

# YYRating 爬虫程序操作手册

## 1 环境配置

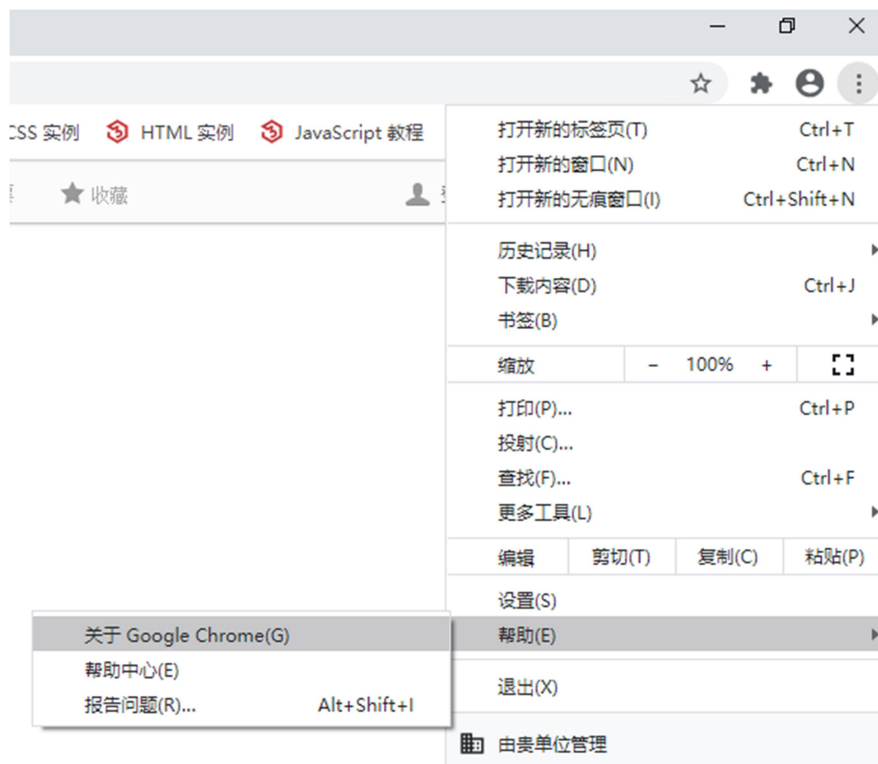
### 1.1 配置 chromedriver.exe

(1) 在 <http://npm.taobao.org/mirrors/chromedriver/> 获取 chromedriver

**注意：需要下载与当前谷歌浏览器版本一致的 chromedriver 版本（向下兼容）**

(2) 如何获得谷歌浏览器版本：

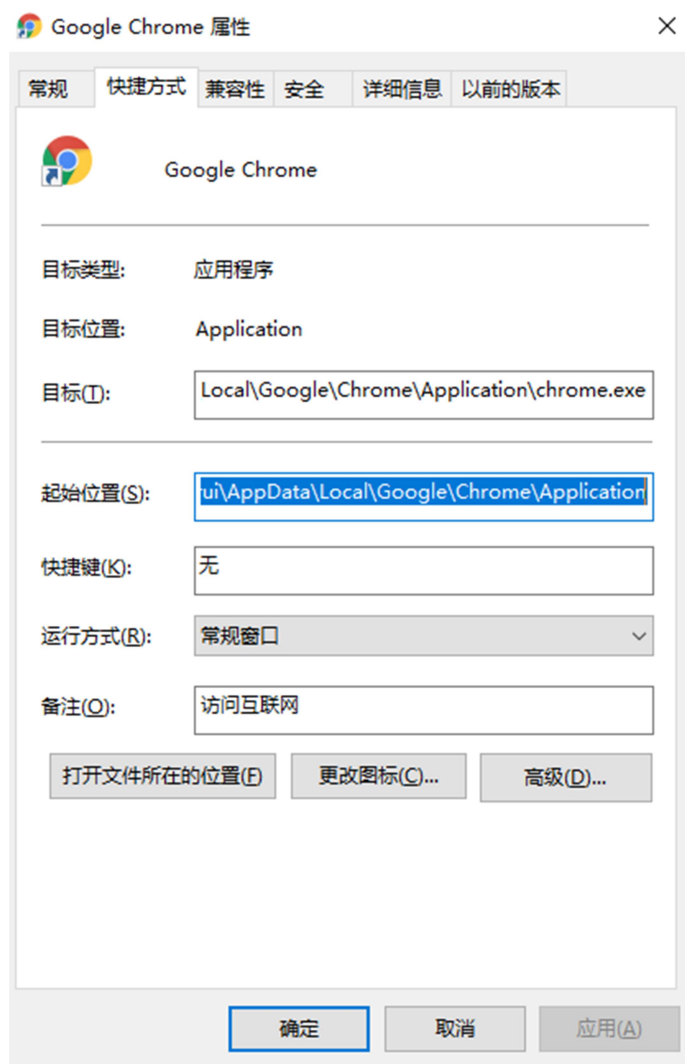
获得谷歌浏览器版本→浏览器右上角三个点→帮助→关于 Chrome→版本



