**实验报告九**

姓名：袁楚 班级：应统一班 学号：117060400114 指导老师：林卫中

实验名称：组合数据类型

实验目的：（1）了解3类基本组合数据类型

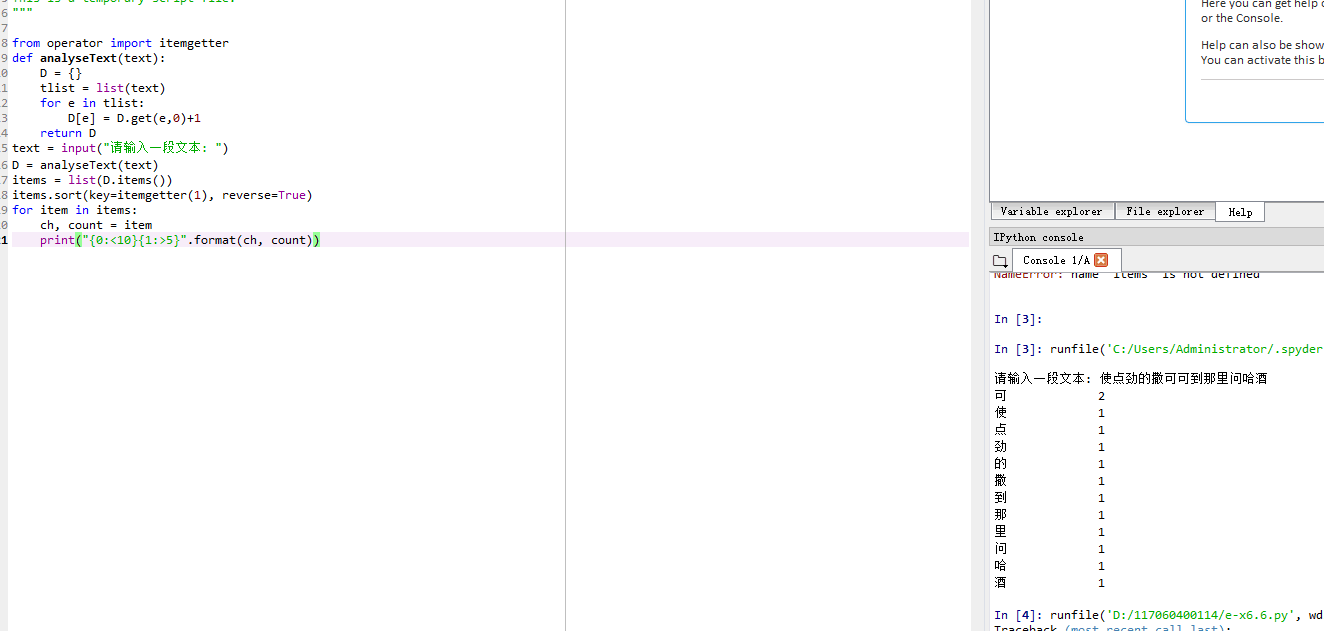
1. 理解列表概念并掌握Python中列表的使用
2. 理解字典概念并掌握Python中字典的使用
3. 运用列表管理采集的信息，构建数据结构
4. 运用字典处理复杂的数据信息
5. 运用组合数字类型进行文本词频的统计

实验题目：（1）文本字符分析

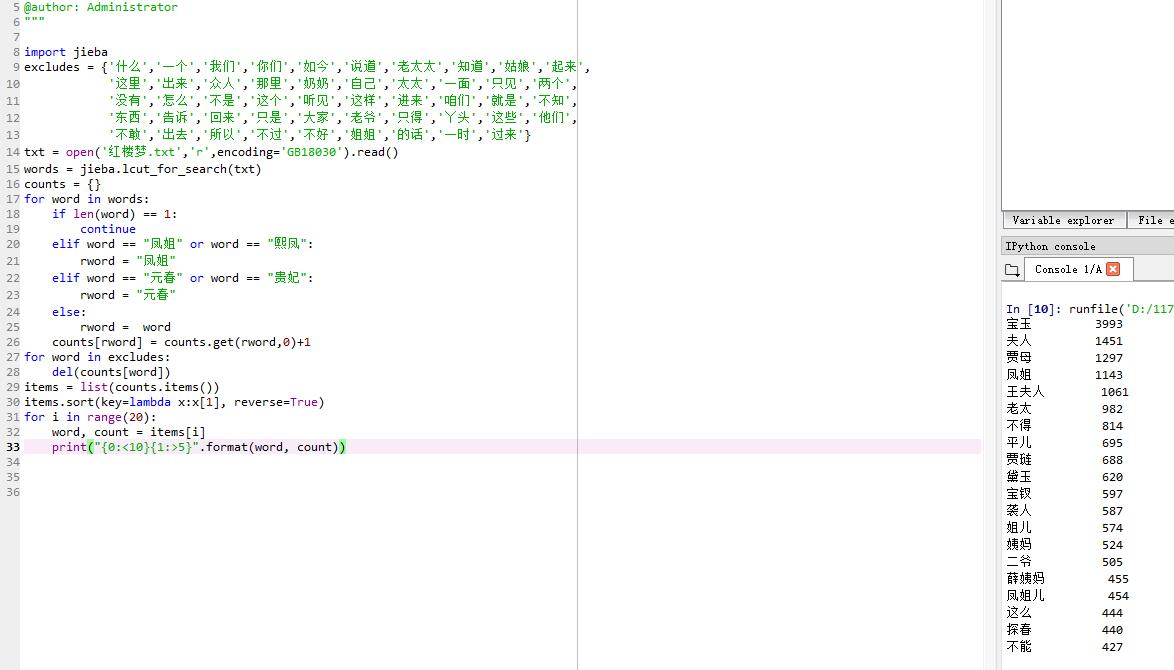
1. 《红楼梦》人物统计分析
2. 生日驳论分析

实现算法：

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)  
   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))



1. 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\_for\_search(txt)  
   counts = {}  
   for word in words:  
       if len(word) == 1:  
           continue  
       elif word == "凤姐" or word == "熙凤":  
           rword = "凤姐"  
       elif word == "元春" or word == "贵妃":  
           rword = "元春"  
       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))



（3）from datetime import datetime

import random

def generateSamples1(n:int):

birthdays = []

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

for i in range(n):

month = random.randint(1,12)

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

someday = (month,day)

birthdays.append(someday)

return birthdays

birthdays = generateSamples1(50)

def isRepetitive(ls):

n = set(ls)

if len(ls) != len(n):

return True

return False

c = 0

for i in range(10000000):

subset = random.sample(birthdays,23)

if isRepetitive(subset):

c = c + 1

a = c/10000000

print("n个随机样本数量下，23个人中生日相同的概率是{:.3f}".format(50,5))

