MySQL数据库设计

目录

[1. 仿冒网站黑名单表 1](#_Toc452407773)

[2. 仿冒网站模板统计表 2](#_Toc452407774)

[3. 系统后台引擎描述表 2](#_Toc452407775)

[4. 可疑网站灰名单表 2](#_Toc452407776)

[5. 重点检测网站灰名单表 3](#_Toc452407777)

[6. 被保护网站白名单表 3](#_Toc452407778)

[7. 引擎存活情况表 3](#_Toc452407779)

[8. 任务信息表 3](#_Toc452407780)

[9. 任务结果表 4](#_Toc452407781)

[10. 被信任网站白名单表 6](#_Toc452407782)

[11. 域名责任用户表 7](#_Toc452407783)

[12. 域名whois信息表 7](#_Toc452407784)

[13. Whois反向查询结果表 7](#_Toc452407785)

1. 仿冒网站黑名单表

物理设计：每个仿冒网站对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| counterfeit\_list | | 仿冒网站列表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| url\_hash | varchar | y | URL的16进制MD5 |
| url | varchar | n | 仿冒网站URL |
| source\_url | varchar | n | 被仿冒网站URL |
| source\_name | varchar | n | 被仿冒网站名称 |
| source\_type | varchar | n | 被仿冒网站类型 |
| discover\_time | char | n | 发现时间 |
| discover\_way | varchar | n | 发现方式 |
| noaccess\_time | char | n | 失效时间 |
| ip | varchar | n | 仿冒网站IP |
| country | varchar | n | 仿冒网站IP所在国家 |
| city | varchar | n | 仿冒网站IP所在城市 |
| template\_num | int | n | 对应的仿冒网站模板id |
| comment | text | n | 备注 |

1. 仿冒网站模板统计表

物理设计：每个模板对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| counterfeit\_template | | 仿冒网站模板统计表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| template\_ID | int | y | 模板ID |
| url\_hash | varchar | n | 被仿冒网站URL的16进制MD5 |
| url | varchar | n | 被仿冒网站URL |
| name | varchar | n | 被仿冒网站名称 |
| type | varchar | n | 被仿冒网站类型 |
| counterfeit\_url | varchar | n | 该模板示例仿冒网站URL |

1. 系统后台引擎描述表

物理设计：每个引擎对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| engine\_describe | | 系统后台引擎描述表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| engine\_type | char | y | 引擎编号 |
| engine\_name | varchar | n | 引擎名称 |
| engine\_code\_name | varchar | n | 引擎代号，在后台对应目录名 |

1. 可疑网站灰名单表

物理设计：每个可疑网站对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| gray\_list | | 可疑网站灰名单表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| url\_hash | varchar | y | URL的16进制MD5 |
| url | varchar | n | URL |
| add\_time | char | n | 添加时间 |
| source | varchar | n | 可疑URL来源 |
| title\_source\_result | varchar | n | 关键字检测结果中发现该可疑网站仿冒的目标网站 |
| title\_counterfeit\_result | varchar | n | 关键字检测结果中与该可疑网站相似的已知仿冒网站 |
| title\_template\_num | int | n | 关键字检测结果中发现的该网站使用的目标网站仿冒模板编号 |
| structure\_source\_result | varchar | n | 结构比对结果（同上） |
| structure\_counterfeit\_result | varchar | n | 结构比对结果（同上） |
| structure\_template\_num | int | n | 结构比对结果（同上） |
| view\_source\_result | varchar | n | 视觉比对结果（同上） |
| view\_counterfeit\_result | varchar | n | 视觉比对结果（同上） |
| view\_template\_num | int | n | 视觉比对结果（同上） |

1. 重点检测网站灰名单表

物理设计：每个可疑网站对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| monitor\_list | | 重点检测网站灰名单表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| id | int | y | 唯一标识 |
| url | varchar | n | URL |
| add\_time | char | n | 添加时间 |
| add\_user | varchar | n | 添加用户 |

1. 被保护网站白名单表

物理设计：每个被保护网站对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| protected\_list | | 被保护网站白名单表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| id | int | y | 唯一标识 |
| url | varchar | n | URL |
| website\_name | varchar | n | 网站名称 |
| add\_time | char | n | 添加时间 |
| add\_user | varchar | n | 添加用户 |

1. 引擎存活情况表

物理设计：每个引擎对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| server\_live | | 引擎存活情况表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| ip | varchar | y | 服务器IP |
| port | varchar | n | 服务器端口号 |
| type | varchar | n | 引擎类型 |
| status | int | n | 引擎状态 |
| time | varchar | n | 当前更新时间 |

1. 任务信息表

物理设计：每个任务对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| task\_info | | 任务信息表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| task\_id | int | y | 唯一标识 |
| task\_name | varchar | n | 任务名称 |
| user\_id | int |  | 任务创建的用户id |
| task\_type | int | n | 任务类型 |
| task\_engine | varchar | n | 任务使用引擎 |
| protected\_id | text | n | 被保护网站在数据库中id |
| gray\_id | text | n | 可疑网站在数据库中id |
| monitor\_id | text | n | 监测的网站在数据库中id |
| counterfeit\_id | text | n | 已知仿冒网站在数据库中id |
| whois\_search\_url | varchar | n | 待查询其whois信息的URL |
| whois\_reverse\_url | varchar | n | 待whois反向查询的URL |
| host\_rule\_id | text | n | 主机域名变换规则id |
| top\_rule\_id | text | n | 顶级域名变换规则id |
| path\_rule\_id | text | n | 高频仿冒路径变换规则id |
| kword\_id | text | n | 敏感关键字id |
| add\_time | datetime | n | 任务添加时间 |
| last\_time | datetime | n | 任务最后一次启动时间 |

1. 任务结果表

物理设计：每个任务的单次启动结果对应一条记录，一个任务可对应多条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| task\_result | | 任务结果表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| task\_id | int | y | 任务id=任务启动时间 |
| start\_time | datetime | y | task启动时间 |
| task\_state | int | n | 任务状态（01：未开始、02：进行中、03：已完成） |
| original\_grayid | varchar | n | 原始灰名单（人工输入和主动探测两种途径得到） |
| e\_domain\_state | int | n | 域名变换探测状态（01：未启动、02：进行中、03：已完成） |
| domain\_changed\_all\_num | int | n | 域名变换生成的所有域名数量 |
| domain\_changed\_exist\_num | int | n | 域名变换存在的域名数量 |
| domain\_detected\_num | int | n | 已变换域名数量 |
| domain\_gray\_url\_num | int | n | 路径拼接后存在的URL数量 |
| domain\_run\_time | int | n | domain引擎运行时间 |
| e\_search\_state | int | n | 元搜索探测状态（01：未启动、02：进行中、03：已完成） |
| search\_kword\_num | int | n | 已搜索完成的关键字数量 |
| search\_url\_num | int | n | 搜索到的url数量 |
| search\_run\_time | int | n | search引擎运行时间 |
| e\_weibo\_state | int | n | 微博探测状态（01：未启动、02：进行中、03：已完成） |
| e\_filtrate\_state | int | n | 黑白名单过滤引擎状态 |
| filtrate\_trusted\_num | int | n | 使用被信任名单过滤掉的数量 |
| filtrate\_counterfeit\_num | int | n | 使用已知仿冒网站黑名单过滤掉的数量 |
| filtrate\_trusted\_objectid | varchar | n | 被信任名单过滤的URL在mongo gary\_list中的objectid |
| filtrate\_counterfeit\_objectid | varchar | n | 已知仿冒网站黑名单过滤的URL在mongo gary\_list中的objectid |
| filtrate\_objectid | varchar | n | 过滤后待检测的灰名单在mongo gray\_list中objectID |
| filtrate\_run\_time | int | n | 黑白名单过滤引擎运行时间 |
| e\_web\_save\_state | int | n | 网页信息保存引擎状态 |
| web\_request\_num | int | n | 已经请求信息的网页数量 |
| web\_save\_num | int | n | 当前已经保存的网页数量（之前已保存过且无更新的同样计算） |
| save\_protected\_objectid | varchar | n | 保存的被保护名单在mongo protected\_list中的objectid |
| save\_gray\_objectid | varchar | n | 保存的可疑网站灰名单在mongo protected\_list中的objectid |
| save\_counterfeit\_objectid | varchar | n | 保存的仿冒网站黑名单在mongo protected\_list中的objectid |
| save\_monitor\_objectid | varchar | n | 保存的监测网站灰名单在mongo protected\_list中的objectid |
| web\_save\_run\_time | int | n | web\_save运行时间 |
| e\_qt\_crawler\_state | int | n | qt爬虫引擎状态（01：未启动、02：进行中、03：已完成） |
| qt\_crawler\_num | int | n | 已经处理完成的url数量 |
| qt\_crawler\_run\_time | int | n | qt\_crawler引擎运行时间 |
| e\_feature\_save\_state | int | n | 网页基本特征抽取引擎状态（01：未启动、02：进行中、03：已完成） |
| feature\_save\_num | int | n | 抽取完特征的网页数量 |
| feature\_save\_run\_time | int | n | 网页基本特征抽取引擎运行时间 |
| e\_whois\_search\_state | int | n | whois查询引擎状态（01：未启动、02：进行中、03：已完成） |
| whois\_search\_run\_time | int | n | whois查询引擎运行时间 |
| e\_title\_state | int | n | title比对状态（01：未启动、02：进行中、03：已完成） |
| title\_check\_num | int | n | title当前已已检测URL个数 |
| title\_find\_num | int | n | title当前已发现仿冒网站个数 |
| title\_result | text | n | 当前关键字比对已发现的仿冒网站URL |
| title\_run\_time | int | n | title引擎运行时间 |
| e\_structure\_state | int | n | structure比对状态（01：未启动、02：进行中、03：已完成） |
| structure\_check\_num | int | n | structure当前已检测URL个数 |
| structure\_find\_num | int | n | structure当前已发现仿冒网站个数 |
| structure\_result | text | n | 当前结构比对已发现的仿冒网站URL |
| structure\_run\_time | int | n | structrre引擎运行时间 |
| e\_view\_collect\_state | int | n | view特征提取状态（01：未启动、02：进行中、03：已完成） |
| view\_collect\_finish\_num | int | n | view当前已提取特征的URL个数 |
| view\_collect\_task\_num | int | n | view特征提取的总任务数目（弃用该字段） |
| view\_collect\_run\_time | int | n | view特征提取的运行时间 |
| e\_view\_emd\_state | int | n | view特征相似度计算的状态（01：未启动、02：进行中、03：已完成） |
| view\_check\_num | int | n | view特征比对检测的数量 |
| view\_find\_num | int | n | view 发现的仿冒网站数量 |
| view\_visual\_result | text | n | 当前视觉比对中图片等特征比对已发现的仿冒网站URL |
| view\_location\_result | text | n | 当前视觉比对中位置比对已发现的仿冒网站URL） |
| view\_emd\_run\_time | int | n | view比对运行的时间 |
| task\_run\_time | int | n | 任务总运行时间 |
| task\_stop\_time | datetime | n | 任务结束时间 |

1. 被信任网站白名单表

物理设计：每个被信任网站对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| trusted\_list | | 被信任网站白名单表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| url\_id | int | y | 唯一标识 |
| url | varchar | n | URL |
| hash | varchar | n | URL哈希值 |
| add\_user | varchar | n | 添加用户 |
| add\_time | datetime | n | 添加时间 |

1. 域名责任用户表

物理设计：每个用户对应一条记录，包括whois信息中的注册人、管理人、技术负责人。

|  |  |  |  |
| --- | --- | --- | --- |
| whois\_contacts | | 顶级域名变换规则表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| contacts\_id | int | y | 唯一标识 |
| name | varchar | n | 姓名 |
| phone | varchar | n | 电话 |
| fax | varchar | n | 传真 |
| email | varchar | n | 邮箱 |
| org | varchar | n | 所属组织 |
| country | varchar | n | 所属国家 |
| city | varchar | n | 城市 |
| street | varchar | n | 街道 |
| postalcode | varchar | n | 邮编 |

1. 域名whois信息表

物理设计：每个网站域名对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| whois\_domain | | 网站whois信息表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| auto\_id | int | y | 唯一标识 |
| domain\_id | varchar | n | 域名的唯一ID（如:3342026\_DOMAIN\_COM-VRSN） |
| name | varchar | n | 主机名 |
| admin | int | n | 管理员id |
| tech | int | n | 所有人id |
| registrant | int | n | 注册商id |
| create\_date | varchar | n | 网站创建时间 |
| update\_date | varchar | n | 网站创建时间 |
| expiration\_date | varchar | n | 网站不可用时间 |

1. Whois反向查询结果表

物理设计：每个规则对应一条记录。

|  |  |  |  |
| --- | --- | --- | --- |
| whois\_reverse | | Whois反向查询结果表 | |
| 字段 | 类型 | 是否主键 | 备注 |
| id | int | y | 唯一标识 |
| domain | varchar | n | 反查所得域名 |
| original\_domain | varchar | n | 反查源域名 |

2016.5