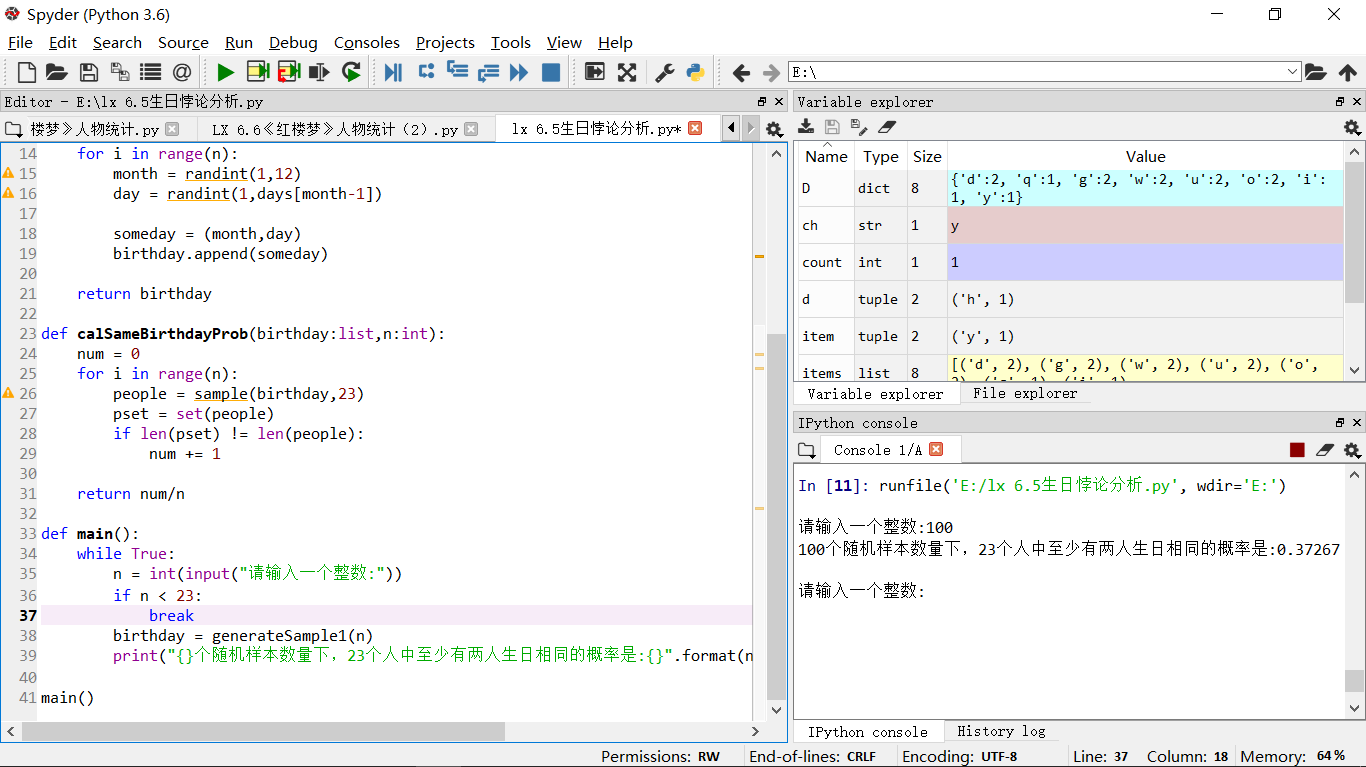
break

birthday = generateSample1(n)

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(n,calSameBirthdayProb(birthday,100000)))

main()

**执行结果：**



# 实验心得：

首先，不得不说Python很好玩也很实用，python已经开始展现它的魅力，确实很喜欢。其次，要学会举一反三，将以前学过或码过的代码要熟练掌握并运用到一起解决更复杂的问题。