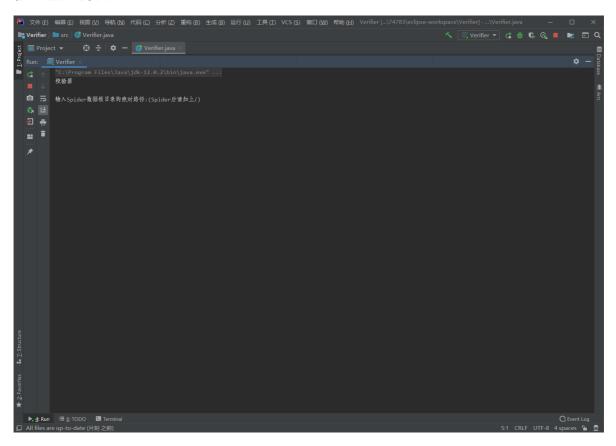
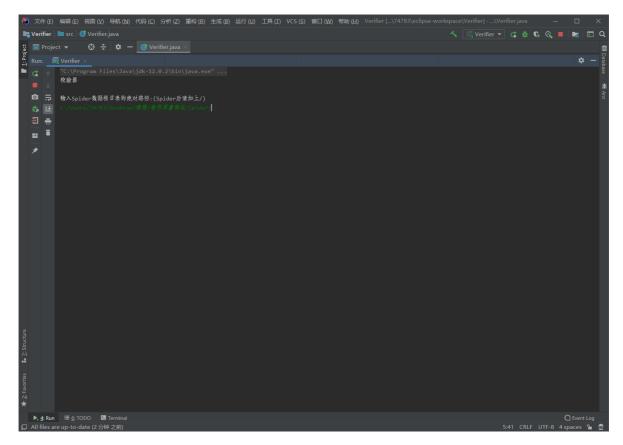
校验器演示

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。

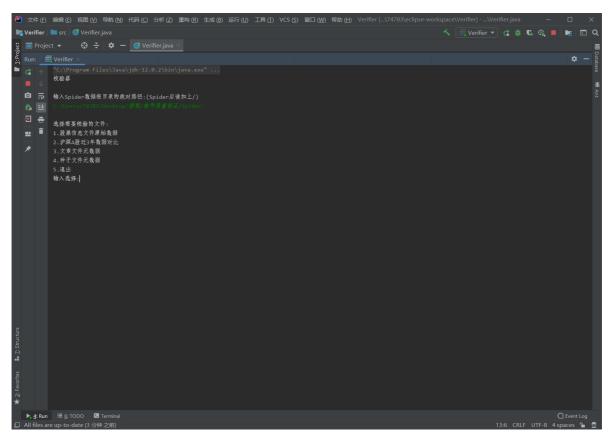
校验器主界面



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:

```
选择需要收租的义件:
1.股票信息文件原始教器
2.沪深A股近3年数据对比
3.文章文件元数器
4.种子文件元数器
5.退出
                     输入选择:
选择需要校验的股票
                     1.沪深A股
2.上证A股
                     2.工业A版
3.深证A股
4.新股
5.中小板
6.创业板
                      7.科创板
   🗓 文件 (E) 编辑 (E) 视图 (M) 号航 (M) 代码 (C) 分析 (Z) 重构 (R) 生成 (B) 运行 (U) 工具 (T) VCS (S) 窗口 (M) 帮助 (H) Verifier [...\
1.股票信息文件原始教括
2.沪深A股近3年教据对比
3.文章文件元教됨
     5. 大平大行人或者
4. 种子文件元数据
5. 退出
输入选择:
包 去接需要校验的股票
     ■ 1.沪深A股
2.上证A股
3.深证A股
                     4.新股
5.中小板
                     6.创业板
7.科创板
                     Loading...
N科安达 002972 checked!
成龙控股 000955 checked!
                     中昌教据 600242 checked!
希努尔 002485 checked!
                     协套集成 002506 checked!
天创时尚 603608 checked!
                     圣龙股份 603178 checked!
云南锖业 002428 checked!
                      长白山 603099 checked!
引力传媒 603598 checked!
                     制力传集 683598 checked!
雪迪克 902558 checked!
广东甘化 999576 checked!
咸唐工业 398797 checked!
住沃股份 399268 checked!
年本基金 902163 checked!
每天股份 399097 checked!
元隆雅图 902878 checked!
                     星期六 002291 checked!
西部矿业 601168 checked!
电声股份 300805 checked!
 ☐ Event Log
612:6 CRLF UTF-8 4 spaces 🚹 👮
 Run: ②Verifier × 東京東達 300341 checked! 東衛地子 002036 checked! 原物地子 002036 checked! ② 市 新春達 300044 checked! 公 付格館 300014 checked! 公 卡斯克達 3000273 checked! 中級股份 000035 checked! 中级股份 000195 checked! * 黄光悠 600195 checked!
                     普元信息 688118 checked!
至纯科技 603690 checked!
                     卓易信息 688258 checked!
賽騰股份 603283 checked!
浙南证券 601878 checked!
勝海证券 300438 checked!
                     三利诺 002876 checked!
星星科技 300256 checked!
东睢股份 600114 checked!
                     三峡水利 600116 checked!
奥联电子 300585 checked!
                      圣邦股份 300661 checked!
华特气体 688268 checked!
香雪制药 300147 checked!
                     万讯自控 300112 checked!
奥普光电 002338 checked!
                      长城影视 002071 checked!
竞人乌 603555 checked!
                     亚玛顿 002623 checked!
教据比对成功!
                     1.股票信息文件原始数据
2.沪深A股近3年数据对比
                      3.文章文件元数据
4.种子文件元数据
       ▶. 4: Run 單 6: TODO ► Terminal All files are up-to-date (11 分钟之前)
```