(3) 将解压后的 chromedriver 放到谷歌浏览器的 Application 目录下

如 C:\Users\dengrui\AppData\Local\Google\Chrome\Application

谷歌浏览器属性→打开文件所在的位置→将 chromedriver.exe 放入



管理

Application

查看

应用程序工具

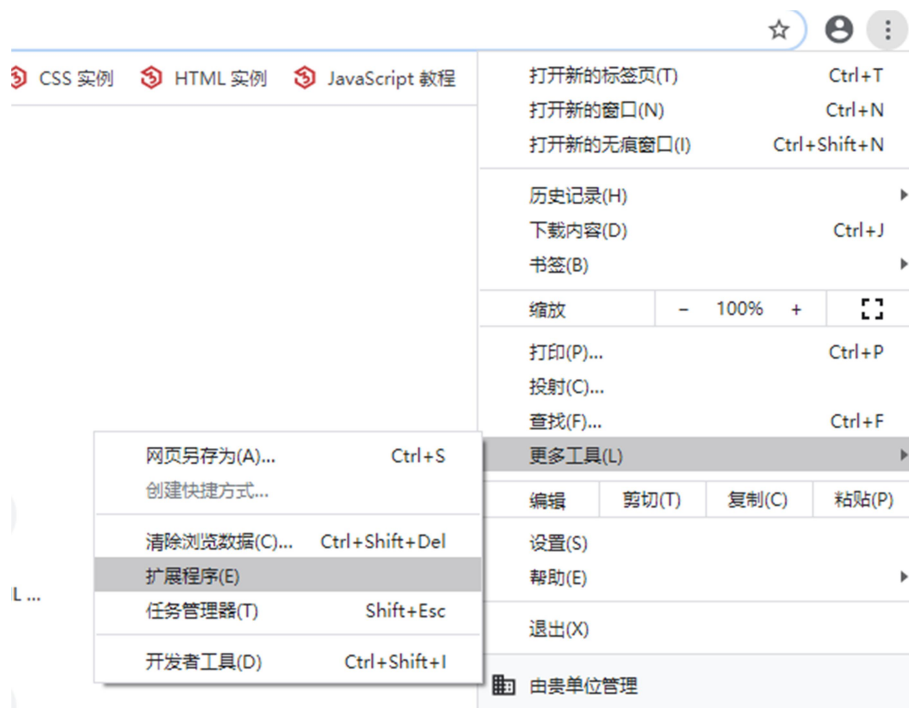
tui Deng
> AppData
> Local
> Google
> Chrome
> Application

