

**【信用长春-行政许可&行政处罚】**

**爬虫数据需求文档**

**版本号V1.0**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 编写 | | 王皓晔 | | 编写时间 | | 20180622 | |
| 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 | 王皓晔 | 20180622 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## 数据源

信用长春：http://cccredit.changchun.gov.cn/web/index/（注：含有个人数据）

## 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

1. 进入网站后，点击查看更多进入列表页



进入新页面后，点击行政许可，然后依次点击每条数据的行政相对人名称进入详情页面（需要翻页抓取全量数据）。

http://cccredit.changchun.gov.cn/web/dou/permit/non/load

1. 抓取字段：行政许可决定书文号、项目名称、审批类别、许可内容、行政相对人名称、行政相对人代码\_1(统一社会信用代码)、行政相对人代码\_2(组织机构代码)、行政相对人代码\_3(工商登记码)、行政相对人代码\_4(税务登记号)、行政相对人代码\_5 (居民身份证号)、法定代表人姓名、许可决定日期、许可截止期、许可机关、当前状态、地方编码、数据更新时间戳、备注。



### 4.2行政处罚

1. 参考行政许可，进入网站后，点击查看更多进入列表页。进入新页面后，点击行政处罚，然后依次点击每条数据的行政相对人名称进入详情页面（需要翻页抓取全量数据）。

http://cccredit.changchun.gov.cn/web/dou/punish/non/load



1. 抓取字段：行政处罚决定书文号、处罚名称、处罚类别1、处罚类别2、处罚事由、处罚依据、行政相对人名称、行政相对人代码\_1(统一社会信用代码)、行政相对人代码\_2(组织机构代码)、行政相对人代码\_3(工商登记码)、行政相对人代码\_4(税务登记号)、行政相对人代码\_5 (居民身份证号)、法定代表人姓名、处罚结果、处罚决定日期、处罚机关、当前状态、地方编码、数据更新时间戳、备注



## 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 |
| 1 | \_id | 唯一id |  | 不可为空 | string |
| 2 | bbd\_source | 数据源 | 信用长春 | 默认值，不可为空 | string |
| 3 | bbd\_table | 最终入库表名 | qyxg\_xzxk | 默认值，不可为空 | string |
| 4 | bbd\_type | 表类型 | credit\_changchun | 默认值，不可为空 | string |
| 5 | uptime | 时间戳 |  | 不可为空 | int |
| 6 | do\_time | 日期 |  | 不可为空 | string |
| 7 | version | 版本号 | 1 | 初始值为1，不可为空 | int |
| 8 | bbd\_seed | 种子对象 |  | 可为空 | string |
| 9 | bbd\_html | 网页原文 |  | 可为空 | string |
| 10 | bbd\_params | 网络提交参数 |  | 可为空 | string |
| 11 | bbd\_url | 网页链接 |  | 不可为空 | string |
| 12 | license\_code | 行政许可决定书文号 | 榆烟专[ 2018 ]第307号 | 可为空 | string |
| 13 | case\_name | 项目名称 | 烟草专卖零售许可证 | 可为空 | string |
| 14 | approval\_category | 审批类别 | 普通 | 可为空 | string |
| 15 | license\_content | 许可内容 | 卷烟、雪茄烟零售 | 可为空 | string |
| 16 | company\_name | 行政相对人名称 | 李洪星 | 不可为空 | string |
| 17 | credit\_code | 行政相对人代码\_1(统一社会信用代码) |  | 可为空 | string |
| 18 | organization\_code | 行政相对人代码\_2(组织机构代码) |  | 可为空 | string |
| 19 | regno | 行政相对人代码\_3(工商登记码) |  | 可为空 | string |
| 20 | tax\_code | 行政相对人代码\_4(税务登记号) |  | 可为空 | string |
| 21 | id\_number | 行政相对人代码\_5 (居民身份证号) | 2201821974\*\*\*\*0644 | 可为空 | string |
| 22 | frname | 法定代表人姓名 | 李洪星 | 可为空 | string |
| 23 | license\_org | 许可机关 | 榆树市烟草专卖局 | 可为空 | string |
| 24 | license\_start\_date | 许可决定日期 | 2018/06/12 | 可为空 | date |
| 25 | license\_end\_date | 许可截止期 | 2018/12/31 | 可为空 | date |
| 26 | administrative\_code | 地方编码 | 220182 | 可为空 | string |
| 27 | license\_status | 当前状态 | 正常 | 可为空 | string |
| 28 | remark | 备注 |  | 可为空 | string |
| 29 | update | 数据更新时间戳 | 2018/06/13 | 可为空 | date |

### 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 |
| 1 | \_id | 唯一id |  | 不可为空 | string |
| 2 | bbd\_source | 数据源 | 信用长春 | 默认值，不可为空 | string |
| 3 | bbd\_table | 最终入库表名 | qyxg\_xzcf | 默认值，不可为空 | string |
| 4 | bbd\_type | 表类型 | credit\_changchun | 默认值，不可为空 | string |
| 5 | uptime | 时间戳 |  | 不可为空 | int |
| 6 | do\_time | 日期 |  | 不可为空 | string |
| 7 | version | 版本号 | 1 | 初始值为1，不可为空 | int |
| 8 | bbd\_seed | 种子对象 |  | 可为空） | string |
| 9 | bbd\_html | 网页原文 |  | 可为空 | string |
| 10 | bbd\_params | 网络提交参数 |  | 可为空 | string |
| 11 | bbd\_url | 网页链接 |  | 不可为空 | string |
| 12 | punish\_code | 行政处罚决定书文号 | 榆烟专处【2018】第68号 | 可为空 | string |
| 13 | case\_name | 处罚名称 | 未在当地烟草批发企业进货 | 可为空 | string |
| 14 | punish\_category\_one | 处罚类别1 | 罚款 | 可为空 | string |
| 15 | punish\_category\_two | 处罚类别2 | 违规真品卷烟予以返还 | 可为空 | string |
| 16 | punish\_type | 处罚事由 | 未在当地烟草专卖批发企业进货 | 可为空 | string |
| 17 | punish\_basis | 处罚依据 | 《中华人民共和国烟草专卖条例》第23条第二款、56条 | 可为空 | string |
| 18 | company\_name | 行政相对人名称 | 刘杨 | 不可为空 | string |
| 19 | credit\_code | 行政相对人代码\_1(统一社会信用代码) |  | 可为空 | string |
| 20 | organization\_code | 行政相对人代码\_2(组织机构代码) |  | 可为空 | string |
| 21 | regno | 行政相对人代码\_3(工商登记码) |  | 可为空 | string |
| 22 | tax\_code | 行政相对人代码\_4(税务登记号) |  | 可为空 | string |
| 23 | id\_number | 行政相对人代码\_5 (居民身份证号) | 2201821988\*\*\*\*0214 | 可为空 | string |
| 24 | frname | 法定代表人姓名 | 刘杨 | 可为空 | string |
| 25 | punish\_content | 处罚结果 | 罚款5.25元，真烟返还。 | 可为空 | string |
| 26 | punish\_date | 处罚决定日期 | 2018/05/31 | 可为空 | date |
| 27 | punish\_org | 处罚机关 | 榆树市烟草专卖局 | 可为空 | string |
| 28 | punish\_status | 当前状态 | 正常 | 可为空 | string |
| 29 | administrative\_code | 地方编码 | 220182 | 可为空 | string |
| 30 | remark | 备注 |  | 可为空 | string |
| 31 | update | 数据更新时间戳 | 2018/06/01 | 可为空 | data |

## 6、队列需求

无队列需求