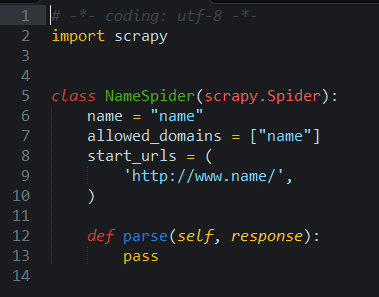
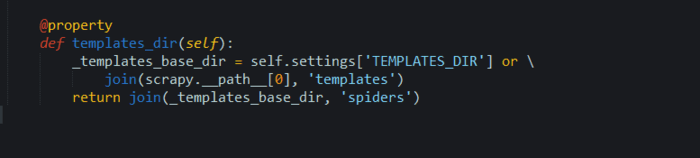
**Python Scrapy 自定义模版文件--理解创建的项目文件**

在编写**scrapy**爬虫的时候，我们很烦每次都是要自己创建一个新的的spider，当然创建完项目的时候开业再次执行 **scrapy genspider name “name”** 来创建一个name.py文件，如图：



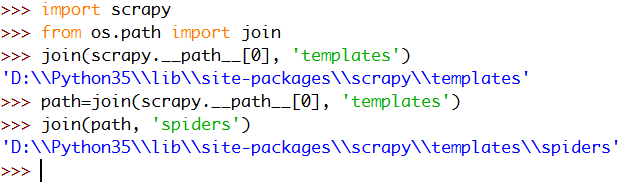
执行scrapy genspider name “name”

现在来简单介绍一下这个命令是如何创建一个spider文件的，先仔细观察一下上面的类命名和函数命名。所有**scrapy**的命令是通过路径：**X:\Python35\Lib\site-packages\scrapy\commands**来执行的，先看看我们的**genspider命令**，在路径下找到genspider.py文件，如图：



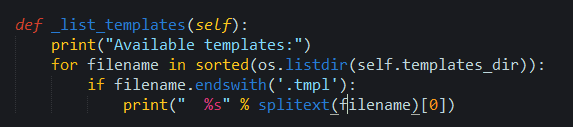
这个是genspider最后的一个函数

仔细观察不难发现这个函数通过**os.path.join**方法组建一个完整的**scrapy**路径，这里演示一下：



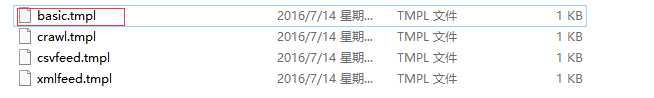
对比两个图的结果

在看看倒数第二个函数：



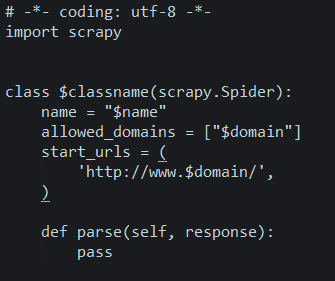
倒数第二个函数

仔细观察发现它是**os.listdir**列出**templates\spiders**文件夹下的所有的文件和文件夹，去寻找扩展名为**.tmpl**的文件，看到这里聪明的你肯定知道了，没错我们去也去需要这个文件**.tmpl**，路径：**D:\Python35\Lib\site-packages\scrapy\templates\spiders**下看到如图：



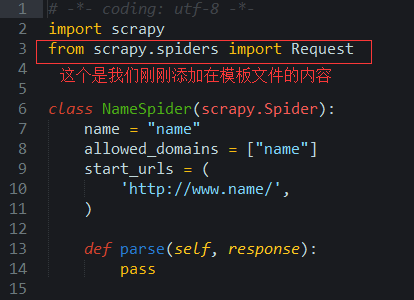
红色框这个文件就是创建spiders.py的模板

记事本、sublime打开看到这样的情况：



对比上面说的

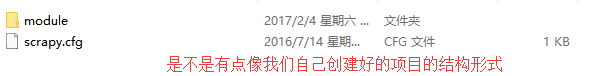
对比上面说的内容，可以知道**$classname**是获取项目名称来创建类名，以此类推就不难知道了，现在我们可以修改这个文件（想对比最开始执行**scrapy genspider name “name”**），我们在最上面添加一行：**from scrapy.spiders import Request**。再次执行**scrapy genspider name “name”**看看结果：



成功添加了

现在介绍第二种情况：每次执行完**scrapy crawl name** 在项目的spiders文件夹下只是生成了一个\_\_init\_\_.py文件，那么我们每次都是要自己去新建一个spider文件来写爬虫，最好是执行生成项目的时候spider文件也生成好了，还像上面一样也写好了部分内容的文件，这里有两种方法

一、先看看路径：**D:\Python35\Lib\site-packages\scrapy\templates\project**



和自己的项目文件对比形式

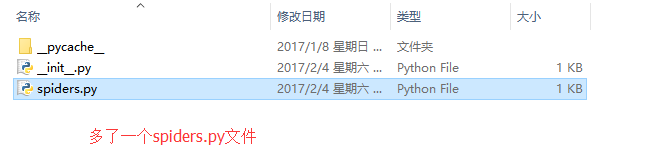


是不是



路径上和文件对应

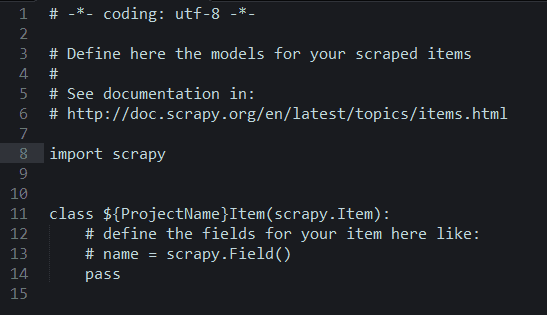
确定是和项目文件的文件个数和名字一样的，那么在上图的spiders文件夹下的**\_\_init\_\_.py**添加内容项目文件的**\_\_init\_\_.py**也是对应上相同的内容（自行操作），那么在spiders这个地方添加一个.py文件，并且添加内容，对应的在项目的文件内spiders文件夹下也会自动新建一个一模一样的.py文件



