

文档转换中心开发文档

1、文档转换方式调研（使用openoffice为佳）

（1）基于Windows系统的office组件

缺点：只能进行本地文件之间的格式转换，提供的转换接口只基于文件对象之间

如果采用此组件，文档转换中心服务必须先从七牛云存储服务器上根据文件链接地址下载源文件到本地路径目录存储，然后调用基于Windows系统的office组件转换成本地路径下的PDF文件，然后将转换后的PDF文件上传到七牛服务器。

1、开销很大

下载源文件到本地 -> 本地转换成PDF文件 -> 上传转换后的文件到七牛 -> 删除本地源文件目标文件

2、系统限制

依赖操作系统(windows)

文档转换中心服务器必须和office组件必须在一台服务器上

（2）基于openoffice的soffice端口监听方式的格式转换服务组件

优点：

1、提供的转换接口既可以基于文件对象之间，又可以基于Socket输入输出流对象之间 如果采用此组件，完全不用根据文件链接地址下载文件到本地，基于输入输出流就可以完成转换

根据文件链接地址获取输入流 -> 转换成PDF输出流 -> 上传PDF输出流到七牛

2、可以集群方式部署openoffice服务，不依赖特定操作系统

两者对于文件转换效果的支持差异不大，都对Excel文件转换效果支持不好，但对DOC和PPT文档支持效果没有明显差异，因此准备采用openoffice进行文档转换。

2、文档转换中心接口说明

（1）异步转换请求接口

请求参数：

docFileUrl: 待转换文档七牛链接地址或者其它在线链接地址

接口说明：

请求来到,生成任务ID (taskId) 将文件链接地址加入到任务队列中，由工作线程依次从队列中取出进行文档转换。转换过程中的状态（WAITING等待，EXECUTING执行中，SUCCESS转换成功，FAILURE转换失败），任务触发某种状态时，将该种状态存入MONGODB数据库种记录。实时返回客户端改文件转换请求的任务ID，后续可以根据任务ID，查询任务执行情况。

（2）异步转换请求执行状态查询接口

请求参数：

taskId: 由上述异步转换请求接口实时返回的任务ID提供

或者

docFileUrl: 待转换文件链接地址（二者取其一，优先级taskId>docFileUrl）

接口说明：

根据任务Id查询任务执行状态

或者

待转换文件链接地址查询任务执行状态（4种）

WAITING:等待

EXECUTING:正在执行

SUCCESS:转换成功

FAILURE:转换失败

转换成功，转换后的PDF文档链接地址一同返回

（3）同步转换请求接口

请求参数：

docFileUrl：待转换文档七牛链接地址或者其它在线链接地址

接口说明：

请求来到，直接进行文档转换，返回转换后的PDF文档地址。

若转换时间长导致请求超时，可根据待转换文件链接地址查询转换情况（参考第二步）。

3、文档转换中心监控管理页面

1、展示总的转换请求数及列表

2、展示等待转换的请求数及列表

3、展示正在执行的请求数及列表

4、展示执行成功的请求数及列表（耗时情况、文档大小...）

5、展示执行失败的请求数及列表（失败原因、文档大小...）