

第十周课外作业

编程题一：

参考上课讲授的案例，在母校合肥工业大学新闻网站中选择某一主题新闻页面(如教学科研、合作交流、工大要闻、综合新闻)，使用 `requests.get()`爬取该主题下面所有页面新闻的新闻标题、超链接和新闻的发布日期，保存到.xlsx 文件中。

编程题二：

- 1、选取一个问题调研的主题，确定一个与主题问题相关的信息数据网站。
- 2、使用 `requests.get()`方法，自动爬取此网站连续 10 个以上的页面数据；
- 3、对爬取的每个页面数据，根据分析问题的要求，使用 `BeautifulSoup` 库进行页面解析，提取出问题描述的相关数据，存在到二维列表中。
- 4、使用 `pandas` 库，将二维列表数据转成 `DataFrame` 数据，设置好列标签，然后分别保存为.csv 和.xlsx 文件。

主题网站可以选择：

- 1) 旅游网站，了解景点、美食的评论因素；
 - 2) 房地产网站，了解房产价格的影响因素；
 - 3) 电商网站，了解商品销售价格、量及用户的评价信息；
 - 4) 岗位职业网站，了解行业岗位的区域、薪水与需求等影响因素
-

要求使用函数结构编程，每条语句后，用#加注释，说明本条语句的功能或目的。作业完成后于本周日（5月16日）前将源码及生成的数据文件放在以自己学号+姓名的文件夹中上传到FTP服务器上指定的文件夹里。