**基于Python的当当网商品信息爬虫系统测试文档**

**测试目的：**

测试出程序依旧存在的不足之处，并以文档的形式记录下来，方便接下来的开发者了解项目历史。

**测试方法：**

因为没有找到关于爬虫程序的测试工具，所以这次的测试是以黑盒测试的方法，也就是直接运行爬虫程序，对比结果来测试的。

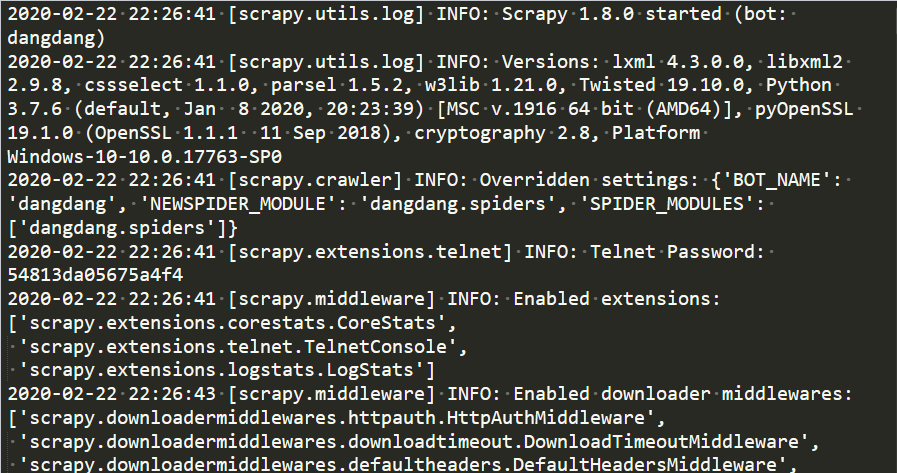
直接打开run.py文件，按“Ctrl+B”运行即可。其中run.py文件中，如果不希望出现scrapy日志文件，想要无日志输出，只需将语句：

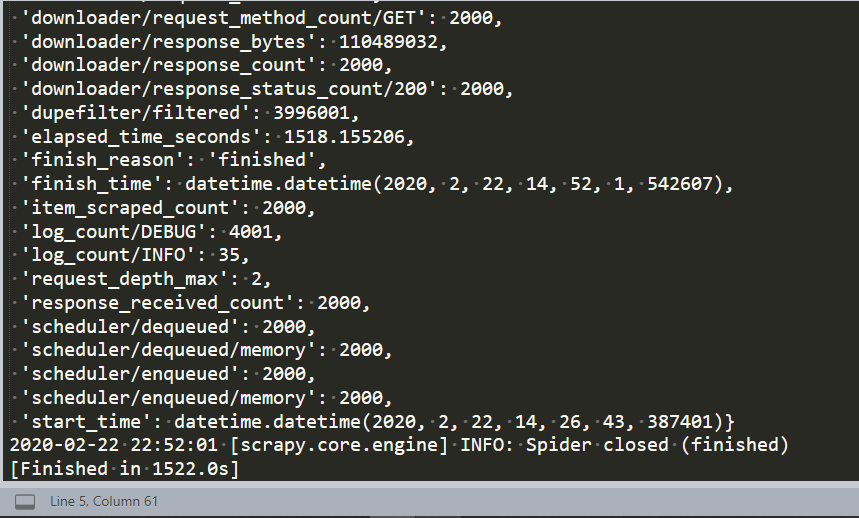
cmdline.execute('scrapy crawl dd'.split())

修改为：

cmdline.execute('scrapy crawl dd --nolog'.split())

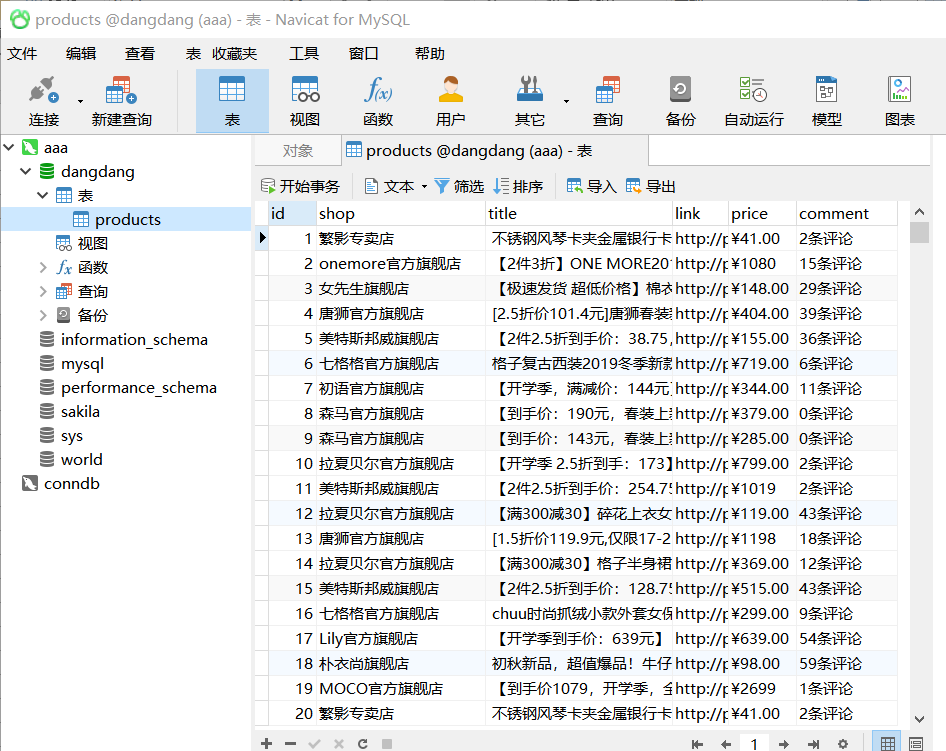
这里，本系统采用的显示日志文件的语句：cmdline.execute('scrapy crawl dd'.split())，测试结果显示如下：

