**系统使用字段**

**V3.0**

**2016-12-8**

# 数据采集

## 1.1 TW数据

### 1.1.1TW账号集合

### tw\_acoounts\_all\_collection（600W）

定义：存储采集到的推特账号信息，主要用于地理空间、ZDRW、实时信息推送等功能模块。推特账号详细信息为人物关联关系、人物数据画像等数据分析提供数据支持。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tw\_accounts\_collection | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| account\_id | 账号ID | int |  |  |
| account\_update\_time | 账号信息最新更新时间 | int |  |  |
| alive | 用户是否存在 | bool |  |  |
| classification | 分类 | string |  |  |
| created\_at | 创建时间 | int |  |  |
| description | 自我介绍 | string |  |  |
| favourites\_count | 账号点赞TW的数量 | int |  |  |
| followers\_count | 粉丝数量 | int |  |  |
| followers\_id | 粉丝账号列表（用户ID） | int list | list里面有些是long int |  |
| friends\_count | 关注账号数量 | int |  |  |
| friends\_ids | 关注账号列表（用户ID | int list | list里面有些是long int |  |
| latest\_tweet\_created\_at | 最新TW时间 | int |  |  |
| listed\_count | 所属公共列表数目 | int | 关注的越多，说明这个人发的TW越有价值 |  |
| location | 地理位置 | string | 用户自己填的 |  |
| name | 用户别名 | string |  |  |
| profile\_storage\_path | 头像存储路径 | string |  |  |
| screen\_name | 屏显姓名 | string |  |  |
| statuses\_count | TW数量 | int |  |  |
| useful | 是否有用 | string |  |  |

### 1.1.2TW集合

### tweets

定义：存储采集到的TW信息，主要用于地理空间、ZDRW、实时信息推送等功能模块。TW详细信息为语义分析、文本分析、影像识别及QB分析提供数据支持。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tweets | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| account\_id | 账户ID | int |  |  |
| captured\_time | TW采集到的时间 | int | UTC时间，在本地时间上少（提前）了8个小时 |  |
| classification | 分类 | string | 现在是人为在分类 |  |
| coordinates | 用户报告的地理位置 | dict |  |  |
| created\_at | 推特创建时间 | int | UTC时间，始于1970年1月1日00:00 （原始TW的时间） |  |
| favorite\_count | TW被点赞的次数 | int | 若是转发TW则为被转发TW的点赞数 |  |
| favorited | TW是否被认证用户点赞 | bool |  |  |
| id | TW序号ID | long int |  |  |
| img\_storage\_path | 图片存在本地的路径 | string list |  |  |
| in\_reply\_to\_screen\_name | 被回复TW作者屏显姓名 | string |  |  |
| in\_reply\_to\_status\_id | 被回复原始TWID | long int |  |  |
| in\_retweet\_to\_screen\_name | 被转发者屏显姓名 | string |  |  |
| in\_retweet\_to\_status\_id | 被转发原始TWID | long int | 以前为long int list |  |
| is\_retweet\_status | 是否为转发TW | bool |  |  |
| original\_text | 原始TW的文字 | string |  |  |
| place | TW关联的地理位置 | dict | 以前为float list |  |
| retweet\_count | TW被转发的次数 | int | 若是转发TW则为被转发TW的转发数 |  |
| retweeted | TW是否被认证用户转推 | bool |  |  |
| screen\_name | 屏显姓名 | string |  |  |
| source | 发布TW的链接 | string |  |  |
| text | TW | string |  |  |
| useful | 是否有用 | bool | 预设的字段，以后可能会用 |  |
| vid\_storage\_path | 视频存在本地的路径 | string list |  |  |

## 1.2基础信息

### 1.3.1ZDRW账号集合

### tw\_acoounts\_collection（4W）

定义：存储重点关注的人物推特账号信息，主要用于地理空间、ZDRW、实时信息推送等功能模块。推特账号详细信息为人物关联关系、人物数据画像等数据分析提供数据支持。

库表结构同1.1.1

### 1.3.3消息来源集合

### attribute

定义：存储TW内容来源分类，主要为TW搜索、TW分类查看模块提供搜索条件选择。消息来源分类将TW按照其内容信息来源进行划分。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **attribute** | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| attribute\_keyword | 属性名称 | String |  |  |
| user\_setting\_id | 用户id | String | 同系统用户配置表的\_id字段 |  |
| user\_setting\_name | 用户名 | String | 同系统用户配置表的user\_name字段 |  |
| created\_at | 创建时间 | Double |  |  |

