# 1、数据权限管理

**功能概述**

该功能主要用于定义用户拥有数据的合规性，即指定用户/组织对指定的分类/分级有合法的数据权限。

13816790280

15300578876

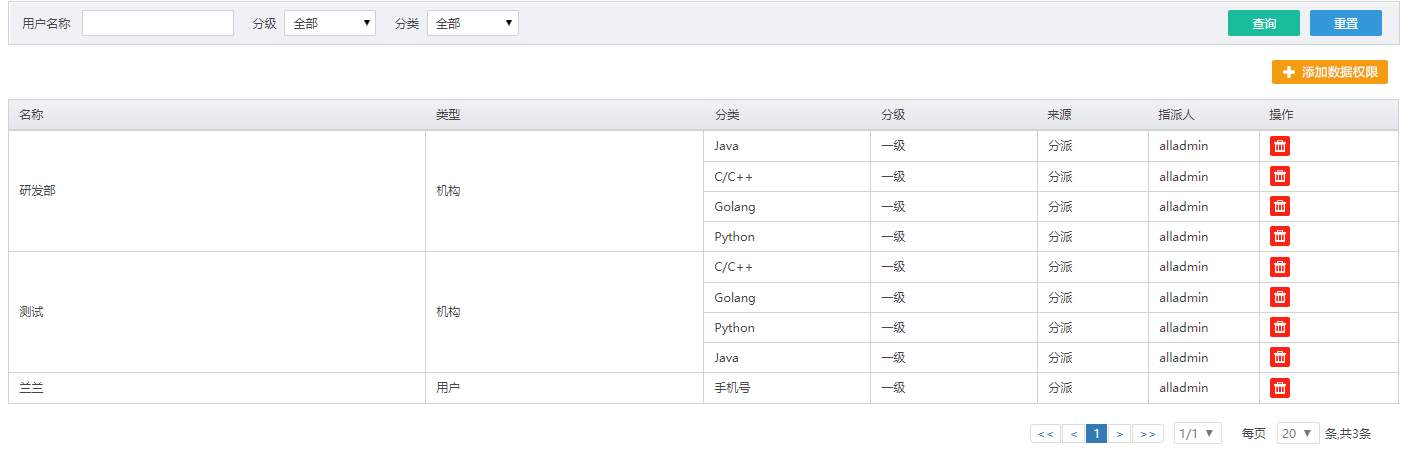
18017080083

18708455754

数据权限指用户对数据具备合法拥有的权力。对于某终端用户，赋予数据权限数据分类不是风险数据，未赋予数据权限的数据为风险数据。

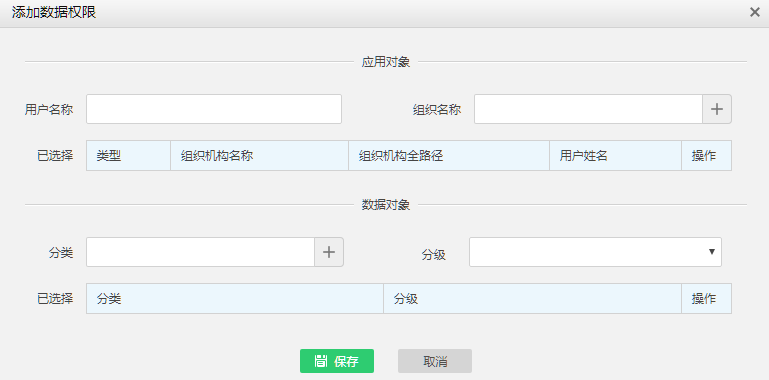
**操作步骤**

数据管理员登录平台，进入“数据分类”>”数据权限管理”界面，如图所示：



添加数据权限

数据权限管理页面，点击“添加数据权限”按钮，弹出新建数据权限页面。如图所示。选择各项信息后，点击“保存”按钮，可保存数据权限信息。

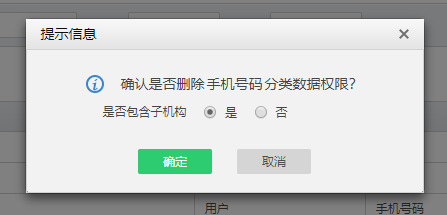


查询数据权限

在“数据权限管理”页面，输入/选择查询条件，点击“查询”按钮，系统在当前页面显示所有符合条件的数据权限信息，也可点击“重置”按钮清空设置的查询条件。

删除数据权限

在数据权限列表中选择任一数据权限记录，点击“删除”按钮，系统提示如图所示，点击 “确定”按钮，可删除数据权限；也可点击“取消”按钮,可取消删除操作。



是否包含子机构：指该组织机构下的所有子机构是否均拥有该数据权限；

# 2、机器学习

**功能概述**

此功能的主要用途有以下两部分：

a) 用于进行数据梳理，通过聚类算法配合人工梳理敏感数据分类。

b) 是添加机器学习条件分类的前置配置。

在此功能模块下，通过上传样本、分析样本，进行聚类，生成相应的机器学习条件，在添加该特征的分类时可从机器学习规则列表中选择指定条件，便于进行后续的监控。

**注意事项**

a) 机器学习样本资源管理页面的“上传样本”只可以是单个文件或ZIP格式压缩包文件。

b) 机器学习样本支持的格式包括：rar、zip、xls、xlsx、doc、docx、txt、html、xml、rtf、pdf、ppt、pptx、wps、wpt、dot、dps、dpt、pot、pps、et、ett、xlt。

c) 添加误报样本之前需要确保样本列表中已有学习样本，只有误报样本无法生成条件。

**机器学习测试用例**

**智能内容识别能力**

**人工智能内容识别**

|  |  |
| --- | --- |
