

**【信用淄博-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 | | 20180621 | |
| 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 | 20180621 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## 数据源

信用淄博：http://www.xinyongzibo.gov.cn/zibo/index.html

## 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.xinyongzibo.gov.cn/zibo/administration.html?tab=1

1. 进入网站后，点击更多进入列表页



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：决定书文号、项目名称、审批类别、行政相对人名称、法定代表人名称、社会信用代码、组织机构代码、工商登记码、税务登记号、身份证号、许可机构、许可决定日期、许可截止期、地方编码、备注、许可内容

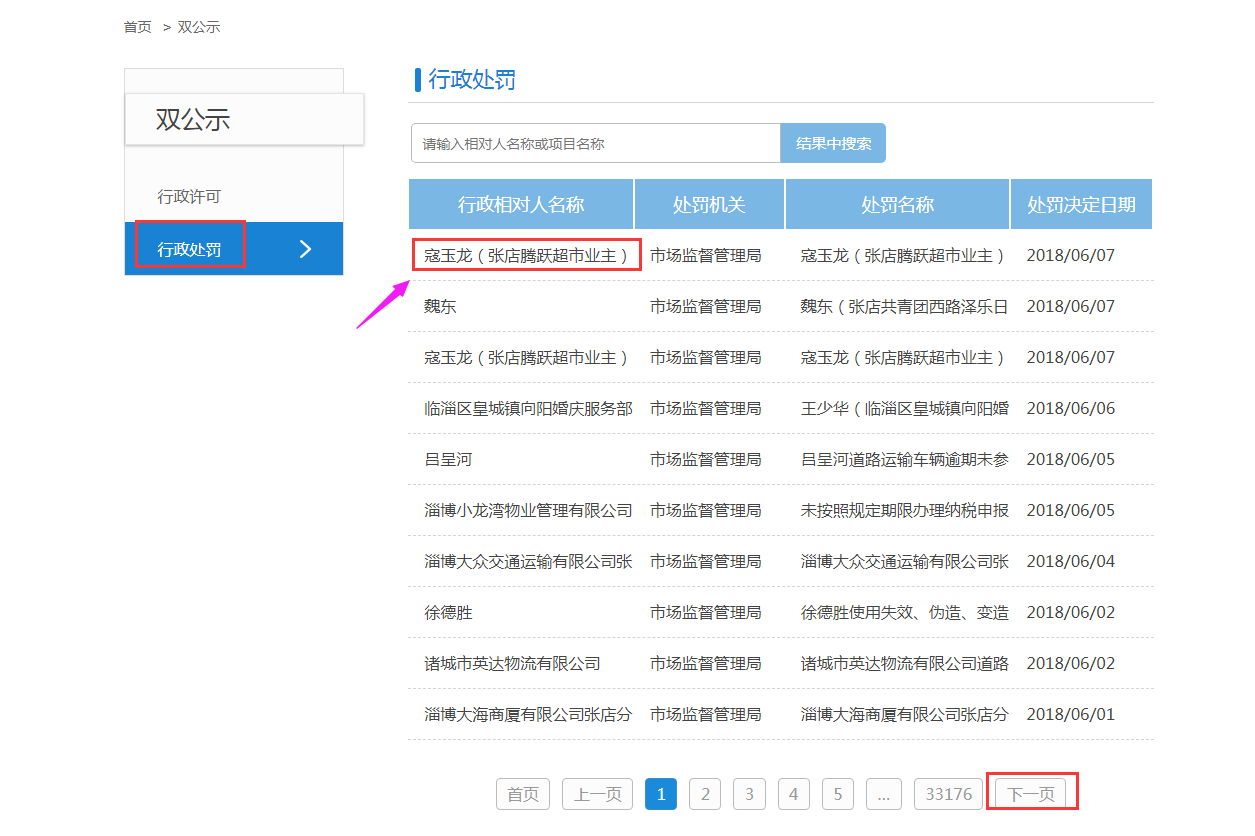
<http://www.xinyongzibo.gov.cn/zibo/administration_details.html?sys_id=1B09394A0A80110F19941890D143B047>



