

foxit 福昕

股票代码：688095

“ IDP DAC

版式文档智能解析

让解析结果更精准

“ 产品概述



PDF转JSON



PDF转Markdown

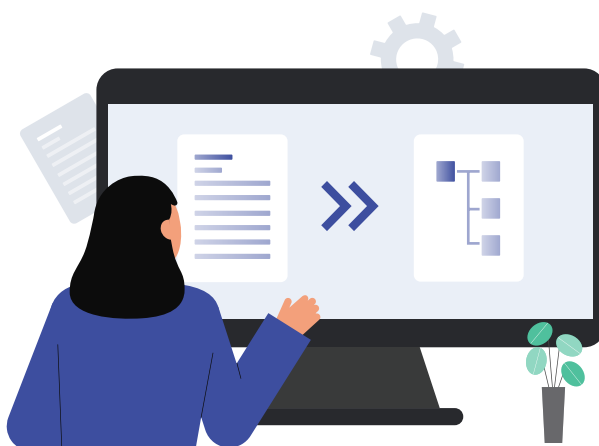


RESTFul APIs

IDP DAC 是IDP产品家族的基础产品，通过对 PDF 页面的版面分析、对象分析、分类提取、内容组合和格式转换，可以为上层应用提供容易识别的 JSON 格式数据或 Markdown 格式数据。

IDP DAC 生成的 JSON 数据包含了标题、段落、表单、表格、页眉、页脚、图形、链接等类型对象的标识和数据，便于上层应用分类使用。

IDP DAC 提供 Web 和 API 两种模式，既可以支持用户通过 Web 交互式操作立即转换，也可以支持其他应用系统通过 API 调用作为上层应用的基础功能。



PRODUCT FEATURES

功能特点



功能一

布局分析，分析识别页面上的标题、段落、表单、表格、页眉、页脚、链接等对象



功能二

OCR，对页面上以图片格式显示的文字、公式图片、矢量图形等进行识别，将识别的结果汇入 PDF 中对应的对象



功能三

数据提取，根据布局分析的结构进行数据提取，在提取过程中根据PDF标准进行校正



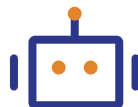
功能四

格式转换，对提取的数据进行格式转换，按照用户需要转成 JSON 或 Markdown 格式，输出给用户或推送到下游系统



功能五

人工介入，对识别不准的地方可以人工介入对参数进行校正，从而得到正确的结果



功能六

智能学习，对于人工介入参数校正的结果进行自动整理，作为样本对识别模型进行训练，让系统越用越好

典型场景

TYPICAL SCENES

○ 场景一

将 PDF 解析后，送入大模型训练平台，训练私域、专有的大模型

○ 场景二

将 PDF 解析后，作为 RAG 的基础数据，构建私域、专有的知识库

○ 场景三

将 PDF 解析后，基于规则进行精准的数据提取

○ 场景四

将 PDF 解析后，基于其语义和阅读顺序，经重构、组合生成新的文档

福建福昕软件开发股份有限公司

电话：010-50951668

邮箱：idp-ml@foxitsoftware.com

网址：www.foxitsoftware.cn



微信扫码
查看产品视频

