景德镇陶瓷大学PYTHON语言程序设计基础实验报告

**学号：**117060400118 **姓名：**梁倩乐 **班级：**17应用统计学1班 **指导老师：**林卫中

**实验名称：实验九**

**实验要求：学习文本词频统计。**

**实验题目：1.《红楼梦》人物统计。编写程序统计《红楼梦》中前20位出场最多的人物。**

1. **生日饽论分析。生日饽论指如果一个房间里有23人或以上，那么至少有两个人生日相同的概率大于50%。编写程序，输出在不同随机样本数量下，23个人中至少两个人生日相同的概率。**

**实验过程：1程序》：**

**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(50):**

**word, count = items[i]**

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

**程序2》**

**from datetime import datetime**

**from random import \***

**def generateSamples1(n:int):**

**birthdays = []**

**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)**

**birthdays.append(someday)**

**return birthdays**

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

**num = 0**

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

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

**pset = set(people)**

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

**num += 1**

**return num/n**

**def main():**

**while True:**

**n = int(input("输入一个整数："))**

**if n < 23:**

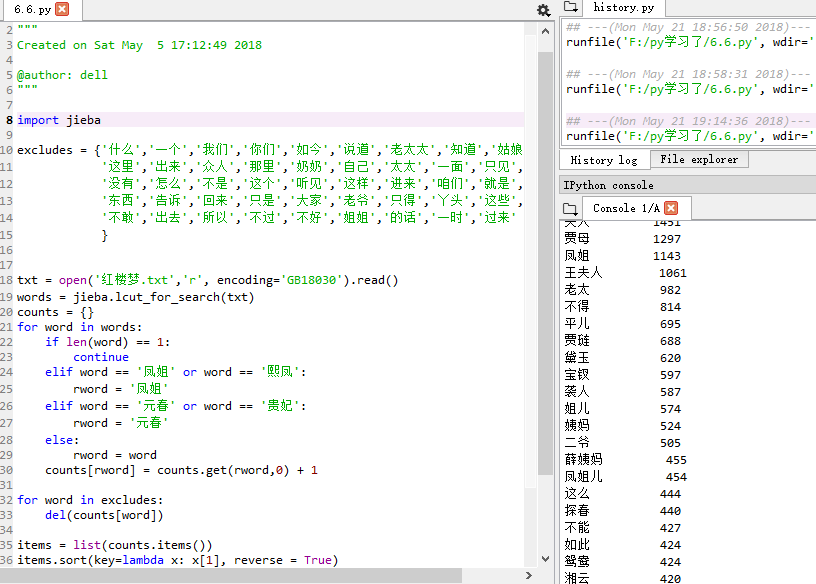
**break**

**birthdays = generateSamples2(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(birthdays, 100000)))**

**main()**

**实验结果：1》**

****

**2.**

**实验小结：经历了很多遍的失败后总算是安装好了jieba库。事实证明，坚持下来就是胜利。好好学习。**