**数据结构实验报告9**

**学号：** 117060400102 **姓名**： 严帝 **班级：**应用统计一班  **指导老师：** 林卫中

**实验名称**： 第六章程序练习题 6.5 6.6

**实验要求：** jieba库的使用 元组

**1：import random**

**def generateSamples(n:int): #生成n个人的生日，并放到一个列表中**

**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) #生成月，日**

**birthdays.append(someday)**

**return birthdays**

**def sameBirthday(birthdays:list, n:int):**

**num = 0**

**for i in range(n):**

**people = random.sample(birthdays, 23)**

**pset = set(people) #将列表变为集合，计算相同生日的人的个数**

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

**num = num + 1**

**return num/n**

**def main():**

**while True:**

**n = int(input("please enter a number: "))**

**if n < 23:**

**break**

**birthdays = generateSamples(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, sameBirthday(birthdays, 100000)))**

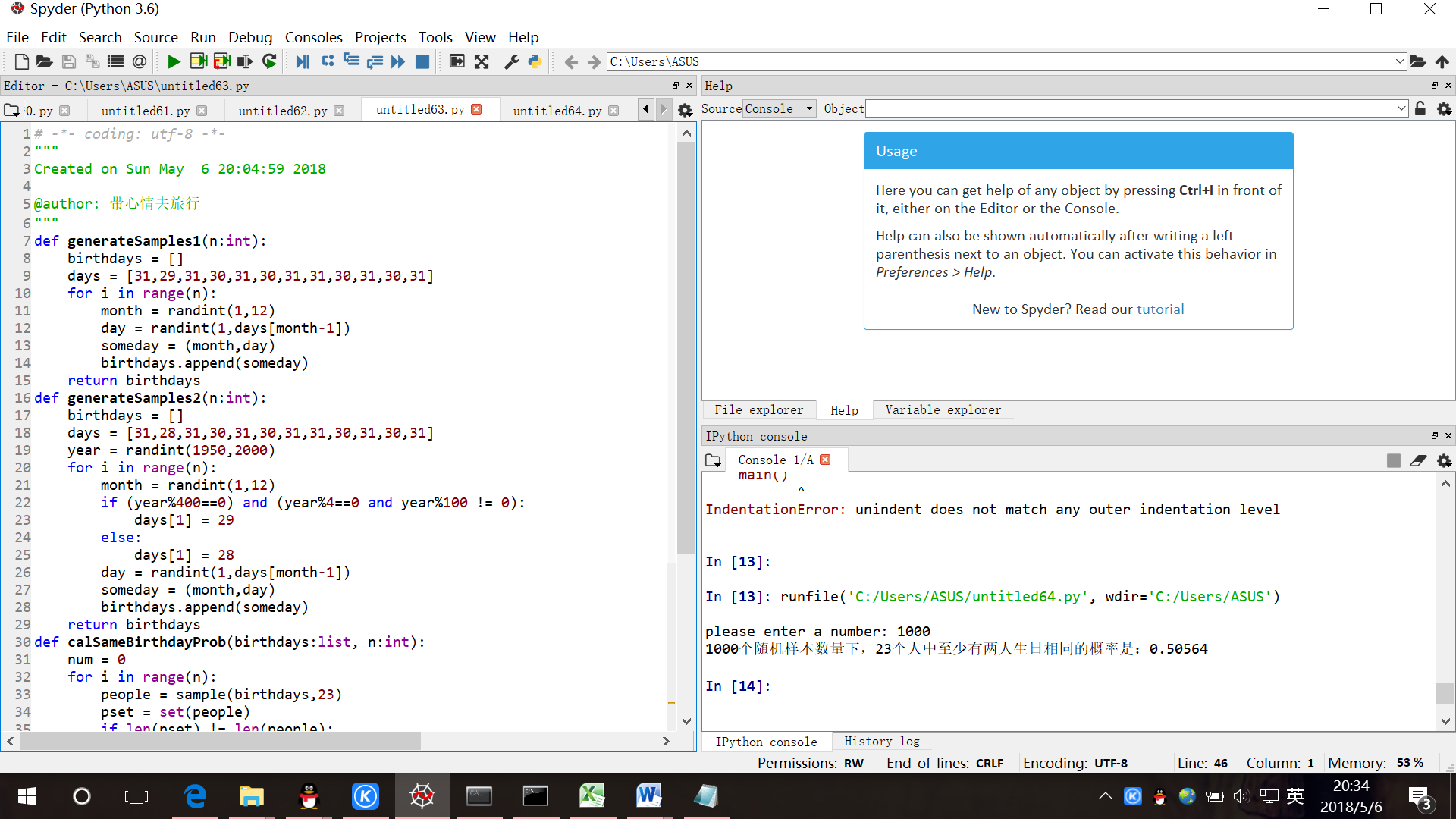
**break**

**main()**

**2：**

import jieba  
txt = open("红楼梦.txt","r",encoding='GB18030').read()  
words = jieba.lcut(txt)  
count = {}  
for word in words:  
    if len(word) == 1:  
        continue  
    else:  
        count[word] = count.get(word,0) + 1  
item = list(count.items())  
item.sort(key=lambda x:x[1],reverse=True)  
for i in range(20):  
    word,count = item[i]  
    print("{0:<10}{1:>5}".format(word,count))  
  
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'  
            }  
for word in  excludes:  
    del(count[word])  
  
includes = {'王夫人','凤姐','贾母','老太太'}  
for word in words:  
    if word in includes:  
        if len(word) == 1:  
            continue  
        elif word == '凤姐' or word == '王夫人'：:  
            rword = '凤姐'  
        elif word == '贾母' or word == '老太太'：:  
            rword = '贾母'  
        count[rword] = count.get(rword,0) + 1

实验结果：

1: 

2222222222

由于