Python网络爬虫进阶实战第九周闯关作业

1. 问题描述：

本周课程主要讲解Scrapy框架的使用。具体有：Scrapy爬虫框架以及内部每个组件的使用（Selector选择器、Spider爬虫类、Downloader和Spider中间件、ItemPipeline管道类等）：

1）.使用Scrapy爬虫框架爬取新浪网的分类导航信息：

网址：http://news.sina.com.cn/guide/



2 ). 使用scrapy模拟完成人人网登陆的登陆操作:

如URL地址：http://www.renren.com/

3）. 爬取当当图书网站中所有关于python关键字的图片信息。

参考URL：http://search.dangdang.com/?key=python&act=input

要求：将图书图片下载存储指定的目录中，而图书信息写入到数据库中。



1. 解题提示：

* 第一道题请参考Selector选择器中实战案例。
* 第二道题参考爬虫基础的人人网登陆，使用scrapy的POST带参数请求
* 第三道题请参考ItemPipeline中的实战案例。

1. 批改标准：
   1. Scrapy新浪网的分类导航信息爬取（20分）
   2. Scrapy模拟登陆人人网（20分）
   3. 当当网的图片信息爬取（50分）（信息爬取20，数据存储15，图片下载15）。
   4. 其他项：(10分)（文件结构清晰，代码整洁，要求适量的注释）。