**实验报告九  
实验目的：**

**了解列表、字典、集合三种组合数据类型  
实验要求：  
1、能运用集合的特性对数据去重。  
2、掌握运用字典概念处理复杂数据信息，及字典的排序。  
3、能运用列表构建数据结构。  
4、综合运用组合数据类型进行文本词频统计。  
实验操作与心得  
1.生日悖论分析  
实验心得：**

**参考代码如下：  
from datetime import datetime  
from random import \*  
  
# 随机生成n个人的生日，返回一个列表，列表中每一个元素的形如（月，日）  
def generateSamples1(n:int):  
    birthdays = []  
    days = [31,29,31,30,31,30,31,31,30,31,30,31]  
      
    for i in range(n): # 随机生成n个人的生日（月，日）  
        month = randint(1,12)  
        day = randint(1,days[month-1])  
          
        someday = (month,day)  
        birthdays.append(someday)  
      
    return birthdays  
  
  
# 随机生成n个人的生日，返回一个列表，列表中每一个元素的形如（月，日）  
def generateSamples2(n:int):  
    birthdays = []  
    days = [31,28,31,30,31,30,31,31,30,31,30,31]  
    year = randint(1950,2000)  
      
    for i in range(n): # 随机生成n个人的生日（月，日）  
        month = randint(1,12)  
        if (year%400==0) and (year%4==0 and year%100 != 0):  
            days[1] = 29  
        else:  
            days[1] = 28  
        day = randint(1,days[month-1])  
          
        someday = (month,day)  
        birthdays.append(someday)  
      
    return birthdays  
  
  
# 计算在给定的样本列表birthdays中，23个人中至少有两人生日相同的概率  
# 在函数中随机取23个人，计算是否有两人生日相同，重复n次来计算概率  
# param birthdays -- 样本列表，列表中存有若干个人的生日数据  
# param n -- 计算概率时，事件的重复次数。n越大，计算的概率越接近真实值  
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：")) # n是人群的数量  
        # 如果输入的人群数量少于23则结束  
        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()      
2.《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。  
实验心得：**

**该实验与教材实例10类似，据此改变即可。  
总结：虽然这次实验只有两个题，但是这两个题目真的好大啊，对词汇的统计，呜呜呜，我要哭了，红楼梦的文本还要下载，但是书上有例子还是相对来说降低了难度，要是全让我自己编写我肯定不行**