**大连理工大学开源软件基础大作业**

**关于OSU玩家的数据统计**

学院（系） \_\_\_\_\_\_ 软件学院\_\_\_\_\_\_\_\_\_

专 业 \_\_\_ 软件开发与测试\_\_\_\_\_\_

组 长 \_宋宏良\_ 学 号 201892371 工作量 50%

组 员 \_邢子涵\_ 学 号 201892152 工作量 20%

组 员 \_邹维海\_ 学 号 201892120 工作量 10%

组 员 \_林佳豪\_ 学 号 201892241 工作量 10%

组 员 \_刘镇硕\_ 学 号 201892146 工作量 10%

大连理工大学

Dalian University of Technology

# 项目背景

本次大作业主要统计了一个开放社区音乐节奏游戏 “OSU！”的玩家分布情况和不同地区玩家的得分情况。代码部分主要先使用request相关函数获得网页源代码，并使用了两种方式以获得目的数据。对于OSU！部分主要使用BeautifulSoup针对标签获得数据。而在F-work.py文件中后加入的“查看热搜”功能则是使用re来获得目的数据

OSU！玩家数据保存于数据库中并生成图片，在F-work.py文件调用数据库与图片并利用Flask在网页上显示出图片与表格。热点显示功能由于在F-work.py中直接进行request和return，并没有经过数据库存储至本地。

通过这个大作业，OSU！玩家可以方便的查看排名以及当前国内的热点新闻。

**目 录**

[项目背景 2](#_Toc60254097)

[**1** **代码说明** 4](#_Toc60254098)

[数据获取 5](#_Toc60254099)

[数据展示 5](#_Toc60254100)

[其他 5](#_Toc60254101)

[**2 项目展示** 5](#_Toc60254102)

[主界面展示 5](#_Toc60254103)

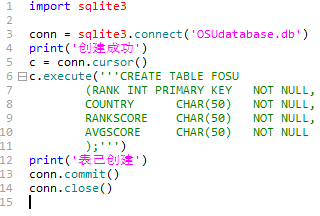
[内容页展示 5](#_Toc60254104)

[控制台 5](#_Toc60254105)

[**3 感想与总结** 5](#_Toc60254106)

# **代码说明**

**数据库的创建**



数据获取

由于代码部分完全整进来有点长，这里就展示具有代表性的代码段。完整代码会以电子版的形式发送。

玩家数量数据的取得主要通过F-osu.py文件

如下图



F-osu.py文件

19—25行完成网页源代码的获取

26—38行完成将整理过的数据存入列表

39—42行实现画图并保存图片

数据展示

这里先以F-work.py文件中的一个展示列表的函数为第一个例子

如图所示



基本上就是简单地读取数据库，将其拼接到html代码中并在最后return拼接好的字符串。通过这种方式就能在网页上看到一个简单地表格。其余显示表格函数同理。

其他

代码在最后初步完成阶段时，听到好友的一句“你怎么不去爬一爬热搜”时灵感迸发，就去爬了。爬热搜的代码直接写入了F-work.py文件中，没有设计数据库的读写，在运行时会稍微有些慢。如图



爬取百度热搜的代码

这里……百度热搜有点问题，在F-work.py文件开头提到了。我的好朋友也写了类似的大作业或许在他那里能看到合适的解决办法。问题简单来说就是：能够爬到的热搜有两份，两份之间有点细微的差别，不知道为什么。