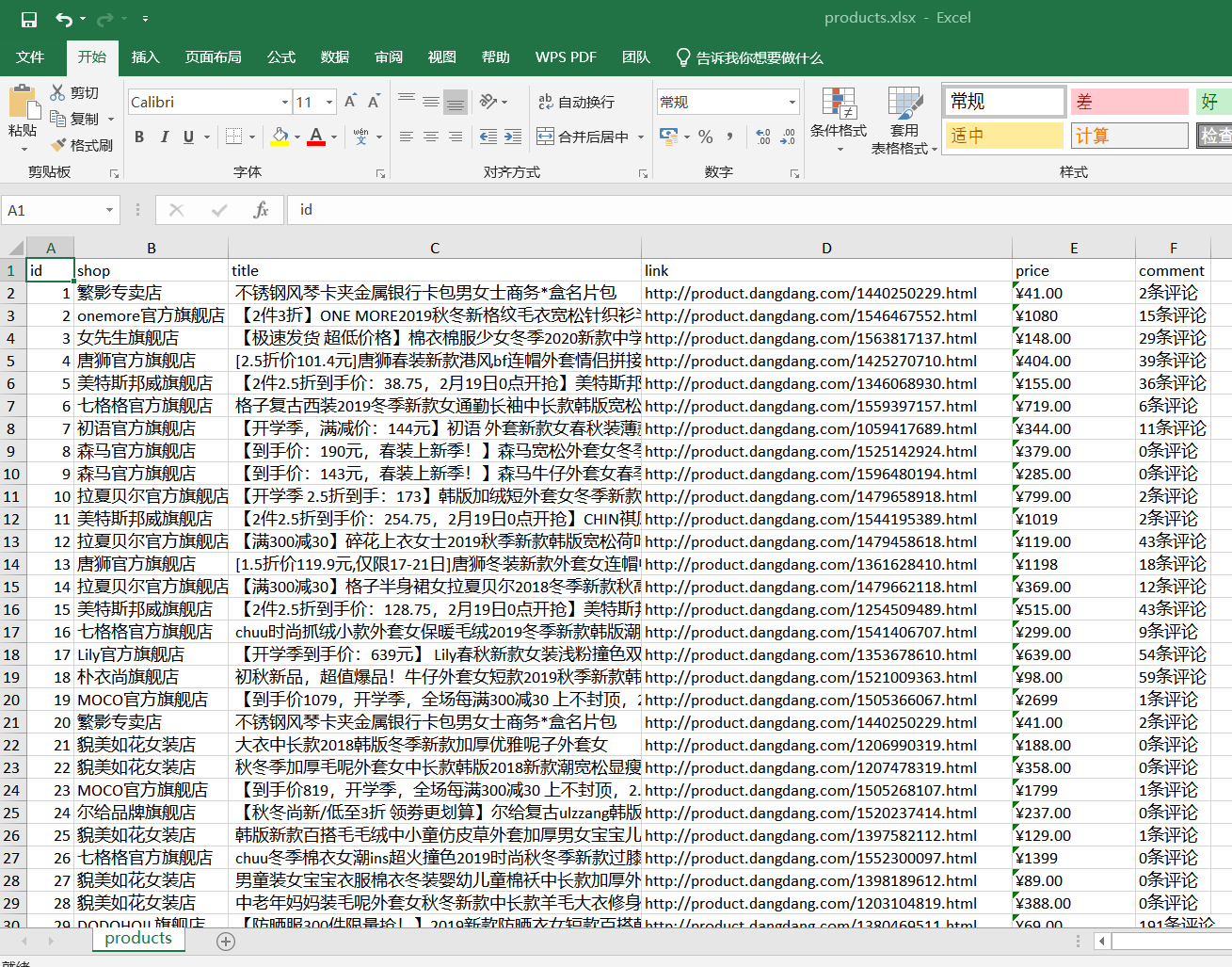


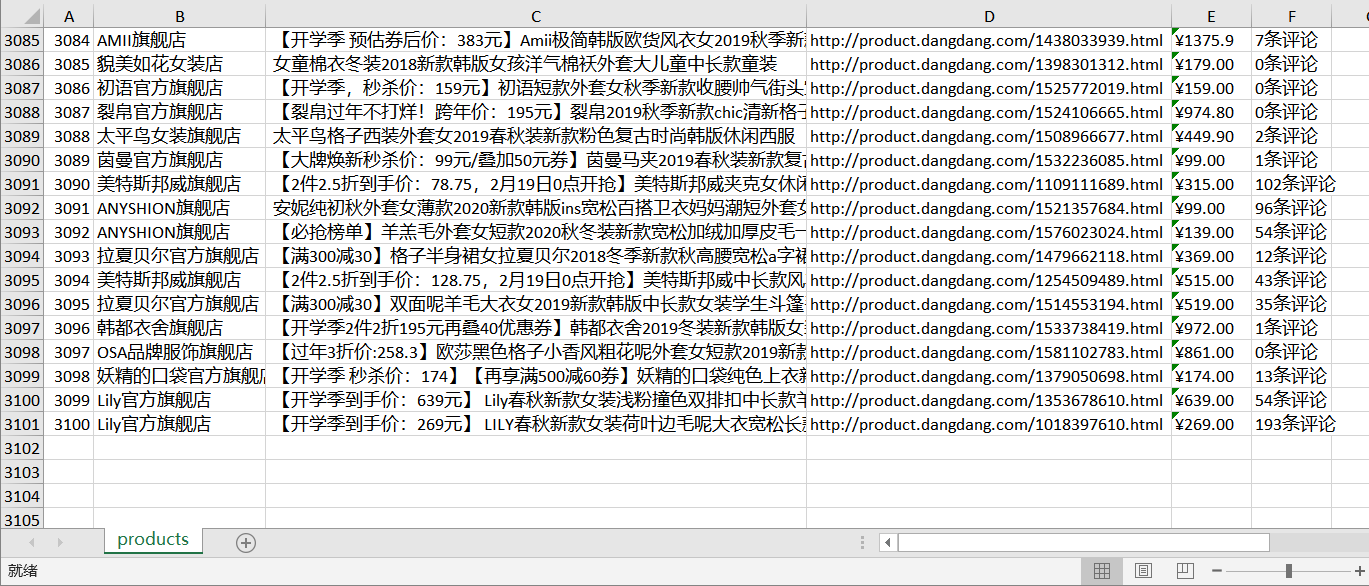


最后爬取到的数据用MySQL数据库保存，将数据导出为Excel文件，名为“products.xlsx”，该文件在dangdang文件夹中可以找到。可以通过修改dangdang文件夹中的dd.py中i的取值范围，来改变爬虫到的数据条数。

爬虫数据结果部分截图如下：







**测试总结：**

通过测试，我发现了爬虫程序依旧存在的不足：爬虫虽然设计成了可以爬取大半的文本格式，但是因为网站文本格式太过繁杂的原因，还是有部分的文本无法爬取，这就导致了部分网页上的信息不能保存到数据库；代码还可以进行优化，在未来的时间里我将对该程序进行改进。