**数据库大作业个人报告**

姓名：陈彬彬

学号：16340015

项目名称：WorldCup\_Database\_Website

1. 个人分工

网页爬取新浪体育源的世界杯相关数据

设计worldcup数据库模型，包括games、playerdatas、players、scoreboard、teaminfo、teams、topscorer、users共8个实体集。

Java-JDBC连接数据库，实现其中两个实体的代码。(playerdatas、teaminfo)

1. 参考资料
2. Ronoob.com菜鸟教程：

菜鸟教程Java 8 新特性之Java My SQL连接：

<http://www.runoob.com/java/java-mysql-connect.html>

1. CSDN技术博客

2018俄罗斯世界杯之Java爬虫：

<https://blog.csdn.net/longyuhome/article/details/80732607>

1. 博客园

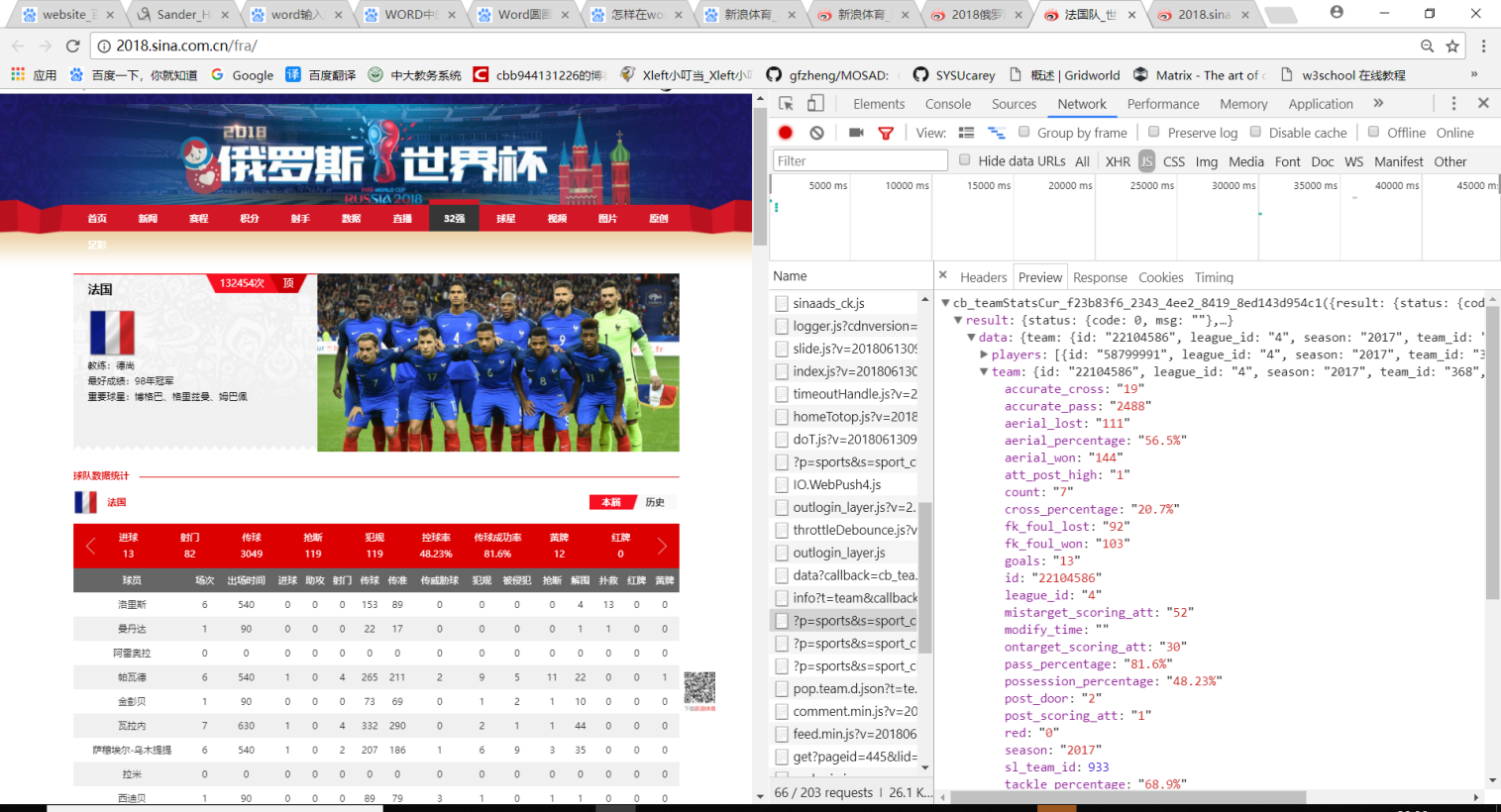
myspl插入表表中乱码的解决方法：

<https://www.cnblogs.com/houqi/p/5713176.html>

1. 实现步骤

**1.网页爬取新浪体育源的世界杯相关数据**

·网页访问新浪体育世界杯栏目网页，检查网页源代码，在Network，传输的JS信息中抓取对应的数据API：



以抓取球队数据统计信息为例子，抓取到球队数据统计信息API为：

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* 请求参数

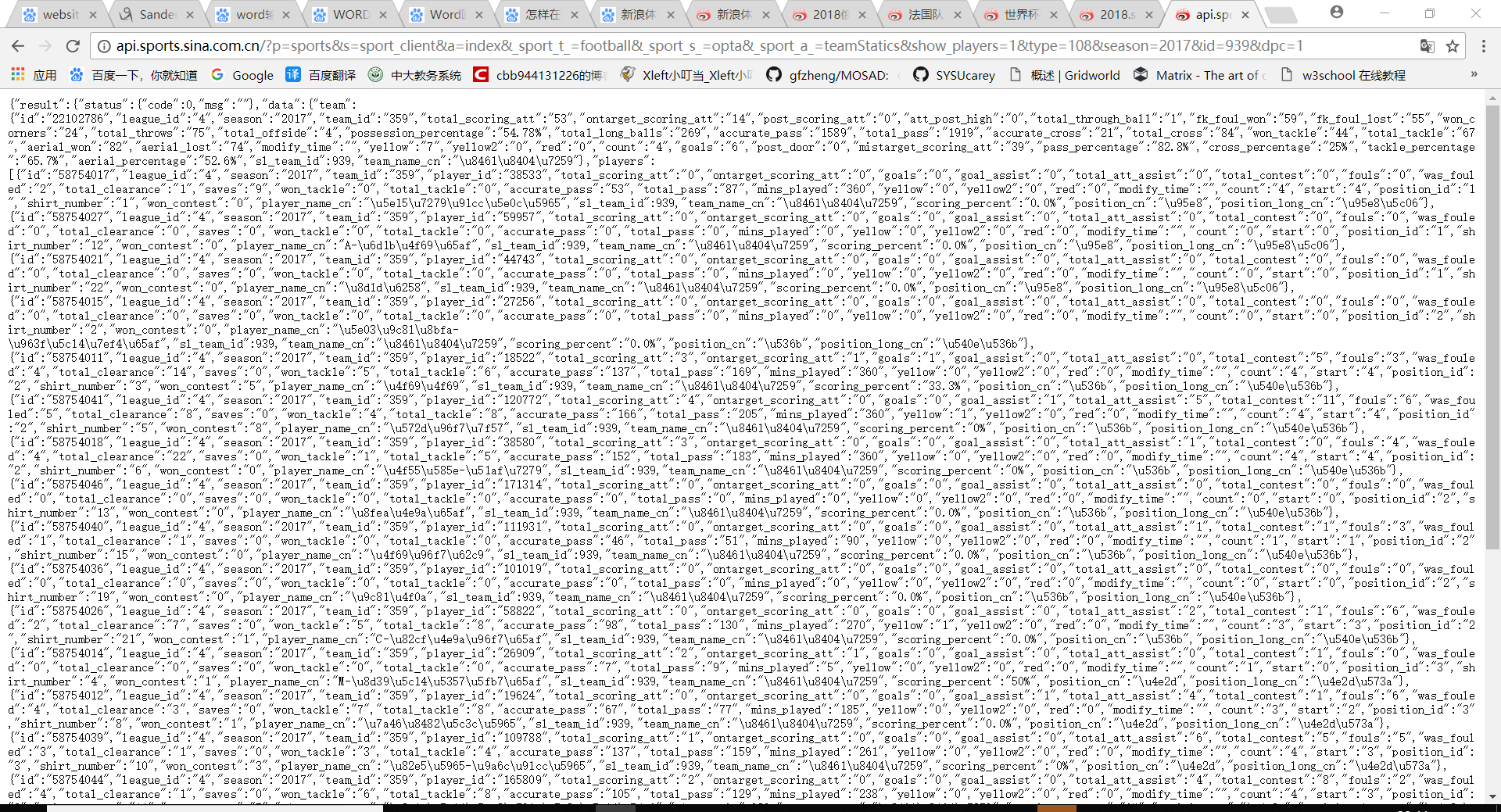
\* id是球队id： 与查询小组积分的的s1\_id属性值一致）

\* season是当届/历史： 2017是当届，2013上一届

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

<http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=teamStatics&show_players=1&type=108&season=2017&id=939&dpc=1>

API访问的是网页GET请求返回的Json数据，包含着球队的数据统计信息和球员的数据信息，大致如下图所示：



以下是抓取到本次项目所要用到的所有API：