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:

```
选择需要校验的文件:
1.股票信息文件原始贵铝
2.涉深A股近3字最揭对比
3.文章文件元贵铝
4.种子文件元贵铝
5.退出
输入选择:
                        輸入选择:
选择需要校验的股票
1.沪深A股
3.深证A股
4.新股
5.中小板
6.创业板
                          7.科创板
8.B股
                                                                                                                                                                                                                                                                                                                        ☐ Event Log
16930:2 CRLF UTF-8 4 spaces 🔒 💆
        ▶. 4: Run 單 6: TODO ☑ Terminal
All files are up-to-date (12 分钟之前)
      型 ■ 《基择需要校验的文件:

1.股票信息文件原始表据

2.沪深A股近3年表据对比

3.文章文件元表据

4.相子文件元表据

5.退出

输入选择:

▲地程用依贴的股票
                        輸入选择:
选择需要校验的股票
1.沪深A股
2.上证A股
3.深证A股
4.新股
5.中小板
6.创业的45
                         7.科创板
8.B股
                         Loading...
中昌教据 600242 checked!
天创时尚 603608 checked!
圣龙政份 603178 checked!
                          长白山 603099 checked!
引力传媒 603598 checked!
                         9 月 7 年 603-398 Checked!
百部矿业 601168 checked!
三季度份 603938 checked!
洛阳玻璃 600876 checked!
浪芬股份 603137 checked!
単街駅乗会 603825 checked!
                          吉翔股份 603399 checked!
五洲新春 603667 checked!
☐ Event Log
1240:6 CRLF UTF-8 4 spaces 🔏 👳
                         新力金融 600318 checked!
南京熊猫 600775 checked!
华贵实业 600191 checked!
柳荫股份 603368 checked!
桂东电力 600310 checked!
                         程东电力 600310 checked!
海床股份 600851 checked!
展率科技 603685 checked!
海信視像 600060 海信电器 600060
存在差异
                         1.股票信息文件原始教据
2.沪深A股近3年教据对比
3.文章文件元教据
4.种子文件元教据
        Event Log
```

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

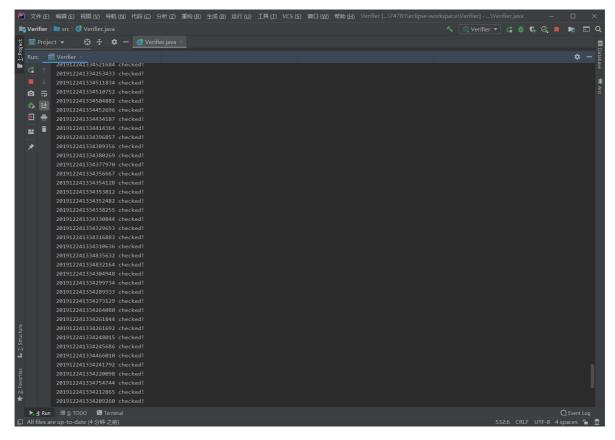
此时将对于3877支沪深A股的进3年的数据进行对比,选择的是自2017年1月1日到2019年12月27日收盘时的数据。验证效果如下:

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。

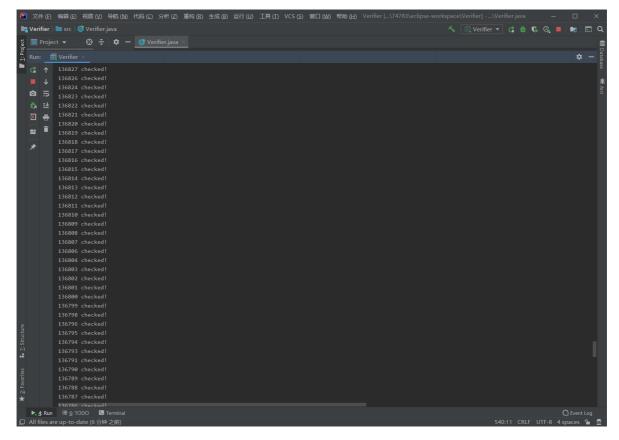
```
201912241334087042 checked!
201912241334057136 checked!
201912241334057136 checked!
201912241334052182 checked!
2019122413340951016 checked!
2019122413340014081 checked!
20191224133400142 checked!
201912241333992727 checked!
2019122413339957732 checked!
2019122413339957382 checked!
2019122413339957382 checked!
2019122413339957912 checked!
201912241333995910 checked!
201912241333995910 checked!
201912241333799911 checked!
201912241333799911 checked!
201912241333799911 checked!
201912241333799911 checked!
```

验证成功输出done。

种子文件元数据校验

```
本部重要収益物文件・
1. 服務意志文序無効本理
2. 产業品度工作業類対比
3. 大学文件元素類
5. 退世
輸入支票:
Louding...
136953不在未地
136953不在未地
136953不在未地
136953不在未地
136953不在未地
136953不在未地
136943不在未地
136943不在未地
136945不在未地
136955不在未地
1369557在年末地
1369557在日本地
1369557在日本地
1369557年日本地
1369557
```

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。

```
15650 Checked |
15650 Checked |
15670 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: