
数据分类引擎

接口说明 v2.0

思睿嘉得（北京）信息技术有限公司

目录

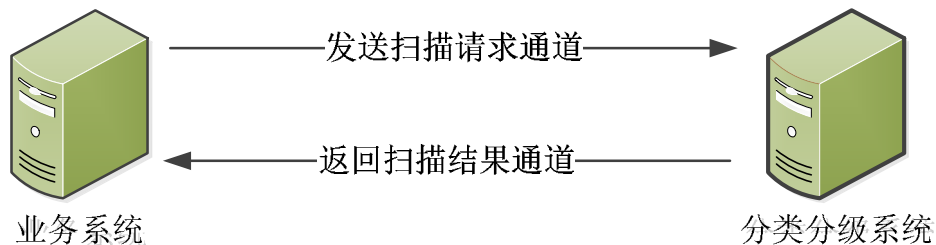
1	交互模型	2
2	接口说明	2
2.1	向分类分级系统提交扫描请求.....	2
2.1.1	发送	2
2.1.2	响应	3
2.2	向业务系统提交扫描结果.....	3
2.2.1	发送	3
2.2.2	响应	6
3	系统配置	6
3.1	mds.ini 文件	6
3.2	mds_mailsrv.ini 文件	7

1 交互模型

为了方便快捷的实现业务系统与分类分级对接,分类分级与业务系统各自监听一个端口:

提供一个 **Socket** 通道, 分别用于接收扫描请求和获取扫描结果。

1. 业务系统连接分类分级系统监听端口, 建立发送扫描请求通道, 用于提交扫描请求;
2. 分类分级系统连接业务系统监听端口, 建立返回扫描结果通道, 用于返回扫描结果。



两个通道中的数据通过业务系统定义的唯一 **ID (TID)** 进行对应关系关联。
建议 **TID** 长度不要超过 **32** 个字节。

两个通道建立连接以后, 都可以持续发送数据 (扫描请求、扫描结果), 直到某一方断开连接。

中文统一按照 **UTF8** 进行编码。

2 接口说明

2.1 向分类分级系统提交扫描请求

2.1.1 发送

业务系统向分类分级系统提交的扫描请求遵循如下的格式:

```
Post / http/1.0\r\n
```

```
Version:0\r\n
```

TID: xxxx\r\n
FileName: xxxxx\r\n
Content-Length: 1*DIGIT\r\n
\r\n
FileContent

具体说明:

- I Post / http: “/” 前后各有一个空格;
- I Content-Length: 实际 HttpBody 的长度, 这里是 size(FileContent);
- I FileName: 请求扫描的文件名称, 经过 Base64 编码的文件名 (UTF8 编码);
- I TID: 由调用者定义的本次扫描的唯一 ID, 供后续识别扫描结果使用, 建议 TID 长度不要超过 32 个字节;
- I FileContent: 需要扫描的二进制文件内容。

2.1.2 响应

分类分级系统向业务系统返回
HTTP/1.0 200 OK\r\n
表示成功接收扫描请求

2.2 向业务系统提交扫描结果

2.2.1 发送

分类分级系统向业务系统提交扫描结果遵循如下的格式:

Post / http/1.0\r\n
Version:0\r\n
TID: xxxxx \r\n
Content-Length: 1*DIGIT\r\n

\r\n

ScanResult

具体说明:

I Content-Length: 实际 HttpBody 的长度, 这里是 size(ScanResult);

ScanResult: 格式例子如下

```
{
  "TID" : "xxxx001",
  "FilsList" : [
    {
      "FileName" : "File1.pdf",
      "riskList" : [
        {
          "Class" : "一级阻断",
          "HitCount" : 3,
          "LevelName" : "一级",
          "SensitveContent" : "火箭 火箭 火箭"
        },
        {
          "Class" : "二级疑似",
          "HitCount" : 2,
          "LevelName" : "二级",
          "SensitveContent" : "发射 发射"
        }
      ]
    }
  ]
}
```

```
"FileName" : "File2.doc",
"riskList" : [
  {
    "Class" : "一级阻断",
    "HitCount" : 2,
    "LevelName" : "一级",
    "SensitveContent" : "长征 长征"
  },
  {
    "Class" : "二级疑似",
    "HitCount" : 2,
    "LevelName" : "二级",
    "SensitveContent" : "燃料 燃料"
  }
]
}
```

其中：

TID: 由业务系统定义的本次扫描的唯一 ID，必须定长，在扫描请求中提供的，分类分级系统原样返回；

FilsList: 分类列表（允许一次扫描同时命中多个分类）

n LevelName: 敏感数据分级（如果为空，则不属于任何预定义分级），由业务系统定义，分类分级系统可配置；

n Class: 敏感数据分类（如果为空，则不属于任何预定义分类），由业务系统定义，分类分级系统可配置；

n HitCount: 如果敏感数据分类方法使用了关键字或者正则表达式方法，则返回命中本次扫描关键字或者正则表达式总命中次数；

n SensitiveContent: 本次命中的典型内容；

当本次扫描没有发现风险,也要返回扫描结果,如下所示:

```
{  
  "TID": "xxxxx",  
  "FilsList": null  
}
```

2.2.2 响应

业务系统向分类分级系统返回

HTTP/1.0 200 OK\r\n

表示成功接收扫描结果。

3 系统配置

配置文件位置:

/mds/config 下的 mds.ini、mds_mailsrv.ini 文件

3.1 mds.ini 文件

IpMailServer 中设置与服务器通讯,发送扫描文件的客户端应用的 IP 和接收扫描结果端口,如 192.168.20.153:6195,

多个 IP 和端口以分号隔开。例:

IpMailServer = 192.168.20.153:6195;192.168.20.153:6195

[global]

ConnectionResetTimes = 10000 接收 1 万次请求后要重连

ReceiveMailTimeoutLimit = 60

CMSNetworkFailLimit = 120

MaxMailServerScanRequest = 100 文件最大不能超过 100M

MDSConnectNumLimit = 64 客户端最多可以有 64 个并发

3.2 mds_mailsrv.ini 文件

srv_ip = 127.0.0.1 #设置 mds 服务器的 IP
srv_port = 55588 #设置 mds 服务器接收待扫描文件的端口

path_eml = /opt/path_sendfile #待扫描文件所在的目录
sleep_ms= 5000 #发送扫描文件间隔的时间, 单位毫秒
SendFileNameEncode=gb2312 #发送的扫描文件的文件名编码方式, 可
以为 gb2312 或 utf8

注意: 当 testClient,testMailSrv 运行在本地时, /mds/config/mds.ini 中配
置的 IpMailServer=127.0.0.1:6195,
IP 需设置为 127.0.0.1

扫描最大的文件为 100M

MaxMailServerScanRequest = 100 表示 100M