**爬虫组协助需求**

**编制:雷邓君**

**项目类型:核心产品**

**本文件修订历史信息:**

|  |  |  |  |
| --- | --- | --- | --- |
| **更新日期** | **修订者** | **批准人** | **主要修订摘要** |
| 2018.7.13 | 雷邓君 | 刘伟 | 新编写文件 |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# 1.需求概况

## 1.1目标

估值项目组请求爬虫项目组协助抓取竞争对手与我方二手车估值价格,以帮助进行数据验证对比分析.

# 2.功能性需求

根据指定的数据来源,交易类型,车况,抓取竞争对手与我方对应的二手车估值的中间值.

## 2.1数据来源

需要爬虫组抓取的数据汇总表 valuate.valuate\_management\_competitor\_origin\_data

**字段说明:**

CREATE TABLE valuate.valuate\_management\_competitor\_origin\_data (

id INT AUTO\_INCREMENT PRIMARY KEY ,

source\_id int(11) DEFAULT NULL COMMENT '数据来源ID',

title varchar(200) DEFAULT NULL COMMENT '标题',

pub\_time datetime NOT NULL COMMENT '发布时间',

brand\_slug varchar(32) DEFAULT NULL COMMENT '品牌slug',

model\_slug varchar(32) DEFAULT NULL COMMENT '型号slug',

model\_detail\_slug varchar(32) DEFAULT NULL COMMENT '款型slug',

control varchar(32) DEFAULT NULL COMMENT '变速箱',

volume decimal(5,1) DEFAULT NULL COMMENT '排量',

body\_style varchar(50) DEFAULT NULL COMMENT '车型样式',

car\_condition\_slug varchar(32) DEFAULT NULL COMMENT '车况全局变量',

`year` int(11) NOT NULL DEFAULT '2000' COMMENT '上牌年份',

month int(11) NOT NULL DEFAULT '6' COMMENT '上牌月份',

mile decimal(5,2) NOT NULL COMMENT '里程(万公里)',

city varchar(50) DEFAULT NULL COMMENT '城市',

province varchar(50) DEFAULT NULL COMMENT '所在省',

popularity enum('A','B','C') NOT NULL DEFAULT 'C' COMMENT '流行程度',

price\_slug varchar(10) DEFAULT '' COMMENT '交易类型',

price decimal(7,2) DEFAULT NULL COMMENT '价格(万元)',

color varchar(32) DEFAULT NULL COMMENT '颜色',

contact varchar(32) DEFAULT NULL COMMENT '联系人',

phone varchar(20) DEFAULT NULL COMMENT '电话',

domain varchar(32) DEFAULT NULL COMMENT '来源网站',

url varchar(200) DEFAULT NULL COMMENT '来源网址',

dealer\_id int(11) DEFAULT NULL COMMENT '车商编号',

use\_way varchar(16) DEFAULT NULL COMMENT '使用方式{非运营:A,运营:B,营转非:C}',

status varchar(10) DEFAULT NULL COMMENT '状态{在售:sale,已售:review}',

process\_status varchar(1) DEFAULT NULL COMMENT '处理状态{新增:A,已爬取:P,删除:D}',

expired\_at datetime DEFAULT NULL COMMENT '过期时间',

sold\_time datetime DEFAULT NULL COMMENT '卖出时间',

update\_time DATETIME DEFAULT NULL comment '更新时间',

create\_time DATETIME DEFAULT NULL comment '创建时间',

KEY i\_model\_detail\_slug(model\_detail\_slug),

KEY i\_model\_slug(model\_slug),

KEY i\_brand\_slug(brand\_slug),

KEY i\_process\_status(process\_status)

) comment '爬虫抓取数据汇总表';

## 2.2 数据输出

抓取各竞争对手估值录入到valuate.valuate\_management\_competitor\_data

**字段说明:**

|  |  |  |
| --- | --- | --- |
| 字段名 | 说明 | 参考相关表 |
| car\_id | 对应数据来源的ID | valuate\_management\_competitor\_origin\_data 字段id |
| competitor\_id | 竞争对手ID | valuate\_management\_competitor |
| valuate\_price | 预测价格 | json格式,包含车况,各类交易价 |
| status | 状态 | 状态:{可用:1,不可用:描述原因} |
| create\_time | 创建时间 |  |

