

**【信用忻州-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

信用忻州：http://www.creditxz.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.creditxz.gov.cn/XzCreditWeb/xzxkList.jspx

1. 进入网站后，点击 行政许可 进入列表页



进入新页面后，依次选择 **法人行政许可公示、自然人行政许可公示**，点击每条数据的查看按钮进入详情页，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、法定代表人姓名、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

**法人行政许可公示**：

<http://www.creditxz.gov.cn/XzCreditWeb/xzxkDetial-5896033c4d3c4cf5a550f1aee6b5ef7b.jspx?sgstype=sgs>



**自然人行政许可公示：**

<http://www.creditxz.gov.cn/XzCreditWeb/xzxkPDetial-328fc86c24264833bfc8d2f4109acfed.jspx?sgstype=sgs>



### 4.2行政处罚

网址：http://www.creditxz.gov.cn/XzCreditWeb/xzcfList.jspx?sgstype=sgs

1. 抓取 法人行政处罚公示 的数据，需要翻页抓取全量



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、组织机构代码、工商登记码、税务登记号、法定代表人居民身份证号、法定代表人姓名、处罚结果、处罚生效期、处罚机关、当前状态、地方编码、备注、信息提供部门、数据报送时间

<http://www.creditxz.gov.cn/XzCreditWeb/xzcfDetial-613b217243f24ce0a2af035d6bc41bec.jspx?sgstype=sgs>



## 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 | 信用忻州 | 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\_yizhou | 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.creditxz.gov.cn/XzCreditWeb/xzxkDetial-5896033c4d3c4cf5a550f1aee6b5ef7b.jspx?sgstype=sgs | 不可为空 | string |
|  | company\_name | 行政相对人名称 | 忻州盛世房地产开发有限公司 | 不可为空 | string |
|  | credit\_code | 统一社会信用代码/登记证号 | 9114090079635281XM | 可为空 | string |
|  | frname | 法定代表人姓名 | 贺换如 | 可为空 | string |
|  | case\_name | 项目名称 | 盛世佳苑新建商品住宅楼 | 可为空 | string |
|  | license\_code | 行政许可决定书文号 | 140900201800008 | 可为空 | string |
|  | approval\_category | 审批类别 | 普通 | 可为空 | string |
|  | license\_content | 许可内容 | 建设工程竣工规划认可证核发 | 可为空 | string |
|  | license\_org | 许可机关 | 忻州市规划勘测局 | 可为空 | string |
|  | license\_start\_date | 许可生效期 | 2018-05-31 | 可为空 | date |
|  | license\_end\_date | 许可截止期 | 2099-12-31 | 可为空 | date |
|  | license\_status | 当前状态 | 正常 | 可为空 | string |
|  | administrative\_code | 地方编码 | 140900 | 可为空 | string |
|  | data\_source | 信息提供部门 | 忻州市发展和改革委员会 | 可为空 | string |
|  | pubdate | 数据报送时间 | 2018-06-14 | 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 | 信用忻州 | 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\_yizhou | 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.creditxz.gov.cn/XzCreditWeb/xzcfDetial-613b217243f24ce0a2af035d6bc41bec.jspx?sgstype=sgs | 不可为空 | string |
|  | punish\_code | 行政处罚决定书文号 | 忻环法〔2018〕002号 | 不可为空 | string |
|  | case\_name | 案件名称 | 行政处罚决定书 | 可为空 | string |
|  | punish\_category\_one | 处罚类别 | 罚款 | 可为空 | string |
|  | punish\_type | 处罚事由 | 1、厂区运输道路未硬化，未采取降尘措施，易对周边大气环境造成扬尘污染。 2、物料堆场… | 可为空 | string |
|  | punish\_basis | 处罚依据 | 《中华人民共和国大气污染防治法》第一百零八条第五项和… | 可为空 | string |
|  | company\_name | 行政相对人名称 | 山西忻州忻博矿业有限公司 | 不可为空 | string |
|  | credit\_code | 统一社会信用代码 | 91140900715986248J | 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 | 法定代表人居民身份证号 |  | 可为空 | string |
|  | frname | 法定代表人姓名 |  | 可为空 | string |
|  | punish\_content | 处罚结果 | 罚款十三万元 | 可为空 | string |
|  | punish\_date | 处罚生效期 | 2018-03-02 | 可为空 | date |
|  | punish\_org | 处罚机关 | 忻州市环境保护局 | 可为空 | string |
|  | punish\_status | 当前状态 | 正常 | 可为空 | string |
|  | administrative\_code | 地方编码 | 140927 | 可为空 | string |
|  | remark | 备注 |  | 可为空 | string |
|  | data\_source | 信息提供部门 | 忻州市发展和改革委员会 | 可为空 | string |
|  | pubdate | 数据报送时间 | 2018-06-14 | 可为空 | date |

## 6、队列需求

无队列需求