

## 文档转 HTML（华为）概述：

### 3 种转换方式：

单一文档转为 HTML，文档是一个下载链接，用 HTTP GET 方式，见：[文档转换 GET](#)

单一文档转为 HTML，文档 POST 到服务器，用 HTTP POST 方式，见：[文档转换 POST](#)

多个图片转为 HTML，文档是多个下载链接，用 HTTP POST 方式，见：[多张图片转换 POST](#)

### 基本用法：

由于转换需要时间，文件越大页数越多，转换越久，故系统采用**异步**的方式获得转换结果。

调用转换接口后会获得 token，随后有 2 种方式查询转换结果：

1. 用 HTTP GET 方式，轮询“查询 query 接口”获得结果。详细见：[查询 QUERY](#)
2. 设置 callbackurl，当转换结束后，系统会回调该 URL 直接推送转换结果。详细见：[回调 URL](#)

API 调用需要签名，详细见[华为官方签名](#)：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

## 文档转换 GET

将文档下载地址 url 转换为 HTML，type 是源文档的 type，比如要把 docx 转为 html，type 就是 docx

### 请求参数：

参数	类型及范围	备注	是否必须发送
url	string	文件 url，支持 http(s)，ftp 开头，需要 URL Encoding	是
type	string	要转换的文档扩展名	是
excelislandscape	int	如果是 Excel 文件，是否横屏，默认 0 否（竖屏），1 是	否
exceliscenter	int	如果是 Excel 文件，是否横竖居中，默认 0 否，1 是	否
excelmargin	int	如果是 Excel 文件，四边的边距，默认 10px，单位是像素	否
excelsheetindex	int	如果是 Excel 文件，指定转换的 Sheet 索引，默认 0：全部，第一个 sheet 就是 1，以此类推	否
excelnotshowgridlines	int	如果是 Excel 文件，不显示	否

参数	类型及范围	备注	是否必须发送
		网格线，默认 0 显示，1 不显示	
exceluseprintarea	int	如果是 Excel 文件，是否使用打印区域，默认 0 不使用，1 使用	否
excelpagesize	int	如果是 Excel 文件，设定页面大小，默认 A4。0:A4，1:A3，2:A4，3:A5，4:B4，5:B5，6:Letter，7:Legal，8:Tabloid，9:Ledger	否
wordshowmarkup	int	如果是 Word 文件，是否显示审阅标记，默认 0 不显示，1 显示	否
powerpointoutputtype	int	如果是 PPT 文件，设定导出样式，默认幻灯片。0:幻灯片，0 以上是讲义模式：1:每页一个幻灯片，2:每页二个幻灯片，3:每页三个幻灯片，4:每页四个幻灯片，5:每页六个幻灯片，6:每页九个幻灯片	否
powerpointhandoutorder	int	如果是 PPT 文件，当设置为讲义时，设定顺序，默认 0 水平，1 垂直	否
powerpointhandoutorientation	int	如果是 PPT 文件，当设置为讲义时，输出文件的方向，默认 0 不改变，1 横向，2 纵向	否
imageocr	int	如果是图片文件，是否识别图中文字并且在 HTML 中可选可搜索文字，默认 1 是，0 否	否
imagedeskew	int	如果是图片文件，是否将斜的文字矫正，默认 0 否，1 是	否
imageclean	int	如果是图片文件，是否清除	否

参数	类型及范围	备注	是否必须发送
		图像背景只显示文字，默认 0 否，1 是	
language	int	OCR 识别语言选项，默认 2 简体中文： 1: 英语 2: 简体中文 3: 繁体中文 4: 法语 5: 德语 6: 意大利语 7: 俄语 8: 日文 9: 韩文 10: 西班牙语 11: 葡萄牙语 12: 丹麦语 13: 荷兰语 14: 芬兰语 15: 挪威语 16: 瑞典语 17: 土耳其语	否
watermark	string	添加水印，字符个数最大 10 个	否
outfilename	string	生成的文件的文件名，默认随机	否
callbackurl	string	回调 URL，转换结束后，会回调该 URL，需要 URL Encoding，详细见 <a href="#">回调 URL</a>	否