### 1.3.4信息类别集合

### classification

定义：存储推特账号人物属性或TW内容分类，主要为TW/账号搜索、TW/账号分类查看模块提供搜索条件选择。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **classfication** | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| classfication\_keyword | 中文分类名称 | String |  |  |
| classfication\_keyword\_en | 英文分类名称 | String |  |  |
| classficationWordsCounts | 分类中的词频统计 | Map<String,int> | key值为词，value为词频 |  |
| created\_at | 创建时间 | Double |  |  |
| user\_setting\_id | 用户id | String | 同系统用户配置表的\_id字段 |  |
| user\_setting\_name | 用户名 | String | 同系统用户配置表的user\_name字段 |  |

# 2.数据分析

## 2.1 ZT信息

### 2.1.1关键词集合

### key\_word

定义：存储用户定义的TW查询关键词，主要为ZT监控、TW搜索等模块提供关键词记录及选择。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| key\_word | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| add\_time | 添加时间 | Double |  |  |
| delete\_flag | 删除标志 | Boolean |  |  |
| keyword | 关键词 | String |  |  |
| status | 是否禁用状态 | Boolean | 启用/禁用 |  |
| user\_setting\_id | 创建此关键词的用户 | String | 用户id |  |
| usei\_setting\_name | 添加用户姓名 | String | 用户名 |  |
| web\_id | id | String | 将ObjectId转换为String |  |

### 2.1.2ZT类型集合

### topic\_type

定义：存储用户设置的ZT类型，主要为ZT监控模块提供ZT类型分类选择。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| topic\_type | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| created\_at | 创建时间 | Double |  |  |
| delete\_flag | 删除标志 | Boolean |  |  |
| keywordList | 关键词列表 | List<KeyWord> | KeyWord见2.1.1 |  |
| topic\_description | ZT描述 | String | 用户自己输入 |  |
| topic\_name | ZT名称 | String |  |  |
| topic\_type | ZT类型 | String | 由界面确定 |  |
| user\_setting\_id | 创建者 | String | 用户id |  |
| user\_setting\_name | 创建者姓名 | String | 用户名 |  |
| web\_id | id | String | 将ObjectId转换为String |  |

### 2.1.3中文TW信息集合（临时表）

### topic\_tweets\_copy

定义：MongoDB数据库全文索引不支持中文，为加快中文TW内容的查询速度，将中文TW从“TW集合”中提取出来。topic\_tweets\_copy集合为中文搜索提供数据支持。

库表结构同1.1.2

## 2.2YJ信息

### 2.2.1重要YJTW集合（临时表）

### significant\_event\_tweets

定义：临时存储非常有价值的YJTW，为ZT监控模块提供数据支持。

库表结构同2.2.2

### 2.2.2YJTW信息集合

### tw\_incident\_topic

定义：存储用户或数据分析系统判定的有YJ价值的TW。主要为ZT监控、信息查询模块提供数据支持。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tw\_incident\_topic | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| account\_id | 账户ID | int |  |  |
| alert\_level | YJ级别 | int | 机器抓取，0=分析未处理，1~4=无用、一般、重要、紧急 |  |
| captured\_method | 采集方式 | int |  |  |
| captured\_time | 采集时间 | int |  |  |
| created\_at | 创建时间 | int |  |  |
| created\_at\_new |  | int |  |  |
| event\_place | 事件位置 | String | 地名(城市？街道?) |  |
| expanded\_url | TW链接 | List<String> |  |  |
| favorite\_count | TW被点赞的次数 | int |  |  |
| favorited | TW是否被认证用户点赞 | Boolean |  |  |
| find\_time | YJ时间 | Double | 程序发现有问题YJ的时间,UTC时间，始于1970年1月1日00:00 |  |
| hashtags | 标签 | List<String> |  |  |
| id | TW序号 | long |  |  |
| in\_reply\_to\_status\_id | 被回复原始TWID | int |  |  |
| in\_retweet\_to\_screen\_name | 被转发者屏显姓名 | String |  |  |
| in\_retweet\_to\_status\_id | 被转发原始TWID | long |  |  |
| insert\_time | 插入时间 | int |  |  |
| is\_quote\_status | 是否为引用TW | Boolean |  |  |
| is\_retweet\_status | 是否为转发TW | Boolean |  |  |
| keywords | 关键词 | List<String> | 事件分析得出的关键词, UTC时间，始于1970年1月1日00:00 |  |
| lang | 语言种类 | String |  |  |
| modify\_time | 修改时间 | Double |  |  |
| msg\_type | 所属类别 | List<String> |  |  |
| original\_text | 原始TW的文字 | String |  |  |
| publish\_place | 发布位置 | String | 地名 |  |
| retweet\_count | TW被转发的次数 | int |  |  |
| retweeted | TW是否被认证用户转推 | Boolean |  |  |
| screen\_name | 屏显姓名 | String |  |  |
| source | 发布TW的链接 | String |  |  |
| text | TW | String |  |  |
| trans\_cn | 原文中文翻译 | String |  |  |
| tw\_type | TW类型 | String | 四种类型：1,普通相关TW、2,YJ相关TW、3,ZT相关TW、4,即是YJ又是ZTTW |  |
| u\_r\_status | 已阅状态 | int | 用户操作，0:未读；1：已读。默认是0 |  |
| usefully |  | int |  |  |
| user\_attribute | 消息来源 | List<String> |  |  |
| user\_info | 人物简要信息 | SimpleUserAccount | 自定义对象 |  |
| user\_setting\_id | 审计人员ID | String | 用户ID |  |
| user\_setting\_name | 审计人员姓名 | String | 用户姓名 |  |
| user\_mentions | TW里@的账户 | List<String> |  |  |

