**《Python玩转大数据》实验报告（第2次实验）**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **学号** | **2017217713** | **姓名** | **王子翔** | **班级** | **计算机科学与技术17-1班** | **日期** | **04.20** |
| **实验名称： python第二次实验** | | | | | | | |
| **实验目的：**  ⒈熟悉Python编程环境  ⒉熟悉jieba、numpy、pandas、matplotlib库的使用  ⒊掌握对文本文件的字、词频率分析方法。  ⒋掌握文件操作。  ⒌掌握函数的绘图方法。 | | | | | | | |
| **实验内容：**  ⒈词频分析  ⑴自己从网上搜索下载若干篇中文或英文文本文件；  ⑵编写Python程序代码，对下载的文档进行词频统计(应去除无关的构词，可增添新创的构词)；  ⑶根据词频统计的结果构建自己的分析模型，分析这篇文档的内涵(如政府工作报告中出现的那些最频繁的词与国家治理发展的关系)；  2.字符数据加、解密  ⑴下载一片英文文章。  ⑵通过文件读入字符串，剔除其中的符号，自保留字母数字。  ⑶通过凯撒加密对文件加密（秘钥为13），并保存到Cy0204.txt中。  ⒊函数绘图   1. 安装numpy、pandas、matplotlib库。 2. 自定义函数，绘制函数图形。 | | | | | | | |
| **实验分析：**  实验一使用的文件为巴黎圣母院着火的新闻报道. 包含了纽约时报, CNN, BBC和RFI的四篇报道. 通过re.spilt正则化提取关键字, 并使用stopwords删去无意义的介词. 最终分析结果显示排行靠前的词为: fire, said, Paris, Cathedral, roof, flames, cathedral, smoke….可以看出巴黎圣母院着火和’教堂’, ‘屋顶(着火)’, ‘火焰’, ‘巴黎’等都有关系.  实验二使用了re.sub删除除了字母数字以外的字符, 再通过chr改变字符的值加密. 最后保存到Cy0204.txt.  实验三使用了numpy作为数学计算工具和matplotlib作为绘图工具. 分别在-5,5取1000个点绘制arcsin(x)/tan(x)的图像. 反复试来试去觉得这个函数最好看就选了这个函数. | | | | | | | |
| **实验结果（运行结果截屏）** | | | | | | | |
| **第1题结果** | |  | | | | | |
| **第2题结果** | |  | | | | | |
| **第3题结果** | |  | | | | | |