北京航空航天大学

2018－2019 学年 第二学期

《大学计算机基础（理科）》

大作业实验报告

班 级\_\_187722\_\_\_\_\_\_学 号 18377333\_

姓 名\_\_张浩远\_\_\_\_\_\_成 绩 \_\_\_\_\_\_\_\_\_

2019年6月

**作业1 文本编辑器的设计**

一、实验目的

描述通过本题目的实现将考察哪些知识点和方法。

1.Tkinter的使用和设计：button的功能和位置 mainloop的使用方法

2.打开并读取txt文件的方式

3.列表的使用

4.正则表达式的使用

5.字典的创建和值的用法

6.停用词表中对txt的读取和元素的筛选

7.绘图:柱状图的绘制

二、实验任务

描述本题目需要完成的内容和要求。

1.打开并读取txt文件

2.去除文本中的空格、符号和数字

3.统计文本中的单词总数

4.统计文本中每个单词的出现频率

5.识别并去除停用词表中的单词

6.统计出现频率最高的6个实词，并使用柱状图绘制

7.使用tkinter设计GUI界面，对以上功能进行交互式设计。

三、总体设计方案

介绍作业完成的功能，实现方案（如程序分为几部分，每部分采用什么方法、解决什么问题）。

（一）tkinter部分

1.创建tkinter父窗口

2.创建功能实现的按钮

3.实现mainloop循环

（二）读取文件和处理部分的函数

1.读取文件txt

2.处理换行中的 /n 符号

3.将去除文本中的符号和数字（使用正则表达式）

4.将文本中的单词作为列表中的元素

5.求列表的长度，得出单词数

（三）求出高频实词部分的函数

1.对处理好的单词列表，通过元素是否在停用词表中进行判断，进行筛选

2.创建字典，对于每个新出现的元素，创建值为1；对于出现的重复元素，值加一

3.按照降序对字典的值进行排序。提取字典的元素和值作为两个列表。

4.取列表中前六个元素，使用matplotlib进行柱状图的绘制

四、设计思路和关键语法

# -\*- coding: utf-8 -\*-

"""

Created on Sun Jun 9 10:27:51 2019

@author: zhy

"""

import re #正则表达书去除空格和标点符号,数字

punctuation = '.!,;:?"\''+"("+")"+"0"+"1"+"2"+"3"+"4"+"5"+"6"+"7"+"8"+"9"

import operator

import matplotlib.pyplot as plt

#加入GUI

from tkinter import \*

def removePunctuation(text):

text = re.sub(r'[{}]+'.format(punctuation),'',text)

return text.strip().lower()

#使用正则表达式re.sub除去符号和数字

def read():

with open(r'test.txt','r') as f1:

print(f1.read())

#读取文件的函数

def numbersta():

with open(r'test.txt','r') as f1:

list1=f1.readlines()

list2=[]

for i in list1:

list1\_1=i.split()

for j in list1\_1:

j1=removePunctuation(j) #正则表达式处理

list2.append(j1)

list3=[]

for i in list2:

if i!="":

list3.append(i)

#除去空格

print("单词的个数为：",len(list3))

#求出单词总数

#获得单词总数的函数

def sta():

with open(r'test.txt','r') as f1:

list1=f1.readlines()

list2=[]

for i in list1:

list1\_1=i.split()

for j in list1\_1:

j1=removePunctuation(j)

list2.append(j1)

list3=[]

for i in list2:

if i!="":

list3.append(i)

#利用字典统计词频

dic = {} #创建字典

for i in list3:

#对于所有的英语单词，如果在字典中，则令值+1，若不在字典中，则创建这个值并赋值为1

if i not in dic:

dic[i] = 1

else:

dic[i] = dic[i] + 1

v1 = sorted(dic.items(),key=operator.itemgetter(1),reverse=True)

#按字典的值进行降序排序

with open(r'stopword.txt','r') as f2:

list11=f2.readlines()

line12=[]

line13=[]

for i in list11:

i2=i.strip()

line12.append(i2)

for i in range(0,len(line12)):

for j in line12[i].split():

line13.append(j)

#利用停表去除

#创建两个空列表，储存字典的元素和对应的值

list2x=[]

list2y=[]

for k,v in v1:

if k not in line13:

print(k,v)

list2x.append(k)

list2y.append(v)

list2m=[]

list2n=[]

#取最高频的前六个函数获得自变量和因变量

for i in range(0,6):

list2m.append(list2x[i])

list2n.append(list2y[i])

plt.bar(list2m,list2n)

plt.show()

#plt函数制作统计表

window=Tk() #创建界面

window.title(u'文本编辑器')

#加入标题

Button(window,text='显示文本',command=read).grid(row=3, column=0, sticky=W, padx=10, pady=5) #输入按钮

Button(window,text=u'统计字数',command=numbersta).grid(row=3, column=1, sticky=E, padx=10, pady=5)

Button(window,text=u'高频词查询',command=sta).grid(row=3, column=2, sticky=W, padx=10, pady=5)

Button(window,text=u'退出程序',command=window.destroy).grid(row=3, column=3, sticky=W, padx=10, pady=5)