具体的API信息，例如请求参数，返回格式请查看 ./Resource/API信息.docx

|  |  |  |
| --- | --- | --- |
| order | API信息 | API请求地址 |
| 1 | 查询分组积分情况 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=teamOrder&type=108&use_type=group> |
| 2 | 查询球队球员当届或历史整体统计数据情况 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=teamStatics&show_players=1&type=108&season=2017&id=939&dpc=1> |
| 3 | 查询射手榜 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=playerorder&type=108&item=13&limit=20&season=2017> |
| 4 | 查询球员基本信息 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=getPlayer&id=14937&type=108&dpc=1> |
| 5.1 | 查询球队赛程信息（按比赛时间查询） | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=livecast&_sport_a_=matchesbytype&type=108&fields=group%2Clivecast_id%2Cdate%2Cstatus%2CTeam1%2CTeam2%2Ctime%2CTeam1Id%2CTeam2Id%2CScore1%2CScore2%2CMatchCity%2CNewsUrl%2CVideoUrl%2CLi> |
| 5.2 | 查询球队赛程信息（按球队查询） | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=livecast&_sport_a_=groupMatchesByType&type=108&fields=group%2Clivecast_id%2Cdate%2Cstatus%2CTeam1%2CTeam2%2Ctime%2CTeam1Id%2CTeam2Id%2CScore1%2CScore2%2CMatchCity%2CNewsUrl%2CVideoUrl> |
| 6 | 查询球队基本信息 | <http://events.sports.sina.com.cn/worldcup/api/team/data?&season=2017&tid=939&dpc=1> |
| 7 | 查询球员各场次数据信息 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=playerMatchesStatics&id=14937&type=108&season=2017&dpc=1> |

**2.worldcup数据库模型设计**

根据爬取的API，我们用所爬取得信息建立我们的worldcup数据库模型，包括games、playerdatas、players、scoreboard、teaminfo、teams、topscorer、users共8个实体集。

具体的各实体集如下表所示:

##### 1.scoreboard

|  |  |  |  |
| --- | --- | --- | --- |
| scoreboard实体集——小组积分信息，查询API1 | | | |
| API1 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=teamOrder&type=108&use_type=group> | | |
| 属性 | 类型 | 属性含义 | 样例信息 |
| sl\_id | char(10) | 球队ID（主码） | 939 |
| team\_cn | char(50) | 球队名称 | 葡萄牙 |
| team\_order | char(5) | 球队排名 | 2 |
| count | char(5) | 场次 | 3 |
| win | char(5) | 胜 | 1 |
| lose | char(5) | 负 | 0 |
| draw | char(5) | 平 | 2 |
| goal | char(5) | 进球 | 5 |
| losegoal | char(5) | 失球 | 4 |
| truegoal | char(5) | 净胜球 | 1 |
| score | char(5) | 积分 | 5 |
| groups | char(5) | 小组 | B |
| logo | varchar(255) | 国家图片 | http://www.sinaimg.cn/lf/sports/logo85/939.png |
| row\_id | int | 记录所在行数，用于排序 | 6 |