### 请求示例：

<http://all2html.apistore.huaweicloud.com/v1/convert?url=https%3a%2f%2fxxx%2fxxx.docx&type=docx>

将所在 url 地址的 docx 文件转为 HTML，type 就是源文件的 type，这个例子里就是 docx

必须签名才能调用成功，签名见华为签名规则：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

支持多种文件格式，具体如下（type 可传入如下格式）：

PDF 文档: pdf

微软 Office 文档: doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx

WPS 文档: wps, wpt, dps, dpt, et, ett

苹果 iWork 文档: pages, key, numbers

开放版式文档: ofd

电子刊物: caj

电子书: epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu

Markdown 格式: md

SVG 格式: svg

CAD 文档: dwg, dxf, dwt, dws

Sketch 文档: sketch

网页文件: html, htm, mht, eml

图片文件: 几乎所有格式例如 png, jpg, jpeg, gif, tif, tiff, bmp, **可以统一传 img, 代表一切图片**

文本文件: txt, rtf, java, js, c, cpp, jsp, css, xml, properties, log, 其他任意文本文件都可以传 txt

### 返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

### 返回示例(成功状态):

```
{
  "code": 10000,
```

```
"msg": "",
"result": {"token": "xxx"}
}
```

### 返回示例(失败状态):

```
{
  "code": 40001,
  "msg": "ParmNotRight"
}
```

## 文档转换 POST

直接将单个文档 POST 到服务器，大小限制 12M

### 请求参数:

参数	类型及范围	备注	是否必须发送
file	file	要转换的文档，Content-Type 使用 multipart/form-data，最大 12M	是
type	string	要转换的文档扩展名	是
excelislandscape	int	如果是 Excel 文件，是否横屏，默认 0 否（竖屏），1 是	否
exceliscenter	int	如果是 Excel 文件，是否横竖居中，默认 0 否，1 是	否
excelmargin	int	如果是 Excel 文件，四边的边距，默认 10px，单位是像素	否
excelsheetindex	int	如果是 Excel 文件，指定转换的 Sheet 索引，默认 0：全部，第一个 sheet 就是 1，以此类推	否
excelnotshowgridlines	int	如果是 Excel 文件，不显示网格线，默认 0 显示，1 不显示	否
exceluseprintarea	int	如果是 Excel 文件，是否使用打印区域，默认 0 不使用，1	否

参数	类型及范围	备注	是否必须发送
		使用	
excelpagesize	int	如果是 Excel 文件，设定页面大小，默认 A4。0:A4，1:A3，2:A4，3:A5，4:B4，5:B5，6:Letter，7:Legal，8:Tabloid，9:Ledger	否
wordshowmarkup	int	如果是 Word 文件，是否显示审阅标记，默认 0 不显示，1 显示	否
powerpointoutputtype	int	如果是 PPT 文件，设定导出样式，默认幻灯片。0:幻灯片，0 以上是讲义模式：1:每页一个幻灯片，2:每页二个幻灯片，3:每页三个幻灯片，4:每页四个幻灯片，5:每页六个幻灯片，6:每页九个幻灯片	否
powerpointhandoutorder	int	如果是 PPT 文件，当设置为讲义时，设定顺序，默认 0 水平，1 垂直	否
powerpointhandoutorientation	int	如果是 PPT 文件，当设置为讲义时，输出文件的方向，默认 0 不改变，1 横向，2 纵向	否
imageocr	int	如果是图片文件，是否识别图中文字并且在 HTML 中可选可搜索文字，默认 1 是，0 否	否
imagedeskew	int	如果是图片文件，是否将斜的文字矫正，默认 0 否，1 是	否
imageclean	int	如果是图片文件，是否清除图像背景只显示文字，默认 0 否，1 是	否
language	int	OCR 识别语言选项，默认 2 简体中文： 1: 英语 2: 简体中文	否

