**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

组 员 \_邢子涵\_ 学 号 201892152

组 员 \_邹维海\_ 学 号 201892120

组 员 \_林佳豪\_ 学 号 201892241

组 员 \_刘镇硕\_ 学 号 201892146

大连理工大学

Dalian University of Technology

# 摘 要

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。

**关键字：request；BeautifulSoup；re；sqlite3；**

**目 录**

[摘 要 2](#_Toc59368709)

[**1** **代码说明** 5](#_Toc59368710)

[数据库的创建 5](#_Toc59368711)

[数据获取 5](#_Toc59368711)

[数据展示 6](#_Toc59368712)

[其他 6](#_Toc59368713)

[**2 项目展示** 7](#_Toc59368714)

[主界面展示 7](#_Toc59368715)

[内容页展示 8](#_Toc59368716)

[控制台 8](#_Toc59368717)

[**3 总结** 9](#_Toc59368718)

**引 言**

本次开源软件基础大作业的内容为“统计OSU玩家的分布”，之所以选择这个内容主要是因为组长个人（也就是写这个大作业文档的我）很喜欢这款游戏，玩了3—4年觉得自己已经有了两把刷子才想到“要不，就把这个当成大作业？”。恰好这个游戏的官方网站有几个页面以表格的形式显示了玩家的分布情况。我一看，很好，很适合用来爬。于是就有了这个项目

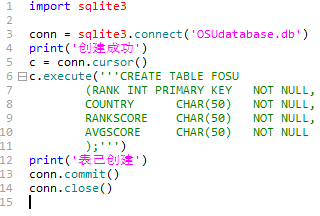
项目中涉及到的大部分库，函数基本都出自课堂内容（我也就记住这几个了，别的也不会用呀……）。大部分问题出在Flask上，这个项目最终显示出来的结果也没有达到预期。不过也有几个小设计是在项目确定时完全没有想到的。

