文档审核

API 文档

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2024 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】

腾讯云

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。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。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默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

文档目录

API 文档
简介
公共请求头部
公共响应头部
请求签名
提交文档审核任务
查询文档审核任务结果
文档审核回调内容
错误码

API 文档
简介

最近更新时间：2021-11-26 17:12:23

文档审核（Document Auditing Service，DAS）基于腾讯云 数据万象（Cloud Infinite，CI），使用 XML API，这是一种轻量级的、无连接状态的接口，调用此接口您可以直接通过 HTTP/HTTPS 发出请求和接受响应，实现与腾讯云文档瘦身后台的交互操作。

术语信息

文中会出现的一些主要概念和术语：

名称	描述
APPID	开发者访问 DAS 服务时拥有的用户维度唯一资源标识，用以标识资源，可在 API 密钥管理 页面获取
SecretId	开发者拥有的项目身份识别 ID，用以身份认证，可在 API 密钥管理 页面获取
SecretKey	开发者拥有的项目身份密钥，可在 API 密钥管理 页面获取
Bucket	存储桶，DAS 中用于存储数据的容器。有关存储桶的进一步说明，请参见 存储桶概述 文档
Object	对象，DAS 中存储的具体文件，是存储的基本实体
ObjectKey	对象键，对象（Object）在存储桶（Bucket）中的唯一标识。有关对象与对象键的进一步说明，请参见 对象概述 文档
Region	地域信息，枚举值可参见 地域和访问域名 文档，例如：ap-beijing、ap-hongkong、eu-frankfurt 等

快速入门

要使用腾讯云文档审核，需要先执行以下步骤：

1. 在腾讯云 数据万象控制台 开通腾讯云数据万象服务。
2. 在腾讯云 数据万象控制台 创建一个存储桶。
3. 在访问管理控制台中的 API 密钥管理 页面里获取 APPID、SecretId、SecretKey 内容。
4. 编写一个请求签名算法程序（或使用任何一种服务端 SDK），详情请参见 请求签名 文档。
5. 计算签名，调用 API 执行操作。

公共请求头部

最近更新时间：2021-11-26 17:12:26

描述

此篇文档将为您介绍在使用 API 时候会使用到的公共请求头部（Request Header），下文提到的头部在之后的具体 API 文档中不再赘述。

请求头部列表

Header 名称	描述	类型	是否必选
Authorization	携带鉴权信息，用以验证请求合法性的签名信息。针对公有读的对象可不携带此头部，如通过请求参数传递鉴权信息也无需携带此头部，详情请参见 请求签名 文档。	string	是。针对公有读的对象或通过请求参数传递鉴权信息，此头部为可选项。
Content-Length	RFC 2616中定义的 HTTP 请求内容长度（字节）。	integer	否。针对 PUT 和 POST 请求（指定 Transfer-Encoding 请求头部的 PUT Object 请求除外），此头部是必选项。
Content-Type	RFC 2616中定义的 HTTP 请求内容类型（MIME）例如 `application/xml` 或 `image/jpeg` 。	string	否。针对有请求体的 PUT 和 POST 请求，此头部是必选项。
Content-MD5	RFC 1864中定义的请求体内容的16字节二进制 MD5 哈希值的 Base64 编码形式，最终的取值长度应为24个字符，请注意在编写代码时使用正确的方法和参数，例如 `ZzD3iDJdrMAAb00lgLLeig==` 。此头部用于完整性检查，验证请求体在传输过程中是否发生变化。针对有请求体的 PUT 和 POST 请求（除 POST Object），强烈建议携带此头部。	string	否。
Date	RFC 1123中定义的 GMT 格式当前时间，例如 `Wed, 29 May 2019 04:10:12 GMT` 。	string	否。

Host	请求的主机，形式为 `<BucketName-APPID>.cos.<Region>.myqcloud.com` 。	string	是。
x-cos-security-token	使用临时安全凭证时需要传入的安全令牌字段，详情请参见 临时安全凭证 相关说明。	string	否。当使用临时密钥并通过 Authorization 携带鉴权信息时，此头部为必选项。

服务端加密专用头部

对于支持服务端加密（Server Side Encryption，SSE）的接口，根据不同的加密方式适用如下请求头部，请参阅具体的接口文档确定是否适用 SSE。下列头部的是否必选仅针对使用 SSE 的场景，如果请求不支持 SSE 的接口或不使用 SSE ，则以下头部均无需携带。详情请参见 服务端加密概述 。

SSE-COS

Header 名称	描述	类型	是否必选
x-cos-server-side-encryption	服务端加密算法，使用 SSE-COS 时指定为：AES256。	string	上传或复制对象（包括简单上传/复制与分块上传/复制）时必选，下载对象时不能指定此头部。

SSE-KMS

Header 名称	描述	类型	是否必选
x-cos-server-side-encryption	服务端加密算法，使用 SSE-KMS 时指定为：cos/kms。	string	上传或复制对象（包括简单上传/复制与分块上传/复制）时必选，下载对象时不能指定此头部。
x-cos-server-side-encryption-cos-kms-key-id	当 x-cos-server-side-encryption 值为 cos/kms 时，用于指定 KMS 的用户主密钥 CMK，如不指定，则使用 COS 默认创建的 CMK，更多详细信息可参见 SSE-KMS 加密 。	string	否。
x-cos-server-side-	当 x-cos-server-side-encryption 值为 cos/kms 时，用于指定加密上下文，值为	strin	否。

encryption-context	JSON 格式加密上下文键值对的 Base64 编码。 例如 `eyJhIjoiYXNkZmEiLCJiIjoiMTIzMzIxIn0=` 。	g	

