

文档转HTML（阿里云市场）

在线试用：

<https://try.dhconvert.com/>

小程序搜索：

度慧文档转换

概述

使用流程

由于转换需要时间，文件越大页数越多，转换越久，故采用**异步**的方式获得转换结果。即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. 定时轮询结果，调用“查询结果接口”。详细见：[查询结果](#)
2. 回调，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](#)

调用转换接口

v1和v2接口功能相同，调用方式不一样，可自选合适的。

v1接口包括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_v1](#)
- 单一文档转为HTML，文档POST到服务器，用HTTP POST Form Data方式，见：[文档转换POST_v1](#)
- 多个图片转为HTML，文档是多个下载链接，用HTTP POST JSON方式，见：[多张图片转换POST_v1](#)

v2接口统一为HTTP POST JSON：

- JSON支持输入：文件url；文件Base64字符串；多张图片的url。见：[文档转换_v2](#)
- v2版本的API除了异步，还支持同步调用，见：[同步调用](#)
- v2版本中的参数如果未出现在v1版本中，v1版本用同样参数也能工作

调用转换API需要签名，详细见文档附录：[阿里签名](#) 调用[查询结果](#)API无需签名。

阿里云支持从OSS内网直接下载文件，节约流量，见：[阿里云独有部分](#)

文档转换_v2

异步url:

```
https://all2html.market.alicloudapi.com/v2/convert_async
```

同步url:

```
https://all2html.market.alicloudapi.com/v2/convert_sync
```

HTTP方式: POST

Header中的Content-Type传入application/json

Body是JSON格式, 支持以下3种输入源文件的方法:

方法1: 文件url, 最大1500M:

```
{"input": ["http://xxx.docx"]}
```

方法2: 文件Base64字符串, 注意需要传入type (源文档类型), Base64字符串最大8M:

```
{"input": ["base64字符串"], "type": "docx"}
```

方法3: 多张图片url, 注意只能是图片.

```
{"input": ["http://xxx.jpg", "http://xxx.png"]}
```

必须签名才能调用成功, 签名见[阿里签名规则](#):

<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/use-cases/call-apis>

支持以下源文件格式:

类型	扩展名 (type取值)	备注
PDF文件	pdf	
微软Office文档	doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx, csv	
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, nh, kdh	
电子书	epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu	
Markdown	md	
SVG	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, webp, ai等	type可以统一传  , 代表一切图片

类型	扩展名 (type取值)	备注
文本文件	txt, rtf, java, js, c, cpp, jsp, css, xml, properties, log,	type可以统一传 txt ，代表一切文本
网址网页	url	网址，例如： https://www.duhuitech.com

自定义参数:

参数	类型	备注	默认值
type	string	要转换的原始文件扩展名，如果不传，则取url中的扩展名， 注意如果无法获取扩展名会导致转换异常或出错	空
输出HTML相关			
outline	int	如果有大纲，是否生成大纲，默认0：不含大纲，1：含大纲	0
embhtml	int	如果是Word文件，生成嵌入式的HTML代码，适用于放入网页编辑器等场景	0
outfilename	string	生成的文件的文件名，默认随机	空
输出水印相关			
watermark	string	添加水印，字符个数最大10个	空
watermarkfontsize	int	水印的字体大小，默认24pt	24
watermarkfontcolor	string	水印的颜色，输入颜色码，默认黑色#000000，必须7位	#000000
watermarkfontalpha	int	水印的透明度，取值范围1-100，越小越透明，默认20	20
watermarkstyle	int	水印的样式，默认0：文档中央一个水印；1：文档铺满水印	0
输入Office文件相关（源文件为Word，PPT，Excel，WPS等）			
excelislandscape	int	如果是Excel文件，是否横屏，默认0否（竖屏），1是	0
exceliscenter	int	如果是Excel文件，是否横竖居中，默认0否，1是	0
excelmargin	int	如果是Excel文件，四边的边距，默认10px，单位是像素	0
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	0
excelnotshowgridlines	int	如果是Excel文件，不显示网格线，默认0显示，1不显示	0
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，2不使用且只显示有内容的区域	0
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	0
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	0
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	0
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	0
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	0
pageorientation	int	如果是Word，TXT，Excel，HTML，设定页面横向还是竖向，此参数会覆盖excelislandscape。默认0不变，1横向，2 竖向	0
输入图片文件相关（源文件jpg，png等）			