##### 2.teams

|  |  |  |  |
| --- | --- | --- | --- |
| teams实体集——球队数据统计信息，查询API2 | | | |
| API2 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=teamStatics&show_players=1&type=108&season=2017&id=939&dpc=1> | | |
| 属性 | 类型 | 属性含义 | 样例信息 |
| sl\_team\_id | char(10) | 球队ID（主码） | 939 |
| team\_name | char(50) | 球队名称（主码） | 葡萄牙 |
| goals | char(5) | 进球 | 6 |
| total\_scoring\_att | char(5) | 射门 | 53 |
| total\_pass | char(10) | 传球 | 1919 |
| total\_tackle | char(5) | 抢断 | 67 |
| possession\_percentage | char(10) | 控球率 | 54.78% |
| pass\_percentage | char(10) | 传球成功率 | 82.8% |
| yellow | char(5) | 黄牌 | 7 |
| red | char(5) | 红牌 | 0 |
| fk\_foul\_won | char(5) | 任意球 | 59 |
| won\_corners | char(5) | 角球 | 24 |
| total\_cross | char(5) | 传中 | 84 |
| total\_offside | char(5) | 越位 | 4 |
| total\_throws | char(5) | 界外球 | 75 |
| post\_door | char(5) | 击中门框 | 0 |

##### 3.topscorer

|  |  |  |  |
| --- | --- | --- | --- |
| topscorer实体集——射手榜统计信息，查询API3 | | | |
| API3 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=playerorder&type=108&item=13&limit=20&season=2017> | | |
| 属性 | 类型 | 属性含义 | 样例信息 |
| row\_id | int | 记录所在行数,用于排序（主码） | 略 |
| num | char(5) | 排名（主码） | 略 |
| player\_name | char(50) | 球员名字 | 略 |
| player\_id | char(10) | 球员ID | 略 |
| team\_name | char(50) | 球队名称 | 略 |
| goals | char(5) | 进球数 | 略 |
| penalty | char(5) | 点球数 | 略 |
| owngoal | char(5) | 乌龙数 | 略 |
| assists | char(5) | 助攻数 | 略 |
| minutes | char(5) | 出场时间 | 略 |
| sl\_team\_id | char(10) | 球队ID | 略 |
| pic | varchar(255) | 球员图片 | 略 |

##### 4.players

|  |  |  |  |
| --- | --- | --- | --- |
| players实体集——球员基本信息，API4 | | | |
| API4 | <http://api.sports.sina.com.cn/?p=sports&s=sport_client&a=index&_sport_t_=football&_sport_s_=opta&_sport_a_=getPlayer&id=14937&type=108&dpc=1> | | |
| 属性 | 类型 | 属性含义 | 样例信息 |
| player\_id | char(10) | 球员id | 14937 |
| team\_id | char(10) | 国家队id | 939 |
| player\_name | char(50) | 球员中文名 | C罗 |
| team\_name | char(50) | 国家队中文名 | 葡萄牙 |
| position\_cn | char(50) | 场上位置 | 前锋 |
| club\_cn | char(50) | 俱乐部名 | 皇家马德里 |
| age | char(20) | 年龄 | 33 |
| birth\_date | char(20) | 出生 | 1985-02-05 |
| weight | char(10) | 体重 | 83 |
| height | char(10) | 身高 | 187 |
| jersey\_num | char(5) | 球衣号码 | 7 |
| pic | varchar(255) | 球员照片url | http:\/\/www.sinaimg.cn\/ty\/opta\/players\/14937.jpg |
| team\_logo | varchar(255) | 国家队Logo url | http:\/\/www.sinaimg.cn\/lf\/sports\/logo85\/939.png |