SSE-C

Header 名称	描述	类型	是否必选
x-cos-server-side-encryption-customer-algorithm	服务端加密算法，支持 AES256和 KMS，值：AES256、cos/kms。	string	是。
x-cos-server-side-encryption-customer-key	服务端加密密钥的 Base64 编码。 例如 `MDEyMzQ1Njc4OUFCQ0RFRjAxMjM0NTY3ODlBQkNERUY=` 。	string	是。
x-cos-server-side-encryption-customer-key-MD5	服务端加密密钥的 MD5 哈希值，使用 Base64 编码。 例如 `U5L61r7jcwdNvT7frmUG8g==` 。	string	是。

公共响应头部

最近更新时间：2021-11-26 17:12:29

描述

此篇文档将为您介绍在使用 API 时会出现的公共响应头部（Response Header），下文提到的头部在后续的具体 API 文档中不再赘述。

响应头部列表

Header 名称	描述	类型
Content-Length	RFC 2616 中定义的 HTTP 响应内容长度（字节）。	string
Content-Type	RFC 2616 中定义的 HTTP 响应内容类型（MIME）。	string
Connection	RFC 2616 中定义，表明响应完成后是否会关闭网络连接。枚举值：keep-alive，close。	Enum
Date	RFC 1123 中定义的 GMT 格式服务端响应时间，例如Wed, 29 May 2019 04:10:12 GMT。	string
ETag	ETag 全称为 Entity Tag，是对象被创建时标识对象内容的信息标签，可用于检查对象的内容是否发生变化，例如 `"8e0b617ca298a564c3331da28dcb50df"` 。此头部并不一定返回对象的 MD5 值，而是根据对象上传和加密方式而有所不同。	string
Last-Modified	对象的最近一次上传的时间，例如 `Fri, 10 Apr 2020 18:17:25 GMT` 。	string
Server	接收请求并返回响应的服务器的名称，默认值： `tencent-cos` 。	string
Transfer-Encoding	RFC 2616 中定义的传输编码格式。	string
x-cos-hash-crc64ecma	对象的 CRC64 值，详情请参见 CRC64 校验 文档。	number

x-cos-request-id	每次请求发送时，服务端将会自动为请求生成一个 ID。	string
x-cos-trace-id	每次请求出错时，服务端将会自动为这个错误生成一个 ID。仅当请求出错时才会在响应中包含此头部。	string

服务端加密专用头部

对于支持服务端加密（Server Side Encryption，SSE）且在请求中使用了服务端加密的接口，将根据具体的加密类型返回以下响应头，详情请参见 服务端加密概述 。

SSE-COS

Header 名称	描述	类型
x-cos-server-side-encryption	当上传对象时使用了 SSE-COS，或者下载经过 SSE-COS 加密的对象时，请求的响应均会返回此头部，用于表明对象上传时使用的服务端加密算法。	string

SSE-KMS

Header 名称	描述	类型
x-cos-server-side-encryption	当上传对象时使用了 SSE-KMS，或者下载经过 SSE-KMS 加密的对象时，请求的响应均会返回此头部，用于表明对象上传时使用的服务端加密算法。	string
x-cos-server-side-encryption-cos-kms-key-id	返回 KMS（密钥管理系统）的用户主密钥 CMK。如未指定，则返回 COS 默认创建的 CMK。	string

SSE-C

Header 名称	描述	类型
x-cos-server-side-encryption-customer-algorithm	当上传对象时使用了 SSE-C，或者下载经过 SSE-C 加密的对象时，请求的响应均会返回此头部，用于表明对象上传时使用的服务端加密算法。	string
x-cos-server-side-encryption-customer-key-MD5	对象上传时使用的服务端加密密钥的 MD5 哈希值，使用 Base64 编码。 例如 `U5L61r7jcwdNvT7frmUG8g==` 。	string

请求签名

最近更新时间：2021-12-20 16:24:48

文档审核由 数据万象（Cloud Infinite，CI） 提供。腾讯云数据万象服务通过签名来验证请求的合法性。开发者通过将签名授权给客户端，使其具备上传下载及管理指定资源的能力。
使用数据万象服务时，可通过 RESTful API 发起 HTTP 匿名请求或 HTTP 签名请求，对于签名请求，服务器端将会对请求发起者进行身份验证。

- 匿名请求：HTTP 请求不携带任何身份标识和鉴权信息，通过 RESTful API 进行 HTTP 请求操作。
- 签名请求：HTTP 请求时携带签名，服务器端收到消息后，即可进行身份验证，验证成功则可接受并执行请求，否则将会返回错误信息并丢弃此请求。

数据万象基于 对象存储（Cloud Object Storage，COS），当前签名算法与 COS 保持一致，基于密钥 HMAC（Hash Message Authentication Code）的自定义方案进行身份验证。