内容和你再上面的文件内容一模一样

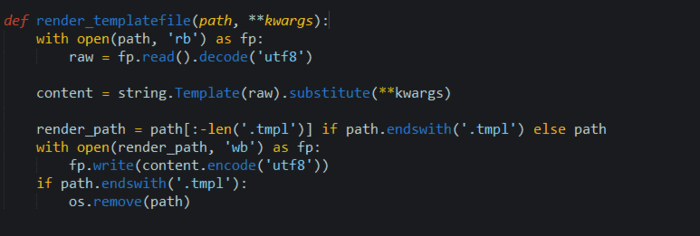
二、自动新建一个和项目文件名一样的spider，并且类名也是和项目名称一样联系起来。

上面的方法发现有个缺陷，就是项目里面的spiders文件夹下的spider.py只能叫做spider 内容也是和自己在上面加的一样，类名完全千篇一律，最好是和项目名称以前变化，有点像项目里面的items.py的类名称（自己去看看）,可以看看**items.py**的模板文件（**items.py.tmpl**）



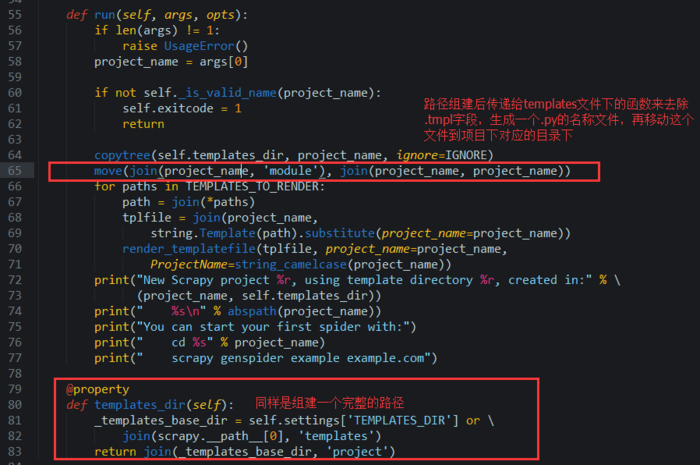
观察

我们的项目的spiders文件夹下也应该这样变化还要添加内容，这里直接给出。查看路径**X:\Python35\Lib\site-packages\scrapy\utils**下的**template.py**文件：



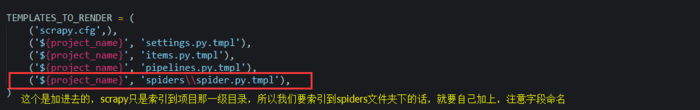
这里是路径作为参数

路径作为参数传递进来，判断是不是.tmpl扩展名文件，而这个路径同样我们去：**X:\Python35\Lib\site-packages\scrapy\commands**下查找**startproject.py**文件，在内容最下面也是一个函数来组建一个完整的scrapy路径（**os.path.join**）



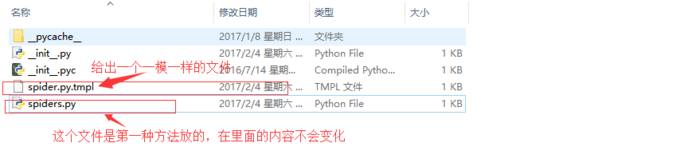
观察

而如何传递给**templates**里面的函数呢，同样看上面这个图的文件的最顶上：



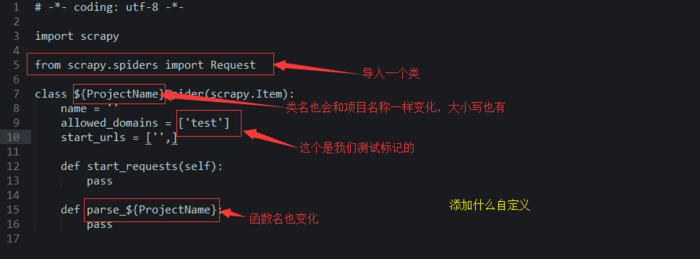
观察

既然这样给出了spiders文件夹下的**spider.py.tmpl**，那么在这个这一级模板文件夹下也也应该给出这个一样文件（名字也是一模一样）放在这里：**X:\Python35\Lib\site-packages\scrapy\templates\project\module\spiders**



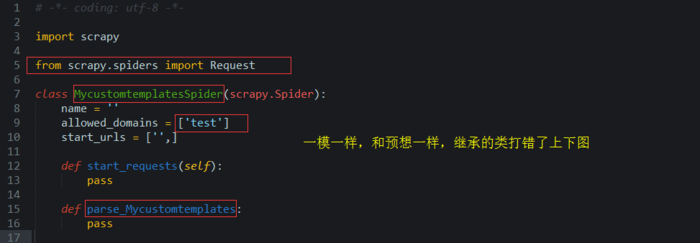
观察

修改spider.py.tmpl文件如下：



观察

我们执行**scrapy startproject mycustomtemplates**:



完全一样

后面还有可以自定义很多，大家一起共勉，博主属于新手，哪里不对的地方还请指点，见谅。