**数据结构实验报告9**

**学号：** 117060400211 **姓名**：王婕冉 **班级：** 应用统计学2班 **指导老师：林卫中**

**实验名称**： 组合数据类型

**实验要求：** 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

**实验题目：文本分析，生日分析 ，文本人物统计**

**算法实现：（核心代码）**

**①**

from operator import itemgetter

def AN(text):

D ={}

tlist = list(text)

for e in tlist:

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

return D

text = input("Enter:")

D = AN(text)

items = list(D.items())

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

for item in items:

e,count = item

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

②

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

def isR(ls):

for x in ls:

a = ls.count(x)

if a >= 2:

return True

return False

birthdays = generateSamples1(50)

c = 0

for i in range(10000):

subset = random.sample(birthdays,23)

if isR(subset):

c = c+1

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

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(txt)

counts = {}

for word in words:

    if len(word) == 1:

  continue

    elif word =="贾母" or word =="老太太":

        rword ="贾母"

    elif word =="凤姐" or word =="熙凤":

        rword ="熙凤"

    else:

        rword = word

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

items = list(counts.items())

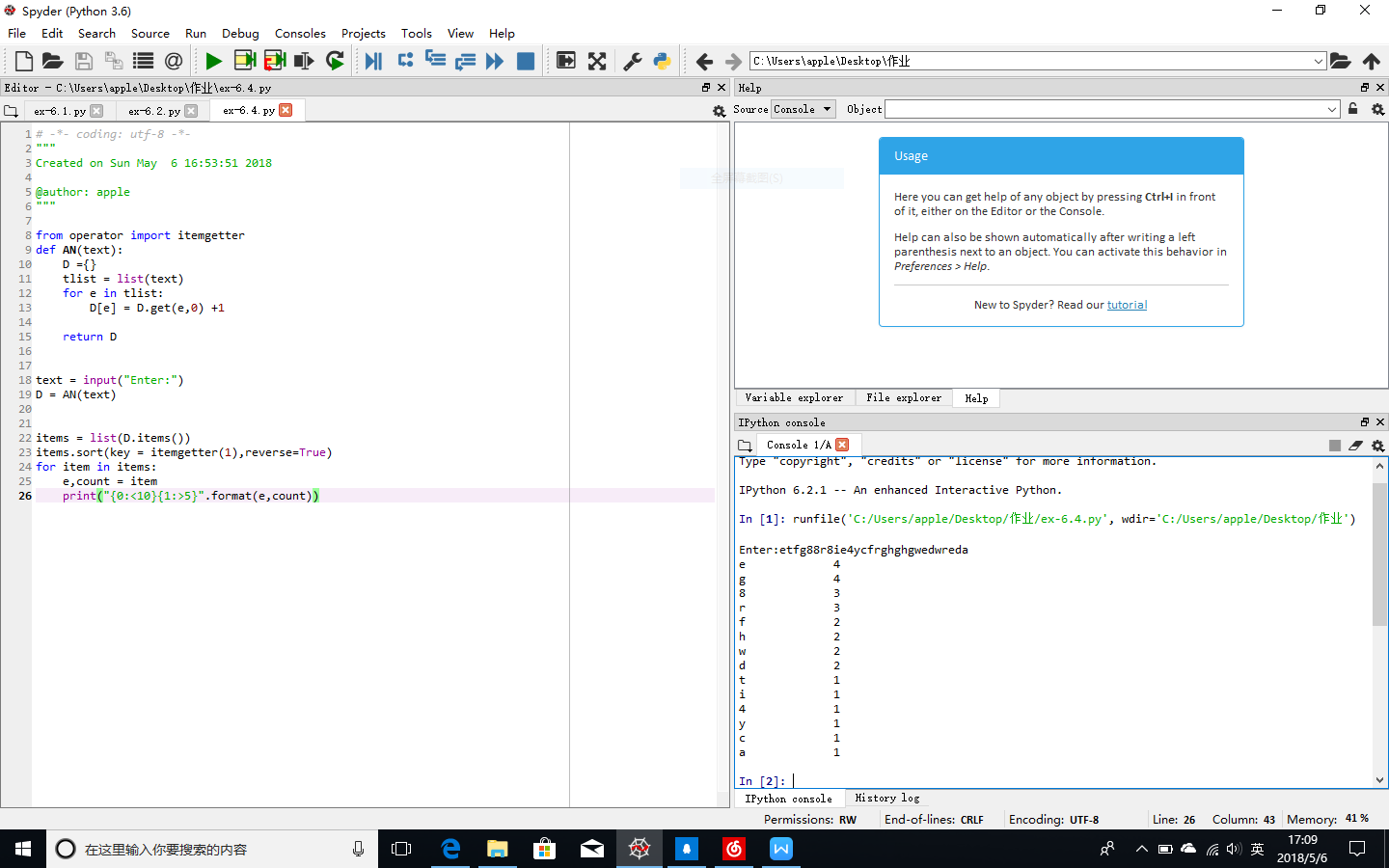
items.sort(key=lambda x:x[1],reverse = True)