| 测试目的 | 验证产品具备基于机器学习和自然语言处理的人工智能内容识别能力：通过分析大量未经标注的原始文档集，自动按照内容进行主题梳理，并可人工干预灵活调整语义相似度，获得满意的聚类效果；将聚类结果作为标注样本，提取短句和组合词作为语义特征，训练分类器，自动生成文本分类规则库，亦可人工干预对提取出来的语义特征进行选择，训练分类器，自动生成文本分类规则；分词支持短语或长组合词，支持对提取的语义特征（短句和长组合词）进行打分排序，提高人工筛选效率 |
| 测试步骤 | 1. 数据管理员登陆管理平台，上传样本， 2. 调节相似度，选择某个聚类结果，如“信息安全”作为学习样本，选择“语义特征分析” 3. 语义特征选择->扩展，勾选语义特征分析出的长词。点击“机器学习”按钮。生成机器学习条件。 |
| 测试结果 | 1. 系统具备机器学习的人工智能内容识别功能，上传未经标注的聚类样本，可通过调节相似度进行聚类   C:\Users\n\Desktop\截图\1314134e1bf189054565068742c5ba9.png   1. 机器学习支持语义特征分析，可选择符合学习样本的语义特征词，并成功生成机器学习条件   C:\Users\n\Desktop\截图\ab8fffa77b4747b5650189e1472d3e7.png  C:\Users\n\Desktop\截图\d6e3328d0e144cd9a5e7b3d8d6e22f9.png |
| 测试结论 | ☑通过 ☐不通过 |

**持续机器学习及调优**

|  |  |
| --- | --- |
| 测试目的 | 验证产品机器学习功能可通过正反项样本的持续机器学习实现数据内容识别规则的在线调优，可持续提高识别准确性 |
| 测试步骤 | 1. 将误报样本进行再次聚类，将误报样本组聚成误报组 2. 在已创建的机器学习条件下添加误报组 3. 选择已有的样本组，点击“语义特征分析”，将误报样本组作为对比组，重新选择语义特征后，点击“机器学习”按钮，生成新的机器学习条件 4. 重新发布分类，更新策略 5. 将原先被审计到的误报文件再次外发 |
| 测试结果 | 1、上传漏报样本  D:\AAAAAAAAA\东航补充截图\截图\漏报调优\上传漏报样本.png  2、选择对照样本组进行学习，生成漏报条件  D:\AAAAAAAAA\东航补充截图\截图\漏报调优\生成漏报条件.png  3、再次发送误报样本，无事件上报 |
| 测试结论 | ☑通过 ☐不通过 |