ocr	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	0
language	int	如果开启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	2
dewarp	int	是否切边矫正，默认0否，1是。 如果打开，每次只允许传入一张图（图片分辨率短边大于20，长边小于10000）	0
deskew	int	如果是图片文件，是否将斜的文字矫正，默认0否，1是	0
clean	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	0
grayimage	int	如果是图片文件，是否把图片变为灰度图，默认0否，1是	0
text	int	如果开启ocr，是否使用矢量文字替换图片内文字，使得即使图片中的文字模糊，HTML放大后文字仍然清晰。默认0否，1是	0
split2p	int	如果开启ocr，横向图片若有左右两部分（常见于试卷），分割为2页。默认0:否，1是	0
imagesize	int	统一每页为固定宽，默认0否，输入数字即每页宽度，最大4096。	0
网址网页相关（type传url, html, md等）			
ismobileurl	int	如果type是url,html,md，是否移动端显示。默认0：桌面端网页，1：移动端网页	0
urldesktopwidth	int	如果type是url,html,md，桌面端显示。设定网页最大的宽度，默认：1440，最大：1920	0
htmlpagemargin	string	如果type是url,html,md，生成的页面边距。输入的是字符串，按照顺序：left top right bottom。单位可以是：px, in, cm, mm。例如左上右下分别是1px, 2px, 3px, 4px，传的字符串就是：1px 2px 3px 4px。默认值：左右0，上下各1cm	空
urlwait	int	如果type是url，停留一段时间再抓取页面，单位秒。默认0：不停留，最大30。比如10就是延迟10秒	0
输入EPUB相关（type传epub等）			
epubfontsize	int	如果type是epub等，输出的字体大小（单位pt，磅）。默认0自动	0
epublineheight	int	如果type是epub等，输出的行高（单位百分比）。例如120表示行高是字体大小的120%。默认0自动	0
epubpagesize	string	如果type是epub等，输出的页面自定义大小（单位厘米）。 此参数如果设置会覆盖pagesize。 默认空自动。例如7.2x15.5	空
epubmargin	string	如果type是epub等，输出的页面边距（单位pt，磅）。默认空自动。页边距格式：左 上 右 下，用空格分隔。例如5 5 5 5	空
输入CAD文件相关（源文件dwg, dwf）			

cadlayer	int	如果是CAD文件，是否生成Layer层，默认0否，1是	0
cadisplay	int	如果是CAD文件，是否按照Display显示，默认0不按照，1按照Display	0
cadquality	int	如果是CAD文件，生成的文件品质，默认3。取值从1-5品质从低到高。	3
caddetectempty	int	如果是CAD文件，是否删除空页，默认0不删除，1删除。	0
其他			
password	string	源文件的密码，支持有密码的PDF，Word，PPT，Excel文件类型，默认空无密码	空
callbackurl	string	回调URL，转换结束后，会回调该URL，详细见 回调URL	空

请求示例:

- 例1: 把word文件转为HTML

```
{"input": ["http://xxx.docx"]}
```

- 例2: 把2张图片转为HTML，并识别图中文字，去除图片，转为矢量文字可选的HTML

```
{"input":["http://xxx.jpg", "http://xxx.png"], "ocr":1, "clean":1, "text":1}
```

- 例3: 把Excel文件转为HTML，并只包含有内容的单元格，加上水印"度慧科技"

```
{"input":["http://xxx.xlsx"], "exceluseprintarea":2, "watermark":"度慧科技"}
```

返回数据结构【异步】：

名称	类型	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	类型	是否必须返回	备注
token	string	是	用于 查询结果 接口

返回示例(成功状态)【异步】：

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

返回示例(失败状态)【异步】：

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

同步调用

同步调用的最大返回时间是60秒，如果60秒内转换结束则直接返回结果。否则会返回token，之后和异步方式一样可以调用[查询结果](#)接口查询该token的转换结果。所以同步调用如果传入的文件过大，无法保证在60秒内结束，则转为异步流程。

