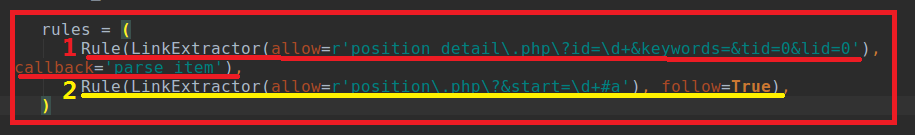
# Tencent\_work\_log

这个爬虫中的重点也就是使用crawl spider, 来帮我们进行获取下一页的url地址的。如图所示：

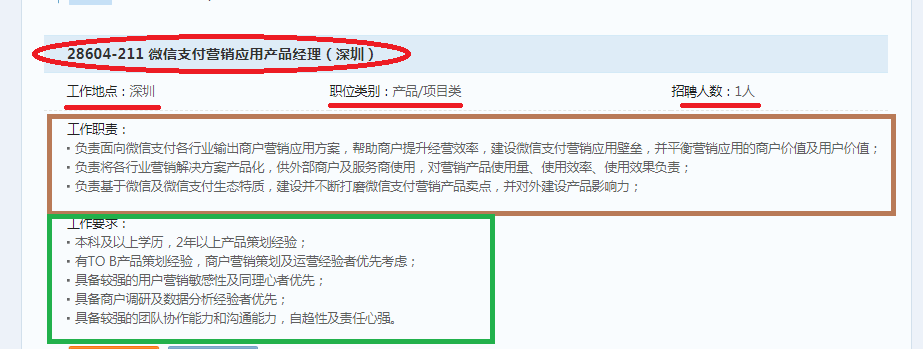
第一个链接是列表中的每个工作的详细页面的链接，这个链接是个不完整的链接，只要在开发工具中分析出所要访问的链接地址即可，crawlspider会自动补全这个url,并交个引擎处理，访问后的response，交给callback中的parse\_item这个方法解析：

第二个链接是下一页的链接，也是在开发工具中分析即可。



## Parse\_item中的处理内容

在parse\_item中，主要获取的是如图所示的内容信息：



也就是对应的：

标题：’title’, 工作地点：’location’, 职位类别：’type’, 招聘人数：’number’,

工作职责：’responsibility’, 工作要求：’require’

把这些数据放入字典中，并对数据进行清洗，然后保存在mongdb中。