# 3.其他资源

* + - * 1. CREATE TABLE valuate.valuate\_management\_price\_category(

1. id INT AUTO\_INCREMENT PRIMARY KEY comment '交易类型主键,供关联',
2. price\_slug VARCHAR(32) DEFAULT NULL comment '全局交易类型',
3. name VARCHAR(32) DEFAULT NULL comment '交易类型名称',
4. spider\_support VARCHAR(1) DEFAULT NULL comment '爬虫是否支持抓取该价格类型{支持:1,不支持:0}',
5. description longtext DEFAULT NULL comment '交易类型说明',
6. update\_time DATETIME DEFAULT NULL comment '更新时间',
7. create\_time DATETIME DEFAULT NULL comment '创建时间',
8. KEY i\_price\_slug(price\_slug)
9. ) comment '提供给后台项目管理系统所需要展示的交易类型';
10. CREATE TABLE valuate.valuate\_management\_competitor(
11. id INT AUTO\_INCREMENT PRIMARY KEY comment '竞争对手主键,供关联',
12. name VARCHAR(32) DEFAULT NULL comment '竞争对手名称',
13. domain VARCHAR(32) DEFAULT NULL comment '竞争对手域名',
14. description longtext DEFAULT NULL comment '竞争对手说明',
15. update\_time DATETIME DEFAULT NULL comment '更新时间',
16. create\_time DATETIME DEFAULT NULL comment '创建时间',
17. KEY i\_id(id)
18. ) comment '提供给后台项目管理系统所需要展示的竞争对手';
19. CREATE TABLE valuate.valuate\_management\_competitor\_data(
20. id INT AUTO\_INCREMENT PRIMARY KEY ,
21. car\_id INT(11) DEFAULT NULL comment '对应数据来源的车源ID',
22. competitor\_id INT(4) DEFAULT NULL comment '竞争对手ID',
23. valuate\_price longtext DEFAULT NULL comment 'json格式,包含车况,各类交易价',
24. status VARCHAR(64) DEFAULT NULL comment '状态:{可用:1,不可用:描述原因}',
25. create\_time DATETIME DEFAULT NULL comment '创建时间',
26. KEY i\_car\_id(car\_id)
27. ) comment '提供给后台项目管理系统所需竞争对手数据,包括公平价';

CREATE TABLE valuate.valuate\_management\_competitor\_origin\_data (

id INT AUTO\_INCREMENT PRIMARY KEY ,

source\_id int(11) DEFAULT NULL COMMENT '数据来源ID',

title varchar(200) DEFAULT NULL COMMENT '标题',

pub\_time datetime NOT NULL COMMENT '发布时间',

brand\_slug varchar(32) DEFAULT NULL COMMENT '品牌slug',

model\_slug varchar(32) DEFAULT NULL COMMENT '型号slug',

model\_detail\_slug varchar(32) DEFAULT NULL COMMENT '款型slug',

control varchar(32) DEFAULT NULL COMMENT '变速箱',

volume decimal(5,1) DEFAULT NULL COMMENT '排量',

body\_style varchar(50) DEFAULT NULL COMMENT '车型样式',

car\_condition\_slug varchar(32) DEFAULT NULL COMMENT '车况全局变量',

`year` int(11) NOT NULL DEFAULT '2000' COMMENT '上牌年份',

month int(11) NOT NULL DEFAULT '6' COMMENT '上牌月份',

mile decimal(5,2) NOT NULL COMMENT '里程(万公里)',

city varchar(50) DEFAULT NULL COMMENT '城市',

province varchar(50) DEFAULT NULL COMMENT '所在省',

popularity enum('A','B','C') NOT NULL DEFAULT 'C' COMMENT '流行程度',

price\_slug varchar(10) DEFAULT '' COMMENT '交易类型',

price decimal(7,2) DEFAULT NULL COMMENT '价格(万元)',

color varchar(32) DEFAULT NULL COMMENT '颜色',

contact varchar(32) DEFAULT NULL COMMENT '联系人',

phone varchar(20) DEFAULT NULL COMMENT '电话',

domain varchar(32) DEFAULT NULL COMMENT '来源网站',

url varchar(200) DEFAULT NULL COMMENT '来源网址',

dealer\_id int(11) DEFAULT NULL COMMENT '车商编号',

use\_way varchar(16) DEFAULT NULL COMMENT '使用方式{非运营:A,运营:B,营转非:C}',

status varchar(10) DEFAULT NULL COMMENT '状态{在售:sale,已售:review}',

process\_status varchar(1) DEFAULT NULL COMMENT '处理状态{新增:A,已爬取:P,删除:D}',

expired\_at datetime DEFAULT NULL COMMENT '过期时间',

sold\_time datetime DEFAULT NULL COMMENT '卖出时间',

update\_time DATETIME DEFAULT NULL comment '更新时间',

create\_time DATETIME DEFAULT NULL comment '创建时间',

KEY i\_model\_detail\_slug(model\_detail\_slug),

KEY i\_model\_slug(model\_slug),

KEY i\_brand\_slug(brand\_slug),

KEY i\_process\_status(process\_status)

) comment '爬虫抓取数据汇总表';

1. CREATE TABLE valuate.valuate\_management\_data\_source(
2. id INT AUTO\_INCREMENT PRIMARY KEY comment '数据来源主键,供关联',
3. name VARCHAR(64) DEFAULT NULL comment '数据来源名称',
4. table\_name VARCHAR(64) DEFAULT NULL comment '车源表名',
5. description longtext DEFAULT NULL comment '数据来源说明',
6. update\_time DATETIME DEFAULT NULL comment '更新时间',
7. create\_time DATETIME DEFAULT NULL comment '创建时间',
8. KEY i\_id(id)
   * + - 1. ) comment '提供给后台项目管理系统所需数据来源,比如爬虫,人工录入等';

CREATE TABLE valuate.valuate\_management\_car\_condition(

id INT AUTO\_INCREMENT PRIMARY KEY comment '车况类型主键,供关联',

car\_condition\_slug VARCHAR(32) DEFAULT NULL comment '全局车况类型',

name VARCHAR(32) DEFAULT NULL comment '车况类型名称',

description longtext DEFAULT NULL comment '车况类型说明',

update\_time DATETIME DEFAULT NULL comment '更新时间',

create\_time DATETIME DEFAULT NULL comment '创建时间',

KEY i\_car\_condition\_slug(car\_condition\_slug)

) comment '提供给后台项目管理系统所需要展示的车况类型';