

个人报告——沈成泽

1、分工任务的解决方案

我的分工任务为数据源的获取以及数据结构设计。

数据源网站对于爬虫程序有一定的防范，因此这里采用了模拟浏览器进行访问的方法，并通过批量访问需求数据所在的页面来获取所需数据。

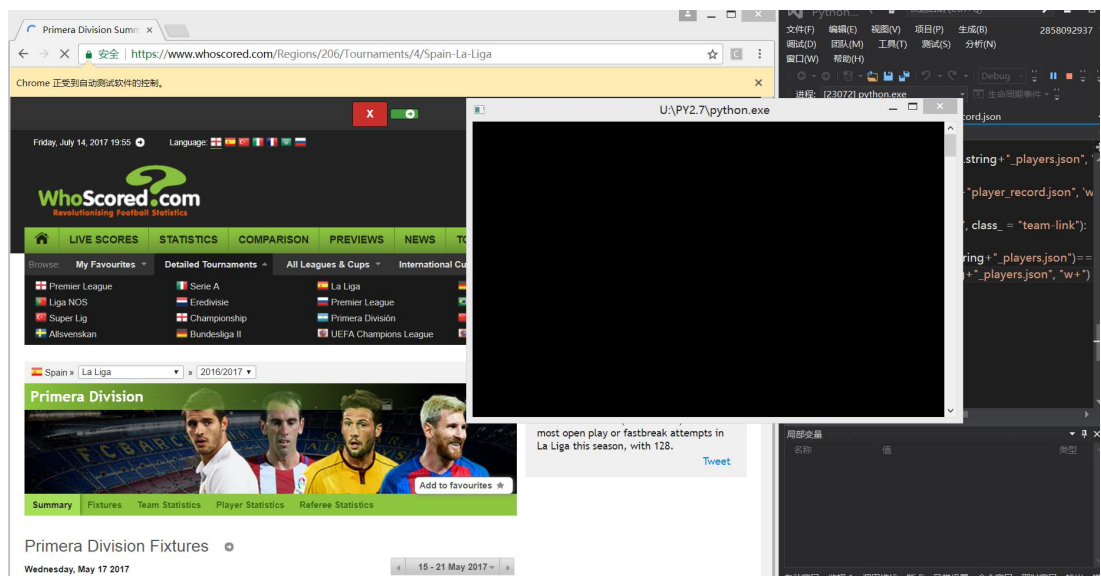
数据的存储结构设计需要考虑到数据类型和数据访问需求，在后文会进行描述。

2、设计思路

既然确定了网站的访问模式，数据获取就成了简单的页面元素索引工作。我设计了几个不同的访问程序，分别获取联赛的积分榜、球队的基本数据、球队的队徽、球队内的球员数据以及最重要的近期比赛数据。之所以未将这几个程序合并，是因为网站的访问过程不甚稳定，过于庞大的数据吞吐可能会频繁的失败。为此我还为几个访问程序分别设计了访问记录，来记录下已经获取成功的场次/球队，避免重复获取。

数据的存储采用了文件格式的存储方式。一开始的想法是直接建立一个总的 json 型数据文件，其中包含了所有我们要访问的数据，这样只需访问一个文件即可。但考虑到数据量和访问方式，我最终将数据按照球队分开保存。如此一来虽然文件增加，但利用球队名索引文件会比较容易，而且数据也更加直观。

3、运行效果图



这里给出数据获取过程中的一个截图，可以看到由程序驱动浏览器自动访问了数据源网页。

DeJia	2017/7/14 8:54	文件夹	
DejiaPlayer	2017/7/14 8:54	文件夹	
teamlist	2017/7/14 8:54	文件夹	
Xijia	2017/7/14 8:54	文件夹	
XijiaPlayer	2017/7/14 8:54	文件夹	
Yijia	2017/7/14 8:54	文件夹	
YijiaPlayer	2017/7/14 8:54	文件夹	
Yinchao	2017/7/14 8:54	文件夹	
YinChaoPlayer	2017/7/14 8:54	文件夹	
DeJia.txt	2017/7/14 8:54	TXT 文件	2 KB
XiJia.txt	2017/7/14 8:54	TXT 文件	2 KB
YiJia.txt	2017/7/14 8:54	TXT 文件	2 KB
YinChao.txt	2017/7/14 8:54	TXT 文件	2 KB

数据的存储结构如图，其中的具体结构在总报告中也有描述，这里不做赘述。

所有数据的处理按标准 json 格式进行。

4、心得及意见

在本次短学期中，我们收获了许多与平时课程所授不同的知识。相对于平时的理论课而言，本课程更加注重编程的实践，很多知识对于实际编程过程的帮助是立竿见影的，这对于我们来说显然是一种新鲜的体验。我们接触到了包括持续集成平台和 mvvm 模型在内的许多新的工具，这些工具也许我们现在还难以完全掌握，但将来一定对我们大有帮助。

意见方面，我认为本课程的授课内容很新颖也很有帮助，问题是通过外部的课程介绍，我们很难了解到本课程所真实讲授的内容，也许应该在课程介绍中丰富这些内容，让更多想要获

得实践经验的同学关注到这门课。