17/6/3

完成功能：当前比赛日结果抓取，前一比赛日结果抓取

已解决问题：bs4抓取的话没什么问题，但是需要点击向前按钮的话，就要用selenium抓取

selenium包含webdriver，职责是抓取和动态地去点按钮，webdriver需要与浏览器结合，来请求网页内容，代码中使用phantomJS，它没有图形界面，是轻量的浏览器。

webdriver的使用中遇到问题。首先是为webdriver绑定phantomJS浏览器。首先弄明白了，executable\_path要输入浏览器exe的路径（包括exe本身），一般方法是把浏览器执行文件目录加入系统变量，或者是把浏览器执行文件拷到python的相关目录下，配置python的时候要加入系统路径变量path的那个。

这里我考虑到最好还是能够在不同的计算机上能够便捷运行的问题，把phantomJS放到脚本同目录下，用path = os.getcwd()+'\phantomjs.exe'来指定phantomJS的位置。

其次是在webdriver中获得相关按钮、相关文本信息。按照bs4的方法指定会报错。Compound class names not permitted。

http://blog.csdn.net/Jywangkeep\_/article/details/45568723 (要注意这个是java代码)

根据上面的文章，发现是webdriver无法识别带空格class，会认为是复合class不支持。解决办法是用CSS\_SELECTOR来迂回，把空格换成点。这个挺不方便的。要区分复合class么让用户传个链表不就好了，不允许有空格很困扰。

driver.find\_element(By.CSS\_SELECTOR,".flecha\_jornada\_anterior.pointer.flecha\_fecha")

之后不小心搞错了按钮位置，让代码点另外个按钮去了，结果抓取输出的没有变，因为这样网页刷新的部分不是抓的部分。

下一步峰作：这样的话干脆就弃用bs4，全部用selenium来抓，因为selenium必要，bs4可以替代。接下来要抓前一比赛日和后一比赛日，前一比赛日和后一比赛日都有可能不存在，所以要考虑到这个情况，加入处理的代码。将抓下来的信息最好放进结构的列表中，方便使用。