参数	类型及范围	备注	是否必须发送
		3: 繁体中文 4: 法语 5: 德语 6: 意大利语 7: 俄语 8: 日文 9: 韩文 10: 西班牙语 11: 葡萄牙语 12: 丹麦语 13: 荷兰语 14: 芬兰语 15: 挪威语 16: 瑞典语 17: 土耳其语	
watermark	string	添加水印，字符个数最大 10 个	否
outfilename	string	生成的文件的文件名，默认随机	否
callbackurl	string	回调 URL，转换结束后，会回调该 URL，详细见 <a href="#">回调 URL</a>	否

### 请求示例：

`http://all2html.apistore.huaweicloud.com/v1/convert_post`

Header 中的 Content-Type 必须是 multipart/form-data

必须签名才能调用成功，签名见华为签名规则：

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

支持多种文件格式，具体如下（**type** 可传入如下格式）：

PDF 文档：pdf

微软 Office 文档：doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx

WPS 文档：wps, wpt, dps, dpt, et, ett

苹果 iWork 文档：pages, key, numbers

开放版式文档：ofd

电子刊物: caj

电子书: epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu

Markdown 格式: md

SVG 格式: svg

CAD 文档: dwg, dxf, dwt, dws

Sketch 文档: sketch

网页文件: html, htm, mht, eml

图片文件: 几乎所有格式例如 png, jpg, jpeg, gif, tif, tiff, bmp, **可以统一传 img, 代表一切图片**

文本文件: txt, rtf, java, js, c, cpp, jsp, css, xml, properties, log, 其他任意文本文件都可以传 txt

### 返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

### 返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "result":{"token":"xxx"}
}
```

### 返回示例(失败状态):

```
{
  "code":40001,
```



```
"msg": "ParmNotRight"
}
```

## 多张图片转换 POST

将多张图片转换为 html，url 传入多张图片的地址，传入非图片格式无效

Header 中的 Content-Type 传入 application/json

POST Body 传入 JSON，格式如下：

### 请求参数 BODY:

参数	类型及范围	备注	是否必须发送
url	string 数组	要转换的图片 URL 数组，支持 http(s)，ftp 开头	是
ocr	int	是否识别图中文字并且在 HTML 中可选可搜索文字，默认 0 否，1 是	否
deskew	int	是否将斜的文字矫正，默认 0 否，1 是	否
clean	int	是否清除图像背景只显示文字，默认 0 否，1 是	否
language	int	OCR 识别语言选项，默认 2 简体中文： 1: 英语 2: 简体中文 3: 繁体中文 4: 法语 5: 德语 6: 意大利语 7: 俄语 8: 日文 9: 韩文 10: 西班牙语 11: 葡萄牙语 12: 丹麦语 13: 荷兰语 14: 芬兰语 15: 挪威语	否

参数	类型及范围	备注	是否必须发送
		16: 瑞典语 17: 土耳其语	
watermark	string	添加水印, 字符个数最大 10 个	否
outfilename	string	生成的文件的文件名, 默认随机	否
callbackurl	string	回调 URL, 转换结束后, 会回调该 URL, 详细见回调 URL	否

### 请求示例:

`http://all2html.apistore.huaweicloud.com/v1/convert`

传入的 Body 为 JSON 格式, 如下:

```
{ "url": [ "http://xxx/xxx1.png", "http://xxx/xxx2.png" ] }
```

必须签名才能调用成功, 签名见华为签名规则:

<https://support.huaweicloud.com/usermanual-apig/apig-ug-0011.html>

例如要把多张图片 OCR, 变为文字可选的 HTML, 并且将斜的文字矫正, 将图片背景去除, 那么 JSON 就是:

