# 广州理工学院实验报告

**学院：计算机科学与工程学院 专业： 数据科学与大数据技术 年级： 2022**

**姓名： 吴恩威 学号： 20220409430251 组 实验时间：2023.12.23**

**指导教师签字： 成绩：**

|  |
| --- |
| **实验项目名称：**  实验12 扩展库应用   * **实验目的和要求：**   1、熟练掌握Python常用的扩展库；  2、了解扩展库Numpy、Pandas、matplotlab等库的功能；  3、了解使用扩展库实现程序设计与应用； |
| * **主要仪器设备：**   PC机一台；Windows操作系统；Python运行环境；开发工具（PyCharm)。 |
| * **实验内容和原理：**   1、扩展库的安装及导入使用；  2、扩展库方法的功能与应用；  3、简单扩展库实现程序功能。 |
| * **操作方法和实验步骤（绘图）：**   1、使用Numpy、PIL等库将本校的校徽LOGO（下图）转换为Ndarray数组，并使用matplotlib重新绘制图像输出显示。    操作方法与步骤：   1. 导入PIL的Image，使用Image模块的open函数，将本地文件读取为图像。 2. 导入numpy库，使用asarray将图像转换为Ndarray。 3. 导入matplotlib，将转换后的Ndarray重新绘制成图像   过程代码截图如下：    2、使用jieba库提取《西游记》中的关键字（50个），并使用wordcloud库实现统计《西游记》人物出现次数，并生孙悟空蒙层词云图。  操作方法与步骤：   1. 导入jieba结巴库，使用该库相关的函数实现《西游记》中关键字（可连同词性）的提取，并打印输出。 2. 调用wordcloud库中的Wordcloud类以及generator方法实现词云类的生成，以及词云的显示，注意使用mask参数实现蒙层的设置。   过程代码截图如下:    3、“集体过马路”是网友对集体闯红灯现象的一种调侃，即“凑够一撮人就可以走了，与红绿灯无关。”出现这种现象的原因之一是很多人认为法不责众，从而不顾交通法规和安全，但这种危险的过马路方式造成了很多不同程度的交通事故和人员伤亡。现通过调查形式对过马路进行测试，在所有参考调查的市民中，“从不闯红灯”、“跟从别人闯红灯”、“带头闯红灯”的人数如下表所示，针对这组调查数据，使用Python扩展库pandas、matplotlib编程实现绘制柱状图进行展示和对比。   |  |  |  |  | | --- | --- | --- | --- | | 性别 | 从不闯红灯 | 跟从别人闯红灯 | 带头闯红灯 | | 男 | 450 | 800 | 200 | | 女 | 150 | 100 | 300 |     操作方法与步骤：   1. 导入库函数matplotlib.font\_manager，设置图例字体，调用系统字体实现。 2. 创建DataFrame结构，存储男女两列的相关数据。 3. 使用plot绘制柱状图。设置X轴刻度和文本，设置y轴只在有数据的位置显示刻度，显示柱状图。   过程代码截图如下： |
| * **实验结果与分析**  1. 实验结果如下：     结果分析：（1）导入模块：from PIL import Image: 导入PIL库中的Image模块，用于处理图像文件。import numpy: 导入NumPy库，它提供对多维数组对象的支持，通常与图像处理一起使用。（2）读取图像：image = Image.open('image/学校Logo.png'): 使用PIL的Image.open()函数打开名为'学校Logo.png'的图像文件，并将其存储在image变量中。（3）转换为NumPy数组：nd = numpy.asarray(image): 将PIL图像对象转换为NumPy数组。这样做可以使得图像数据以NumPy数组的形式进行处理，方便进行进一步的操作和分析。（4）打印数组：print(nd): 打印NumPy数组nd的内容。这会将图像转换为数组并显示其像素值。（5）使用Matplotlib展示图像：from matplotlib import pyplot: 导入Matplotlib库中的pyplot模块，用于绘图和数据可视化。pyplot.imshow(nd): 使用Matplotlib的imshow()函数显示NumPy数组nd中的图像数据。pyplot.show(): 显示绘制的图像。   1. 实验结果如下：     结果分析：（1）文本处理部分：使用open()函数打开名为“西游记.txt”的文本文件，并使用GBK编码读取文件内容。使用jieba.analyse.set\_stop\_words()设置停用词路径，并利用jieba.analyse.textrank()提取50个关键词，限制词性为人名（'nr'）。（2）词性标注和词频统计：使用jieba.posseg.lcut()进行分词和词性标注。根据词性将人名词语放入name\_list列表中，将其他词语作为停用词放入stop\_word列表中，并统计词频。（3）生成词云：将人名词语列表连接成一个字符串。使用imageio.imread()读取一张名为“孙悟空.jpg”的图片作为词云的形状。使用wordcloud.WordCloud()初始化词云对象，设置字体、背景颜色、停用词、词云形状等参数。最后使用cloud.to\_file()保存生成的词云图像到指定路径。在这段代码中，可能出现问题的地方包括：（1）路径问题：图片路径（'image\孙悟空.jpg'）中的反斜杠可能引发转义字符问题。确保路径正确，并尝试使用原始字符串表示（加上r前缀）或者使用正斜杠来代替反斜杠，例如 'image/孙悟空.jpg'。（2）词云保存问题：保存词云图像时，请确保文件路径包含文件名和正确的文件格式，比如 .png 或 .jpg。  3、实验结果如下：    结果分析：（1）导入库：import pandas as pd: 导入 Pandas 库并使用 pd 别名。import matplotlib.pyplot as plt: 导入 Matplotlib 库中的 pyplot 模块并使用 plt 别名。import matplotlib.font\_manager as fm: 导入 Matplotlib 的字体管理器模块并使用 fm 别名。  （2）设置中文字体：myfont=fm.FontProperties(fname=r'C:\Windows\Fonts\SimSun.ttc')：设置字体为宋体，用于中文显示。   1. 创建数据框：df = pd.DataFrame({'男':(450,800,200),'女':(150,100,300)})：创建了一个包含男女人数的数据框。 2. 绘制柱状图：df.plot(kind='bar', color=['red', 'blue'])：利用 Pandas 的绘图功能绘制柱状图，男女人数以红色和蓝色进行区分。 3. 设置坐标轴和标签：plt.xticks([0,1,2],['从不闯红灯','跟从别人闯红灯','带头闯红灯'],color='black',fontproperties=myfont)：设置 x 轴刻度标签。plt.yticks(list(df['男'].values)+list(df['女'].values))：设置 y 轴刻度标签。plt.ylabel('人数', fontproperties=myfont, fontsize=14)：设置 y 轴标签。plt.title('集体过马路方式', fontproperties=myfont, fontsize=20)：设置图表标题。   （7）显示图表：plt.show()：显示绘制的柱状图。 |

**批阅评语：**