签名算法

当前存在 XML 和 JSON 两个签名版本：

- **下载时**操作需使用 **JSON** 版本签名，可参阅 JSON 版本 请求签名 。
- **上传时**操作和 **Bucket 接口**等操作需使用 **XML** 版本签名，可参阅 XML 版本 请求签名 。

> **注意**
> - 下载时请求头部 Host 形式为 `<BucketName-APPID>.<picture region>.myqcloud.com/<picture name>` ，例如 `examplebucket-1250000000.picsh.myqcloud.com/picture.jpeg`
> - 上传时请求头部 Host 形式为 `<BucketName-APPID>.pic.<Region>.myqcloud.com` ，例如 `examplebucket-1250000000.pic.ap-shanghai.myqcloud.com/picture.jpeg`

签名适用场景

场景		适用签名
下载时数据处理	无防盗链	不验证签名
	开启防盗链	JSON 版本签名
上传时数据处理	持久化处理	XML 版本签名
Bucket 接口操作	查询、开通、删除等	XML 版本签名
内容识别	鉴黄、违法违规	XML 版本签名

签名工具

生成签名所需信息包括 APPID（例如1250000000），Bucket 名称（例如 examplebucket-ci），项目的 SecretID 和 SecretKey。
获取这些信息的方法如下：

1. 登录 数据万象控制台 。
2. 在左侧导航栏中，单击**存储桶管理**，进入存储桶管理页面。
3. 单击需要操作的存储桶名称，进入该存储桶管理页面。
4. 在左侧导航栏中，单击**存储桶配置**，即可查看存储桶名称和 Bucket ID 。若当前项目无存储桶，您可参考 创建存储桶 文档进行创建。
5. 进入 访问管理 API 密钥管理页面，获取 SecretId 和 SecretKey。

数据万象签名计算过程与 COS 一致，您可以根据**签名适用场景**，使用 COS 提供的 在线签名工具 生成所需版本签名。

签名使用

通过 RESTful API 发起的 HTTP 签名请求，可以通过以下几种方式传递签名：

1. 通过标准的 HTTP Authorization 头，如
```
Authorization: q-sign-algorithm=sha1&q-ak=...&q-sign-time=1557989753;1557996953&...&q-signature=...
```
2. 作为 HTTP 请求参数，请注意 UrlEncode，如
```
/exampleobject?q-sign-algorithm=sha1&q-ak=...&q-sign-time=1557989753%3B1557996953&...&q-signature=...
```

> **注意**
> 上述示例中使用 `...` 省略了部分具体签名内容。

提交文档审核任务

最近更新时间：2021-12-21 15:37:10

功能描述

本接口用于提交一个文档审核任务，可审核您的文档文件是否存在敏感违规信息。文档审核结合了对象存储（Cloud Object Storage，COS）文档预览功能，通过预先将文档转成图片，结合图片内容审核、图片 OCR 审核等方式，进行文档审核。

> **说明**
> - 目前支持的输入文件类型包含如下格式：
  - 演示文件：pptx、ppt、pot、potx、pps、ppsx、dps、dpt、pptm、potm、ppsm。
  - 文字文件：doc、dot、wps、wpt、docx、dotx、docm、dotm。
  - 表格文件：xls、xlt、et、ett、xlsx、xltx、csv、xlsb、xlsm、xltm、ets。
  - 其他格式文件：pdf。
- 表格文件，一张表可能分割为多页转换，生成多张图片。
- 输入文件大小限制在200MB之内。
- 输入文件页数限制在5000页之内。

请求

请求示例

```
POST /document/auditing HTTP/1.1
Host: <BucketName-APPID>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: <Auth String>
Content-Length: <length>
Content-Type: application/xml

<body>
```

> **说明**
> Authorization: Auth String（详情请参见 请求签名 文档）。

请求头

此接口仅使用公共请求头部，详情请参见 公共请求头部 文档。

请求体

该请求操作的实现需要有如下请求体。

```
<Request>
      <Input>
         <Url></Url>
         <Type></Type>
      </Input>
      <Conf>
         <DetectType></DetectType>
         <Callback></Callback>
         <BizType></BizType>
      </Conf>
</Request>
```

具体的数据描述如下：

节点名称（关键字）	父节点	描述	类型	是否必选
Request	无	文档审核的具体配置项	Container	是

Container 类型 Request 的具体数据描述如下：

节点名称（关键字）	父节点	描述	类型	是否必选
Input	Request	需要审核的内容	Container	是
Conf	Request	审核规则配置	Container	是

Container 类型 Input 的具体数据描述如下：

节点名称（关键字）	父节点	描述	类型	是否必选
Url	Request.Input	文档文件的链接地址，例如 http://www.example.com/doctest.doc	String	是
Type	Request.Input	指定文档文件的类型，如未指定则默认以文件的后缀为类型。 如果文件没有后缀，该字段必须指定，否则会审核失败。例如：doc、docx、ppt、pptx 等	String	否

Container 类型 Conf 的具体数据描述如下：

节点名称（关键字）	父节点	描述	类型	是否必选
BizType	Request.Conf	审核策略，不填写则使用默认策略。可在控制台进行配置，详情请参见 设置审核策略	String	否
DetectType	Request.Conf	审核的场景类型，有效值：Porn（涉黄）、Ads（广告），可以传入多种类型，不同类型以逗号分隔，例如：Porn,Ads	String	否
Callback	Request.Conf	审核结果可以回调形式发送至您的回调地址，支持以 `http://` 或者 `https://` 开头的地址，例如：`http://www.callback.com`	String	否

