

**【信用中国（东营）-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 | | 20180702 | |
| 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

信用中国（东营）：http://credit.dongying.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

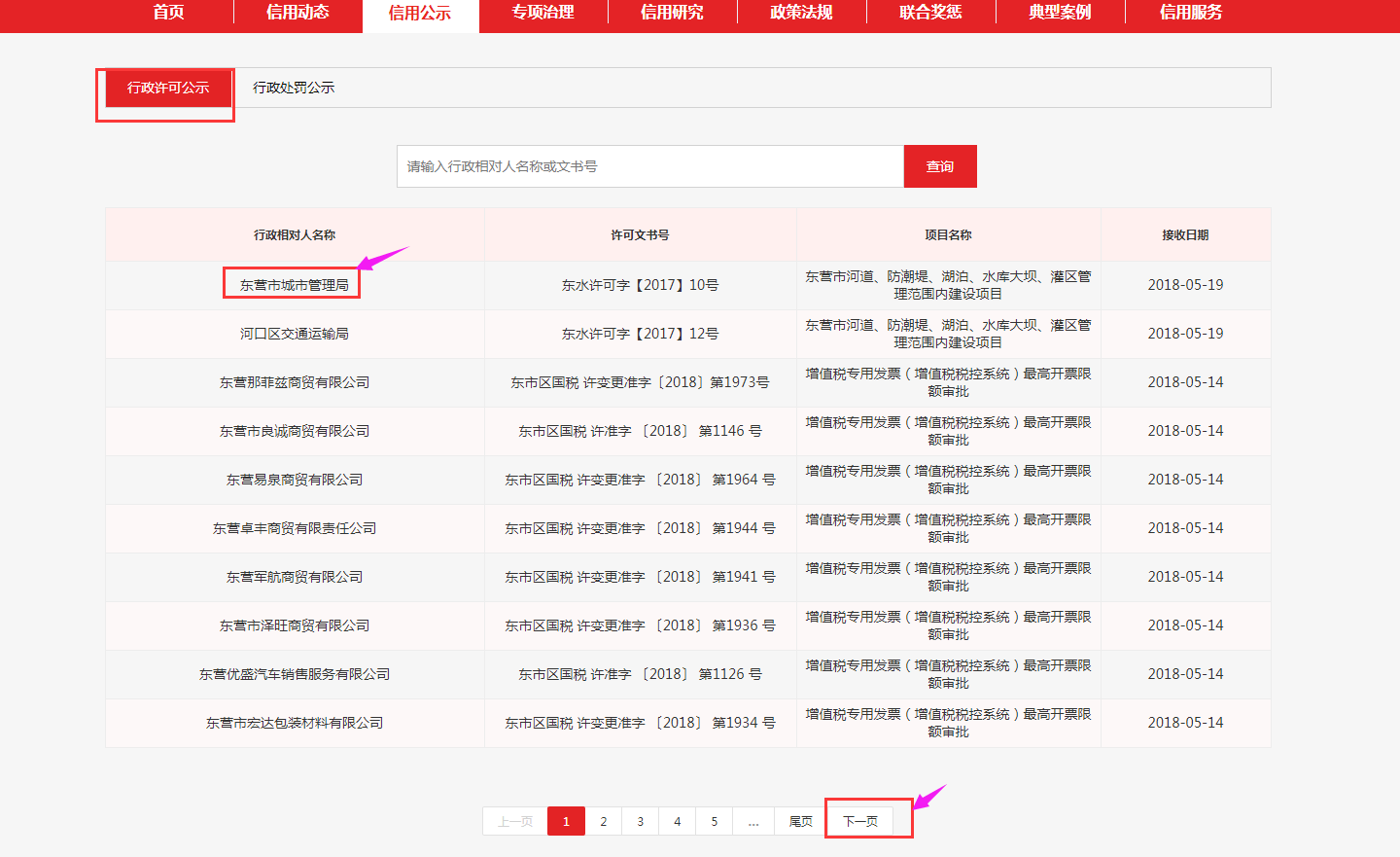
### 4.1行政许可

网址：http://credit.dongying.gov.cn/html/infoPublicity/admPermit.html

1. 进入网站后，点击更多进入列表页



进入新页面后，依次点击每条数据的详情，需要翻页抓取全量数据



1. 抓取字段：行政相对人名称、许可机关、行政许可决定文书号、项目名称、审批类别、许可内容、统一社会信用代码、组织机构代码、工商登记码、税务登记号、居民身份证号、法定代表人姓名、许可生效期、许可截止期、备注