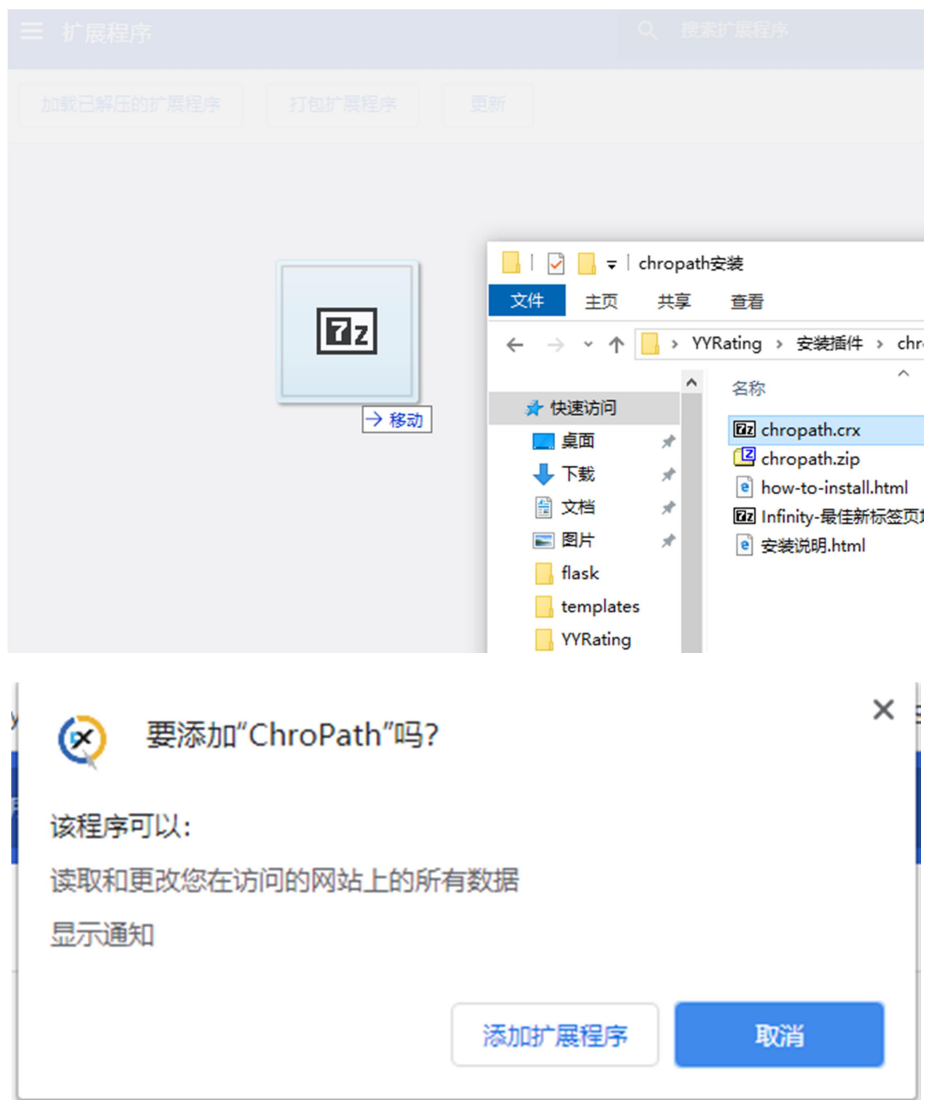
名称	修改日期	类型	大小
83.0.4103.97	2020/7/6 13:21	文件夹	
Dictionaries	2020/7/6 13:21	文件夹	
SetupMetrics	2020/7/6 13:21	文件夹	
chrome.exe	2020/6/2 9:07	应用程序	1,757 KB
chrome.VisualElementsManifest.xml	2020/7/6 13:21	XML 文档	1 KB
chrome_proxy.exe	2020/6/2 9:07	应用程序	676 KB
chromedriver.exe	2020/4/13 18:03	应用程序	8,637 KB
debug.log	2020/7/27 13:19	文本文档	20 KB