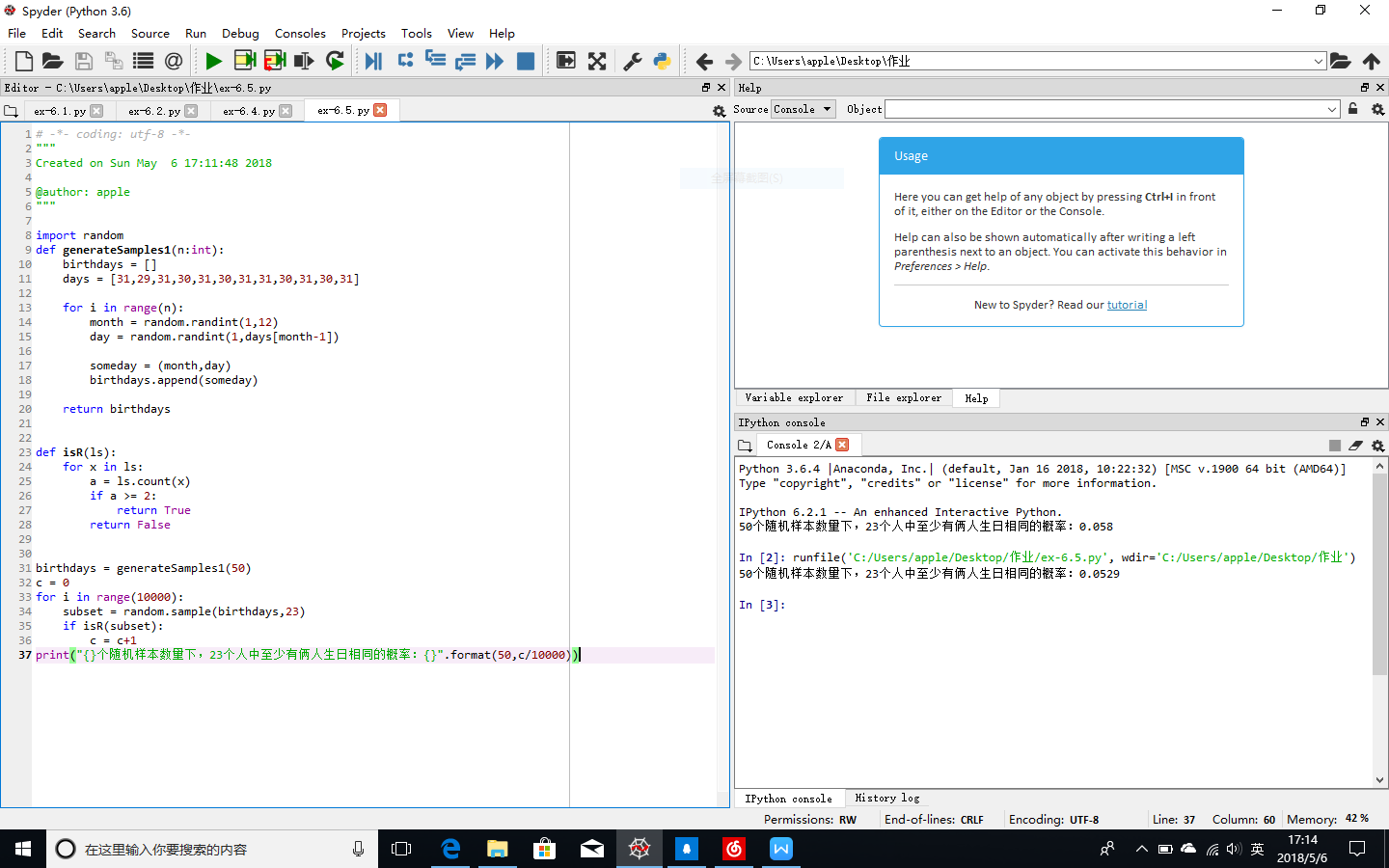
for i in range(100):

    word, count = items[i]

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

**实验结果：（要有截图，白底黑字）**

****

****