##### 5.games

略查看代码

##### 6.users

略查看代码

##### 7.teaminfo

|  |  |  |  |
| --- | --- | --- | --- |
| teaminfo实体集——球队基本信息，API5 | | | |
| API | <http://events.sports.sina.com.cn/worldcup/api/team/data?&season=2017&tid=939&dpc=1> | | |
| 属性 | 类型 | 属性含义 | 样例信息 |
| tid | char(10) | 球队id | 939 |
| team\_name | char(50) | 球队名称 | 葡萄牙 |
| coach | char(50) | 教练名 | 桑托斯 |
| family | char(50) | 球队合照 | http://n.sinaimg.cn/sports/83051191/20180606/PuTaoYa740.jpg |
| king | char(100) | 重要球星 | C罗、穆蒂尼奥、佩佩 |
| honor | char(100) | 最好成绩 | 第三名 |

##### 8.playerdatas

|  |  |  |  |
| --- | --- | --- | --- |
| playerdatas实体集——球员具体各场次数据信息，API6 | | | |
| API5 | <http://events.sports.sina.com.cn/worldcup/api/team/data?&season=2017&tid=939&dpc=1> | | |
| 属性 | 类型 | 属性含义 | 样例信息 |
| player\_id | char(10) | 球员id（主码） | 14937 |
| player\_name\_cn | char(10) | 球员姓名 | C罗 |
| home\_name\_cn | char(50) | 主队名称（主码） | 葡萄牙 |
| away\_name\_cn | char(50) | 客队名称（主码） | 西班牙 |
| home\_score | char(5) | 主队得分 | 3 |
| away\_score | char(5) | 客队得分 | 3 |
| mins\_played | char(5) | 出场时间 | 90 |
| goals | char(5) | 进球数 | 3 |
| goal\_assist | char(5) | 助攻 | 0 |
| total\_scoring\_att | char(5) | 射门数 | 4 |
| total\_pass | char(5) | 传球 | 36 |
| accurate\_pass | char(5) | 传准 | 34 |
| total\_att\_assist | char(5) | 传威胁球 | 1 |
| fouls | char(5) | 犯规 | 1 |
| was\_fouled | char(5) | 被侵犯 | 4 |
| won\_tackle | char(5) | 抢断 | 0 |
| total\_clearance | char(5) | 解围 | 1 |
| saves | char(5) | 扑救 | 0 |
| red | char(5) | 红牌 | 0 |
| yellow | char(5) | 黄牌 | 0 |
| status\_cn | char(5) | 比赛状态（主码） | 结束 |

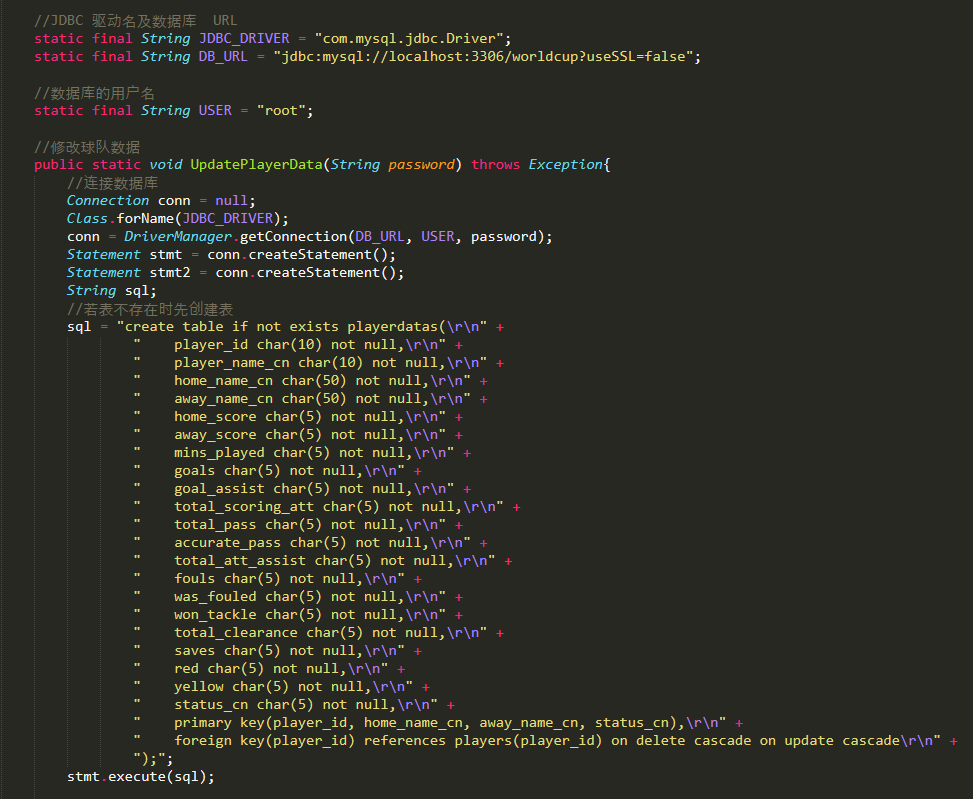
**3.playerdatas、teaminfo Java-JDBC代码实现**