## 1.2 安装 chropath 插件 (如果仅运行程序, 不需要安装, 若需开发其他的爬虫程序, 使用 selenium 包将用到 chropath 插件)

(1) 将 chropath 加到谷歌浏览器扩展程序中

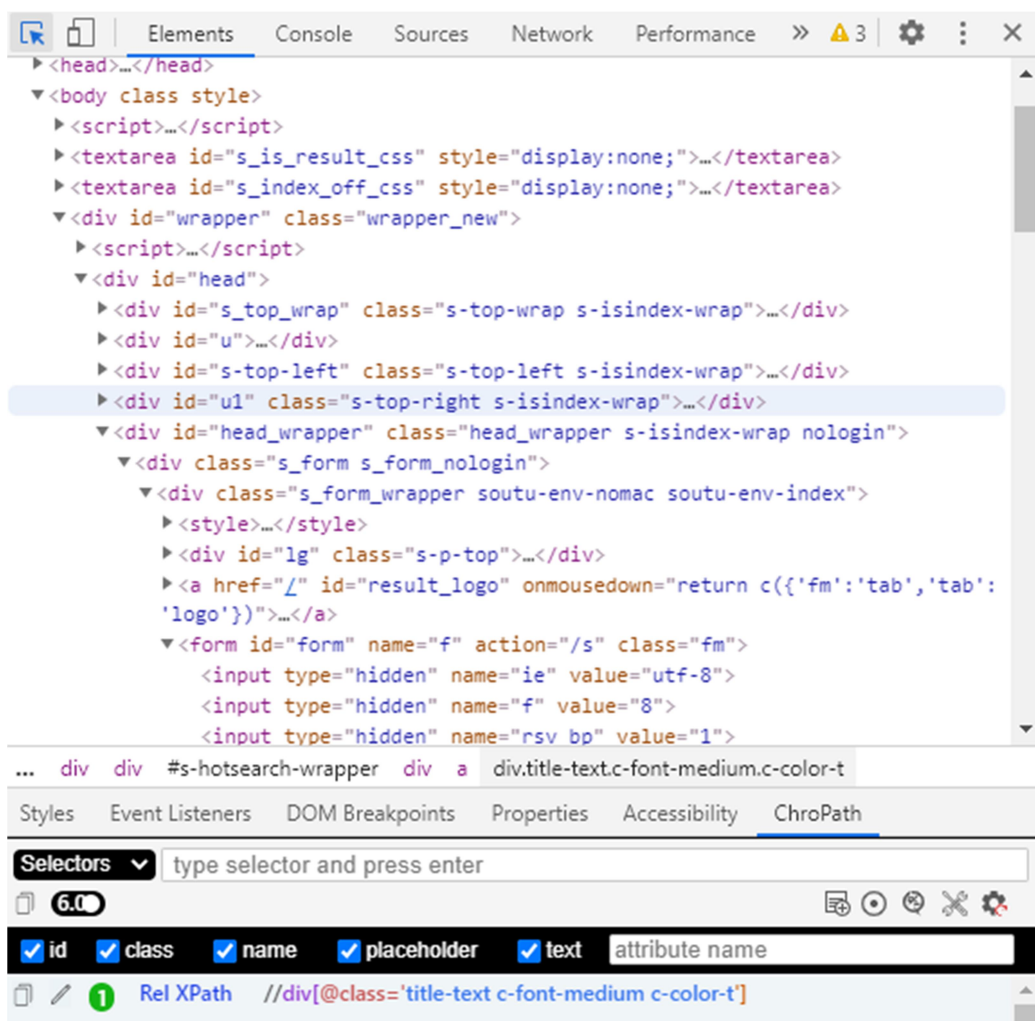
打开扩展程序界面→将 chropath.crx 拖拽进去→添加扩展程序





## (2) 使用 chropath

栏目选择 Elements, 然后点击最左边的箭头, 再选中网页的栏目, 最下方的 Rel Xpath 就是所需要的 Xpath



### 1.3 配置 config

将数据库连接的配置和 Chromedriver 的路径添加进 config.ini

**注意：**只需要更改等号右边的内容

config.ini - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

[msdb\_main]

msdb\_host = mainsql.ciccfund.com:1433

msdb\_user = ciccfund\123

msdb\_pwd = 456

msdb\_db = riskgroup

[Chromedriver]

Chromedriver\_path = C:\Users\dengrui\AppData\Local\Google\Chrome\Application

## 2 操作运行

### 2.1 事先确认数据库的各种表

(1) 确认查询的发行人表 **riskgroup.IA.bond\_issuer\_YYRating**，并将需要查询的  
发行人插入到该表中

(2) 确认插入的表，在程序中为 **riskgroup.IA.bond\_trans\_YYRating\_test**

```
try:
    # 删除数据库 起止时间 和 发行人
    delete_start = re.sub('\D', '', start)
    delete_end = re.sub('\D', '', end)
    sSql1 = '''delete from [riskgroup].[IA].[bond_trans_YYRating_test] where 成交日期 >= '%s' and
               delete_start, delete_end, current_issuer)
    cursor.execute(sSql1)
    connect.commit()

    issuer(current_issuer)
    time.sleep(st)
    issuer_data = get_data(start, end, current_issuer, gap, st) # 得到当前发行人的dataframe
```

```
# 插入数据库
sSql2 = '''insert into [riskgroup].[IA].[bond_trans_YYRating_test](成交日期,
                           债券简称,
                           发行人,
                           剩余期限,
                           成交价,
                           收益率,
                           成交量,
                           成交额,
                           YY等级,
                           成交隐含,
                           YY偏离,
```

(3) 确认纠错的表，在程序中为 **riskgroup.IA.bond\_trans\_YYRating\_error**

```
while status:
    #logging.info('!+1')
    sSql4 = '''select count(*) from [riskgroup].[IA].[bond_trans_YYRating_error] w
    cursor = connect.cursor()
    cursor.execute(sSql4)
    if cursor.fetchall()[0][0] == 0:
        status = False
    else:
        sSql5 = '''select * from [riskgroup].[IA].[bond_trans_YYRating_error]'''
        error_issuer_db = pd.read_sql(sSql5, connect).values.tolist()
        for i in error_issuer_db:
            if i[5] == 0:
```

## 2.2 程序运行

(1) 运行的程序是 **YYRating\_new.py**

(2) 在文件夹的路径下打开 cmd 输入参数 start (开始日期), end (结束日期), gap (间隔) 例如:

**YYRating\_new.py --start=2020-01-28 --end=2020-07-28 --gap=30**



```
C:\Windows\System32\cmd.exe
Microsoft Windows [版本 10.0.17763.379]
(c) 2018 Microsoft Corporation。保留所有权利。

C:\Users\dengrui\Desktop\YYRating>YYRating_new.py --start=2020-01-28 --end=2020-07-28 --gap=30_
```

**注意:** start 和 end 的输入均需要输入 10 位日期格式, gap 的取值依据经验应小于等于 30, 否则会造成错误过多, 使纠错过程冗长