

文本内容安全

快速入门

产品文档



腾讯云

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

快速入门

最近更新时间：2024-01-22 10:27:31

步骤1：登录注册

已注册腾讯云账号。详情请参见 [账号注册教程](#)。

说明：

如果您已拥有腾讯云账号，可略过此步骤。

步骤2：创建子账号（可选）

在 [注册腾讯云](#) 账号时，所生成的账号为主账号，拥有该主账号下所有云资源的管理权限。如果您需要其他用户来协助您一起管理账号下的云资源，可以通过使用访问管理功能创建子账号，并为其绑定权限策略。该功能可以创建一个或多个子账号提供给团队成员，详情请参见 [新建子用户](#)。

子账号刚创建时，默认是没有任何权限的，需要主账号或管理员为其绑定策略，才能使子账号拥有某些云资源的操作权限。您可以为子账号配置内容安全 CAM 授权，详情请参见 [CAM 授权指引](#)。

步骤3：开通服务

前提条件

已获取腾讯云主账号或子账号。

登录账号已拥有图片内容安全服务的读写权限。

操作步骤

登录 [内容安全控制台](#)，在左侧导航栏中选择文本内容安全下的任一菜单。

步骤4：配置策略（可选）

建议您使用配置任务策略，可根据业务需求配置审核策略，用于个性化服务体验。

说明：

腾讯云内容安全服务已预设有默认策略，如使用默认策略，可略过此步骤。

默认策略为天御多行业模型沉淀的策略配置，适用于大部分的内容安全需求。

前提条件

已获取腾讯云主账号或子账号。

登录账号已拥有文本内容安全服务的读写权限。

登录账号已开通文本内容安全服务。

操作步骤

1. 登录 [内容安全控制台](#)，在左侧导航栏中，选择**文本内容安全** > **策略管理**。
2. 在策略管理页面，单击**创建策略**，进入创建策略页面。



