

第十次课外作业

编程题：

1、 参考上课讲授的案例，在母校安徽医科大学网站中选择某一主题页面如(通知公告、医大要闻、校园动态等)，使用 `requests.get()` 爬取该主题下面连续20个以上页面的源码，再使用正则表达式筛选出每页中的标题、标题超链接和标题内容的发布日期，保存到`ahmu_news.csv`文件中。

2、 参考上课讲授的案例，在百度图片搜索引擎中输入任意一主题关键字(如华山、西湖、喀纳斯等)，编程从百度网站上下载300幅以上该主题的图片，并保存在指定的文件夹中。

编程分析题：

1、参考上课讲授的案例，在豆瓣网站中选择某一主题(如图书，电影，音乐等)内容，爬取20个以上的页面，获取500条以上的网友评论，再对评论数据先进行有效性数据清洗处理，再进行词频统计，制作词云图，根据统计的结果，分析网友对该主题的评论，将统计的结果，制作的词云图及分析的结论一同保存在以“**学号姓名**”命名的word文档中。

注意：本次作业要求采用结构化模块设计，每个功能模块通过定义函数来实现，使用函数结构编程。

如爬取网页源代码函数，从源代码中用正则提取数据函数，将数据保存为文件的函数，程序运行主函数等。每条语句后，用`#`加注释，说明本条语句的功能或目的。

本次作业，除要上交程序源代码，程序运行生成的数据文件(保存在指定路径下的图片)，及分析的word文档也需一同上交。