Crawler系统数据库设计

# 规则系统表

## 分类表 category

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| category | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | VARCHAR (36) | id |  |
|  | name | VARCHAR (20) | 分类名称 |  |
|  | parent\_id | VARCHAR (36) | 父分类 |  |
|  | description | VARCHAR (200) | 描述 |  |
|  | orderby | VARCHAR (2) | 排序 |  |
|  | isvalid | CHAR (1) | 是否显示(0:删除1:正常) |  |
|  | creatdate | DATETIME | 创建时间 |  |
|  | modifydate | DATETIME | 修改时间 |  |
|  | creator | VARCHAR (36) | 创建用户 | 引用basic.employee表userid |
|  | modifier | VARCHAR (36) | 修改用户 |  |

## 标签表 tag

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tag | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | VARCHAR (36) | id |  |
|  | name | VARCHAR (20) | 标签名称 |  |
|  | parent\_id | VARCHAR (36) | 父标签 |  |
|  | description | VARCHAR (200) | 描述 |  |
|  | orderby | VARCHAR (2) | 排序 |  |
|  | isvalid | CHAR (1) | 是否显示(0:删除1:正常) |  |
|  | creatdate | DATETIME | 创建时间 |  |
|  | modifydate | DATETIME | 修改时间 |  |
|  | creator | VARCHAR (36) | 创建用户 | 引用basic.employee表userid |
|  | modifier | VARCHAR (36) | 修改用户 |  |

## 采集的新闻原表origin\_news

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| origin\_news | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | INT | id |  |
|  | title | VARCHAR (150) | 标题 |  |
|  | content | LONGTEXT | 新闻内容 |  |
|  | posttime | DATETIME | 原文发布时间 |  |
|  | fname | VARCHAR (50) | 栏目名称 |  |
|  | author | VARCHAR (50) | 作者 |  |
|  | copyfrom | VARCHAR (100) | 来源地址 |  |
|  | copyfromurl | VARCHAR (150) | 来源url |  |
|  | tags | VARCHAR (400) | 标签(存储的是标签id,用","号隔开) |  |
|  | status | VARCHAR (1) | 状态(0:未自动打标签1:已自动打标签) |  |
|  | creatdate | DATETIME | 采集时间 |  |
|  | modifydate | DATETIME | 修改时间 |  |
|  | modifier | VARCHAR (36) | 修改用户 | 引用basic.employee表userid |

## 新闻表(操作的新闻表) news

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| news | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | VARCHAR (36) | id |  |
|  | title | VARCHAR (150) | 标题 |  |
|  | content | LONGTEXT | 新闻内容 |  |
|  | posttime | DATETIME | 原文发布时间 |  |
|  | fname | VARCHAR (50) | 栏目名称 |  |
|  | author | VARCHAR (50) | 作者 |  |
|  | copyfrom | VARCHAR (100) | 来源地址(新采集过来的值是web\_site,代码转换) |  |
|  | copyfromurl | VARCHAR (150) | 来源url |  |
|  | category\_id | VARCHAR (36) | 分类id | 引用basic.category表id |
|  | status | VARCHAR (2) | 状态(0:删除1:待发布) |  |
|  | issue | VARCHAR (1) | 是否发布(0:未发布1:发布) |  |
|  | hits | NUMBER | 点击数 |  |
|  | isrecommend | CHAR (1) | 是否推荐新闻 |  |
|  | istop | CHAR (1) | 置顶 |  |
|  | isautotag | CHAR (1) | 自动标签 |  |
|  | summary | LONGTEXT | 英文摘要 |  |
|  | c\_summary | LONGTEXT | 中文摘要 |  |
|  | importance | VARCHAR (36) | 重要性 |  |
|  | urgent | TINYINT | 紧急程序 |  |
|  | sort | VARCHAR (36) | 种类(knowledge,info等) |  |
|  | his\_news\_id | VARCHAR (36) | 原新闻id |  |
|  | likecount | INT | 喜欢数 |  |
|  | dislikecount | INT | 不喜欢数 |  |
|  | clickmails | INT |  |  |
|  | tags | VARCHAR (1000) | 标签名,用逗号隔开 |  |
|  | tag\_ids | VARCHAR (1000) | 标签名,用逗号隔开 |  |
|  | posted | CHAR (1) |  |  |
|  | weekposted | CHAR (1) |  |  |
|  | web\_url | VARCHAR (150) | 网站域名 |  |
|  | catalog | VARCHAR (100) | 频道标签 |  |
|  | media\_type | VARCHAR (100) | 媒体分类标签 |  |
|  | area | VARCHAR (100) | 地区标签 |  |
|  | lang | VARCHAR (10) | 语言 |  |
|  | web\_summary | VARCHAR (200) | 网站内容摘要 |  |
|  | from\_tags | VARCHAR (200) | 网站标签 |  |
|  | creatdate | DATETIME | 采集时间 |  |
|  | creator | VARCHAR (36) | 创建用户 | 引用basic.employee表userid |
|  | modifydate | DATETIME | 修改时间 |  |
|  | modifier | VARCHAR (36) | 修改用户 | 引用basic.employee表userid |

