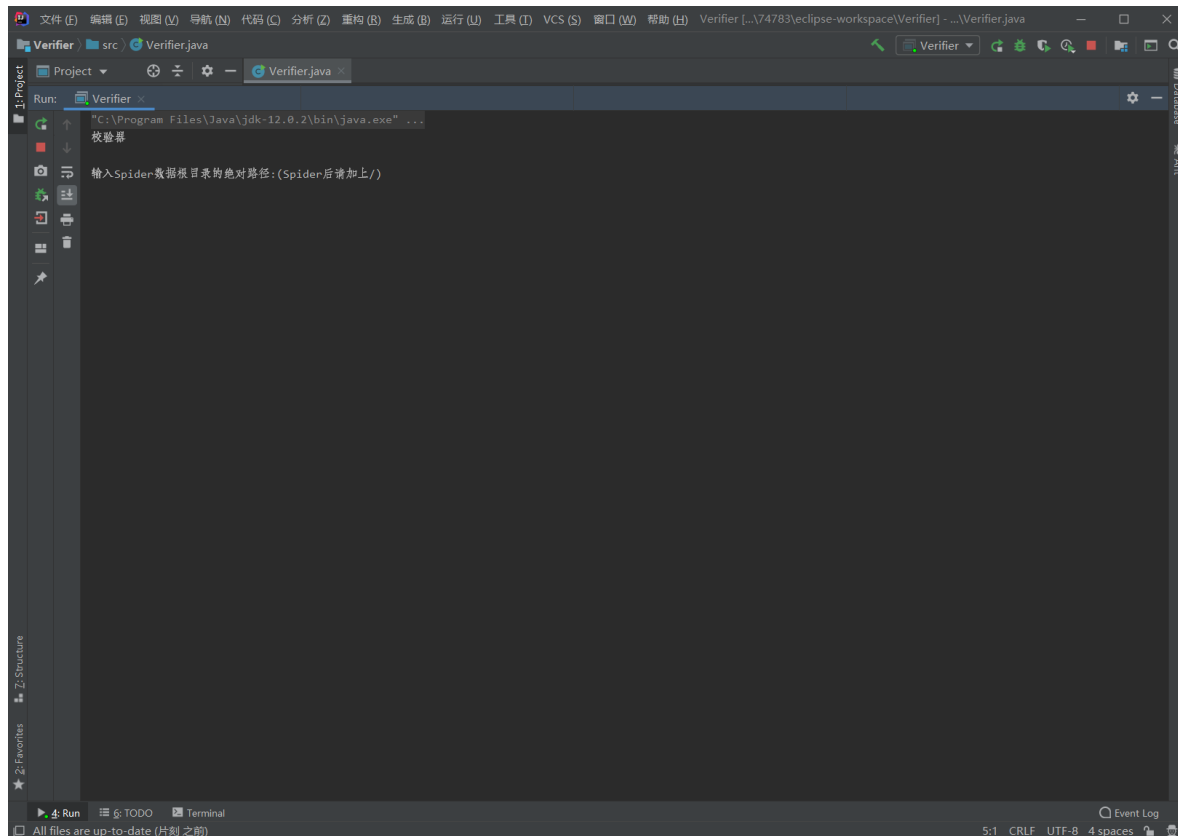


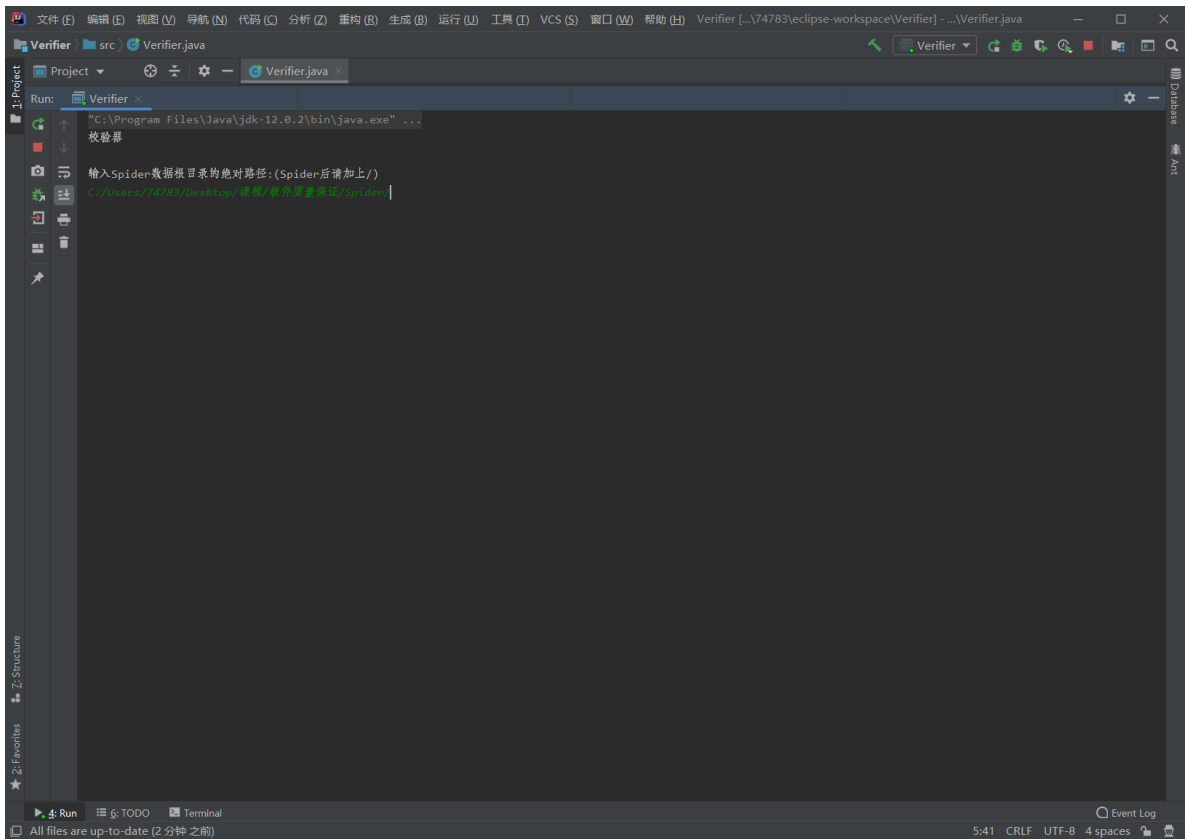
