|  |  |  |
| --- | --- | --- |
| 项目日志 | 工作内容 | 工作难点（问题） |
| 7.9 | 1. 提取数据 2. 网页加载模拟 3. 爬取用户名 设计爬取结构 4. 设计userinfo表 5. 获取了用户匹配对局信息 6. 讨论了具体的分析内容与所需要的数据 7. 学习echart相关 | 1. 可用信息的筛选 2. 数据获取问题 3. “平台区域”分片获取数据时，数据有时间限制 |

7.9 ：爬取用户名工作方面一直研究怎么提取可用和想要的数据，基本可以完成整个网页的加载模拟，但是在信息筛选上遇到了一些小问题--只能获取到整个网页（如图13），经小组成员讨论改正后可以提取所需用户数据（如图14）。其次是对指定的网址进行爬取，并设计了爬取的结构，初步计划利用队列对用户进行2层爬取，并对重复数据进行清理，数据保存到本地且通过数据库插入，设计userinfo表，储存用户名。通过事务的方法，首先是将所有数据合并为一条sql语句，并进行提交，发生错误后进行回滚，并重新提交。

详细数据获取和整理工作方面发现在“平台区域”分片获取数据时，数据有时间限制，发现只能搜索2018年12月份前的数据，如：shard="pc-as"，为亚洲服,现在PUBG官方优化了搜索范围，直接使用steam进行搜索，成功获取到玩家近期14天内的全部对局id。起初有遇到过获取不到数据的情况，查阅官方文件可知18年后 分片选项不需要加分区，最后能得到的数据如图22。

数据分析与整理展示工作方面讨论了具体的分析内容与所需要的数据，并给出了小结（如图23），学习了echart相关。

以下是项目过程中的图值：