#创建按钮，确定其位置和相应的功能

mainloop() #执行循环

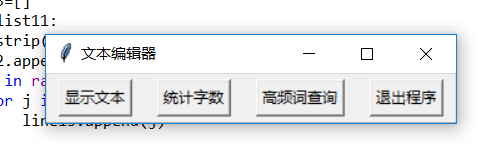
标红色为关键代码

注意：这是本报告的重点！

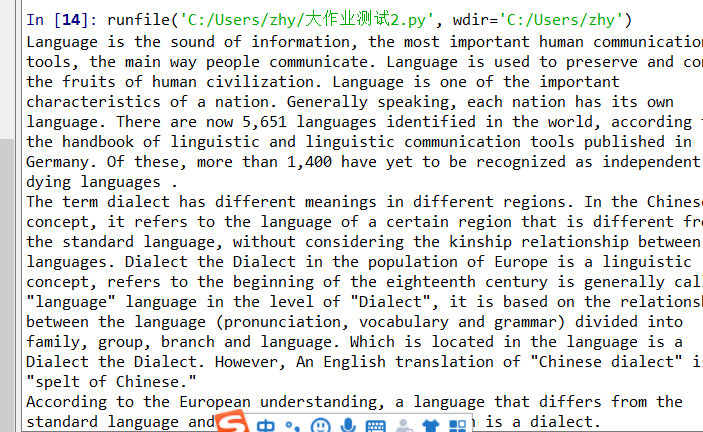
详细分析实验的实现方法，介绍你的设计用到了哪些函数或者涉及哪些知识点；详述每个模块的关键语法及其实现。并**给出关键代码**。

五、程序运行效果

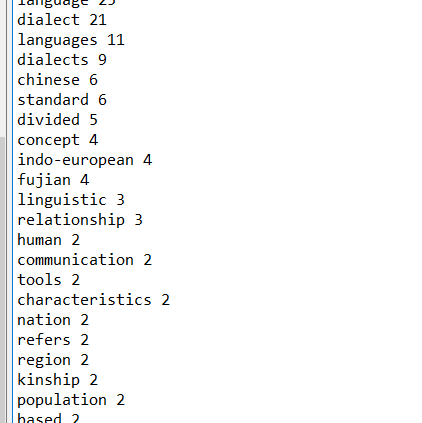
GUI截图

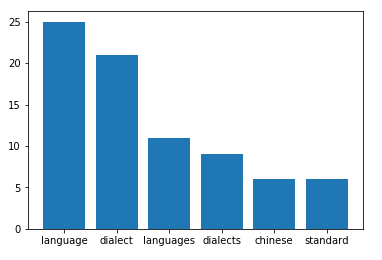


GUI有四项功能，显示文本、统计单词数、词频显示与高频词查询、退出程序

显示文本，可显示txt文件内容

统计单词数

按照降序给出单词频率

给出最高频率实词的柱状统计图

六、创新之处

使用了正则表达式除去符号和数字

七、实验总结

在处理文本中的符号时，曾经希望用split（）再合并的形式，但是操作太过麻烦，所以通过网上搜索，使用正则表达除去符号和数字

Tkiner未加入mainloop时使用按键后父窗口会直接消失，加入mainloop则正常使用

在使用字典的频率排序时，学会使用了operator函数

2、有何收获和体会？你认为大作业题目还可以从哪些角度出题？并说明理由。

复习了列表，字典，绘图，tkinter的使用方法。综合使用，看到自己的GUI出现，很有成就感。

我觉得可以设计一些实用的GUI程序， 比如最小二乘法，lagrange插值的计算。这些程序锻炼了我们对于绘图和函数处理方面的知识在我们日常的学习生活中也会用到。使用GUI封装会方便使用

八、课程学习总结

**1、课程收获和难点分析**

通过学习本课程你收获了什么？你认为现有内容中的难点在于哪些章节？

我学会了计算机的基本组成和基本原理、算法的设计和优化、数据储存结构、绘图和GUI设计的相关知识。我认为难点在于问题的抽象化，设计算法（尤其是动态规划方面的内容）

**2、课程评价**

从课程体系的先进性和合理性、课程内容、实验内容、考核环节、课程管理等方面进行评价。

课程设置较为紧凑，学习内容也很丰富。实验题考察全面，但是存在着部分偏题怪题。有时候会使用QQ群考核，对于不经常使用QQ的人可能不太友好，建议改进。最后的大作业可以考虑提前布置，考期时间较为紧凑。

总体来看，这门课对于学生很有意义。大学生计算机课程内容值得传承，而在课程的一些相关考核上可以进行改进。

**3、教师授课评价**

从工作态度、责任心、是否认真备课、讲课水平、教学质量、是否注重课堂互动、重点难点是否突出、逻辑是否清晰、授课效果、是否关心学生、是否耐心及时解答学生问题等方面进行评价。

万老师非常认真负责。万老师每次是进行精心的备课的，上课的时候也特别认真，让我们有回到高中课堂的感觉。

**4、助教评价**

从工作态度和责任心、是否耐心进行实验辅导、是否认真批改作业并及时反馈、是否及时解答学生问题等方面进行评价。

助教是友善而负责的，对于实验上机有问必答，并且能进行启发式的教学。课后作业反馈及时

**5、课程进一步改进建议**

欢迎同学们对于课程体系、教学内容、实验环节、考核方式、大作业形式等各个方面，提出各种建议！我们将认真听取同学们的建议，不断改进课程。

大学生计算机基础是一门很棒的课程，我们也在其中学到很多有用的知识，它甚至是很多同学的计算机启蒙。不过大一下课程有一些紧凑，大计学分也偏少。这样看大作业和日常作业就比较占用时间。希望能合理安排课程管理和学分，让大计的课程效果更好！

九、主要参考资料

<https://www.cnblogs.com/zywscq/p/5441145.html>

https://blog.csdn.net/Dongfnag\_HU/article/details/85076819