实验十  网络爬虫和自动化  
学习目标：  
1.掌握网络爬虫的基本方法。  
2.运用request库编写基本URL访问过程。  
3.运用beautifulsoup4库解析和处理HTML。  
4.掌握向搜索引擎自动提交关键词并获取返回结果的方法。  
  
实验习题：  
1.按照省份输出中国大学排名。  
2.实现USNEWS美国大学排名的爬虫，并打印结果。  
3.编写视频网站视频播放链接的爬虫。  
4.编写爬取robots.txt文件的爬虫并分析内容。  
5.编写根据robots.txt文件内容爬取网站的程序。  
6.分析百度图片搜索返回结果的HTML代码，编写爬虫抓取图片并下载成专题图片库。  
7.编写程序测量30秒内成功获得百度主页的次数。  
  
小结：本章主要介绍了设计并实现网络爬虫的基本方法，结合requests和beautifulsoup4两个库的使用。讲述了如何处理HTTP协议和解析网页HTML和XML页面信息的方法。本章通过中国大学排名爬虫和搜索关键词提交两个实例，展示了现代网络社会中快速抓取定向网页数据的重要价值。