一、

mkvirtualenv article\_spider创见环境下的项目并且自动进入

创建项目

scrapy startproject ArticleSpider

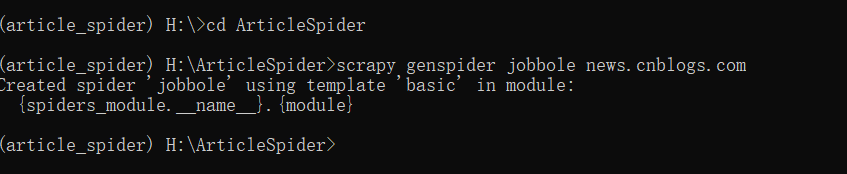


之后在pycharm里面打开open这个ArticleSpider（注意虚拟环境为envs里的之前创建的）

之后cmd

cd ArticleSpider

scrapy genspider jobbole news.cnblogs.com后面是域名

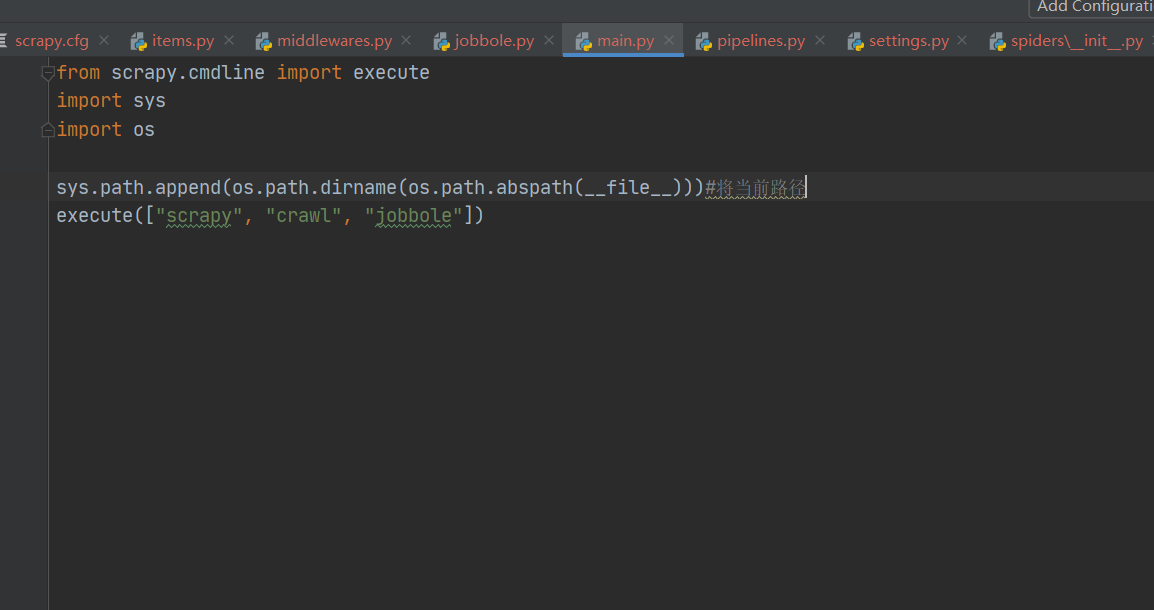


启动运行爬虫

为scrapy crawl jobbole须在pycharm用

步骤：1、新建一个main.py

2、在main文件写启动代码



创建好了可以在jobbole.py里pass打断点验证是否成功