### 4.2行政处罚

网址：http://www.xinyongzibo.gov.cn/zibo/administration.html?tab=1

1. 参考行政许可



1. 抓取字段：决定书文号、处罚名称、处罚类别1、处罚类别2、行政相对人名称、法定代表人姓名、社会信用代码、组织机构代码、工商登记码、税务登记号、身份证号、处罚机构、决定日期、地方编码、状态、备注、处罚事由、处罚依据、处罚结果

<http://www.xinyongzibo.gov.cn/zibo/CfDetails.html?sys_id=16F17A980A80110F19941890F373704A>



## 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 | 信用淄博 | 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\_zibo | 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 | http://www.xinyongzibo.gov.cn/zibo/administration\_details.html?sys\_id=1B09394A0A80110F19941890D143B047 | 不可为空 | string |
|  | license\_code | 决定书文号 | 空 | 可为空 | string |
|  | case\_name | 项目名称 | 公司变更 | 可为空 | string |
|  | approval\_category | 审批类别 | 登记 | 可为空 | string |
|  | company\_name | 行政相对人名称 | 山东紫罗兰新材料科技有限公司 | 不可为空 | string |
|  | frname | 法定代表人姓名 | 董顺卿 | 可为空 | string |
|  | credit\_code | 社会信用代码 | 91370304MA3F81LM69 | 可为空 | string |
|  | organization\_code | 组织机构代码 |  | 可为空 | string |
|  | regno | 工商登记码 |  | 可为空 | string |
|  | tax\_code | 税务登记号 |  | 可为空 | string |
|  | id\_number | 身份证号 |  | 可为空 | string |
|  | license\_org | 许可机构 | 博山区工商行政管理局 | 可为空 | string |
|  | license\_start\_date | 许可决定日期 | 2018-06-19 | 可为空 | date |
|  | license\_end\_date | 许可截止期 | 2099-12-31 | 可为空 | date |
|  | administrative\_code | 地方编码 | 370304 | 可为空 | string |
|  | remark | 备注 |  | 可为空 | string |
|  | license\_content | 许可内容 | 公司变更 | 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 | 信用淄博 | 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\_zibo | 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 | http://www.xinyongzibo.gov.cn/zibo/CfDetails.html?sys\_id=16F17A980A80110F19941890F373704A | 不可为空 | string |
|  | punish\_code | 决定书文号 | 淄高新地税 罚 〔2018〕 4 号 | 可为空 | string |
|  | case\_name | 处罚名称 | 未按照规定期限办理纳税申报和报送纳税资料 | 可为空 | string |
|  | punish\_category\_one | 处罚类别1 | 罚款 | 可为空 | string |
|  | punish\_category\_two | 处罚类别2 |  | 可为空 | string |
|  | company\_name | 行政相对人名称 | 淄博小龙湾物业管理有限公司 | 不可为空 | string |
|  | frname | 法定代表人姓名 | 张宇树 | 可为空 | string |
|  | credit\_code | 社会信用代码 | 370303765754816 | 可为空 | string |
|  | organization\_code | 组织机构代码 |  | 可为空 | string |
|  | regno | 工商登记码 |  | 可为空 | string |
|  | tax\_code | 税务登记号 |  | 可为空 | string |
|  | id\_number | 身份证号 | 370303\*\*\*\*\*\*\*43 | 可为空 | string |
|  | punish\_org | 处罚机构 | 淄博市地方税务局高新区分局 | 可为空 | string |
|  | punish\_date | 决定日期 | 2018-06-05 | 可为空 | date |
|  | administrative\_code | 地方编码 | 370303 | 可为空 | string |
|  | punish\_status | 状态 | 正常 | 可为空 | string |
|  | remark | 备注 |  | 可为空 | string |
|  | punish\_type | 处罚事由 | 未按照规定期限办理纳税申报和报送纳税资料 | 可为空 | string |
|  | punish\_basis | 处罚依据 | 《中华人民共和国税收征收管理法》第六十二条 | 可为空 | string |
|  | punish\_content | 处罚结果 | 3000 | 可为空 | double |

## 6、队列需求

无队列需求