黔南民族职业技术学院

**实 验 报 告**

实验课程 数据挖掘技术

实验项目 Python数据挖掘与分析

实验地点 408实训室

指导教师 姜永成

班 级 20大数据技术与应用(5)班

学生姓名 何儒林

学 号 202003170519

教师评分

日 期

**期末实验 Python数据挖掘与分析**

**一、实验目的**

1. 掌握数据挖掘的基础知识，能够深刻理解并熟练运用Python进行数据处理的过程。

2. 掌握数据获取与收集方法。

3. 学会数据清洗和整理的方法。

4. 了解数据统计方法。

5. 能够利用可视化工具进行数据的展示。

1. **实训任务**

**任务1：【分析网页url】**

爬取的网站名称为：交通旅游网站

爬取内容：交通旅游的排行榜

网页地址为：https://top.chinaz.com/hangye/index\_jiaotonglvyou.html

网址规律：

第一页网址为：https://top.chinaz.com/hangye/index\_jiaotonglvyou\_2.html

第二页的网址为：https://top.chinaz.com/hangye/index\_jiaotonglvyou\_3.html

以此类推可以发现该网址分为三部分，每翻一页网址的变化都为N+1，html

|  |
| --- |
| 源码如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获15.PNG捕获15  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获17.PNG捕获17 |

**任务2：【利用scrapy框架方法爬取交通旅游排行榜据】**

从网页的分析来看，所需爬取的数据为：

“名字”，“周排名”，“网站”，“反链数”，“得分”，

|  |
| --- |
| 源码如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获18.PNG捕获18  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获19.PNG捕获19 |

**任务3：【数据存储】**

将所爬取的相关数据存入到csv

|  |
| --- |
| 源码如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获15.PNG捕获15 |

**任务4：【数据清洗和导入MYSQL】**

将所爬取的相关数据导入到mysql

|  |
| --- |
| 源码如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获20.PNG捕获20  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获21.PNG捕获21  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获22.PNG捕获22 |

**任务5：【数据分析与可视化展示】**

**可视化（1）:**

使用柱形图可以得知以下那个交通旅游得分高

|  |
| --- |
| 源码及图形如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获7.PNG捕获7  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获6.PNG捕获6 |

**可视化（2）:**

利用雷达图表达交通旅游分类的情况，可以准确得知旅游类占比较多

|  |
| --- |
| 源码及图形如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获13.PNG捕获13  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获12.PNG捕获12 |

**可视化（3）:**

利用饼图可看出在在交通旅游排行榜中反链数的占比最高

|  |
| --- |
| 源码及图形如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获11.PNG捕获11  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获10.PNG捕获10 |

**可视化（4）:**

用散点图分析交通旅游那家推荐指数高

|  |
| --- |
| 源码及图形如下 |
| C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获9.PNG捕获9  C:\Users\deins\Desktop\何儒林\WebsiteLeaderboards\截图\捕获8.PNG捕获8 |

。