

**【信用晋中-行政许可&行政处罚】**

**爬虫数据需求文档**

**版本号V1.1**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 编写 | | 成艳琼 | | 编写时间 | | 20180619 | |
| 审核 | |  | | 审核时间 | |  | |
| 审批 | |  | | 审批时间 | |  | |
| 文档管理 | 修订次数 | | 0 | | 版本 | | V1.0 |

文档修改历史

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 版本 | 编写/修订说明 | 修订人 | 修订日期 | 备注 |
| 1 | V1.0 | 初稿 | 成艳琼 | 20180619 |  |
| 2 | V1.1 | 修改行政许可、处罚抓取字段 | 成艳琼 | 20180806 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## 数据源

信用晋中：http://www.creditsxjz.gov.cn/

## 抓取方式

遍历

## 更新周期

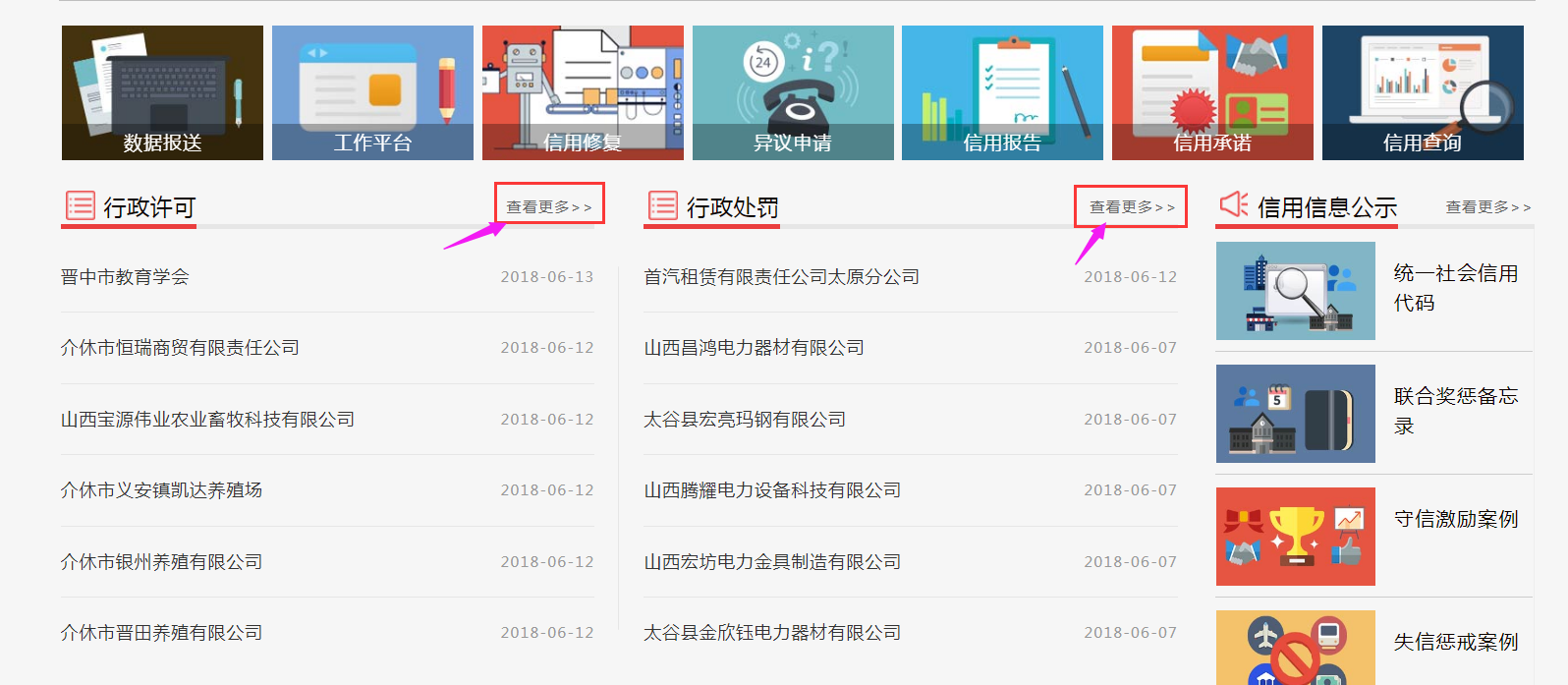
第一次抓取全量，后面每周一更新增量

## 详细抓取方法

### 4.1行政许可

网址：http://www.creditsxjz.gov.cn/xzxkListNew.jspx

1. 进入网站后，点击查看更多进入列表页

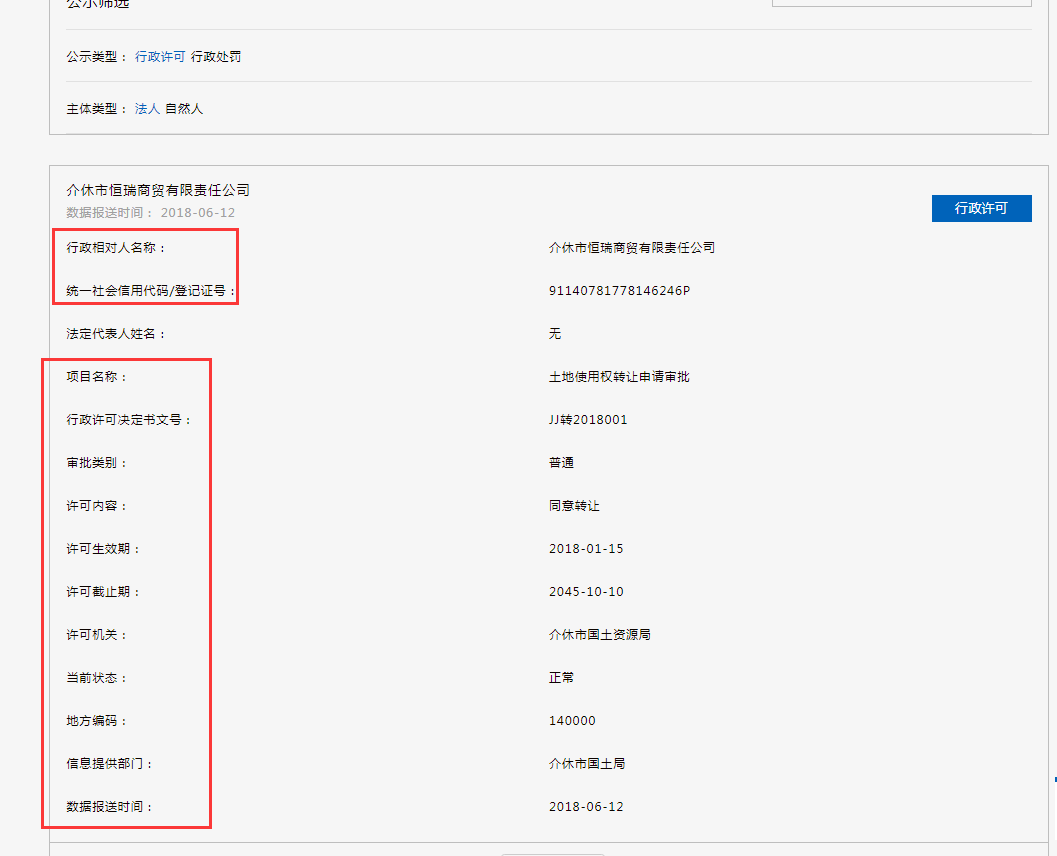


进入新页面后，依次点击每条数据，需要点击“下一页”抓取全量数据



1. 抓取字段：行政相对人名称、统一社会信用代码/登记证号、项目名称、行政许可决定书文号、审批类别、许可内容、许可生效期、许可截止期、许可机关、当前状态、地方编码、信息提供部门、数据报送时间、居民身份证号