响应

响应头

此接口仅返回公共响应头部，详情请参见 公共响应头部 文档。

响应体

该响应体返回为 application/xml 数据，包含完整节点数据的内容展示如下：

```
<Request>
    <JobsDetail>
        <JobId></JobId>
        <State></State>
        <CreationTime></CreationTime>
    </JobsDetail>
</Request>
```

具体的数据内容如下：

节点名称（关键字）	父节点	描述	类型
Response	无	提交文档审核任务返回的具体响应内容	Container

Container 节点 Response 的内容：

节点名称（关键字）	父节点	描述	类型

JobsDetail	Response	文档审核任务的详细信息	Container

Container 节点 JobsDetail 的内容：

节点名称（关键字）	父节点	描述	类型
JobId	Response.JobsDetail	本次文本审核任务的 ID	String
State	Response.JobsDetail	文档审核任务的状态，值为 Submitted（已提交审核）、Success（审核成功）、Failed（审核失败）、Auditing（审核中）其中一个	String
Creation Time	Response.JobsDetail	文档审核任务的创建时间	String

错误码

该请求操作无特殊错误信息，常见的错误信息请参见 错误码 文档。

实际案例

请求

```
POST /document/auditing HTTP/1.1
Authorization: q-sign-algorithm=sha1&q-ak=AKIDZfbOAo7cllgPvF9cXFrJD0a1ICvR****&q-sign-time=1497530202;1497610202&q-key-time=1497530202;1497610202&q-header-list=&q-url-param-list=&q-signature=28e9a4986df11bed0255e97ff90500557e0e****
Host: examplebucket-1250000000.ci.ap-beijing.myqcloud.com
Content-Length: 166
Content-Type: application/xml
<Request>
      <Input>
         <Url>http://www.example.com/doctest.doc</Url>
         <Type></Type>
      </Input>
      <Conf>
         <DetectType>Porn,Ads</DetectType>
         <Callback></Callback>
         <BizType>b81d45f94b91a683255e9a9506f4****</BizType>
```

```
        </Conf>
</Request>
```

响应

```
HTTP/1.1 200 OK
Content-Type: application/xml
Content-Length: 230
Connection: keep-alive
Date: Thu, 15 Jun 2017 12:37:29 GMT
Server: tencent-ci
x-ci-request-id: NTk0MjdmODlfMjQ4OGY3XzYzYzhf****

<Response>
    <JobsDetail>
        <JobId>vab1ca9fc8a3ed11ea834c52540086****</JobId>
        <State>Submitted</State>
        <CreationTime>2019-07-07T12:12:12+0800</CreationTime>
    </JobsDetail>
</Response>
```

查询文档审核任务结果

最近更新时间：2021-12-21 15:33:14

功能描述

本接口用于查询指定的文档审核任务结果。

请求

请求示例

```
GET /document/auditing/<jobId> HTTP/1.1
Host: <BucketName-APPID>.ci.<Region>.myqcloud.com
Date: <GMT Date>
Authorization: <Auth String>
```

> **说明**
> Authorization: Auth String（详情请参见 请求签名 文档）。

请求头

此接口仅使用公共请求头部，详情请参见 公共请求头部 文档。

请求体

该请求无请求体。

响应

响应头

此接口仅返回公共响应头部，详情请参见 公共响应头部 文档。

响应体

该响应体返回为 application/xml 数据，包含完整节点数据的内容展示如下：

```
<Response>
    <JobsDetail>
      <JobId>aab1ca9fc8a3ed11ea834c525400863904</JobId>
      <State>Success</State>
      <Code>Success</Code>
      <Message>Success</Message>
```

```
      <Suggestion>1</Suggestion>
      <CreationTime>2019-07-07T12:12:12+0800</CreationTime>
      <Url>http://www.test.com/123</Url>
      <PageCount>3</PageCount>
      <Labels>
        <PornInfo>
          <HitFlag>1</HitFlag>
          <Score>96</Score>
        </PornInfo>
      </Labels>
      <PageSegment>
        <Results>
          <Url></Url>
          <Text></Text>
          <PageNumber>1</PageNumber>
          <SheetNumber>1</SheetNumber>
          <PornInfo>
            <HitFlag></HitFlag>
            <SubLabel></SubLabel>
            <Score></Score>
            <OcrResults>
              <Text></Text>
              <Keywords></Keywords>
            </OcrResults>
          </PornInfo>
        </Results>
      </PageSegment>
    </JobsDetail>
</Response>
```

