Scrapy框架课程介绍：

1. 框架的简介和基础使用
2. 持久化存储
3. 代理和cookie
4. 日志等级和请求传参
5. CrawlSpider
6. 基于redis的分布式爬虫

CrawlSpider：

问题：如果我们想要对某一个网站的全站数据进行爬取？

解决方案：

1. 手动请求的发送
2. CrawlSpider（推荐）

CrawlSpider概念：CrawlSpider其实就是Spider的一个子类。CrawlSpider功能更加强大（链接提取器，规则解析器）。

代码：

1. 创建一个基于CrawlSpider的爬虫文件
   1. scrapy genspider –t crawl 爬虫名称 起始url