校验器演示

本文档将对校验器进行一些演示，因为网络爬虫难以演示，且远离比较简单，所以这里只选择对于校验器作出演示。

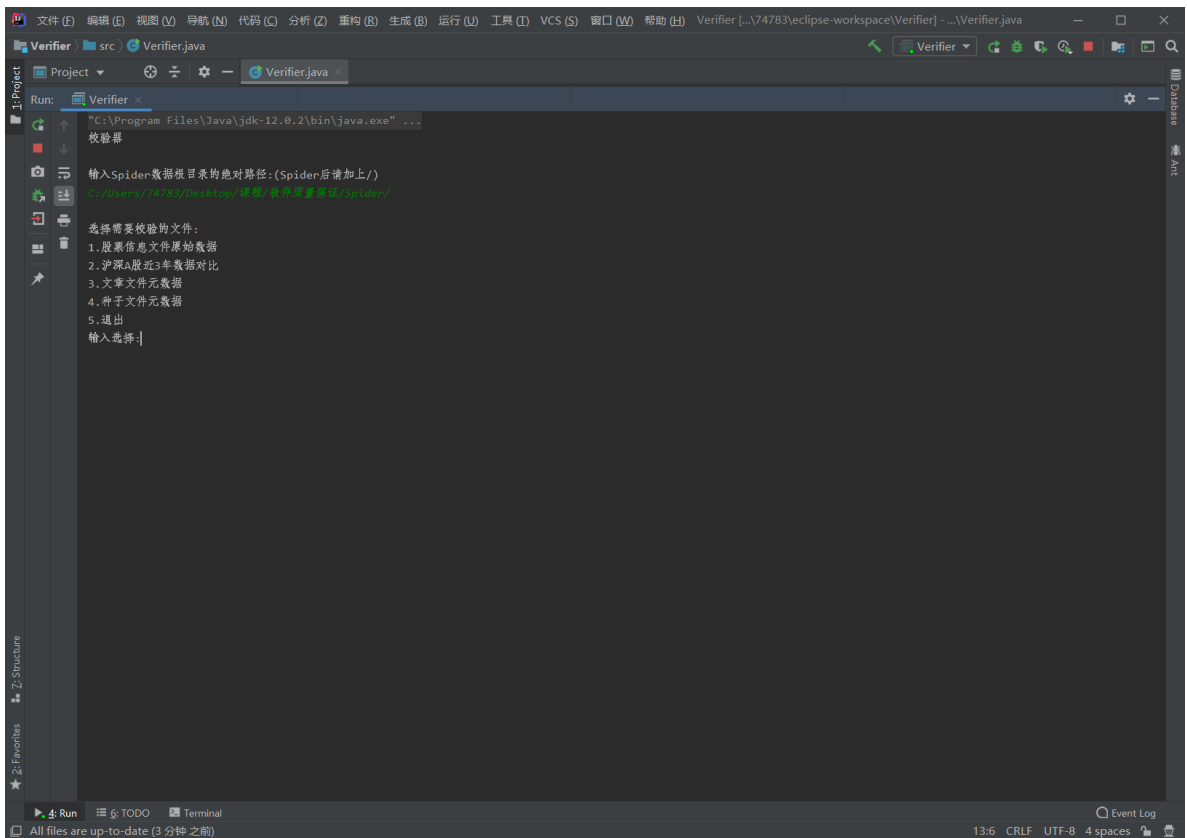
