## Lecture2 Challenge

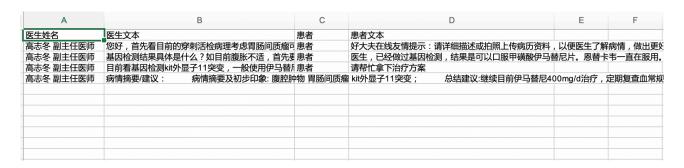
本部分为已经掌握requests爬虫的同学们设计。主要为Selenium爬虫方法的实践,难度跨度较高。

#### 问题描述:

好大夫在线是中国著名的在线医疗问诊网站,存有大量病人和专家团队沟通的对话记录,下面是一个网页例子: https://m.haodf.com//doctorteam/flow\_team\_6480090760.htm



现在需要你抓取这个网页上的对话信息,内容病人/医师的对话文本、不同回答医师的名称。要求以消息记录的形式存储到本地,如:



注意,该网站并不会一次性显示所有对话内容,会有一个"加载更多"的按钮,点击后才会显示一部分新对话,需要不断点击,直到该按钮消失,页面上才会显示完整的对话内容:



请编写代码,完成该项任务。

# Lecture2 Challenge

### Tips:

1.由于需要点击"加载更多"按钮,因此我们需要和网页进行交互。requests库无法解决这一问题,需要我们使用selenium模块,驱动浏览器打开网页,真正地点击按钮,完成交互。

selenium模块的使用教程: https://blog.csdn.net/weixin\_36279318/article/details/79475388

2.我们需要定位多个页面元素(因为要抓取多句对话),那么,如何使用xpath,在selenium中定位 多个元素?

xpath的进阶教程: https://cuiqingcai.com/2621.html

xpath+selenium的一个简短实例: https://www.cnblogs.com/liuliu-word/p/8057998.html

3.如何将所抓取到的对话信息,存储到excel表格中?

我们建议使用xlwt模块,尝试自行百度寻找该模块的使用方法。

#### Answer:

我会在10.6公布详细的注释代码。