**实验报告8**

**学号：**117060400123 **姓名**：黄茜洋 **班级：**应统一班**指导老师：**林卫中**实验名称**：函数和代码复用

**实验要求：**1.了解3类基本组合数据类型

2.理解列表概念并掌握python中列表的使用

3.理解字典概念并掌握python中字典的使用

4.运用字典处理复杂的数据信息

**实验题目：**1. 文本字符分析

2.生日悖论分析

3.红楼梦人物统计

**算法实现：**

**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 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| D = analyseText(text) |
|  |

|  |
| --- |
|  |

|  |
| --- |
| items = list(D.items()) |
|  |

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

|  |
| --- |
|  |
|  |

|  |
| --- |
| for item in items: |
|  |

|  |
| --- |
| ch, count = item |
|  |

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

**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): # 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 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |

|  |
| --- |
| 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 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |

|  |
| --- |
| 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 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| 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()

**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(50): |
|  |

|  |
| --- |
| word, count = items[i] |
|  |

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

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| includes = {'宝钗','黛玉','凤姐','元春','探春','湘云','妙玉','迎春','惜春','巧姐','李纨','可卿' |
|  |

|  |
| --- |
| ,'香菱','宝琴','二姐','三姐','岫烟','李纹','李绮','宝珠','宝蟾','平儿','娇杏','瑞珠','袭人','紫鹃','晴雯'} |
|  |

|  |
| --- |
| txt = open('红楼梦.txt','r', encoding='GB18030').read() |
|  |

|  |
| --- |
| words = jieba.lcut(txt) |
|  |

|  |
| --- |
| counts = {} |
|  |

|  |
| --- |
| for word in words: |
|  |

|  |
| --- |
| if word in includes: |
|  |

|  |
| --- |
| if word == '凤姐' or word == '熙凤': |
|  |

|  |
| --- |
| rword = '凤姐' |
|  |

|  |
| --- |
| elif word == '元春' or word == '贵妃': |
|  |

|  |
| --- |
| rword = '元春' |
|  |

|  |
| --- |
| elif word == '宝姑娘' or word == '宝钗': |
|  |

|  |
| --- |
| rword = '宝钗' |
|  |

|  |
| --- |
| elif word == '林姑娘' or word == '黛玉': |
|  |

|  |
| --- |
| rword = '黛玉' |
|  |

|  |
| --- |
| else: |
|  |

|  |
| --- |
| rword = word |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| counts[rword] = counts.get(rword,0) + 1 |
|  |

|  |
| --- |
|  |
|  |

|  |
| --- |
| 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)) |
|  |

实验结果：

****

****