<http://credit.dongying.gov.cn/html/infoPublicity/permitDetails.html?ID=102665>



### 4.2行政处罚

网址：http://credit.dongying.gov.cn/html/infoPublicity/admPunish.html

1. 参考行政许可
2. 抓取字段：行政相对人名称、处罚机关、处罚名称、行政处罚决定文书号、处罚类别、处罚事由、处罚依据、统一社会信用代码、组织机构代码、工商登记码、税务登记号、居民身份证号、法定代表人姓名、处罚结果、处罚时间、备注

http://credit.dongying.gov.cn/html/infoPublicity/punishDetails.html?ID=101545



## 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 | 信用中国（东营） | 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\_dongying | 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://credit.dongying.gov.cn/html/infoPublicity/permitDetails.html?ID=102665 | 不可为空 | string |
|  | company\_name | 行政相对人名称 | 东营那菲兹商贸有限公司 | 不可为空 | string |
|  | license\_org | 许可机关 | 东营市市区国家税务局胜利经济开发区税务分局 | 可为空 | string |
|  | license\_code | 行政许可决定文书号 | 东市区国税 许变更准字〔2018〕第1973号 | 可为空 | string |
|  | case\_name | 项目名称 | 增值税专用发票（增值税税控系统）最高开票限额审批 | 可为空 | string |
|  | approval\_category | 审批类别 | 普通 | 可为空 | string |
|  | license\_content | 许可内容 | 根据《中华人民共和国行政许可法》第四十九条规定，决定准予你（你单位）变更对增值税防伪税控系统最高开票限额的审批的行政许可 | 可为空 | string |
|  | credit\_code | 统一社会信用代码 | 91370502MA3ENEQT1T | 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 | 91370502MA3ENEQT1T | 可为空 | string |
|  | id\_number | 居民身份证号 | 330324\*\*\*\*\*\*\*\*1052 | 可为空 | string |
|  | frname | 法定代表人姓名 | 林博正 | 可为空 | string |
|  | license\_start\_date | 许可生效期 | 2018-05-11 | 可为空 | date |
|  | license\_end\_date | 许可截止期 | 2099-12-31 | 可为空 | date |
|  | 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 | 信用中国（东营） | 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\_dongying | 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://credit.dongying.gov.cn/html/infoPublicity/punishDetails.html?ID=101545 | 不可为空 | string |
|  | company\_name | 行政相对人名称 | 广饶恒丰健身中心 | 不可为空 | string |
|  | punish\_org | 处罚机关 | 广饶县国家税务局城区税务分局 | 可为空 | string |
|  | case\_name | 处罚名称 | 广饶恒丰健身中心-违反税收管理 | 可为空 | string |
|  | punish\_code | 行政处罚决定文书号 | 广国税 罚 〔2018〕 57 号 | 可为空 | string |
|  | punish\_category\_one | 处罚类别 | 罚款 | 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 |
|  | credit\_code | 统一社会信用代码 | 92370523MA3FBND28T | 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 | 92370523MA3FBND28T | 可为空 | date |
|  | id\_number | 居民身份证号 | 370523\*\*\*\*\*\*\*\*2018 | 可为空 | string |
|  | frname | 法定代表人姓名 | 李德帅 | 可为空 | string |
|  | punish\_content | 处罚结果 | 处以罚款200元 | 可为空 | string |
|  | punish\_date | 处罚时间 | 2018-05-07 | 可为空 | date |
|  | remark | 备注 |  | 可为空 | string |

## 6、队列需求

无队列需求