# UiPath从PDF中提取数据

1. **读取PDF文字（Read PDF Text）的介绍**

Read PDF Text: 从指定的PDF文件中读取相应的字符，并将它们存储在字符串变量中。

**二、方法一: Read PDF Text在UiPath中的使用**

1. 打开设计器, 在设计库中新建一个Sequence，为序列命名及设置Sequence存放的路径, 在Activities中搜索Read PDF Text (如果无法搜索到, 表示你的设计器没有安装PDF相关活动, 在Design选项卡中, 点击”Manage Packages”, 在All Pakcages界面中搜索”Uipath.PDF.”, 选择UiPath.PDF.Activities, 点击右侧”install”, 之后点击”Save”保存即可)，并将其拖至设计区, 在FileName属性中输入要读取的PDF文件路径, 在Range属性值中输入你要阅读的的PDF页面范围, 如果未指定范围，则读取整个文件。你可以指定单个页面（例如“ 7”），页面范围（例如“ 7-12”）或复杂范围（例如“ 2-5、7、15-End”或“All”）。本案例使用默认范围, 即读取整个PDF文件。在Text属性中使用快捷键Ctrl+ K自定义字符串类型变量PDFText, 用来保存读取到的PDF字符串







1. 在Activities中搜索Message Box，并将其拖至设计区, 在Text属性中输入步骤1定义的字符串变量PDFText, 用来查看阅读到的PDF文本字符串. 读取到文本后可以使用Write Text File活动把它写入一个txt文档里，或者进行其他的文本操作



1. **方法二: 使用Screen Scraping工具**

1.在顶部工具栏里选择Screen Scraping工具，在PDF中指定想要抓取的文本元素，UiPath会显示一个预览界面







使用Screen Scraping工具的Full Text方法，最终生成的活动是一个Attach Window容器和一个Get Full Text活动, 同样可以使用Message Box活动输出上述步骤读取的文本内容, 用于后续操作. 注意, 使用Screen Scraping工具必须先打开PDF文件，且不能在后台运行。



方法一方法二的PDF源文件如下图, 高亮部分为图片格式, 其他为PDF可选取的文本



1. **方法三: 使用Get Text工具**

1.如果PDF数据类型是Native Text (本地文本)。那么这种文本元素可以直接被UiPath识别和选中。与之对应的是扫描版的PDF，文本是图片型的文本。使用Get Text活动可提取某个特定文本. 比如, 我们想要对下路所示的PDF文件, 获取PDF文件中Grand Total的值.



2. 在Activities中搜索Get Text，并将其拖至设计区，点击”Indicate on screen”, 选取对应区域的值, 在Value属性中使用快捷键Ctrl+K自定义字符串变量grandTotalValue, 之后使用Message Box活动打印grandTotalValue的值









#运行完成，你会看到UiPath从PDF文件中提取到相应的数据

总结：以上流程就用到了Read PDF Text、Screen Scraping、Get Text从PDF中提取数据的操作，大家可以动手试试