# 3、结构化指纹管理

**功能概述**

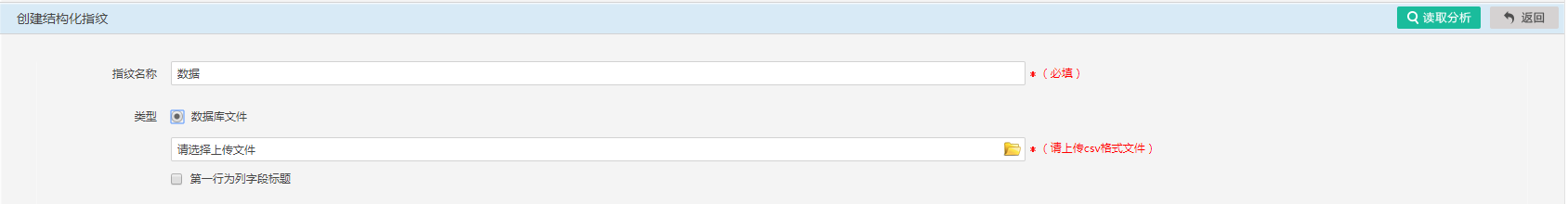
结构化指纹管理是添加结构化指纹条件分类的前置配置。当创建指纹类型为数据库文件时，在结构化指纹管理中可对csv格式文件进行指纹学习。当创建指纹类型为数据库时，在结构化指纹界面可添加数据库并进行参数配置，创建非结构化指纹。

