**《Python高级应用》实验报告**

实验名称 ： 实验3 文本分析及可视化

实验日期 ： 　 2024/10/14

姓 名 ： 　高心阳

学 号 ： 　 084623237

班 级 ： 　计算机232

成 绩 ：

**人工智能与信息技术学院**

**南京中医药大学**

|  |
| --- |
| **实验目的：** |
| 1. 熟悉字典操作以及文本分析库函数 2. 掌握数据可视化的技术 |
| **实验内容和要求** |
| 1. **字典练习**   ① 创建字典D={“张三”:88,“李四”:90,“王五”:73，“赵六”:82}  ② 向字典中添加键值对“钱七”:90”  ③ 修改“王五”对应的值为93  ④ 删除“赵六”对应的键值对   1. **文本分析及可视化**   ① 统计出《三国演义》排名前十五的人物以及其出现的次数。使用词云展示出现的人物。  ② （选做）设置词云的形状。  ‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬‬   1. **社交网络图(P226)**   ① 根据Jazz.txt文件，绘制音乐人社交网络图。  ② 由于前图过于拥挤，因此请在此基础上，改进代码，基于Jazz\_test.txt，生成音乐人网络图。 |
| **运行结果（写清题号）** |
| 1.    2.   |  | | --- | | 2.py | | import re import wordcloud import matplotlib.pyplot as plt import numpy as np from PIL import Image  text = open("../第三次实验材料/三国演义utf8.txt","r",encoding="utf-8").read()  text = re.sub("[^\u4e00-\u9fa5]", "", text)  names=[]# 隐藏，内容为三国演义中所有的人名  names\_count = {name: 0 for name in names}  for name in names:  count = len(re.findall(name, text))  names\_count[name] = count  names\_count = sorted(names\_count.items(), key=lambda x: x[1], reverse=True) print(names\_count[:15])  mask = np.array(Image.open("shape.png")) wc = wordcloud.WordCloud(font\_path="C:/Windows/Fonts/simhei.ttf", width=800, height=600, background\_color="white",mask=mask) wc.generate\_from\_frequencies(dict(names\_count[:30]))  plt.imshow(wc, interpolation='bilinear') plt.axis('off') plt.show() |   输出结果：[('曹操', 945), ('吕布', 365), ('魏延', 322), ('姜维', 313), ('刘备', 297), ('袁绍', 251), ('周瑜', 241), ('马超', 219), ('邓艾', 194), ('黄忠', 189), ('张郃', 184), ('孟获', 161), ('张辽', 147), ('曹仁', 145), ('徐晃', 139)]      3.1   |  | | --- | | 3.py | | import networkx as nx import matplotlib.pyplot as plt import numpy as np import re  G=nx.Graph() node\_list=[] lnum=0 with open(r"../第三次实验材料/Jazz.txt","r") as file:  while True:  lines=file.readline()  if not lines:  break  lnum+=1  if lnum>=4:  temp=' '.join(re.split(' +|\n+',lines)).strip()  line=re.split(' ',temp.strip())  first=line[0]  second=line[1]  node\_list.append(np.append(first,second))  print(len(node\_list))  for i in range(len(node\_list)):  G.add\_edge(node\_list[i][0],node\_list[i][1])  nx.draw(G,node\_color='r',node\_size=50) plt.show() |       3.2   |  | | --- | | 3.py | | import networkx as nx import matplotlib.pyplot as plt import numpy as np import re  G=nx.Graph() node\_list=[] lnum=0 with open(r"../第三次实验材料/Jazz\_test.txt","r") as file:  while True:  lines=file.readline()  if not lines:  break  lnum+=1  if lnum>=4:  temp=' '.join(re.split(' +|\n+',lines)).strip()  line=re.split(' ',temp.strip())  first=line[0]  second=line[1]  node\_list.append(np.append(first,second))  print(len(node\_list))  for i in range(len(node\_list)):  G.add\_edge(node\_list[i][0],node\_list[i][1])  pos = nx.spring\_layout(G) plt.figure(figsize=(10, 10)) nx.draw(G, pos, node\_color='r', node\_size=20, edge\_color='gray', alpha=0.5, with\_labels=False) plt.show() | |
| **实验的体会与建议** |
|  |