两者的实现原理是差不多的，都是根据设计的worldcup数据库模型和对应的API，在Java代码中先进行网页访问API，得到返回的Json文件，然后解析得到的Json文件，获得对应的各项数据。

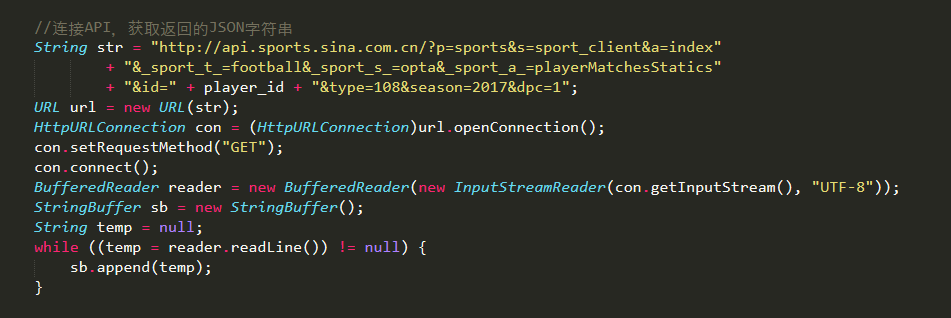
然后创建对应的实体集，用JDBC连接MySQL，实现对worldcup数据库中，创建playerdatas与teaminfo两个实体集，并且导入解析出来的各项数据。其中teaminfo共32项数据；playerdatas共2850条数据。

**关键代码：**

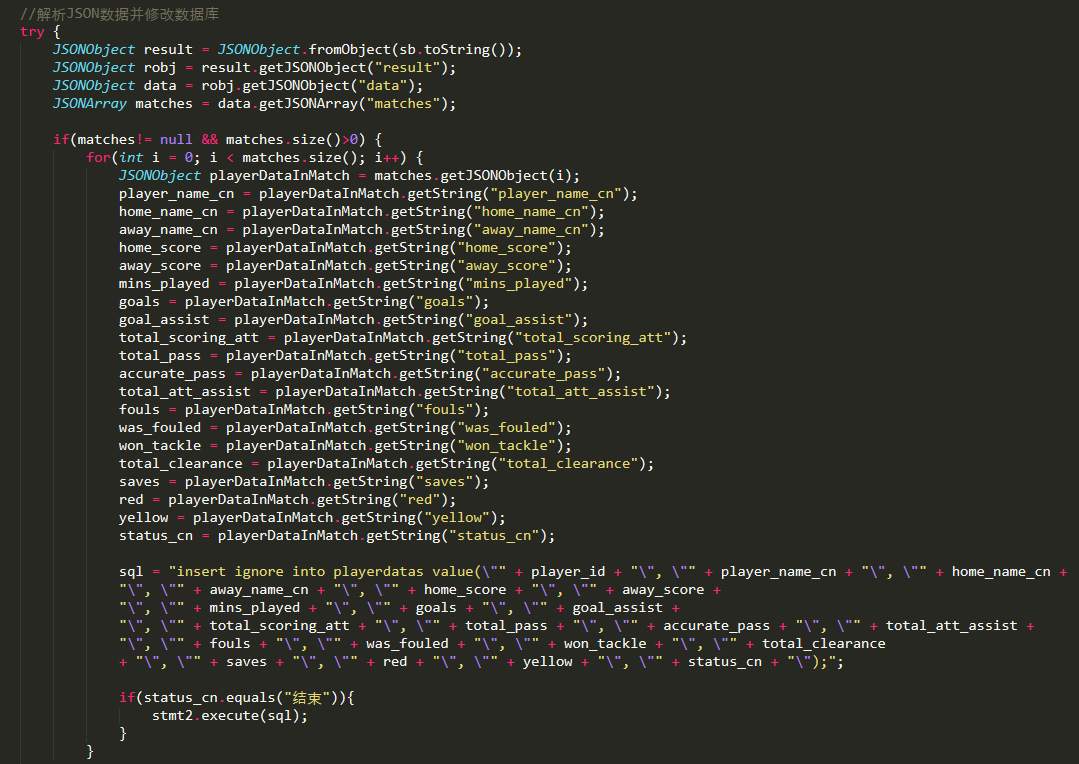
**1.JDBC连接数据库，创建playerdatas实体，设置实体拥有的各属性，主码关系，约束码关系：**