```
{ "url": [ "http://xxx/xxx1.png", "http://xxx/xxx2.png" ], "ocr": 1, "deskew": 1, "clean": 1 }
```

支持几乎所有图片格式

### 返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
token		string	是	用于 query 接口

### 返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "result":{"token":"xxx"}
}
```

### 返回示例(失败状态):

```
{
  "code":40001,
  "msg":"ParmNotRight"
}
```

## 查询 QUERY

### 请求参数:

参数	类型及范围	备注	是否必须发送
token	string		是

### 请求示例:

<https://all2pdf.apistore.huaweicloud.com/v1/query?token=7b799e09e0838919d3ae63d0566683a2>

无需签名, 无调用次数限制

由于转换需要时间, 文件越大页数越多, 转换越久, 故需要**轮询**查询接口来获得结果。查询频率可以是 1s 一次, 也可以更长一些。查询后先看 status, 如果是 Done 或 Failed, 则转换结束, 停止轮询。如果是 Doing 或 Pending, 则继续轮询。

### 返回数据结构:

名称	含义	类型及范围	是否必须返回	备注
code		number	是	10000: 请求成功
msg		string	是	
token		string	是	请求的 token
result		Dictionary	否	成功后返回

result:

名称	含义	类型及范围	是否必须返回	备注
status	状态	string	是	Pending: 还未开始 Doing: 正在转换

名称	含义	类型及范围	是否必须返回	备注
				Done: 转换成功 Failed: 转换失败
progress	进度	number(0.00 - 1.00)	否 (status 为 Doing 时返回)	比如 0.88 表示 88%
fileurl	html 文件地址	string	否 (status 为 Done 时返回)	转换出来的 HTML 地址, http 和 https 都支持。有效期 60 秒。
count	总页数	integer	否 (status 为 Done 时返回)	HTML 页面总数
reason	失败原因	string	否 (status 为 Failed 时可能返回)	转换失败的原因

#### 返回示例(成功状态):

```
{
  "code":10000,
  "msg": "",
  "token": "xxx",
  "result":
  {
    "progress":0.02,
    "status":"Doing"
  }
}
```

```
{
  "code":10000,
  "msg": "",
  "token": "xxx",
  "result":
  {
    "status":"Done",
    "fileurl":"https://file.duhuitech.com/o/7b799e09e0838919d3ae63d0566683a2/cc9c7f1e-03f8-4742-bc37-aab9da191c26.html",
    "count":10
  }
}
```

### 返回示例(失败状态):

```
{  
  "code":40000,  
  "msg":"No such token"  
}
```

注意:

- 上传文件大小 **不能超过 1000M**。
- 转换完成后, 在 **1 小时** 内下载文件。

## 回调 URL:

用途: 客户可以自行部署服务器, 系统转换结束后会调用客户提供的回调 URL, 直接发送转换结果, 从而无需再轮询 Query。

当设置了回调 URL, 转换结束后 (无论成功失败), 系统都会尝试调用该 URL, 具体如下:

以 POST 方式调用该 URL, Header 头中 Content-Type: application/json

Body 为 JSON 格式, 内容和 Query 的结果相同, 例如:

```
{"code":10000,"msg":"","token":"xxx","result":{"status":"Done","fileurl":"https://file.duhuitech.com/o/7b799e09e0838919d3ae63d0566683a2/cc9c7f1e-03f8-4742-bc37-aab9da191c26.html","count":10}}
```

服务端收到该 POST 后需在 10 秒内返回 HTTP STATUS CODE 200, 视为调用成功, 否则系统认为回调失败, 会再次尝试。规则如下:

系统共计最多会调用 3 次回调 URL, 如果第一次失败, 则等待 3 秒后尝试第二次, 如果第二次失败, 则等待 5 秒后尝试第三次, 如果第三次失败, 则不再尝试。

回调 URL 超时时间 10 秒。

## 错误码表:

JSON 里返回的 code	错误信息
40000	通用错误
40001	参数错误
40002	参数不符合规范