**操作步骤**

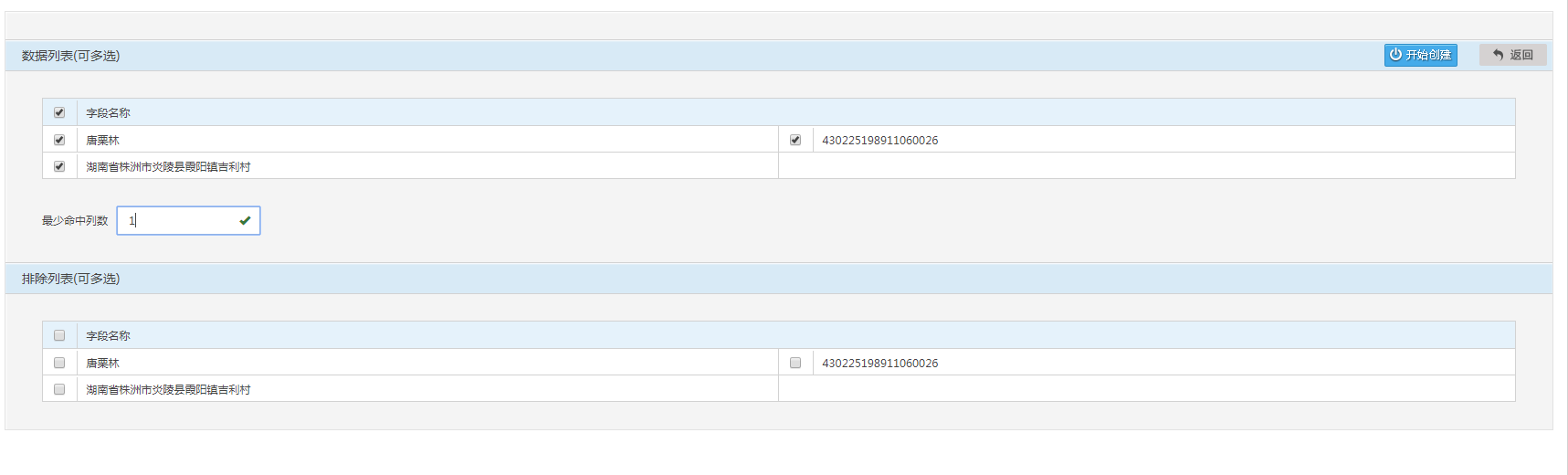
创建数据库文件指纹

数据管理员进入“数据分类”>“指纹管理”>“结构化指纹管理”页面，点击“创建指纹”按钮，进入创建指纹页面。

输入指纹名称，类型选择数据库文件，选择csv文件，可选择第一行为标题（这样指纹不读取第一行信息），如图所示：



点击“读取分析”按钮，如图所示；

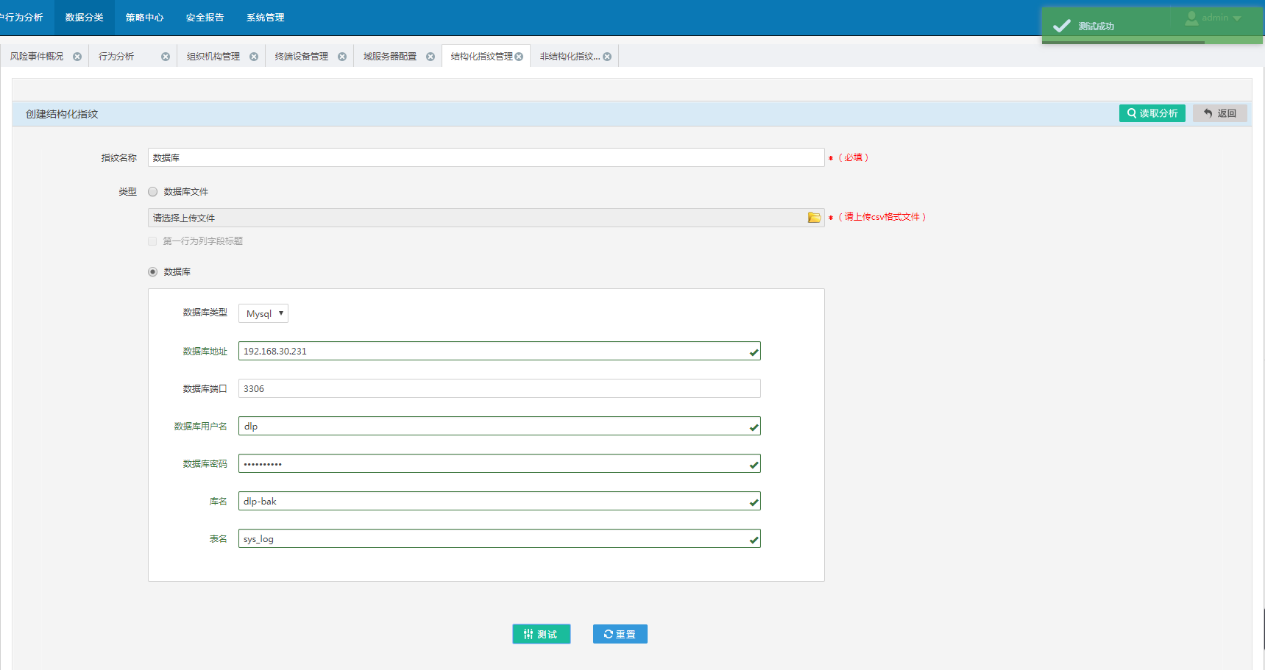


选择指纹所需的数据列，输入至少命中列，点击“开始创建”按钮，创建结构化指纹成功，可对下载、删除指纹样本。如图所示：

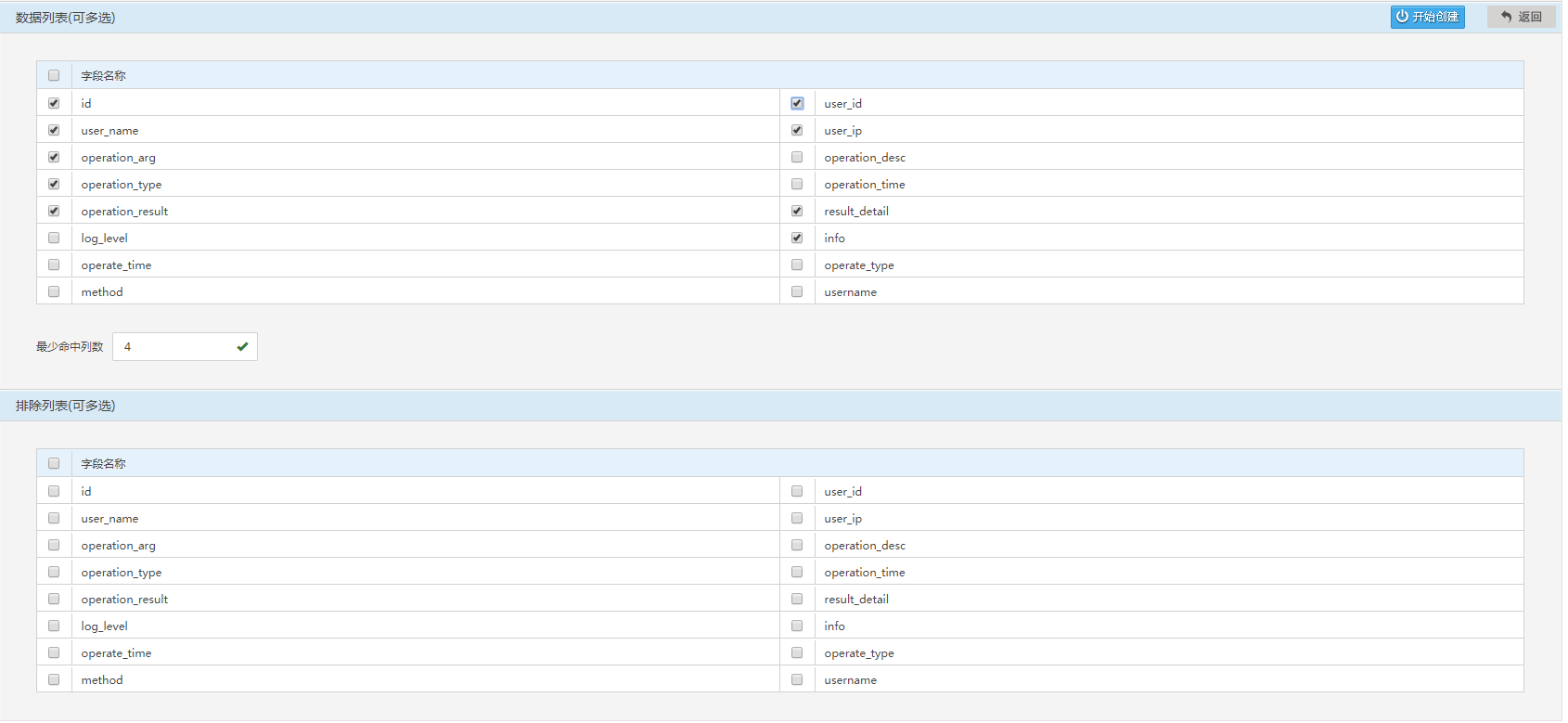


添加数据库类型指纹

配置数据库指纹：输入指纹名，类型选择数据库，选择数据库类型，输入数据库地址和端口、用户名和密码、库名和表名。配置完成后，点击测试，测试通过提示成功后方可进行下一步读取分析。如图所示：



点击“读取分析”按钮，进行读取分析，如图所示：