## 新闻表(历史表) his\_news

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| his\_news | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | VARCHAR (36) | id |  |
|  | title | VARCHAR (150) | 标题 |  |
|  | content | LONGTEXT | 新闻内容 |  |
|  | posttime | DATETIME | 采集时间 |  |
|  | fname | VARCHAR (50) | 栏目名称 |  |
|  | author | VARCHAR (50) | 作者 |  |
|  | copyfrom | VARCHAR (100) | 来源地址 |  |
|  | copyfromurl | VARCHAR (150) | 来源url |  |
|  | category\_id | VARCHAR (36) | 分类id | 引用basic.category表id |
|  | status | VARCHAR (2) | 状态(2:删除1:待发布) |  |
|  | web\_url | VARCHAR (150) | 网站域名 |  |
|  | catalog | VARCHAR (100) | 频道标签 |  |
|  | media\_type | VARCHAR (100) | 媒体分类标签 |  |
|  | area | VARCHAR (100) | 地区标签 |  |
|  | lang | VARCHAR (10) | 语言 |  |
|  | web\_summary | VARCHAR (200) | 网站内容摘要 |  |
|  | from\_tags | VARCHAR (200) | 网站标签 |  |
|  | creatdate | DATETIME | 创建时间 |  |
|  | creator | VARCHAR (36) | 创建用户 | 引用basic.employee表userid |
|  | modifydate | DATETIME | 修改时间 |  |
|  | modifier | VARCHAR (36) | 修改用户 | 引用basic.employee表userid |

## 单据\_标签表 re\_bill\_tag

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| re\_bill\_tag | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | V ARCHAR (36) | id | 主键 |
|  | bill\_id | V ARCHAR (36) | 单据id(如新闻,微博) |  |
|  | tag\_id | VARCHAR (36) | 标签id |  |
|  | createdate | DATETIME | 创建时间 |  |
|  | modifydate | DATETIME | 修改时间 |  |
|  | creator | VARCHAR (36) | 创建用户 |  |
|  | modifier | VARCHAR (36) | 修改用户 | 引用basic.employee表userid |

## 标签规则表tag\_rule

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tag\_rule | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | V ARCHAR (36) | id | 主键 |
|  | type | V ARCHAR (1) | 1:标题或内容0:url |  |
|  | keyword | VARCHAR (100) | 包含的关键字 |  |
|  | tag\_id | VARCHAR (1000) | 标签id,用逗号隔开 |  |
|  | tag\_name | VARCHAR (1000) | 标签NAME,用逗号隔开 |  |
|  | createdate | DATE | 创建时间 |  |
|  | modifydate | DATE | 修改时间 |  |
|  | creator | VARCHAR (36) | 创建用户 | 引用basic.employee表userid |
|  | modifier | VARCHAR (36) | 修改用户 | 引用basic.employee表userid |

## 备忘表newcomments

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| newcomments | | | |  |
| 主键 | 字段名 | 字段类型 | 含义 | 外键 |
| ★ | id | V ARCHAR (36) | id | 主键 |
|  | content | LONGTEXT | 备忘内容 |  |
|  | posttime | DATETIME | 备忘时间 |  |
|  | userid | VARCHAR (36) | 备忘人 |  |
|  | aid | VARCHAR (36) | 关联单据ID |  |
|  | atype | VARCHAR (36) | 关联单据类型 |  |
|  | createdate | DATETIME | 创建时间 |  |
|  | modifydate | DATETIME | 修改时间 |  |
|  | creator | VARCHAR (36) | 创建用户 |  |
|  | modifier | VARCHAR (36) | 修改用户 | 引用basic.employee表userid |