Xpath

十分重要

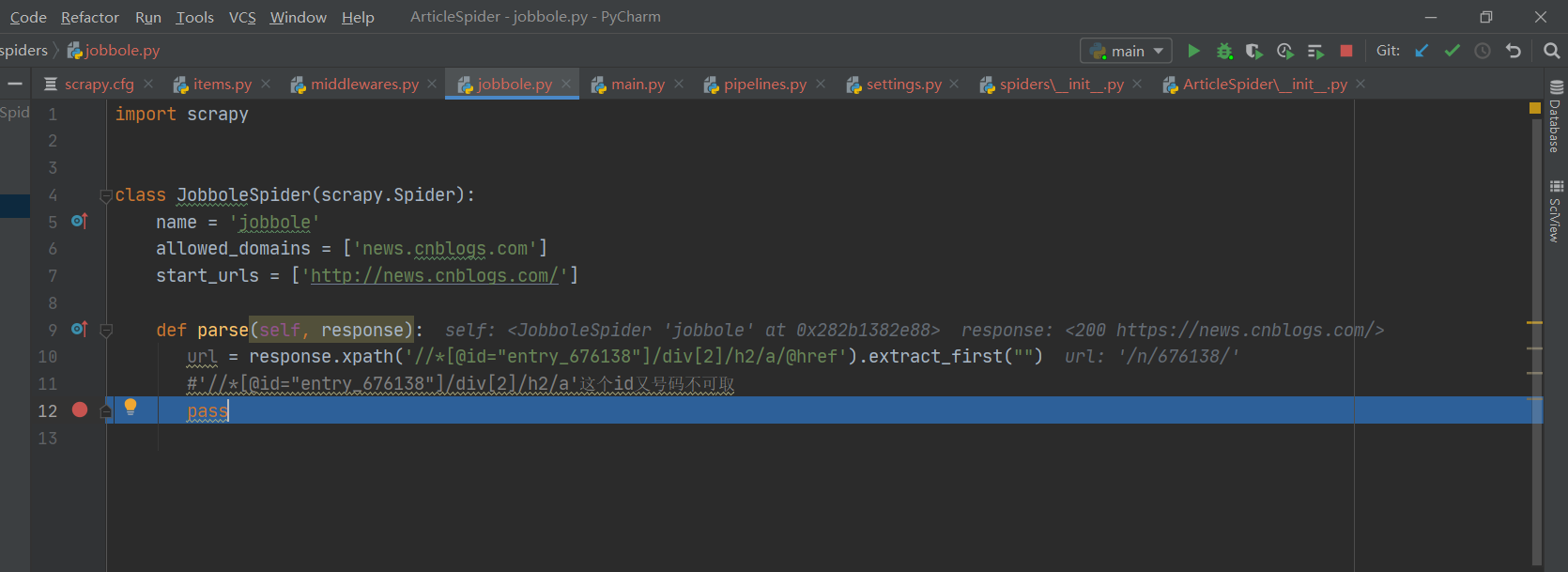




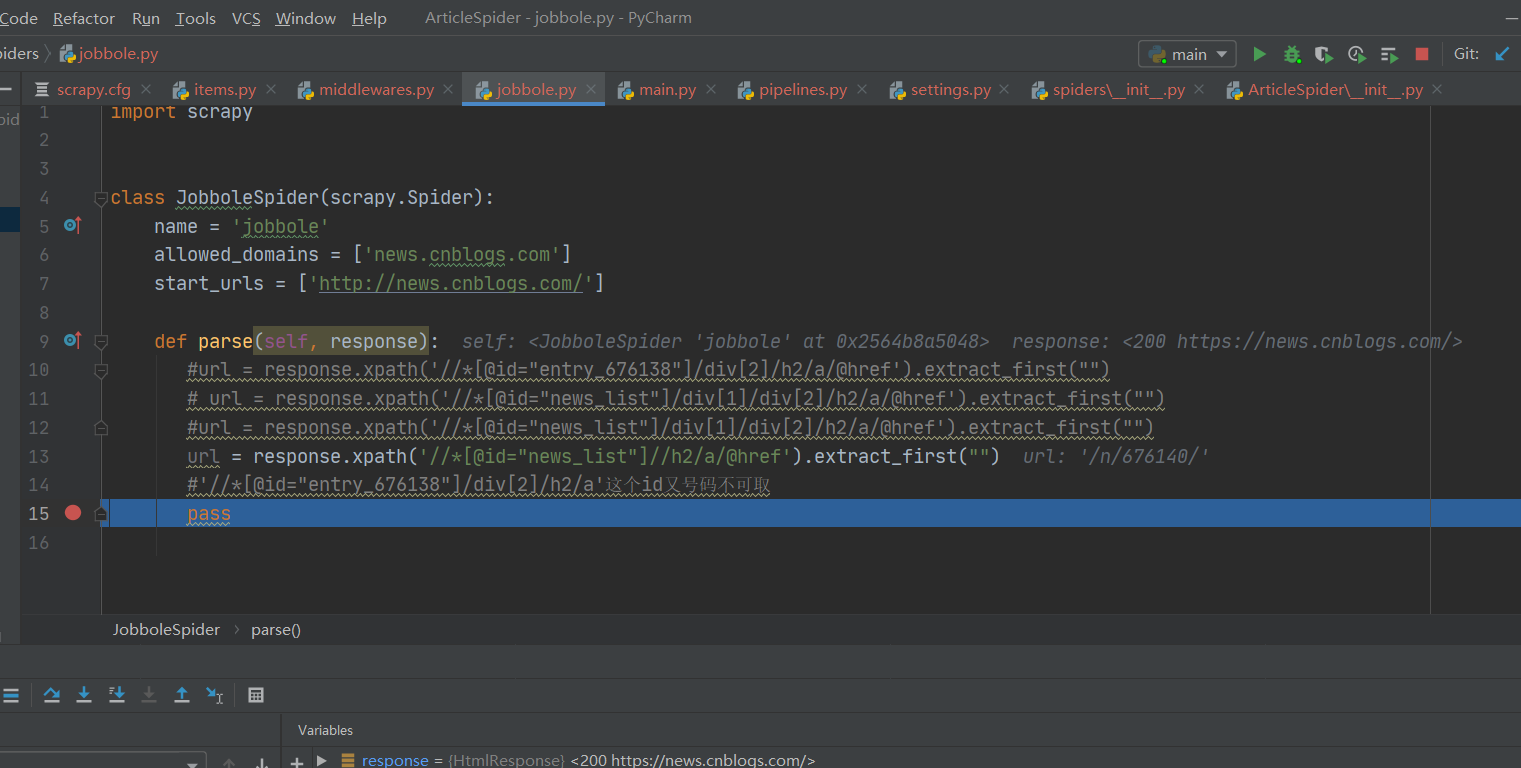


若想深入理解则用语法大全

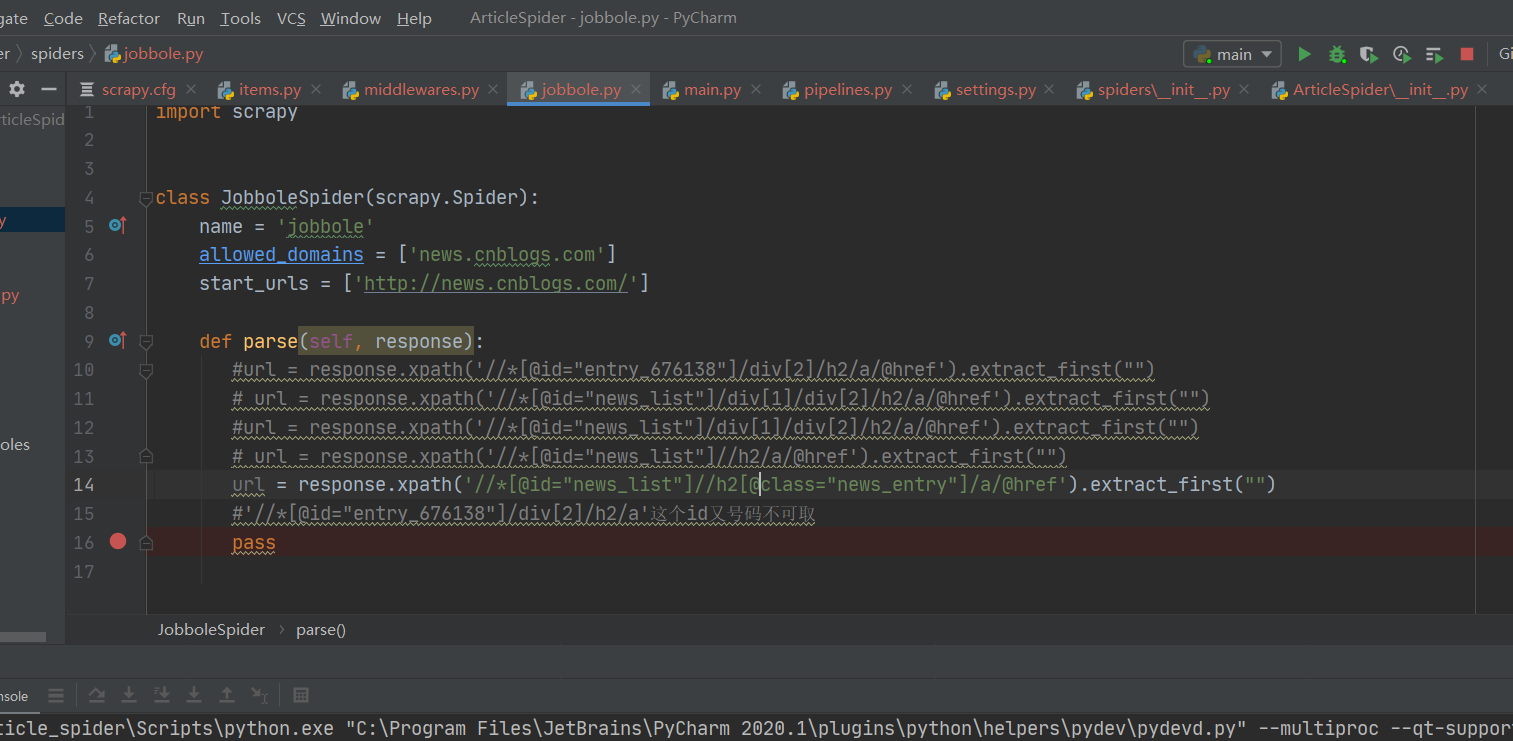
使用在scrapy的使用



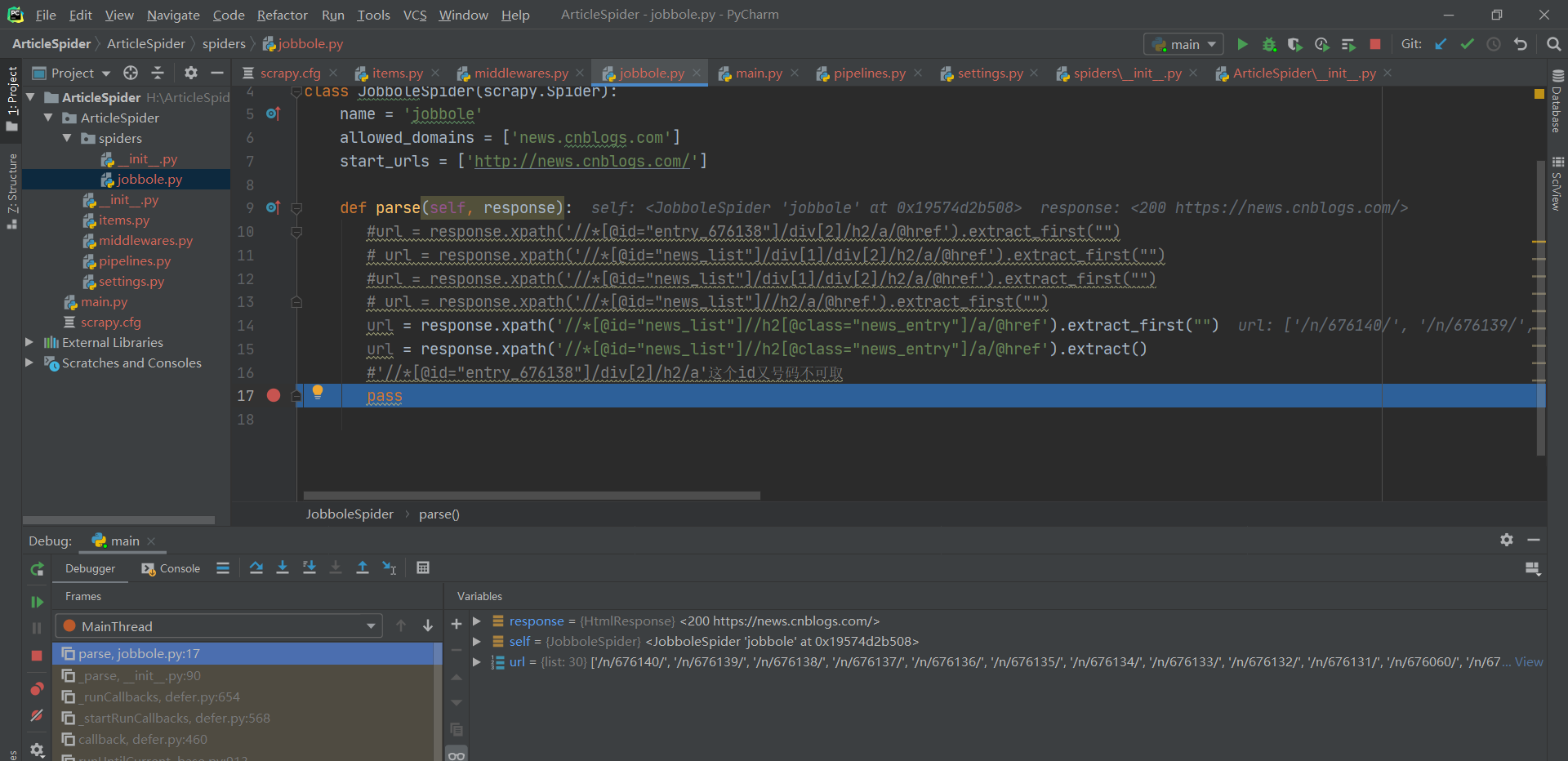
使用extract()因为只有这样才能抽取使用的值



//为后代结点也可以统一标签添加限定元素



若不用firs则出钱全部url如下

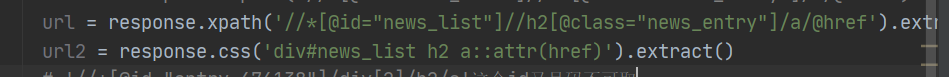


**Css选择器**









为了独立使用selector来解析页面可以导一个selector包来使用（还未使用）

加:还有新技术webkit可以模拟解决js问题