选择数据列，输入至少命中列，点击“开始创建”按钮，创建结构化指纹成功。数据库指纹条件可自动更新/手动更新、可删除。



**注意事项**

创建结构化指纹的时候建议将表中无用列排除掉，避免上报误报事件。

# 4、非结构化指纹管理

**功能概述**

非结构化指纹管理是添加非结构化指纹条件分类的前置配置，在非结构化指纹管理中，可添加、编辑、删除文件指纹；对指纹样本可进行导入、远程读取、删除等。

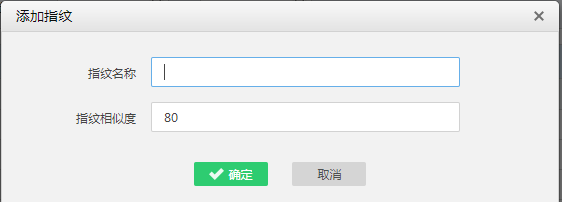
注意：非结构化指纹文件不支持图片格式的指纹样本。

**操作步骤**

添加指纹

数据管理员进入“数据分类”>“指纹管理”>“非结构化指纹”页面。点击“添加”按钮，输入指纹名和指纹相似度，点击“确定”按钮，指纹添加成功。





编辑指纹

选择需要编辑的指纹点击“编辑”按钮，可编辑该指纹



导入指纹样本

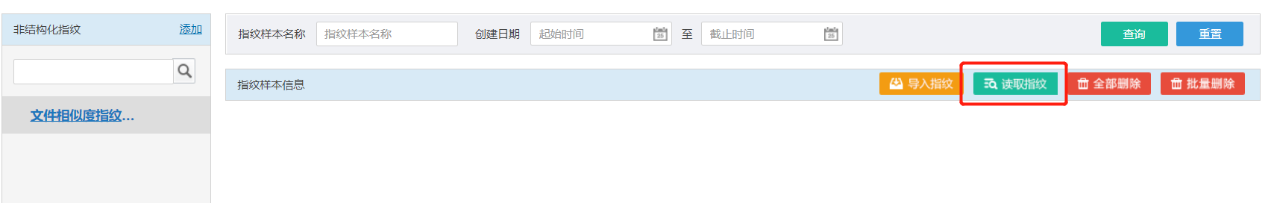
选中指纹条件命中，点击右侧的“导入”按钮，选择导入单一文件或批量文件，如图所示：





