1. 使用Scrapy Shell的控制台（命令scrapy shell进入），和Scrapy进行交流，可用于测试Scrapy代码语句，而不需要写一个整体的爬虫程序。

例如: Scrapy shell 网页url地址，则进入shell后，response对象会对被加载成所指定的 URL地址。

1. 浏览器可以自动生成选定节点的Xpath。
2. response.xpath(‘/html/’).extract() 用于查看下载页面的所有内容。可以查看是否成功的下载页面，而没有被拒绝403访问等等。只有在成功下载页面以后，Xpath的后续过滤才有效果。如果请求被服务器403拒绝访问，那么下载的内容只有403信息。
3. 服务器端为了反爬虫，会对请求的头部进行验证。根据在服务器端设置的验证不同，对请求头部的验证也会不同，有些会验证User-Agent, 有些会验证host， 有些需要用户登录等. 如果这请求中，这些条目没有设置或者设置错误或者没有登陆，则服务器端会拒绝该请求访问。
4. 我们可以尝试使用浏览器去访问被爬网页，然后在浏览器中找到network的tab, 可以查看我们浏览器所发送的请求头。