法人（法定代表人姓名：不存该字段）



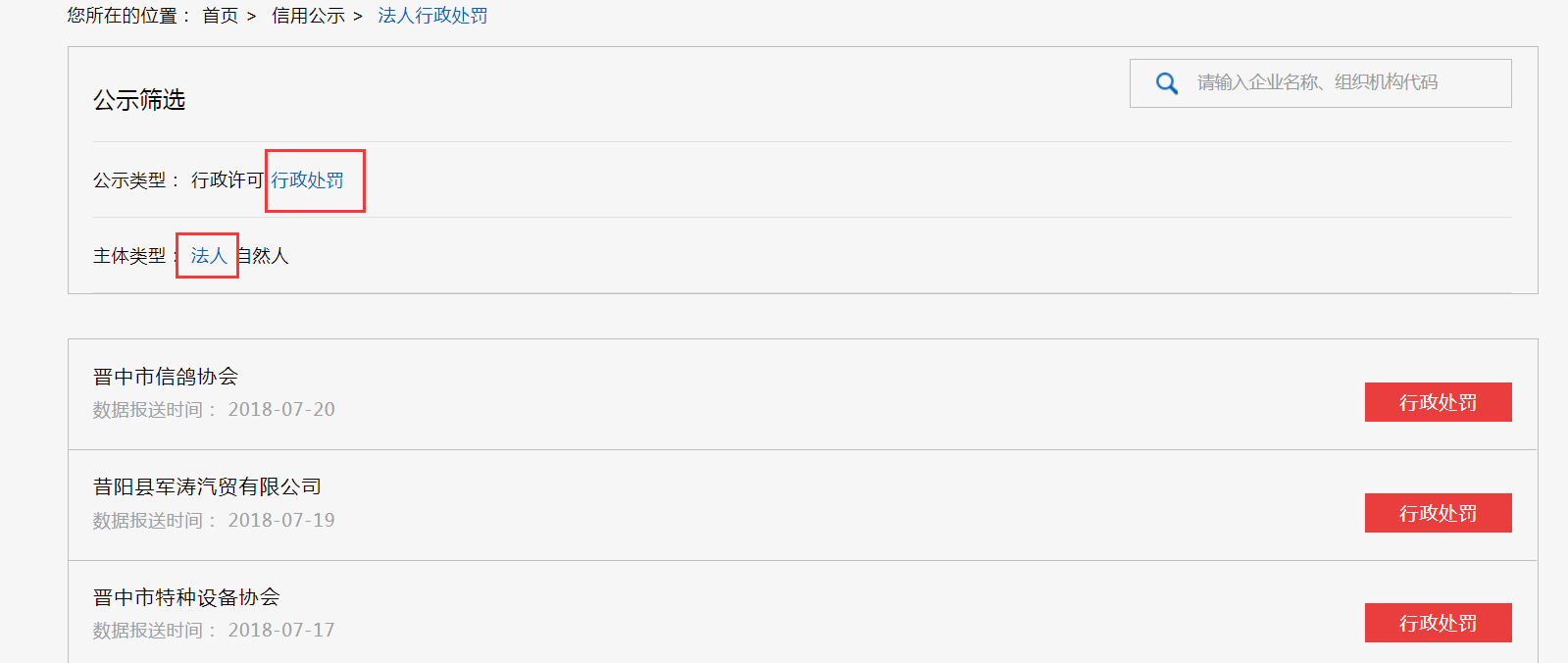
自然人（审批类别保留字段名，置空）



### 4.2行政处罚

网址：http://www.creditsxjz.gov.cn/xzcfListNew.jspx

1. 参考行政许可



1. 抓取字段：行政处罚决定书文号、案件名称、处罚类别、处罚事由、处罚依据、行政相对人名称、统一社会信用代码、处罚结果、处罚生效期、处罚机关、当前状态、地方编码、信息提供部门、数据报送时间