具体的数据内容如下：

节点名称（关键字）	父节点	描述	类型
Response	无	文档审核的全部信息	Container

Container 节点 Response 的内容：

节点名称（关键字）	父节点	描述	类型
JobsDetail	Response	文档审核任务的详细信息	Container

Container 节点 JobsDetail 的内容：

节点名称（关键字）	父节点	描述	类型
Code	Response.JobsDetail	错误码，只有 State 为 Failed 时有意义	String
Message	Response.JobsDetail	错误描述，只有 State 为 Failed 时有意义	String
JobId	Response.JobsDetail	文档审核任务的 ID	String
State	Response.JobsDetail	任务的状态，值为 Submitted（已提交审核）、Success（审核成功）、Failed（审核失败）、Auditing（审核中）其中一个	String
Suggestion	Response.JobsDetail	• 该字段表示本次判定的审核结果，您可以根据该结果，进行后续的操作；建议您按照业务所需，对不同的审核结果进行相应处理。 • 有效值：0（审核正常），1（判定为违规敏感文件），2（疑似敏感，建议人工复核）	String
CreationTime	Response.JobsDetail	任务的创建时间	String
Url	Response.JobsDetail	审核的文件链接	String
PageCount	Response.JobsDetail	文档审核会将文档转换为图片进行审核，该字段表示转换的图片总数量	String
Labels	Response.JobsDetail	该字段用于返回命中的审核场景及对应的结果	Container
PageSegment	Response.JobsDetail	文档转换为图片后，具体每张图片的审核结果信息，只返回带有违规结果的图片	Container

Container 节点 Labels 的内容：

节点名称（关键字）	父节点	描述	类型
PornInfo	Response.JobsDetail.Labels	涉黄信息，包含审核的结果信息	Container

AdsInfo	Response.JobsDetail.Labels	广告引导信息，包含审核的结果信息	Container

Container 节点 PornInfo，AdsInfo 的内容：

节点名称（关键字）	父节点	描述	类型
HitFlag	Response.JobsDetail.Labels.*Info	是否命中该审核分类，0表示未命中，1表示命中，2表示疑似	String
Score	Response.JobsDetail.Labels.*Info	该字段表示审核结果命中审核信息的置信度，取值范围：0（**置信度最低**）-100（**置信度最高**），越高代表该内容越有可能属于当前返回审核信息 例如：色情 99，则表明该内容非常有可能属于色情内容	String

Container 节点 PageSegment 的内容：

节点名称（关键字）	父节点	描述	类型
Results	Response.JobsDetail.PageSegment	文档转换为图片后，每张图片的详细审核结果信息	Container

Container 节点 Results 的内容：

节点名称（关键字）	父节点	描述	类型
Url	Response.JobsDetail.PageSegment.Results	文档转换成图片后，该张图片的 Url	String
Text	Response.JobsDetail.PageSegment.Results	文档审核会使用 OCR 方式进行审核，该字段表示图片 OCR 文本识别出的全部文本内容，识别**上限在5000字节内**	String
PageNumber	Response.JobsDetail.PageSegment.Results	该图片的页码，通常是文档的页码	String
SheetNumber	Response.JobsDetail.PageSegment.Res	如果审核的为表格文件，该字段表示表格内的 Sheet 页码	String

	ults		
PornInfo	Response.JobsDetail.PageSegment.Results	涉黄信息，包含审核的结果信息	Container
AdsInfo	Response.JobsDetail.PageSegment.Results	广告引导信息，包含审核的结果信息	Container

Container 节点 PornInfo，AdsInfo 的内容：

节点名称（关键字）	父节点	描述	类型
HitFlag	Response.JobsDetail.PageSegment.Results.*Info	是否命中该审核分类，0表示未命中，1表示命中，2表示疑似	String
SubLabel	Response.JobsDetail.PageSegment.Results.*Info	• 该字段表示审核命中的具体子标签，例如：Porn 下的 SexBehavior 子标签 • 注意：该字段可能返回空，表示未命中具体的子标签	String
Score	Response.JobsDetail.PageSegment.Results.*Info	• 该字段表示审核结果命中审核信息的置信度，取值范围：0（**置信度最低**）-100（**置信度最高**），越高代表该内容越有可能属于当前返回审核信息 • 例如：色情 99，则表明该内容非常有可能属于色情内容	String
OcrResults	Response.JobsDetail.PageSegment.Results.*Info	该字段表示 OCR 文本识别的详细检测结果；包括：文本识别结果、命中的关键词等信息	Container

Container 节点 OcrResults 的内容：

节点名称（关键字）	父节点	描述	类型
Text	Response.JobsDetail.PageSegment.Results.*Info.OcrResults	图片 OCR 文本识别出的具体文本内容	String
Keywords	Response.JobsDetail.PageSegment.Results.*Info.OcrResults	在当前审核场景下命中的关键词	String

错误码

此接口遵循统一的错误响应和错误码，详情请参见 错误码 文档。

实际案例

请求

```
GET /document/auditing/stb1ca9fc8a34c525400863904**** HTTP/1.1
Accept: */*
Authorization: q-sign-algorithm=sha1&q-
ak=AKIDZfbOAo7cllgPvF9cXFrJD0a1ICvR****&q-sign-
time=1497530202;1497610202&q-key-time=1497530202;1497610202&q-header-
list=&q-url-param-list=&q-
signature=28e9a4986df11bed0255e97ff90500557e0e****
Host: examplebucket-1250000000.ci.ap-beijing.myqcloud.com
```

