实验报告9

学号：117060400115 姓名：魏莹 班级：17应统1班 指导老师：林卫中

实验名称：python第六章程序练习题

实验要求：1.运用列表管理采集的信息，构建数据结构

2.运用字典处理复杂的数据信息

3.运用组合数据类型进行文本词频统计

实验心得：需要辨别出字典与集合的大括号，掌握jieba库，对小说中人名的词频进行统计。

实验结果：

1.

from operator import itemgetter

def analyseText(text):

D = {}

tlist = list(text)

for e in tlist:

if e in D:

D[e] = D[e] + 1

else:

D[e] = 1

return D

text = "python/textbook/English"

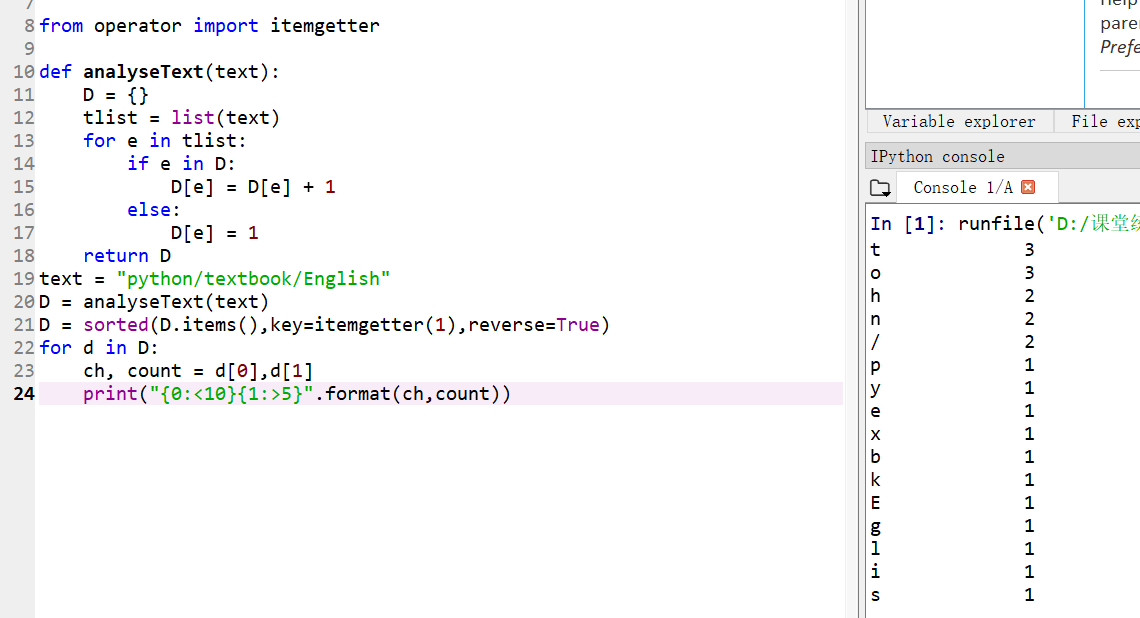
D = analyseText(text)

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))



2.

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):

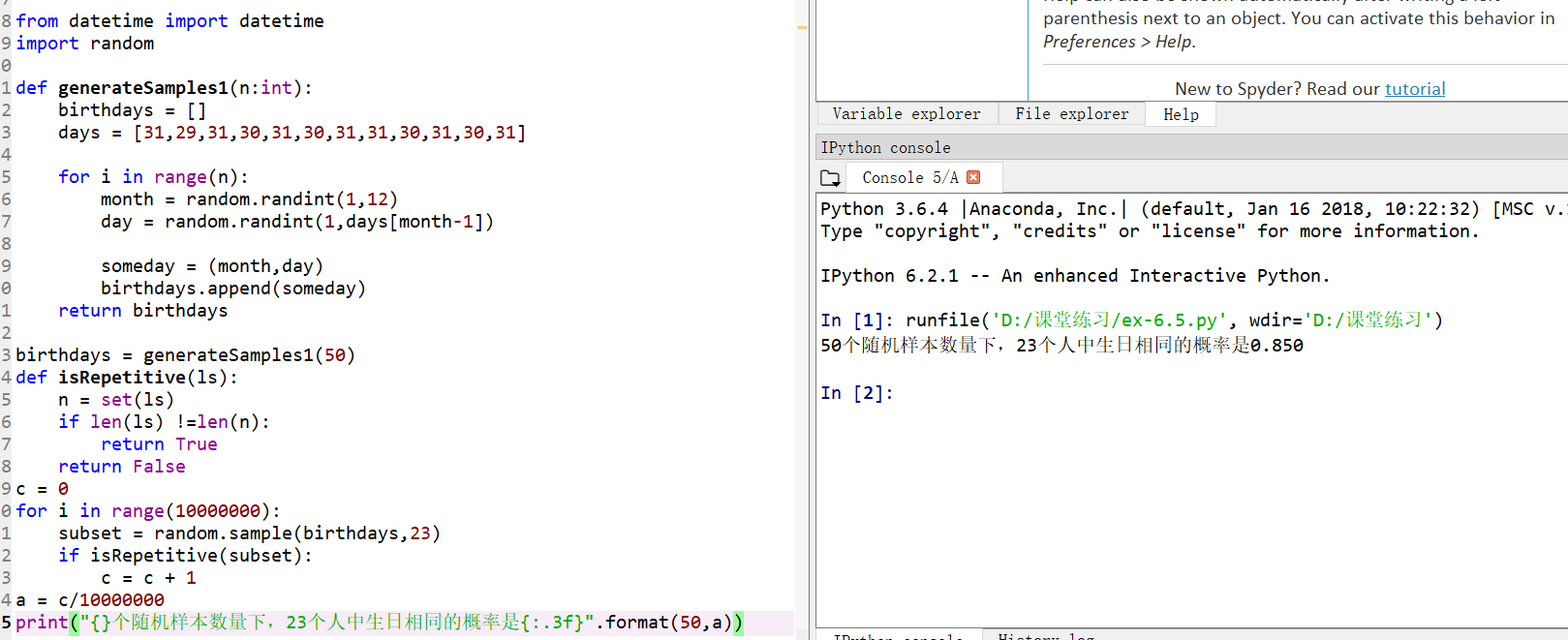
subset = random.sample(birthdays,23)

if isRepetitive(subset):

c = c + 1

a = c/10000000

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



3.

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))