远程读取指纹

选择指纹条件，点击右侧“读取指纹”，配置远程指纹库的基本信息，点击“读取”即可远程读取指纹文件。





查询指纹样本

查询条件包括文档名、创建时间，如图所示，点击“重置”按钮，可清空查询条件。



删除指纹样本

点击“全部删除”按钮，删除所选规则下所有指纹样本。

点击“批量删除”按钮，删除所选规则下所有已选中的单个/多个指纹样本。

注意事项

1. 导入样本为常用格式的文档，如txt/pdf/office文档。
2. 非结构化指纹支持0%-100%的相似度匹配，若选择相似度为100%，则仅支持单一指纹文件。

# 5、代码相似度指纹管理

**功能概述**

代码相似度指纹主要针对代码文件制作相似度指纹，并对这些指纹进行管理。

**操作步骤**

添加指纹

进入“分类管理”- “代码相似度指纹管理”菜单，点击左侧的“添加”指纹条件，填写指纹条件名称、相似度：





导入指纹样本

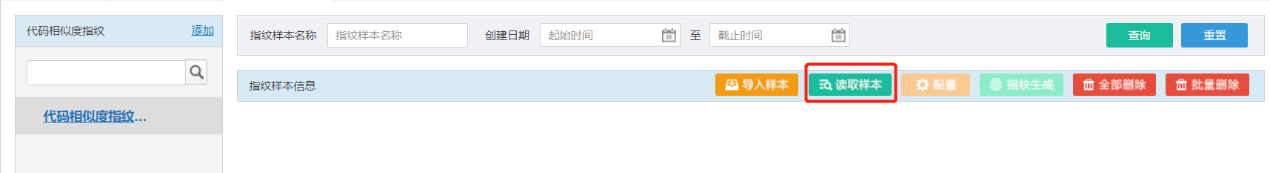
选中指纹条件命中，点击右侧的“导入”按钮，选择导入批量样本，如图所示：





读取指纹样本

选择指纹条件，点击右侧“读取指纹”，配置远程指纹库的基本信息，点击“读取”即可远程读取指纹文件。





指纹学习配置

选择指纹条件，点击右侧“配置”，可对代码指纹的学习进一步配置规则：

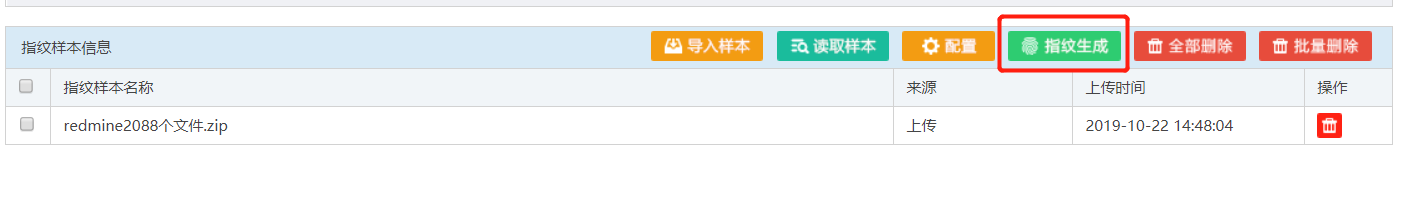


其中：

* 包含：表示指纹中必须包含代码中的内容特征
* 排除：表示指纹中排除部分干扰的内容特征

代码指纹生成

在读取指纹并配置后，点击“指纹生成”方可生成代码相似度指纹。



删除指纹样本

点击“全部删除”按钮，删除所选规则下所有指纹样本。

点击“批量删除”按钮，删除所选规则下所有已选中的单个/多个指纹样本。