返回数据结构【同步】：

名称	类型	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	是否必须返回	备注
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
count	总页数	integer	否（status为Done时返回）	页面总数
filesize	文件大小	integer	否（status为Done时返回）	文件大小
status	状态	string	是	Done：转换成功 Failed：转换失败

返回示例(成功状态)【同步】：

```
{
  "code": 10000,
  "msg": "",
  "result": {
    "count": 1,
    "filesize": 17747,
    "fileurl": "https://file.duhuitech.com/o/xxx/xxx.html",
    "status": "Done"
  },
  "token": "xxx"
}
```

返回示例(超时状态)【同步】：

```
{
  "code": 40500,
  "msg": "Timeout, query token later",
  "token": "xxx"
}
```

查询结果

请求参数:

参数	类型	备注	是否必须发送
token	string	调用转换接口拿到的token	是

请求示例:

```
https://api.duhuitech.com/q?token=xxx
```

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	类型	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	是否必须返回	备注
status	状态	string	是	Pending：还未开始 Doing：正在转换 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
filesize	文件大小	integer	否（status为Done时返回）	输出文件大小
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/xxx/xxx.html",
    "count":10,
    "filesize":17747
  }
}
```

返回示例(失败状态):

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

文档转换GET_v1

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	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	要转换的原始文件扩展名, 如果不传, 则取url中的扩展名, 注意如果获取失败会导致转换异常或出错	否
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使用, 2不使用且只显示有内容的区域	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 16: 瑞典语 17: 土耳其语	否

参数	类型	备注	是否必须发送
watermark	string	添加水印，字符个数最大10个，需要URL Encoding	否
watermarkfontsize	int	水印的字体大小，默认24pt	否
watermarkfontcolor	string	水印的颜色，输入颜色码，默认黑色#000000，必须7位，需要URL Encoding	否
watermarkfontalpha	int	水印的透明度，取值范围1-100，越小越透明，默认20	否
watermarkstyle	int	水印的样式，默认0：文档中央一个水印；1：文档铺满水印	否
password	string	源文件的密码，支持有密码的PDF，Word，PPT，Excel文件类型，默认空无密码	否
outline	int	如果有大纲，是否生成大纲，默认0：不含大纲，1：含大纲	否
embhtml	int	如果是Word文件，生成嵌入式的HTML代码，适用于放入网页编辑器等场景	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，详见见回调URL	否

请求示例:

```
https://all2html.market.alicloudapi.com/v1/convert?url=https%3a%2f%2fxxx%2fxxx.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[阿里签名规则](#):

<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/use-cases/call-apis>

支持多种文件格式，具体如下（type可传入如下格式）：

支持多种文件格式，具体如下（type可传入如下格式）：

类型	扩展名 (type取值)	备注
PDF文件	pdf	
微软Office文档	doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx, csv	
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, nh, kdh	
电子书	epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu	
Markdown	md	
SVG	svg	
CAD文档	dwg, dxf, dwt, dws	
Sketch文档	sketch	
网页文件	html, htm, mhtml, eml	

类型	扩展名 (type取值)	备注
图片文件	几乎所有格式例如png, jpg, jpeg, gif, tif, tiff, bmp, webp, ai等	type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,	type可以统一传 txt , 代表一切文本
网址网页	url	网址, 例如: https://www.duhuitech.com

返回数据结构:

名称	类型	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	类型	是否必须返回	备注
token	string	是	用于 查询结果 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_v1

直接将单个文档POST到服务器，大小限制**8M**

请求参数:

参数	类型	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8M	是
type	string	要转换的原始文件扩展名，如果不传，则取file中的文件扩展名， 注意如果获取失败会导致转换异常或出错	否
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使用，2不使用且只显示有内容的区域	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 16：瑞典语 17：土耳其语	否

