Atitit html数据提取法总结

程中，我们会使用两种方式来提取，一种是 Jsoup 的方式，另一种是 httpclient + 正则表达式的方式，这也是 Java 网络爬虫常用的两种方式，你不了解这两种方式没关系，后面会有相应的使用手册。在正式编写提取程序之前，我先交代一下 Java 爬虫系列博文的环境，该系列博文所有的 demo 都是使用 SpringBoot 搭建的，不管你使用哪种环境，只需要正确的导入相应的包即可。

Jsoup 方式提取信息

对于复杂页面，还需要引入浏览器内核解析js代码，才能完全获取页面

Webview 提取