

**【信用太原-行政许可&行政处罚】**

**爬虫数据需求文档**

**版本号V1.0**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## 数据源

信用太原：http://www.taiyuan.gov.cn/ztzl/ztzw/xyty/index.shtml

## 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：

1. 进入网站后，在页面中部点击 行政许可 进入列表页



进入新页面后，依次点击每条数据进入详情页，需要翻页抓取全量数据



1. 抓取字段：企业名称、统一社会信用代码、等级号证、行政许可决定文号、项目名称、审批类别、法定代表人（负责人）姓名、许可内容、许可生效期、许可截止期、许可机关、地方编码、当前状态、备注

<http://www.taiyuan.gov.cn/doc/2018/01/16/193578.shtml>



### 4.2行政处罚

网址：

1. 点击行政处罚进入列表页



1. 抓取字段：企业名称、统一社会信用代码、组织机构代码、行政处罚决定书文号、处罚名称、处罚类别、法定代表人（负责人）姓名、处罚事由、处罚依据、处罚生效期、处罚机关、处罚结果、当前状态、地方编码、备注

<http://www.taiyuan.gov.cn/doc/2018/01/16/192641.shtml>



## 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 | 信用太原 | 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\_taiyuan | 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 | 企业名称 | 中车置业（太原）有限公司 | 不可为空 | string |
|  | credit\_code | 统一社会信用代码 | 空 | 可为空 | string |
|  | rank\_code | 等级号证 | 91140107MAOGRYBK5D | 可为空 | string |
|  | case\_name | 项目名称 | 中车·国际广场 | 可为空 | string |
|  | license\_code | 行政许可决定文号 | 并发改审核发〔2016〕89号 | 可为空 | string |
|  | approval\_category | 审批类别 | 核准 | 可为空 | string |
|  | frname | 法定代表人（负责人）姓名 | 金磊 | 可为空 | string |
|  | license\_content | 许可内容 | 项目总建筑面积213488平方米，其中：地上建筑面积169338平方米，地下建筑面积44150平方米。 | 可为空 | string |
|  | license\_org | 许可机关 | 太原市发展和改革委员会 | 可为空 | string |
|  | license\_start\_date | 许可生效期 | 2016/3/31 | 可为空 | date |
|  | license\_end\_date | 许可截止期 | 2018/3/31 | 可为空 | date |
|  | license\_status | 当前状态 | 正常 | 可为空 | string |
|  | administrative\_code | 地方编码 | 140100 | 可为空 | string |
|  | remark | 备注 |  | 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 | 信用太原 | 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\_taiyuan | 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 | 企业名称 | 上海市建筑装饰工程集团有限公司 | 不可为空 | string |
|  | credit\_code | 统一社会信用代码 | 无 | 可为空 | string |
|  | organization\_code | 组织机构代码 |  | 可为空 | string |
|  | case\_name | 处罚名称 | 《山西省减少污染物排放条例》第四十七条 | 可为空 | string |
|  | punish\_code | 行政处罚决定书文号 | 并城乡执罚字2015第295号 | 可为空 | string |
|  | punish\_category | 处罚类别 | 罚款 | 可为空 | string |
|  | frname | 法定代表人（负责人）姓名 | 《山西省减少污染物排放条例》第四十七条 | 可为空 | string |
|  | punish\_type | 处罚事由 | 太原万达广场有限公司太原万达广场B2区1#楼工程在中考期间从事影响考生考试的施工作业 | 可为空 | string |
|  | punish\_content | 处罚结果 | 结案 | 可为空 | string |
|  | punish\_basis | 处罚依据 | 《山西省减少污染物排放条例》第五十八条（二）项 | 可为空 | string |
|  | punish\_org | 处罚机关 | 太原市城乡管理行政执法局 | 可为空 | string |
|  | punish\_date | 处罚生效期 | 2015/12/17 | 可为空 | date |
|  | punish\_status | 当前状态 | 正常 | 可为空 | string |
|  | administrative\_code | 地方编码 | 140100 | 可为空 | string |
|  | remark | 备注 |  | 可为空 | string |

## 6、队列需求

无队列需求