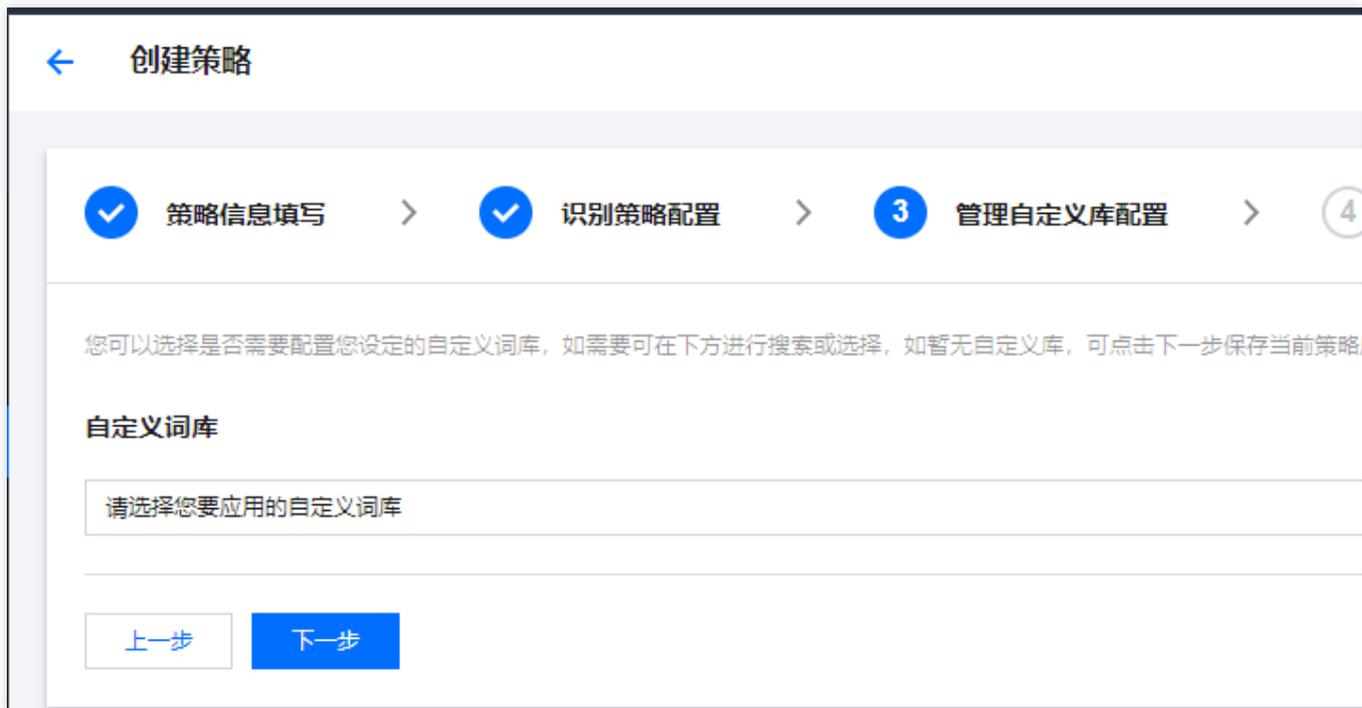
3. 在策略信息填写页面，填写策略的相关信息，完成后单击**下一步**。



参数说明

参数名称	描述
策略名称	策略的文字描述，可使用中文、英文、数字及下划线组合，长度不超过30个字符
Biztype 名称	策略的具体编号，以便用于接口调用，可使用英文、数字及下划线组合，长度为3-32个字符。 注：Biztype 名称必须唯一，不可重复
行业分类	策略涉及的行业场景分类
使用行业模板	仅 行业分类 有设置时显示。选择是否使用腾讯云预设的行业模板进行审核

- 在识别策略配置页面，根据业务需求，选择是否需要审核不同类型的识别内容，单击**下一步**。
- 在管理自定义库配置页面，在自定义词库下拉列表中选择是否需要使用自定义词库用于内容审核，如暂无自定义库，可单击**下一步**跳过，或保存当前策略后，前往 [步骤5：配置自定义词库](#)。