校验器主界面



首先要求输入Spider文件夹的绝对路径，这里需要使用“/”并且最后也要加上，输入如下：



按下回车进入校验选择界面：

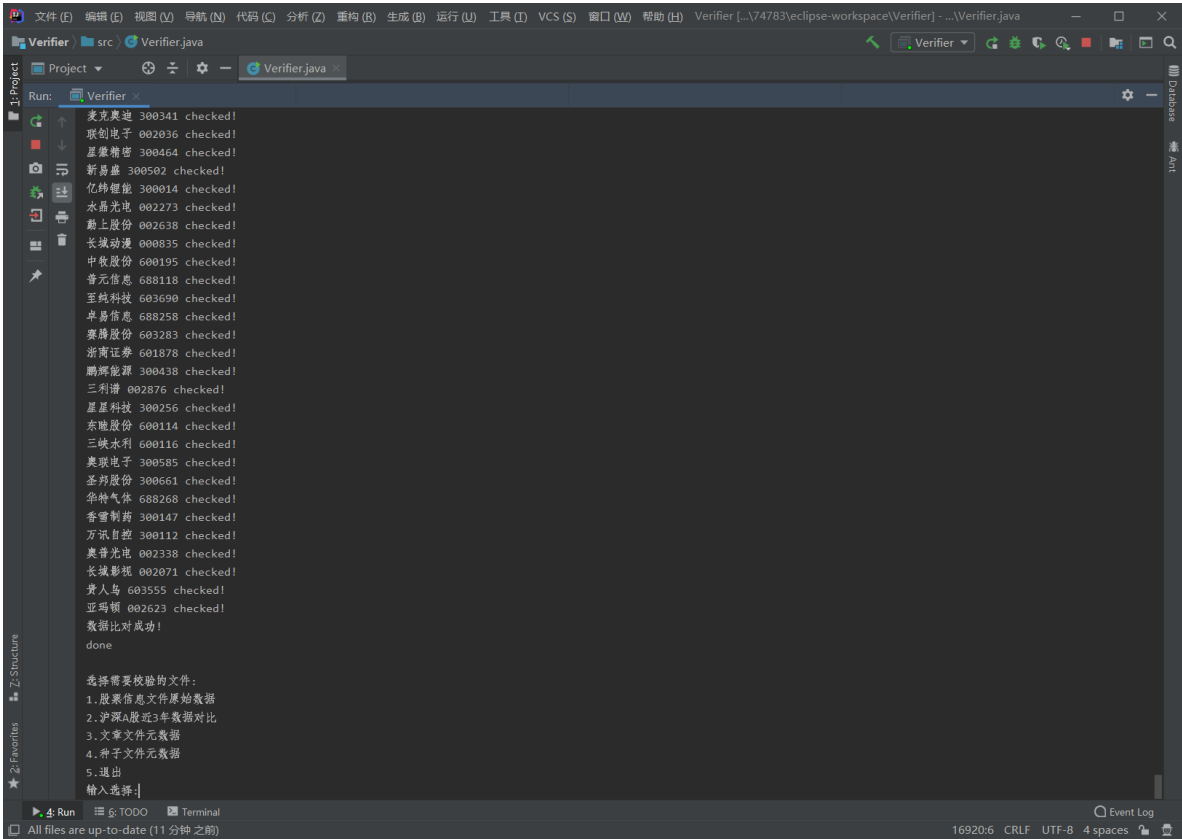
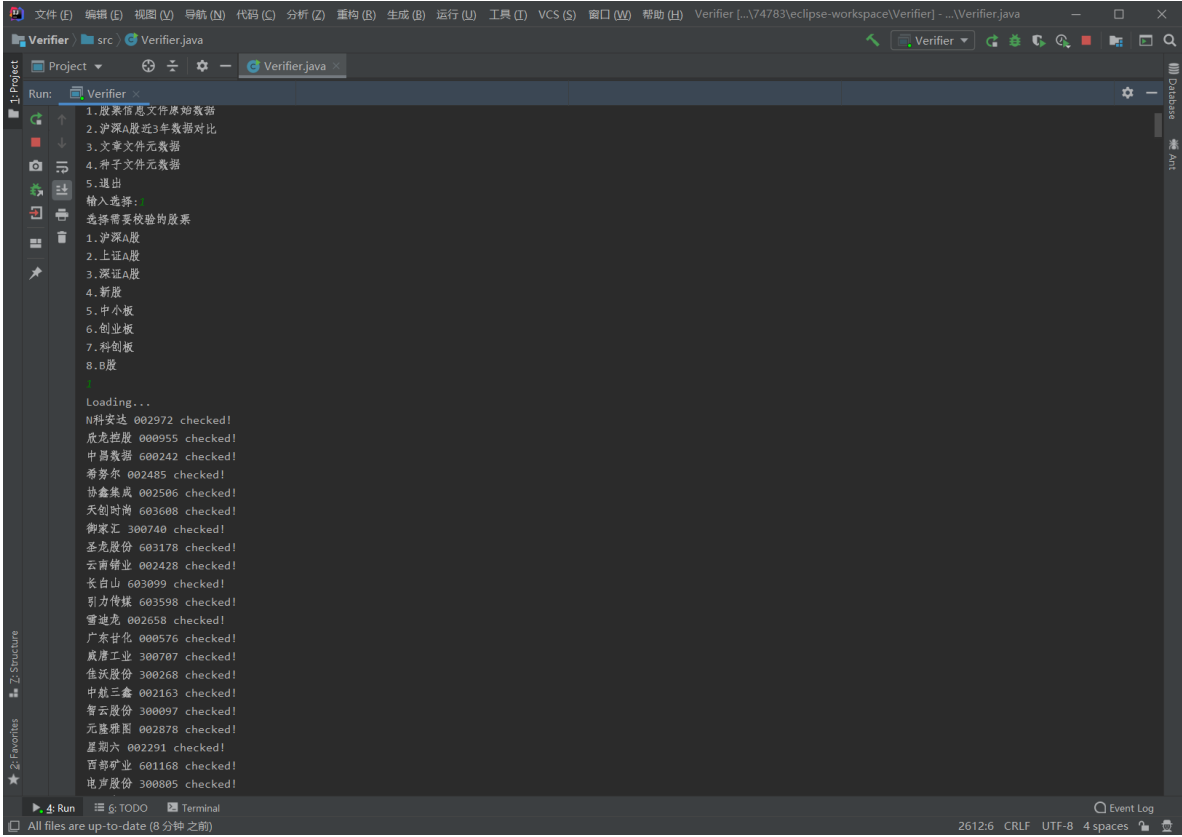