响应

```
HTTP/1.1 200 OK
Content-Type: application/xml
Content-Length: 666
Connection: keep-alive
Date: Thu, 15 Jun 2017 12:37:29 GMT
Server: tencent-ci
x-ci-request-id: NTk0MjdmODlfMjQ4OGY3XzYzYzhf****
<Response>
    <JobsDetail>
      <JobId>stb1ca9fc8a34c525400863904****</JobId>
      <State>Success</State>
      <Code>Success</Code>
      <Message>Success</Message>
      <Suggestion>1</Suggestion>
      <CreationTime>2019-07-07T12:12:12+0800</CreationTime>
      <Url>http://www.test.com/123</Url>
      <PageCount>3</PageCount>
      <Labels>
        <PornInfo>
          <HitFlag>1</HitFlag>
          <Score>96</Score>
        </PornInfo>
      </Labels>
      <PageSegment>
        <Results>
```

```
          <Url>https://examplebucket-1250000000.cos.ap-
nanjing.myqcloud.com/audit-1.jpg</Url>
          <Text>哈哈哈，您好</Text>
          <PageNumber>1</PageNumber>
          <PornInfo>
            <HitFlag>0</HitFlag>
            <SubLabel>sex</SubLabel>
            <Score>0</Score>
            <OcrResults>
              <Text>哈哈哈，您好</Text>
              <Keywords>您好</Keywords>
            </OcrResults>
          </PornInfo>
        </Results>
        <Results>
          <Url>https://examplebucket-1250000000.cos.ap-
nanjing.myqcloud.com/audit-2.jpg</Url>
          <Text>哈哈哈，您好</Text>
          <PageNumber>1</PageNumber>
          <PornInfo>
            <HitFlag>0</HitFlag>
            <SubLabel>sex</SubLabel>
            <Score>0</Score>
            <OcrResults>
              <Text>哈哈哈，您好</Text>
              <Keywords>您好</Keywords>
            </OcrResults>
          </PornInfo>
        </Results>
      </PageSegment>
    </JobsDetail>
</Response>
```

文档审核回调内容

最近更新时间：2021-12-21 15:39:38

功能描述

若您配置了审核回调地址，在文档审核结束后，后台会将审核结果以 json 形式回调至您的回调地址。您可以通过回调内容，进行后续的文件处理操作。

回调内容说明

该响应体返回为 application/json 数据，包含完整节点数据的内容展示如下：

```
{
  "EventName": "ReviewDocument",
  "JobsDetail": {
    "JobId": "6666666666666666666666666666666666",
    "State": "Success",
    "CreationTime": "",
    "Url": "",
    "Label": "Normal",
    "Suggestion": 0,
    "PageCount": 1,
    "Labels": {
      "PornInfo": {
        "HitFlag": 0,
        "Score": 0
      },
      "AdsInfo": {
        "HitFlag": 0,
        "Score": 0
      }
    },
    "PageSegment": {
      "Results": [
      {
        "Url": "",
        "Text": "",
        "PageNumber": 0,
        "SheetNumber": 0,
        "Label": "Normal",
        "Suggestion": 0,
        "PornInfo": {
```

```
            "HitFlag": 0,
            "Score": 0,
            "SubLabel": "",
            "OcrResults": [{
              "Text": "",
              "Keywords":["",""],
              "Location": {
                "X": 0,
                "Y": 0,
                "Width": 0,
                "Height": 0,
                "Rotate": 0
              }
            }]
          },
          "AdsInfo": {
            "HitFlag": 0,
            "Score": 0,
            "SubLabel": ""
          }
        }
        ]
      },
      "BucketId": "examplebucket-1250000000",
      "Region": "ap-chongqing",
      "ForbidState": 0,
      "CosHeaders": {
          "x-cos-meta-id": "666666"
      }
    }
  }
```


具体的数据内容如下：

节点名称（关键字）	描述	类型
JobsDetail	任务的详细结果信息	Object
EventName	任务的类型，为 `ReviewDocument`	String

JobsDetail 的内容如下：

节点名称	描述	类型

（关键字）		
Code	错误码，值为0时表示审核成功，非0表示审核失败。详情请查看 内容审核错误码 。	String
Message	错误描述，当 code 为0时该字段无内容。	String
JobId	审核任务的 ID。	String
State	审核任务的状态，值为 Submitted（已提交审核）、Success（审核成功）、Failed（审核失败）、Auditing（审核中）其中一个。	String
Label	• 该字段用于返回检测结果中所对应的**优先级最高的恶意标签**，表示模型推荐的审核结果，建议您按照业务所需，对不同违规类型与建议值进行处理。 • 返回值：**Normal**：正常，**Porn**：色情，**Ads**：广告。	String
Suggestion	• 该字段表示本次判定的审核结果，您可以根据该结果，进行后续的操作；建议您按照业务所需，对不同的审核结果进行相应处理。 • 有效值：**0**（审核正常），**1**（判定为违规敏感文件），**2**（疑似敏感，建议人工复核）。	Integer
Creation Time	审核任务的创建时间。	String
Url	审核的文件链接。	String
PageCount	文档审核会将文档转换为图片进行审核，该字段表示转换的图片总数量。	Integer
Labels	该字段用于返回命中的审核场景及对应的结果。	Object
PageSegment	文档转换为图片后，具体每张图片的审核结果信息，只返回带有违规结果的图片。	Object

Labels 的内容如下：

节点名称（关键字）	描述	类型
PornInfo	审核场景为**涉黄**的审核结果信息。	Object
AdsInfo	审核场景为**广告引导**的审核结果信息。	Object