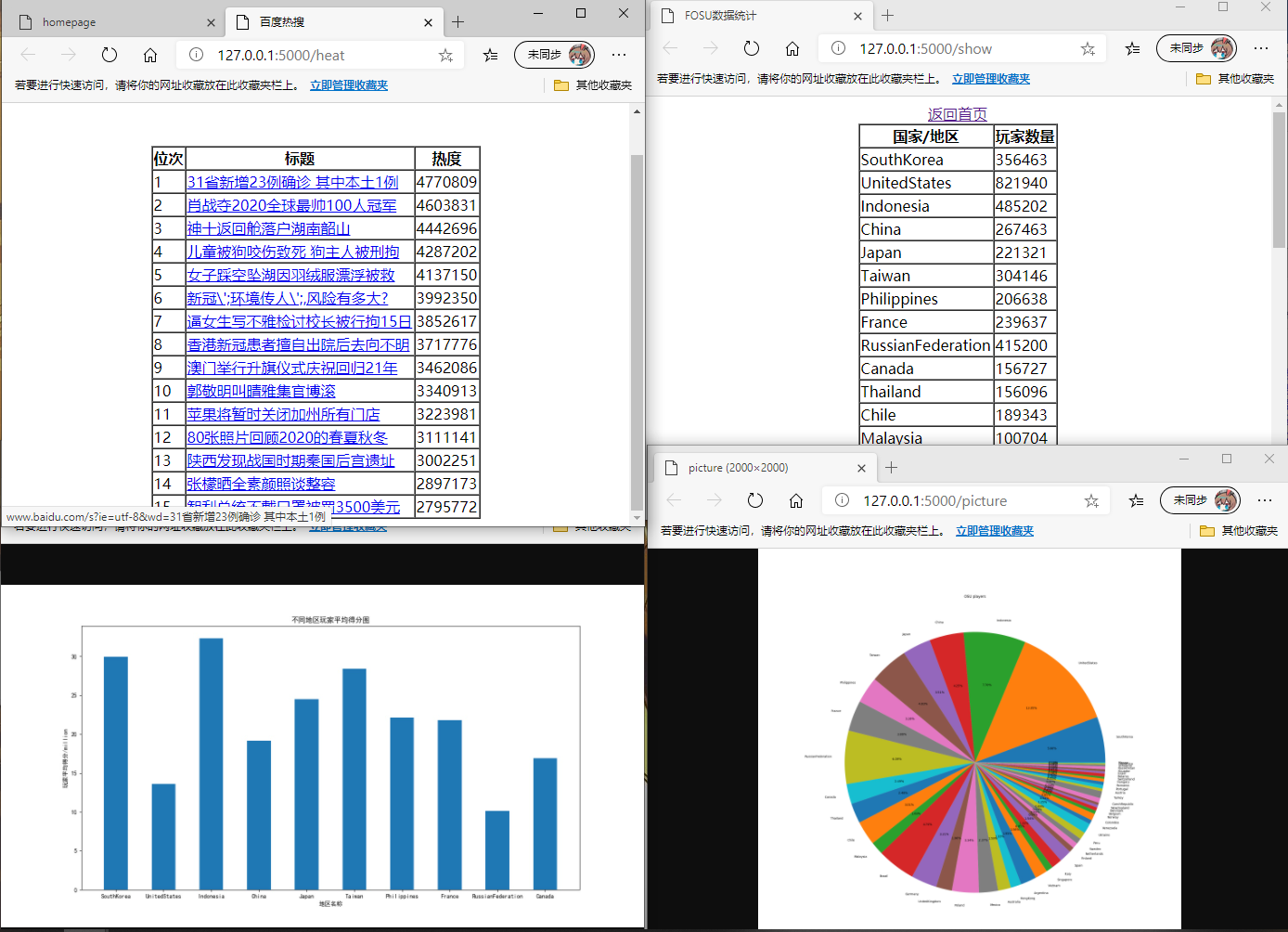
另外，展示百度热搜页面设置了超链接，可以通过点击就能很方便的直接跳转到对应页面。此外还对主页进行了设计？总之比原来的一堆不友好的文字提示友好了一点。会在下面的项目展示部分进行展示。

# **2 项目展示**

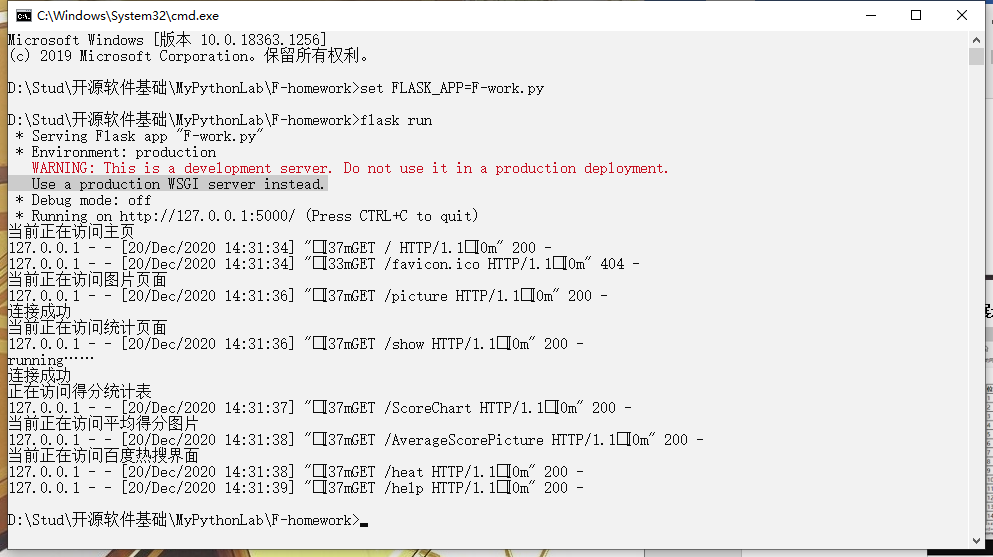
主界面展示



内容页展示



控制台



# **3 感想与总结**

十分感谢百度的力量，让我解决了遇到的一些问题。这个项目第一个大问题就是图片展示了。之前写别的代码时，把文件路径一些，就好了。就像是这样子

<img src="文件路径" style="height:280px;width:560px;">

然而python这边的Flask似乎不能这么干。上网搜了好久后来发现似乎是flask的问题不能直接return。最后借助以下文章实现了图片显示……我找不到那篇文章了……

现在项目写完了一看，其实内容不是很多。虽然后来组员又跟着补了一点其他的内容但是还是感觉不够丰满。写来写去也就是那么几个东西。而且最后实现的时候，说句实在话页面有点丑。HTML没怎么学还想用Flask就会导致这个结果吧……越说越觉得自己写出来的是一堆垃圾了。以后有机会的话，会在这个项目的基础上修修改改再整点花样出来吧，但是现在只能走到这里了

——宋宏良

通过本次大作业，我对于Python爬虫爬取网站数据有了一个更加深刻的理解，并且结合gitlab和github工具完成了组员之间的合作交流,对于网站的结构都了一个大致的了解。另一个就是通过课程和大作业学习，我学习了python的各种知识，丰富了自己的编程语言，并且深刻认识到Python的简洁。

——林佳豪