股票原始数据校验

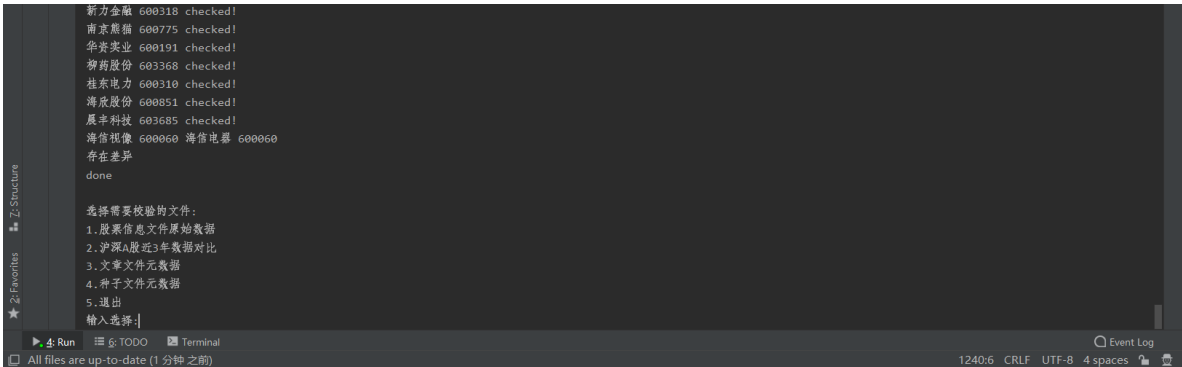
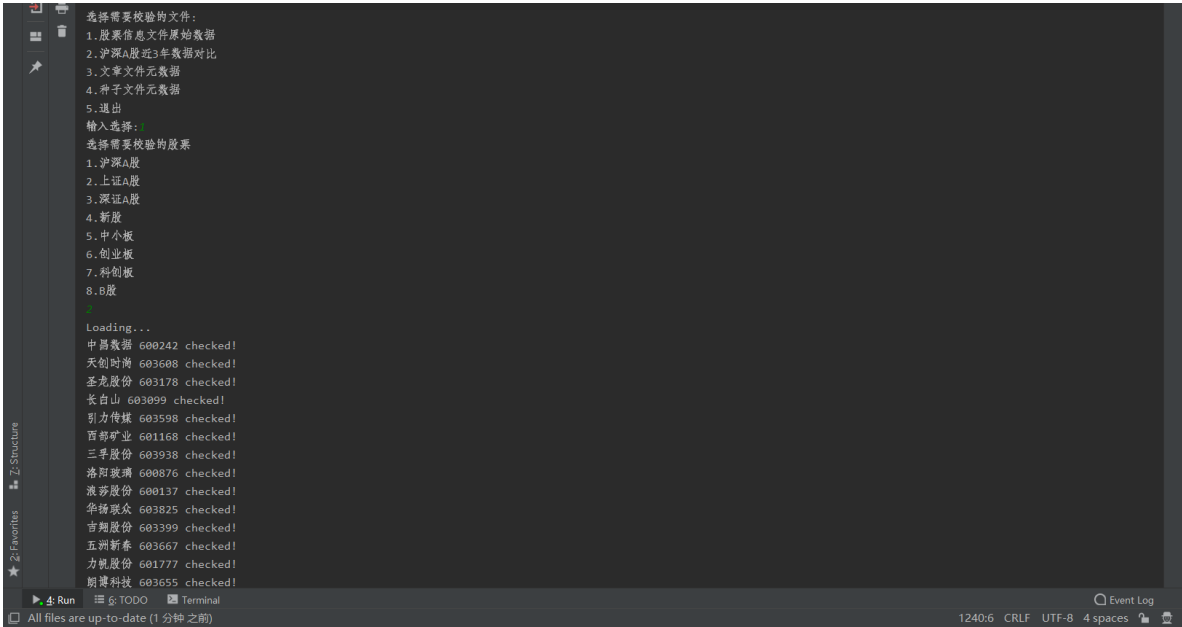
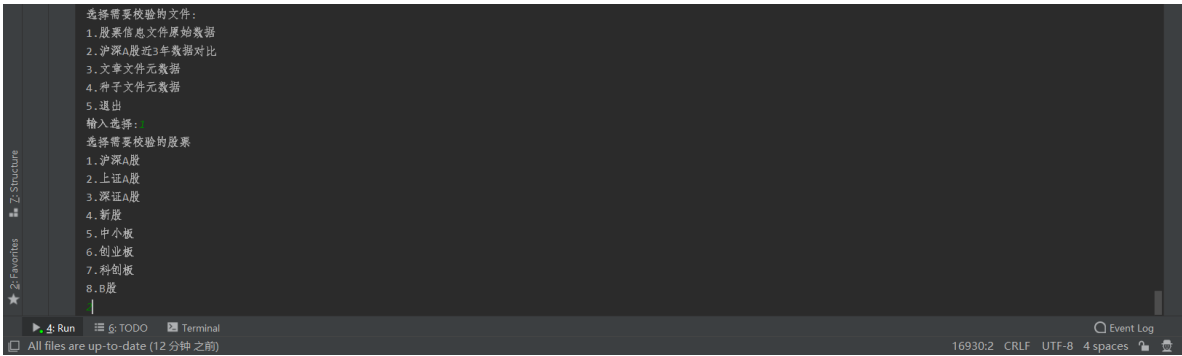
因为一开始爬取的数据是沪深个股当日收盘时的数据，没有太好的数据对比意义，所以在后面重新加入了沪深A股3877支股票近3年的数据。所以，这个数据基本只用来提取沪深个股列表。