## 2.3 TW状态

### 2.3.1TWYJ状态记录集合

### tw\_status

定义：存储多用户修改TWYJ等级记录。主要用于信息查询、ZT监控等模块，记录各个用户操作TWYJ等级信息。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tw\_status | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| tw\_id | TW序号 | Long | 同Tweet存储表的id |  |
| twStatusChanges | TW状态更改 | List<TwStatusChange> | 自定义：List<TwStatusChange> |  |

# 3.界面（业务）呈现

## 3.1 地理信息

### 3.2.1地名-层级映射集合

### location\_level\_mapper

定义：存储地名与地名所在地理层级的映射关系。主要为地理空间QB信息模块提供推特账号分层级地理位置定位。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| location\_level\_mapper | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| \_class | 类名 | String |  |  |
| id | 城市标号 | int |  |  |
| level | 城市等级 | int | 1：七大洲；2：国家；3：省会（特大城市）… |  |
| name | 中文名 | String |  |  |
| upid | 上一级地理位置标号 | int |  |  |

### 3.2.2中文-英文地名映射集合

### location\_mapping

定义：存储中文与英文地名映射关系。将TW及推特账号中的英文地名转换成中文地名。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| location\_mapping | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| location\_cn | 中文地名 | String |  |  |
| location\_en | 英文地名 | String |  |  |

### 3.2.3国家名称中英文及简写对照集合

### national\_encn\_mapper

定义：存储国家地名中文与英文及英文简写映射关系。为地理空间QB模块提供位置定位辅助。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| national\_encn\_mapper | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| \_class | 类名 | String |  |  |
| nation\_cn | 中文国名 | String |  |  |
| nation\_en | 英文国名 | String |  |  |
| nation\_en\_simple | 英文国名简写 | String |  |  |

### 3.2.4地名地理信息数据集合

### place\_name\_data

定义：存储中文地名与地理位置坐标信息映射关系。主要为地理空间QB、TW及推特账号详细信息模块信息内容地图定位功能数据支持。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| place\_name\_data | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| place\_name\_id | 地名标识 | int |  |  |
| place\_name\_level | 地名层级编码 | String |  |  |
| describe | 描述 | String |  |  |
| longitude | 地名经度 | Double |  |  |
| latitude | 地名纬度 | Double |  |  |
| geography\_code | 地理编码 | Double |  |  |
| name\_length | 地名名称长度 | int |  |  |
| place\_name | 地名名称 | String |  |  |
| plcae\_name\_abridge | 地名简拼 | String |  |  |
| place\_name\_spelling | 地名全拼 | String |  |  |
| place\_name\_alias | 地名别名 | String |  |  |
| local\_language | 本地语 | String |  |  |
| english\_name | 英文名称 | String |  |  |
| place\_name\_address | 地名地址 | String |  |  |
| place\_name\_elevation | 地名高程 | int |  |  |

