**学习流程：**

1. 环境配置和基础知识铺垫

2. 爬取真实数据

3. scrapy 突破反爬虫技术

4. scrapy 进阶

5. scrapy redis 分布式爬虫

6. elasticsearch Django 实现搜索引擎

**爬虫基础知识**

1. 正则表达式

2. 深度优先和广度优先遍历算法

3. URL去重的常见策略

4. xpath+css =》Data

5. 模拟登陆

6. 图片验证码

7. IP访问频率限制

8. user-agent随机切换

**scrapy进阶开发：**

1. scrapy的原理

2. 基于scrapy的中间件开发

3. 动态网站的抓取处理

4. 将selenium和phantomjs集成到scrapy

5. scrapy log配置

5. email发送

6. scrapy信号

**实战：**

Redis分布爬虫（elasticsearch Django 实现搜索引擎）