## 存储字段&存储结构

### 5.1行政许可（存储表名： qyxg\_xzxk）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 英文key | 字段名 | Value | 规则 | 数据类型 |
|  | \_id | 唯一id |  | 不可为空 | string |
|  | bbd\_source | 数据源 | 信用晋中 | 默认值，不可为空 | string |
|  | bbd\_table | 最终入库表名 | qyxg\_xzxk | 默认值，不可为空 | string |
|  | bbd\_type | 表类型 | credit\_jinzhong | 默认值，不可为空 | string |
|  | uptime | 时间戳 |  | 不可为空 | int |
|  | do\_time | 日期 |  | 不可为空 | string |
|  | version | 版本号 | 1 | 初始值为1，不可为空 | int |
|  | bbd\_seed | 种子对象 |  | 可为空 | string |
|  | bbd\_html | 网页原文 |  | 可为空 | string |
|  | bbd\_params | 网络提交参数 |  | 可为空 | string |
|  | bbd\_url | 网页链接 |  | 不可为空 | string |
|  | company\_name | 行政相对人名称 | 介休市恒瑞商贸有限责任公司 | 不可为空 | string |
|  | credit\_code | 统一社会信用代码/登记证号 | 91140781778146246P | 可为空 | string |
|  | case\_name | 项目名称 | 土地使用权转让申请审批 | 可为空 | string |
|  | license\_code | 行政许可决定书文号 | JJ转2018001 | 可为空 | string |
|  | approval\_category | 审批类别 | 普通 | 可为空 | string |
|  | license\_content | 许可内容 | 同意转让 | 可为空 | string |
|  | license\_org | 许可机关 | 介休市国土资源局 | 可为空 | string |
|  | license\_start\_date | 许可生效期 | 2018-01-15 | 可为空 | date |
|  | license\_end\_date | 许可截止期 | 2045-10-10 | 可为空 | date |
|  | license\_status | 当前状态 | 正常 | 可为空 | string |
|  | administrative\_code | 地方编码 | 140000 | 可为空 | string |
|  | data\_source | 信息提供部门 | 介休市国土局 | 可为空 | string |
|  | pubdate | 数据报送时间 | 2018-06-12 | 可为空 | date |
|  | id\_number | 居民身份证号 |  | 可为空 | string |

### 5.2行政处罚（存储表名： qyxg\_xzcf）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 英文key | 字段名 | Value | 规则 | 数据类型 |
|  | \_id | 唯一id |  | 不可为空 | string |
|  | bbd\_source | 数据源 | 信用晋中 | 默认值，不可为空 | string |
|  | bbd\_table | 最终入库表名 | qyxg\_xzcf | 默认值，不可为空 | string |
|  | bbd\_type | 表类型 | credit\_jinzhong | 默认值，不可为空 | string |
|  | uptime | 时间戳 |  | 不可为空 | int |
|  | do\_time | 日期 |  | 不可为空 | string |
|  | version | 版本号 | 1 | 初始值为1，不可为空 | int |
|  | bbd\_seed | 种子对象 |  | 可为空） | string |
|  | bbd\_html | 网页原文 |  | 可为空 | string |
|  | bbd\_params | 网络提交参数 |  | 可为空 | string |
|  | bbd\_url | 网页链接 |  | 不可为空 | string |
|  | punish\_code | 行政处罚决定书文号 | 《网络预约出租汽车经营服务管理暂行办法》 | 不可为空 | string |
|  | case\_name | 案件名称 | 首汽租赁有限责任公司太原分公司未取得经营许可，擅自从事网约车经营活动案 | 可为空 | string |
|  | punish\_category\_one | 处罚类别 | 罚款 | 可为空 | string |
|  | punish\_type | 处罚事由 | 未取得经营许可，擅自从事网约车经营活动 | 可为空 | string |
|  | punish\_basis | 处罚依据 | 《网络预约出租汽车经营服务管理暂行办法》 | 可为空 | string |
|  | company\_name | 行政相对人名称 | 首汽租赁有限责任公司太原分公司 | 不可为空 | string |
|  | credit\_code | 统一社会信用代码 | 91140105083745314Y | 可为空 | string |
|  | punish\_content | 处罚结果 | 已缴纳罚款 | 可为空 | string |
|  | punish\_date | 处罚生效期 | 2018-06-05 | 可为空 | date |
|  | punish\_org | 处罚机关 | 晋中市交通运输局 | 可为空 | string |
|  | punish\_status | 当前状态 | 正常 | 可为空 | string |
|  | administrative\_code | 地方编码 | 140700 | 可为空 | string |
|  | data\_source | 信息提供部门 | 晋中市交通运输局 | 可为空 | string |
|  | pubdate | 数据报送时间 | 2018-06-12 | 可为空 | date |

## 6、队列需求

无队列需求