校验器将校验股票的编号与名字是否对应。

校验沪深A股：



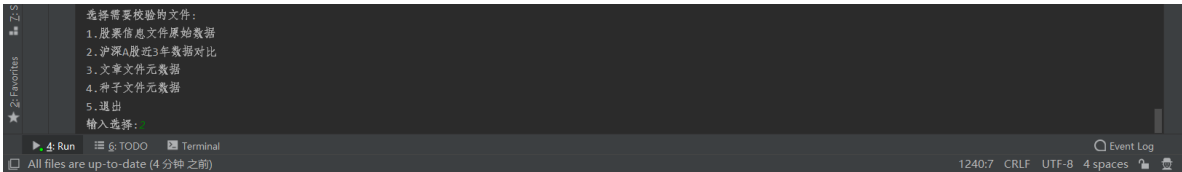
可以看到沪深A股对比成功，当然，有些股票会出现一些变化，以上证A股为例：



可以发现存在差异，也就是文件中为海信电器，现在网站上为海信视像。

沪深A股3年数据校验

此时将对于3877支沪深A股的进3年的数据进行对比，选择的是自2017年1月1日到2019年12月27日收盘时的数据。验证效果如下：

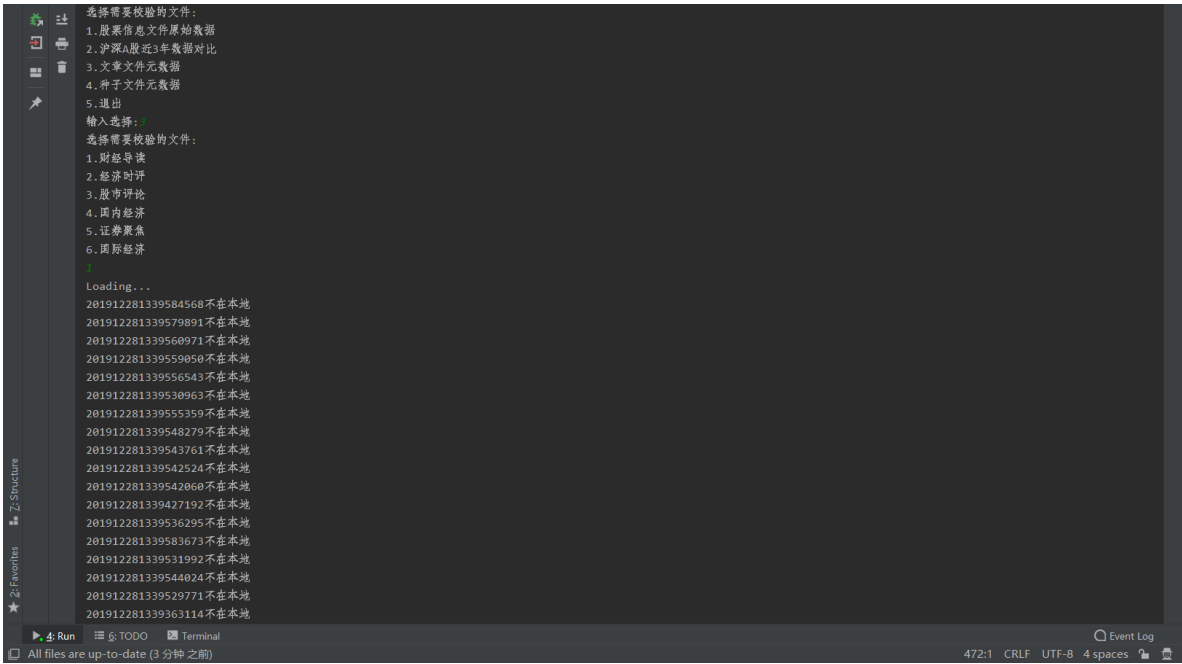