**2.网络API GET类型访问，获得返回JSON文件，进行Json解析，获得各项数据并插入数据库。**



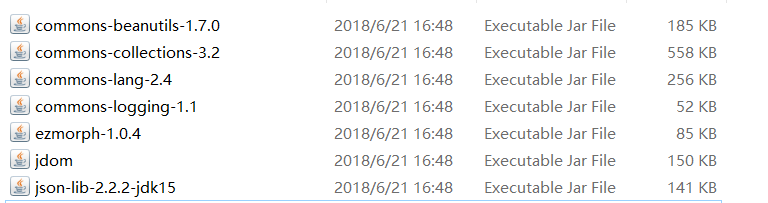
**3.将解析获得的数据，插入到worldcup数据库playerdatas实体中。**



1. 、遇到的问题

**1.Json解析问题**

一开始不知道如何处理返回得到的JSON文件，后来上博客查攻略：



引用这些外部库，可以用库里面的函数来进行Json解析，更容易地获得数据。

另外Json文件里面的信息内容过多过杂，在分解Json结构也浪费了很多时间。

**2.Eclipse Kepler Service Release 1的版本不兼容mysql-connector-java-8.0.11-bin.jar**

一开始用的JDBC链接器是**mysql-connector-java-8.0.11-bin.jar**这个版本，但是最后导入数据运行的时候一直是报错的，一开始是自己代码的问题，连接错误。修改代码后发现还是报错，报错位置是**mysql-connector-java-8.0.11-bin.jar**内部的源代码。后来上网查询错误信息时发现是与自己这里的Eclipse版本不匹配。于是降低JDBC链接器的版本至**mysql-connector-java-5.1.46-bin.jar**问题得以解决。

**3.导入2850条playdatas数据时，网络不稳定+用时略长导致顶层代码运行会报错。**

解决方法是对顶层代码文件导入数据时进行异常处理，抛出异常处理，避免报错。另外为了更好的响应，新开线程进行输出提示。

1. 思考与总结

因为自己和组员们都是足球迷，正好也趁着世界杯的热度，想做一个世界杯数据库。想法也就这么应运而生。

其实最大的难度还是一开始的时候，因为小组的成员都不知道如何进行爬虫爬取网页的数据，于是我自己上网查询了相关博客和信息，找到一个相对较为容易实现的方法，爬取网页传送的.js文件，然后得到里面的API，通过自己实现的代码，用API进行网络GET访问获得数据。然后就是自然而然地进行解析JSON数据，然后JDBC链接数据库，导入数据。任务量还是蛮大的。

这次项目遗憾的可能是因为自己没有学过WEB的相关知识，没有办法帮助到组员完成VUE结构的前端设计，感觉应该本身是可以帮得上忙的。

最后，这项大作业也集合了小组成员的心血，用心地陆续做了4个星期，现在数据库也考完试了，感觉也终于可以稍微放下了，但是学习数据库的路还没有结束，在这次项目中，我意识到了自己WEB方面还没有过多的积累知识，想在暑假丰富一下JAVA WEB相关的知识。

以上，致谢数据库老师和数据库TA，感谢！