Labels 中 PornInfo，AdsInfo 的内容如下：

节点名称（关键字）	描述	类型
HitFlag	供参考的识别结果，0（正常），1（敏感），2（疑似敏感）。	Integer
Score	• 该字段表示审核结果命中审核信息的置信度，取值范围：0（**置信度最低**）-100（**置信度最高** ），越高代表该内容越有可能属于当前返回审核信息 • 例如：色情 99，则表明该内容非常有可能属于色情内容。	Integer

PageSegment 的内容如下：

节点名称（关键字）	描述	类型
Results	文档转换为图片后，每张图片的详细审核结果信息。	Array

Results 的内容如下：

节点名称（关键字）	描述	类型
Url	文档转换成图片后，该张图片的 Url。	String
Text	文档审核会使用 OCR 方式进行审核，该字段表示图片 OCR 文本识别出的全部文本内容，识别**上限在5000字节内**。	String
PageNumber	该图片的页码，通常是文档的页码。	Integer
SheetNumber	如果审核的为表格文件，该字段表示表格内的 Sheet 页码。	Integer
Label	• 该字段用于返回检测结果中所对应的**优先级最高的恶意标签**，表示模型推荐的审核结果，建议您按照业务所需，对不同违规类型与建议值进行处理。 • 返回值：**Normal**：正常，**Porn**：色情，**Ads**：广告。	String
Suggestion	• 该字段表示本次判定的审核结果，您可以根据该结果，进行后续的操作；建议您按照业务所需，对不同的审核结果进行相应处理。 • 有效值：**0**（审核正常），**1**（判定为违规敏感文件），**2**（疑似敏感，建议人工复核）	Integer
PornInfo	审核场景为**涉黄**的审核结果信息。	Object

AdsInfo	审核场景为**广告引导**的审核结果信息。	Object

Results 中 PornInfo，AdsInfo 的内容如下：

节点名称（关键字）	描述	类型
HitFlag	是否命中该审核分类，0表示未命中，1表示命中，2表示疑似。	Integer
Score	• 该字段表示审核结果命中审核信息的置信度，取值范围：0（**置信度最低**）-100（**置信度最高**），越高代表该内容越有可能属于当前返回审核信息。通常[0,60]分可以认为图片属于正常，(60,90]分表示图片疑似敏感，(90,100]分表示图片为敏感的。 • 例如：色情 99，则表明该内容非常有可能属于色情内容。	Integer
SubLabel	• 该字段表示审核命中的具体子标签，例如：Porn 下的 SexBehavior 子标签。 • 注意：该字段可能返回空，表示未命中具体的子标签。	String
OcrResults	该字段表示 OCR 文本识别的详细检测结果，包括文本坐标信息、文本识别结果等信息。	Object

OcrResults 的具体内容如下：

节点名称（关键字）	描述	类型
Text	图片 OCR 文本识别出的具体文本内容。	String
Keywords	在当前审核场景下命中的关键词。	Array
Location	该参数用于返回 OCR 检测框在图片中的位置（左上角 xy 坐标、长宽、旋转角度），以方便快速定位识别文字的相关信息。	Object

ObjectResults 的具体内容如下：

节点名称（关键字）	描述	类型
Name	• 该标签用于返回所识别出的实体名称，例如人名。 • 注意：此字段可能返回 null，表示取不到有效值。	String

Location 的具体内容如下：

节点名称（关键字）	描述	类型
X	该参数用于返回检测框**左上角位置的横坐标**（x）所在的像素位置，结合剩余参数可唯一确定检测框的大小和位置。	Integer
Y	该参数用于返回检测框**左上角位置的纵坐标**（y）所在的像素位置，结合剩余参数可唯一确定检测框的大小和位置。	Integer
Height	该参数用于返回**检测框的高度**（由左上角出发在 y 轴向下延伸的长度），结合剩余参数可唯一确定检测框的大小和位置。	Integer
Width	该参数用于返回**检测框的宽度**（由左上角出发在 x 轴向右延伸的长度），结合剩余参数可唯一确定检测框的大小和位置。	Integer
Rotate	该参数用于返回**检测框的旋转角度**，该参数结合 X 和 Y 两个坐标参数可唯一确定检测框的具体位置；取值：**0-360（角度制）**，方向为**逆时针旋转**。	Integer

实际案例


```
{
  "EventName": "ReviewDocument",
  "JobsDetail": {
    "JobId": "xxxxxx",
    "State": "Success",
    "CreationTime": "2021-08-10T21:01:10+08:00",
    "Url": "http://test.com/test.doc",
    "Label": "Normal",
    "Suggestion": 0,
    "PageCount": 1,
    "Labels": {
      "PornInfo": {
        "HitFlag": 0,
        "Score": 0
      },
      "AdsInfo": {
        "HitFlag": 0,
        "Score": 0
      }
    },
    "PageSegment": {
      "Results": [
```

```
      {
        "Url": "http://audit-125000000.cos.ap-
chongqing.myqcloud.com/1.jpg",
        "Text": "",
        "PageNumber": 1,
        "SheetNumber": 0,
        "Label": "Normal",
        "Suggestion": 0,
        "PornInfo": {
          "HitFlag": 0,
          "Score": 0,
          "SubLabel": ""
        },
        "AdsInfo": {
          "HitFlag": 0,
          "Score": 0,
          "SubLabel": ""
        }
      }
      ]
    },
    "BucketId": "examplebucket-1250000000",
    "Region": "ap-chongqing",
    "ForbidState": 0,
    "CosHeaders": {
        "x-cos-meta-id": "xxxx"
    }
  }
}
```