对比过程如下：

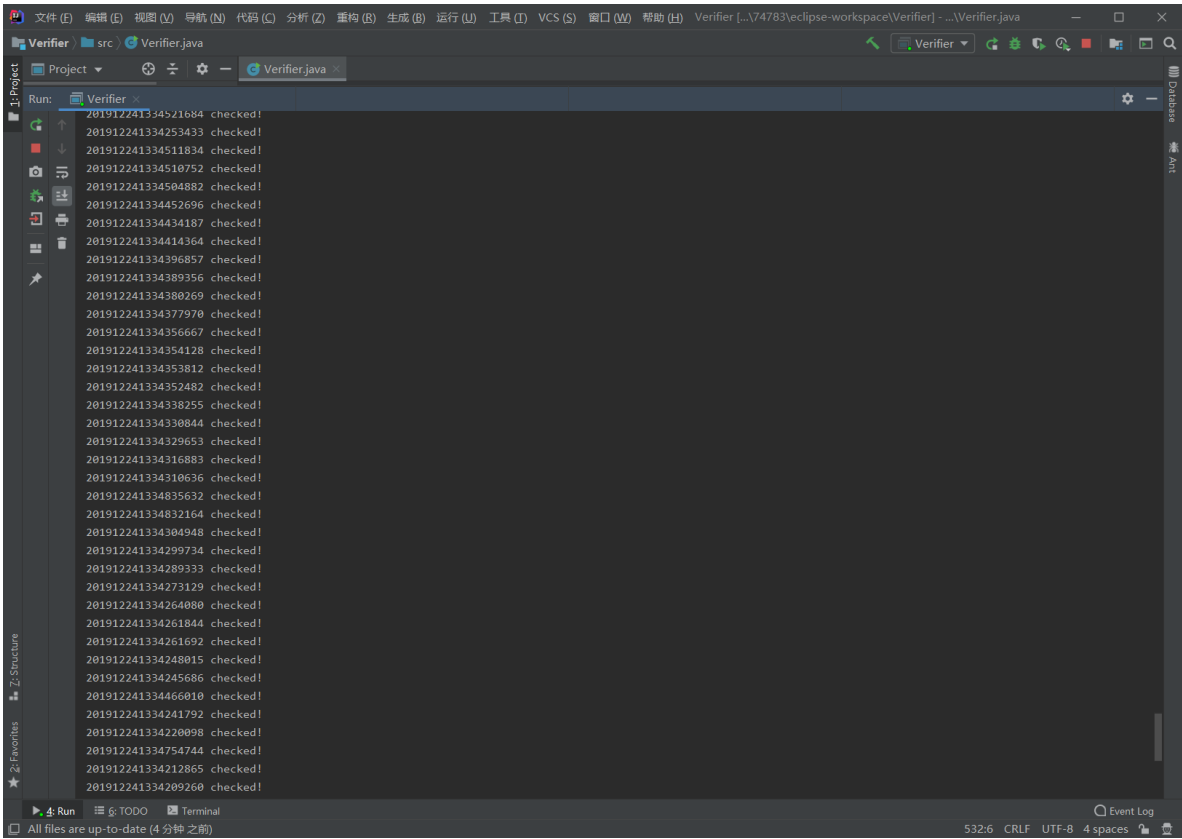


文章元数据校验

文章元数据对比演示如下，本次以财经导读板块的文章元数据为例：



一开始出现不在本地，表示一开始爬取的数据是最新发布的文章，这些文章在之前爬的时候并不存在，所以没办法对比。

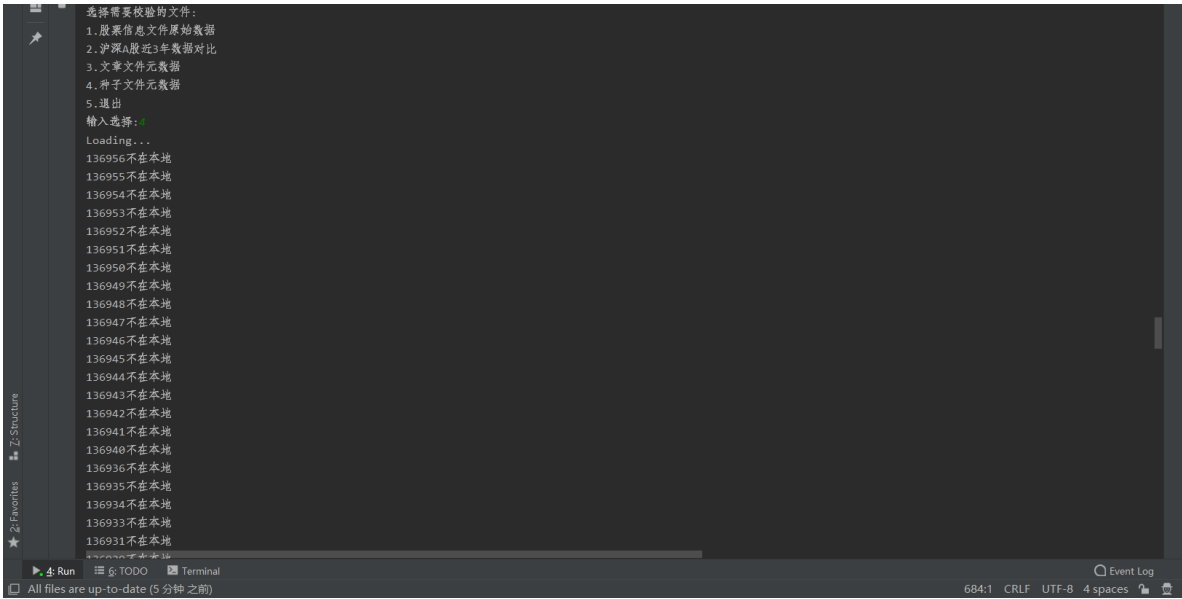


中间存在的文章，验证成功将会显示checked。

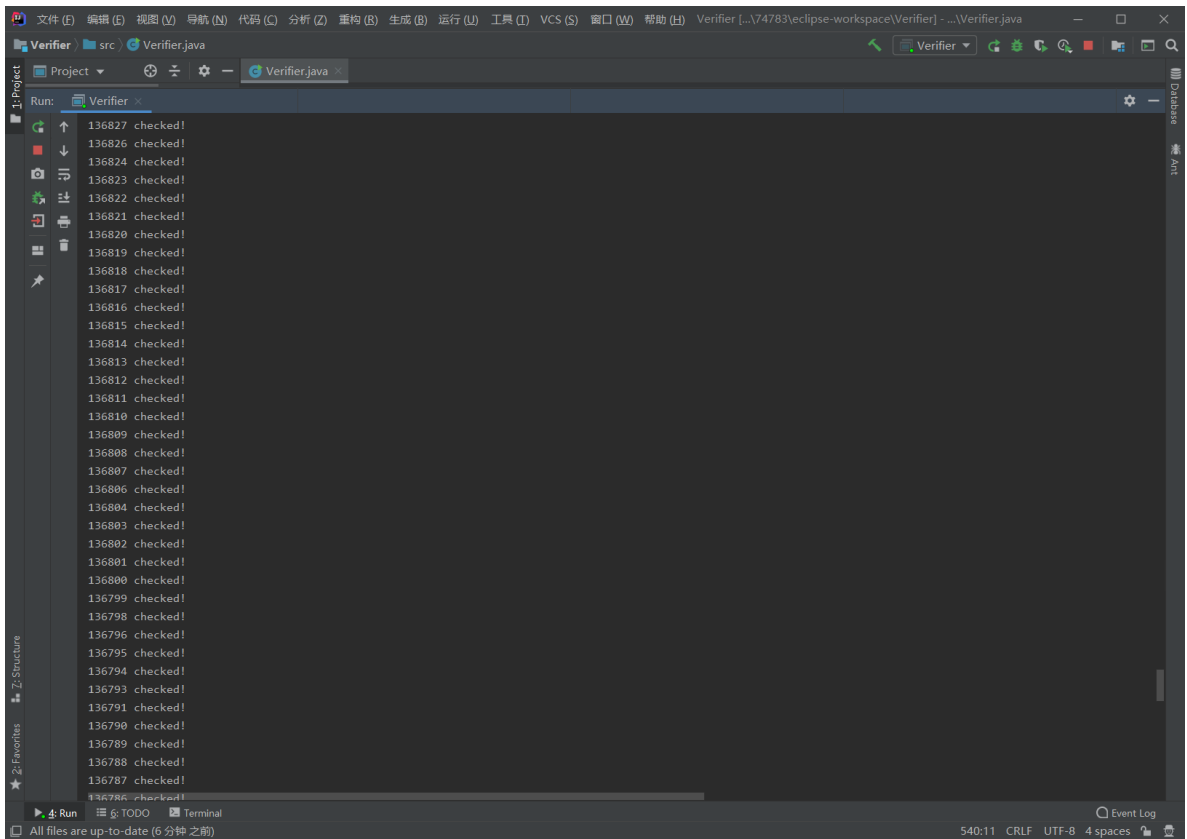


验证成功输出done。