代码部分不长，基本上都是在重复类似的内容。这里就不再过多赘述了。引言部分就写这么多吧。关键还是要看看正文嘛。

啊，对了，文档写完之后代码可能会发生细微的更改，因此文档中的演示的代码或许与实际代码略有不同。

# **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。

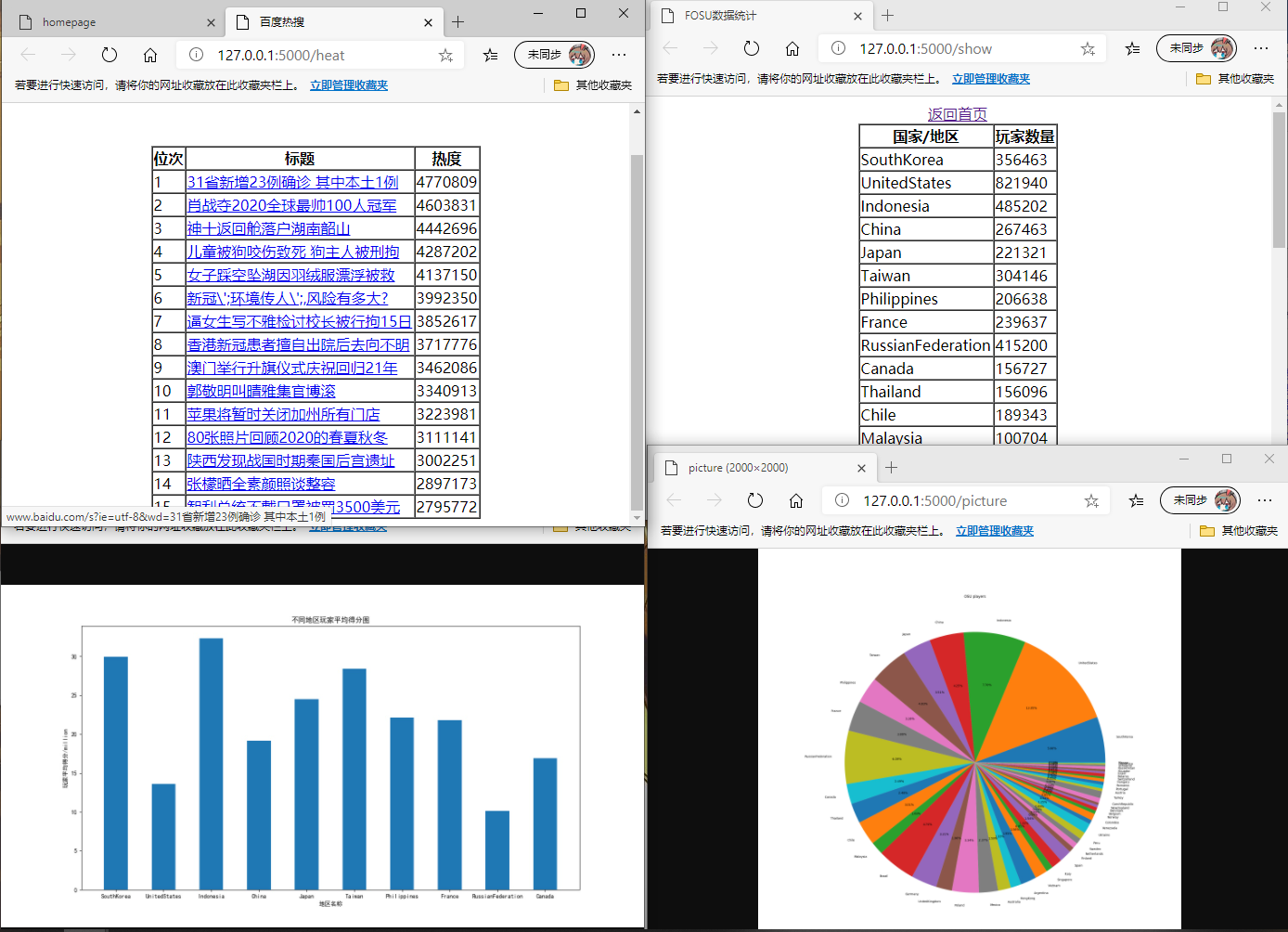
另外，展示百度热搜页面设置了超链接，可以通过点击就能很方便的直接跳转到对应页面。此外还对主页进行了设计？总之比原来的一堆不友好的文字提示友好了一点。会在下面的项目展示部分进行展示。

# **2 项目展示**

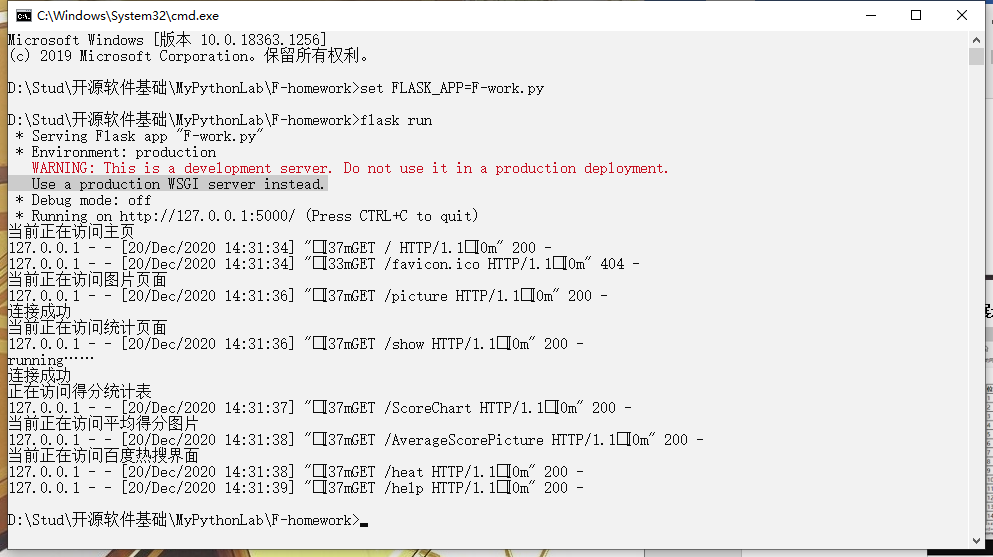
主界面展示



内容页展示



控制台



# **3 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……