参数	类型	备注	是否必须发送
watermark	string	添加水印，字符个数最大10个，需要URL Encoding	否
watermarkfontsize	int	水印的字体大小，默认24pt	否
watermarkfontcolor	string	水印的颜色，输入颜色码，默认黑色#000000，必须7位，需要URL Encoding	否
watermarkfontalpha	int	水印的透明度，取值范围1-100，越小越透明，默认20	否
watermarkstyle	int	水印的样式，默认0：文档中央一个水印；1：文档铺满水印	否
password	string	源文件的密码，支持有密码的PDF，Word，PPT，Excel文件类型，默认空无密码	否
outline	int	如果有大纲，是否生成大纲，默认1含大纲，0：不含大纲	否
embhtml	int	如果是Word文件，生成嵌入式的HTML代码，适用于放入网页编辑器等场景	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:

https://all2html.market.alicloudapi.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[阿里签名规则](#):

<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/use-cases/call-apis>

支持多种文件格式，type可传入如下格式：

类型	扩展名 (type取值)	备注
PDF文件	pdf	
微软Office文档	doc, docx, ppt, pptx, xls, xlsx, pot, pps, ppsx, csv	
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, nh, kdh	
电子书	epub, chm, mobi, azw, azw3, fb2, cbr, cbz, djvu	
Markdown	md	
SVG	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, webp, ai等	type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,	type可以统一传 txt ，代表一切文本

返回数据结构:

名称	类型	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	类型	是否必须返回	备注
token	string	是	用于 查询结果 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_v1

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	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 16：瑞典语 17：土耳其语	否
watermark	string	添加水印，字符个数最大10个， 需要URL Encoding	否
watermarkfontsize	int	水印的字体大小，默认24pt	否
watermarkfontcolor	string	水印的颜色，输入颜色码，默认黑色#000000，必须7位，需要URL Encoding	否
watermarkfontalpha	int	水印的透明度，取值范围1-100，越小越透明，默认20	否
watermarkstyle	int	水印的样式，默认0：文档中央一个水印；1：文档铺满水印	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：

```
https://all2html.market.alicloudapi.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[阿里签名规则](#)：

<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/use-cases/call-apis>

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	类型	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	类型	是否必须返回	备注
token	string	是	用于 查询结果 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"
}
```

备注：

- 文件url方式支持文件大小 1500M。
- 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[查询结果](#)。

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[查询结果](#)的结果相同，例如：

```
{
  "code":10000,
  "msg": "",
  "token": "xxx",
  "result": {
    "status": "Done",
    "fileurl": "https://file.duhuitech.com/o/xxx/xxx.html",
    "count": 10,
    "filesize": 17747
  }
}
```

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。

阿里云独有部分：

支持从阿里云OSS内网直接下载文件，目前支持的是上海地区的阿里云OSS内网：

oss-cn-shanghai-internal.aliyuncs.com

文档转换GET或多张图片转换POST里的url地址包含上述域名则自动支持

错误码表：

返回的code如果是10000，代表成功，其余是失败

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
40500	同步调用超时

附录：阿里签名方式

参考链接：

<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/use-cases/call-apis>

在调用API商品时，首先您需要了解采用哪种API认证方式，云市场API商品的认证方式主要有以下两种方式。两种方式可同时使用，您可以根据不同情况来选择。

- 简单身份认证（AppCode）
- 签名认证

简单身份认证（AppCode）

简单认证（AppCode）调用API，有两种方式，一种是将AppCode放在Header中进行调用，一种是将AppCode放在Query参数中进行调用。

方式一：将AppCode放在Header中

在请求Header中添加一个Authorization参数。

Authorization字段的值的格式为APPCODE + 半角空格 + APPCODE值。格式如下：

Authorization:APPCODE AppCode值

示例：

Authorization:APPCODE 3F2504E04F8911D39A0C0305E82C3301

方式二：将AppCode放在Query中

在请求Query中添加AppCode参数（同时支持appcode , appCode , APPCODE , APPCode四种写法）。

AppCode参数的值为AppCode的值。

示例：

<http://www.aliyum.com?AppCode=3F2504E04F8911D39A0C0305E82C3301>

参考链接：<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/user-guide/call-an-api-operation-by-using-an-appcode>

签名认证

比较复杂，推荐用阿里自己的SDK来调用，参考链接：<https://help.aliyun.com/zh/api-gateway/traditional-api-gateway/user-guide/use-digest-authentication-to-call-an-api>