01/29/2020

此类网页标签树特点

<tbody>

<tr>

<td>

<tr>

<tr>

<tr>

.

.

.

.

抓取方法：

第一步：

直接soup找到所有tr标签

第二部：

使用try命令，以防止中间出现不满足条件的标签导致程序无法正常运行

第三部：

使用for循环在所有tr中找到所有符合条件的td标签（注意，一定要找到data数据对应

标签的特定且通用属性）。

第四部：

将所有满足条件的td标签赋值给tds。这是tds是一个列表，后续提取数据，可按照数据

在列表中的位置用tds【0】（0位第几位的数据）抓取特定数据

第五步：

使用df=df.append({ })将所得数据放入dataframe字典中。

第六步：

打印并输出df数据库

注意：输出csv文档是，若文档内容有中文，需使用utf-8-sig编码输出才能保证简体中文

不乱码

