

Selenium 获取天气网乡镇编码

作者: [liumeijian](#)

原文链接: <https://ld246.com/article/1560092755182>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

爬取天气网 <http://www.weather.com.cn> 乡镇编码

暂且只完成固定省份，可自行扩展到全国

尚存bug, 仅做参考

```
from time import sleep
from selenium import webdriver

#[https://npm.taobao.org/mirrors/chromedriver/](https://npm.taobao.org/mirrors/chromedriver/)
#下载与浏览器对应的平台和版本（本文采用版本号：75.0.3770.8）
chromepath = "~/Downloads/chromedriver"
driver = webdriver.Chrome(executable_path=chromepath);

# 最终结果
resultFile = r"~/Downloads/resultSuccess.sql"
# 事先新建数据 把需要查找的乡镇名称列表放入resultFailed.sql
# 运行完成后 自行查看resultFailed.sql 以及日志 对相应行作出修正
badResultSqlFile = r"~/Downloads/resultFailed.sql"

townNames = []
resultSql = []
badResultSql = []

def readFile():
    reader = open('~\\Downloads\\resultFailed.sql')
    #文件resultFailed.sql格式：其中" 为天气编码字段
    #INSERT INTO `dict_def` VALUES (1002, '行政区划', 1004, '乡镇', '350102001000', '鼓东街道', ''
    '350102000000');
    #INSERT INTO `dict_def` VALUES (1002, '行政区划', 1004, '乡镇', '350102002000', '鼓西街道', ''
    '350102000000');
    #INSERT INTO `dict_def` VALUES (1002, '行政区划', 1004, '乡镇', '350102003000', '温泉街道', ''
    '350102000000');
    #INSERT INTO `dict_def` VALUES (1002, '行政区划', 1004, '乡镇', '350102004000', '东街街道', ''
    '350102000000');
    #INSERT INTO `dict_def` VALUES (1002, '行政区划', 1004, '乡镇', '350102005000', '南街街道', ''
    '350102000000');
    #INSERT INTO `dict_def` VALUES (1002, '行政区划', 1004, '乡镇', '350102006000', '安泰街道', ''
    '350102000000');
    lines = reader.readlines()
    for i in lines:
        townNames.append({'name': i.split("\\")[7], 'sql': i})

def getCode():
    driver.get("http://forecast.weather.com.cn/town/weather1dn/101230101017.shtml#input")
    sleep(1)
    for i in townNames:
        num = ""
        try:
            driver.refresh()
        except Exception:
            #一般为网络状况差，导致刷新超时
            print(i['name'], ", ", '没有找到1->页面刷新异常')
```

```

        badResultSql.append(i)
        continue
    driver.find_element_by_xpath("//*[@id='txtZip']").clear()
    driver.find_element_by_xpath("//*[@id='txtZip']").send_keys(i['name']) # 输入搜索内容
    try:
        #可以根据网络状况, 自行削减时间
        sleep(2)
        li_list = driver.find_elements_by_xpath("//*[@id='txtZip']/../following-sibling::div[2]/ul/li")
    except Exception:
        print(i['name'], ", '没有找到2->不存在'")
        badResultSql.append(i)
        continue
    for x in li_list:
        if "福建" in x.text:
            num = x.get_attribute('num')
            break
    except Exception:
        print(i['name'], ", '没有找到3->未知异常'")
        badResultSql.append(i)
        continue
    result = i['sql'].replace("''", "\"" + num + "\"")
    resultSql.append(result)
    print(result)

def writeFile():
    fp = open(resultFile, 'a+')
    fp.write("\n".join(resultSql))
    fp.close()
    fp = open(badResultSqlFile, 'w+')
    fp.write("\n".join(badResultSql))
    fp.close()

if __name__ == '__main__':
    readFile()
    getCode()
    writeFile()

```