错误码

最近更新时间：2021-11-26 17:12:44

功能说明

图片下载失败时，返回包 Headers 中的 X-ErrNo 为错误码字段，请参考下表进行处理。

错误码列表

下载时处理错误码

错误码	说明	处理措施
-103	无效的请求；请求报文无法识别。	请检查请求内容。
-104	无效的 APPID，URL 中包含的 APPID 无效，或者域名没有和 APPID 绑定。	请检查 APPID 是否正确。
-105	无效的样式名，URL 中指定的样式名或者别名没有配置。	请检查样式配置或样式名内容。
-106	无效的 URL，URL 格式不符合格式要求。	请检查 URL 格式。
-107	无效的 Host 头域。	请检查 Host 是否正确。
-108	无效的 Referer。	请检查相关 Referer 配置。
-109	无效的样式名 ID。没有找到该样式名对应的图片。可能是上传该图片后，新增的样式名， 因此图片上传时不能生成该样式名对应的数据。	-
-110	该图片在黑名单中。	-
-120	回源到源站获取数据时，源站返回的数据有异常，无法正常获取到图片数据。	请检查源站数据。
-124	下载偏移错误。HTTP 请求的 Range 断点续传偏移量可能设置错误。	请检查断点续传偏移量的相关配置。

-154	原图保护机制禁止该 URL 下载。	请使用样式访问。
-156	强制执行302流程。	-
-162	业务配置为欠费，禁止下载图片。	请及时充值。
-163	业务配置不存在。	请检查相关配置项。
-164	热图下载，限制。	请降低请求频率，或 提交工单 联系我们处理。
-165	攻击性下载请求，限制。	-
-166	消息格式错误。	-
-167	文件太大，无法支持图片下载。	请降低图片大小。
-168	文件太大，无法支持图片下载。	请降低图片大小。
-441	图片格式不符，无法进行压缩。	请使用数据万象支持的图片格式进行图片处理操作。具体支持格式为 jpg、bmp、gif、png、webp。
-442	图片超过限制大小。	请将图片长宽限制在9999像素以内。
-443	图片格式不满足处理限制	请使用支持的图片格式进行处理，详情请参见 规则与限制 。
-447	图片分辨率超出限制或动图帧数过多。	请将图片长宽限制在9999像素以内（动图请限制帧数）。
-901	压缩超时。	-
-902	回源超时，镜像存储功能把请求转发到开发商源站，但没有收到响应，超时了。	请检查源站数据或回源配置。
-50	图片涉嫌违禁。	请检查图片内容是否违规。

62		
-6101	图片不存在、图片没有上传或者图片已经删除。	请检查请求源数据。
-29033	下载请求无有效的 range。	-
-29034	下载的 offset 大于文件 size。	-
-29214	触发频控。	请降低请求频率，或 提交工单 联系我们处理。
-40168	原图为空数据。	请检查请求源数据。
-46152	非法的 bucketName。	请检查存储桶名称是否正确。
-46154	非法的 APPID。	请检查 APPID 是否正确。
-46614	文件尚未上传完成，禁止下载。	请等待文件上传成功后请求。
-46617	命中黑名单。	请检查图片内容是否违规。
-46618	签名校验失败。	请检查签名是否准确。
-46619	签名过期。	请更新签名。
-46620	Bucket 被封禁。	您的 Bucket 已被封禁，请 提交工单 联系我们处理。
-46621	文件被封禁。	您的资源已被封禁，请 提交工单 联系我们处理。
-46627	用户目前在黑名单中。	您的账号已被封禁，请 提交工单 联系我们处理。
-46628	文件不存在。	请确认请求下载的文件是否真实存在。

-46642	APPID 不存在。	请检查 APPID 是否正确。
-46646	文件已封禁。	文件已被封禁，请检查文件内容是否违规。

持久化处理错误码

错误码	说明	处理措施
-60998	图片处理参数无效。	请检查图片处理规则是否正确。
-60997	图片处理规则过多。	持久化处理当前最多支持五条规则，请控制在五条以内。
-60996	文件太大，无法处理。	请将文件控制在32MB以内。
-60987	图片获取失败。	请检查图片 URL。
-60983	无效的 Host。	请检查 Host。
-60972	无效地域。	请检查请求地域是否支持数据万象服务。数据万象现有地域详见 地域与域名 。
-60967	图片样式不存在。	请检查图片样式。
-60957	样式数量超过上限。	当前样式数量上限为100条，如需更多样式请联系我们。
-60955	无效的 URL。	请检查 URL 是否正确。
-60950	未指定文件。	请在请求中指定需要处理的文件。
-60949	请求频率过高。	请放缓请求频率。如有需要请联系我们提升频率限制。
-60948	账号涉嫌违规。	您的 APPID 已被封禁，请 提交工单 联系我们处理。

-60938	签名信息不完整。	请检查签名信息中的必选项。
-60936	请求被拒绝。	AccessDenied。
-60932	签名错误。	签名无效，请检查签名信息是否匹配。
-62999	无效的图片格式。	请检查图片格式是否正确。

其他透传自 COS 下载侧错误码详见 COS 错误码 。