### 3.2.5地名层级信息集合

### place\_name\_layer

定义：存储地名与地名所在层级映射关系。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| place\_name\_layer | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| pnl\_id | 地名图层标识 | int |  |  |
| pnl\_code | 地名图层编码 | int |  |  |
| layer\_name | 图层名称 | String |  |  |
| layer\_style | 图层样式 | String |  |  |
| level | 显示层级 | int |  |  |
| layer\_describe | 图层描述 | String |  |  |

### 3.2.6世界城市地理位置集合

### world\_cities\_pop（英文地名，300W）

定义：存储英文地名与地理位置坐标信息映射关系。主要为“地名地理信息数据集合”提供地名对应地理位置坐标的补充。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| world\_cities\_pop | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| accent\_city | 标准地名 | String |  |  |
| city | 城市名称 | String |  |  |
| country | 国家名称 | String |  |  |
| latitude | 地理纬度 | String |  |  |
| longitude | 地理经度 | String |  |  |
| population | 人口 | String |  |  |
| region | 区域 | String |  |  |

## 3.3综合信息

### 3.3.1地区账号统计集合

### area\_account

定义：存储各个区域推特账号数量。主要为地理空间QB模块区域推特账号统计信息、区域信息定位功能提供数据支持。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| area\_account | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| properties | 配置 | Properties | 自定义：Properties |  |
| geometry | 地理位置 | Geometry | 自定义：Geometry |  |

### 3.3.2综合信息统计集合

### summarized\_infor

定义：存储本系统采集TW情况统计信息，主要用于导航页面系统运行状态显示。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| summarized\_infor | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| collect\_tw | 采集TW数（一分钟） | Long | 自定义：Properties |  |
| language\_count | 语言统计 | Map<String,Double> |  |  |

### 3.3.3TW区域分布趋势曲线集合

### tw\_distribute\_timeline

定义：存储各区域TW数量分布。主要用于地理空间QB模块实时TW区域分布趋势图表显示。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tw\_distribute\_timeline | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| area | 地域 | String |  |  |
| timeLines | 语言统计 | List<TimeLine> | 自定义：List<TimeLine> |  |

## 3.4 时间轴

### 3.4.1推特账户发文趋势曲线集合

### account\_publish\_time\_line

定义：存储各推特账号发布TW频数。横轴为时间，纵轴为TW数。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| account\_publish\_time\_line | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| accountId | TW账户ID | int |  |  |
| timeLines | 时间线 | List<TimeLine> | 用户自定对象 |  |

### 3.4.2TW被转发趋势曲线集合

### tw\_retweeted\_time\_line

定义：存储TW被转发的时间趋势。横轴为时间，纵轴为转发量。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tw\_retweeted\_time\_line | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | id | ObjectId | 唯一，mongodb系统自带 |  |
| timeLines | 时间线 | List<TimeLine> | 自定义对象 |  |
| twId | TWid | long |  |  |

# 4.系统管理

## 4.1 系统用户配置

### 4.1.1用户信息集合

### user\_setting

定义：存储系统用户信息，包括用户名、密码及用户个性化设置。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| user\_setting | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| userName | 用户名 | String |  |  |
| pwd | 用户密码 | String |  |  |
| createTime | 创建时间 | Long |  |  |
| userType | 用户类型 | String | 管理员、普通用户、维护用户 |  |
| refreshTime\_page | XXX网页刷新间隔，单位为秒 | int | 默认值=XX，可修改 |  |
| listNum | 滚动条数 | int | 默认值，可修改 |  |
| web\_id | ObjectId的String类型 | String | Mongodb系统自带，唯一 |  |
| isActive | 会话是否活跃 | String |  |  |
| delete\_flag | 逻辑删除标记 | Boolean |  |  |

### 4.1.2用户操作记录集合

### history\_modify

定义：存储登录系统的用户操作。目前主要记录各个用户修改关键词操作。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| history\_modify | | | | |
| **字段** | **字段说明** | **类型** | **备注** | **来源** |
| \_id | 用户ID | ObjectId | Mongodb系统自带，唯一 |  |
| add\_topic\_record | 添加ZT记录信息 | List<TopicType> | TopicType见2.1.2 |  |
| delete\_topic\_record | 删除ZT记录信息 | List<TopicType> | TopicType见2.1.2 |  |
| modify\_topic\_record | 修改ZT记录信息 | List<TopicType> | TopicType见2.1.2 |  |
| user\_setting\_id | 修改配置的系统用户id | String | 用户id |  |
| user\_setting\_name | 修改配置的系统用户名 | String | 用户名 |  |