6. 在创建完成页面，可总览策略配置信息，确认后单击**完成**，完成审核策略配置。

7. 在策略管理页面右侧列表中，将显示刚创建的这一条策略。

策略名称	Biztype名称 ①	行业分类	关联服
策略名称	-	-	无
eee	eee	社... 行业	无

共 2 条

步骤5：配置自定义词库（可选）

自定义词库用于添加个性化的审核词库，识别图片上是否含有违规文本。

说明：

如无需配置自定义词库，可略过此步骤。

前提条件

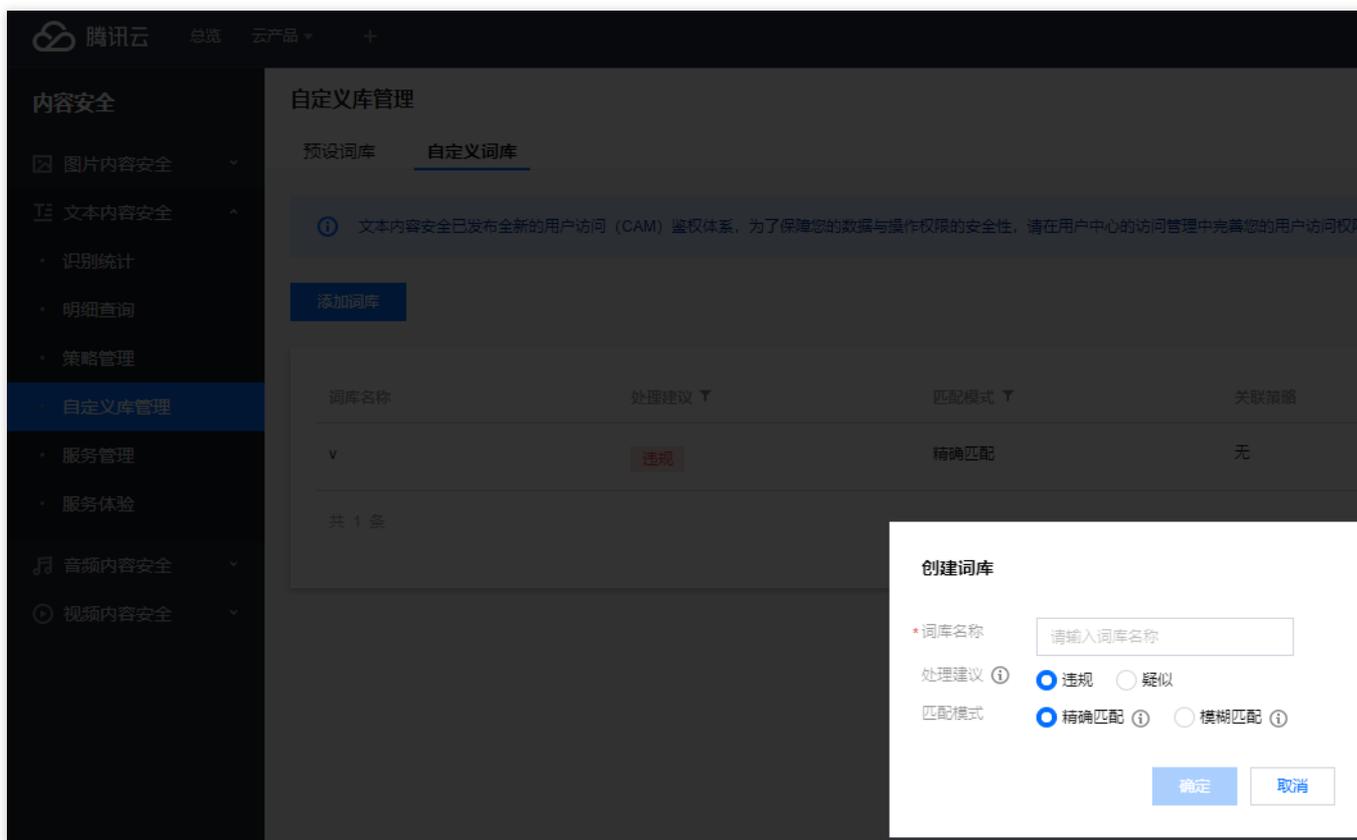
已获取腾讯云主账号或子账号。

登录账号已拥有文本内容安全服务的读写权限。

登录账号已开通文本内容安全服务。

操作步骤

1. 登录 [内容安全控制台](#)，在左侧导航栏中，选择**文本内容安全** > **自定义库管理** > **自定义词库**。
2. 在自定义词库页面，单击**添加词库**，弹出创建词库弹窗。



3. 在创建词库弹窗，填写自定义库的相关信息，结合业务需求配置自定义库。

创建词库
✕

*词库名称

处理建议 ⓘ 违规 疑似

匹配模式 精确匹配 ⓘ 模糊匹配 ⓘ

确定
取消

参数说明

参数名称	描述
词库名称	词库的文字描述，可使用中文、英文、数字及下划线组合，长度不超过32个字符
处理建议	可选择违规或疑似 违规：确认为不良信息 疑似：可能为不良信息，需要人工审核
匹配模式	中文支持精确匹配或模糊匹配，英文支持精确匹配 精确匹配：对输入文本进行匹配，匹配对象需完全一致 模糊匹配：可检测变体后的输入词，支持拆分字、形似字、音似字、简繁体、大小写、大写数字等形式的相似词进行匹配

4. 单击**确认**，即可创建自定义库。
5. 在自定义词库页面下方的列表中，将显示刚创建的词库。

说明：

自定义词库的不同颜色代表不同的屏蔽逻辑，红色代表违规，橙色代表疑似。

词库名称	处理建议 ▼	匹配模式 ▼	关联策略	最近修改时间
999	违规	精确匹配	无	2024-01-22 10:22:53
v	违规	精确匹配	无	2024-01-22 10:17:19

共 2 条

6. 在自定义词库页面，选择刚创建的词库，单击操作栏下的**管理**，进入管理词库页面。



7. 在管理词库页面，单击**添加样本**，弹出添加样本弹窗。

8. 在添加样本弹窗，选择处理建议，输入关键词，单击**确定**，即可保存关键词至当前词库。



参数说明

参数名称	描述
处理建议	识别模型对应的违规类型
关键词	单个关键词以换行来确定，长度在20个汉字或字母以内 关键词支持批量添加，单次最多可批量添加500个关键词 关键词添加个数上限为2000个

说明：

自定义词库配置完成后，可以在 [步骤4：配置策略](#) 时将自定义词库与策略关联使用。

步骤6：体验服务

完成以上步骤后，您可以选择创建的审核策略，通过立即审核图片，体验文本内容安全服务。

前提条件

已获取腾讯云主账号或子账号。

登录账号已拥有文本内容安全服务的读写权限。

登录账号已开通文本内容安全服务。

操作步骤

1. 登录 [内容安全控制台](#)，在左侧导航栏中，选择**文本内容安全 > 服务体验**，进入服务体验页面。
2. 在服务体验页面，选择所需的审核策略，输入需要审核的文本。

说明：

文本内容需满足以下要求：

文本内容大小：文本长度不能超过10000，按 unicode 编码长度计算。

文本审核语言：目前仅支持中文的检测。

3. 单击**立即审核**，在审核内容下方会显示该文本的审核结果。

服务体验

i 文本内容安全已发布全新的用户访问（CAM）鉴权体系，为了保障您的数据与操作权限的安全性，请在用户中心的访问管理

应用策略 **i**

审核内容

立即审核

步骤7：接入服务

如果您需要接入文本内容安全服务，需通过调用 API 接口完成接入，详情请参考 [API 接入指引](#)。

说明：

在调用 API 接口之前，您需要获取腾讯云 API 访问密钥。腾讯云通过 SecretId 和 SecretKey 对开发者的身份和权限进行校验，您可以参考以下步骤获取腾讯云 API 访问密钥：

进入 [云 API 密钥管理](#) 页面，在左侧导航栏选择 **访问密钥 > API 密钥管理**，单击 **新建密钥** 创建密钥，保存 SecretId 和 SecretKey 供后续 API 调用传参。