

7- 存储我们爬取到的数据

存储数据

常用的存储数据方式有很多种：数据存储在文件中 例如csv格式文件或者Excel文件，或者可以将数据存储在数据库中，例如mysql。

“csv” 格式文件和Excel文件的区别？

csv是一种字符串文件的格式，它组织数据的语法就是在字符串之间加分隔符——行与行之间是加换行符，同列之间是加逗号分隔。csv格式文件可以用任意的文本编辑器打开（如记事本），也可以用Excel打开，还可以通过Excel把文件另存为csv格式（因为Excel支持csv格式文件）。

对Excel文件写入

存储成csv格式文件，操作csv文件使用csv模块；

存储成Excel文件，操作Excel文件使用openpyxl模块。

安装openpyxl包命令 `pip install openpyxl`

Excel文档的基本概念：

一个Excel文档也称为一个工作簿（workbook），每个工作簿里可以有多个工作表（worksheet），当前打开的工作表又叫活动表。每个工作表里有行和列，特定的行与列相交的方格称为单元格（cell）。比如上图第A列和第1行相交的方格我们可以直接表示为A1单元格。

代码片段：

```
1 import openpyxl
2 # 引用openpyxl
3 wb = openpyxl.Workbook()
4 # 利用openpyxl.Workbook()函数创建新的workbook（工作簿）对象，就是创建新的空的Excel文件。
5 sheet = wb.active
6 # wb.active就是获取这个工作簿的活动表，通常就是第一个工作簿，也就是我们在上面的图片中看到的sheet1。
7 sheet.title = 'kaikeba'
8 # 可以用.title给工作表重命名。现在第一个工作表的名称就会由原来默认的“sheet1”改为“kaikeba”。
```

```

9  sheet['A1'] = 'kaikeba'
10 # 向单个单元格写入数据
11 score1 = ['math', 95]
12 sheet.append(score1)
13 # 写入整行的数据，变量类型是一个列表
14 wb.save('score.xlsx')
15 # 保存修改的Excel
16 wb.close()
17 # 关闭Excel

```

解读代码：

1. 引用openpyxl
2. 通过openpyxl.Workbook()函数就可以创建新的工作簿，紧接着获取工作表
3. 将这个工作表的名称改为“kaikeba”
4. 像表格中写入数据sheet['A1'] = 'kaikeba' 就是向A1（根据Excel的坐标，A1代表第一列第一行相交的单元格）的单元格写入内容。
5. 用append方法向Excel写入一行内容，要写入多个内容只需添加一个循环即可
6. 保存写入的内容，最后关闭这个文件，避免程序占用

Excel文件的读取

```

1  import openpyxl
2  wb = openpyxl.load_workbook('score.xlsx')
3  # 打开的指定的工作簿
4  sheet = wb['kaikeba']
5  # 指定读取的工作表的名称
6  A1_value = sheet['A1'].value
7  print(A1_value)
8  # 获取

```

1. 调用openpyxl中的load_workbook方法，打开“score.xlsx”文件。
2. 获取“score.xlsx”工作簿中名为“kaikeba”的工作表。
3. 获取“kaikeba”工作表中A1单元格的值，并把它赋值给A1_value，最后A1_value的值打印出来。

关于文件读写的深入了解我们可以到官网上<https://openpyxl.readthedocs.io/en/stable/>