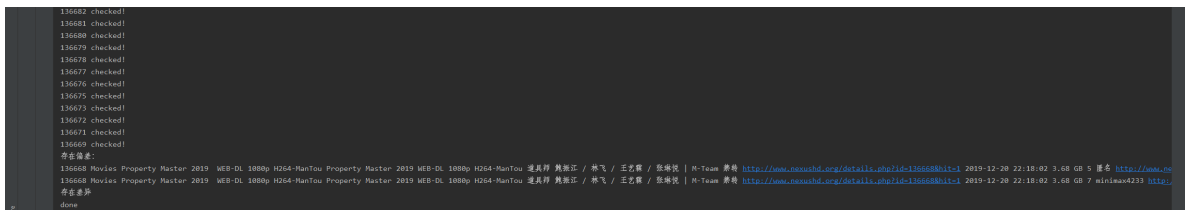
种子文件元数据校验



一开始出现不在本地表示最新网站上的文件在之前爬取时并不存在。



中间存在的文件，验证成功将会显示checked。



在验证的最后，发现出现差异，这时将会显示文件中的元数据与网站上的元数据，可以发现，种子文件的发布者将种子数从7个减少为5个，同时原来显示的发布者名称，作者改为了匿名。

对于存在差异时，将会给出网站新数据与文件数据的对比结果，并且这次对比的结果为存在差异。

程序退出

最后，退出程序如下：

