

通用资料提取平台根据用户定义的规则集，提取非结构化文档中的信息

### 关键词：

- 金融科技、文档处理、数据提取、工作流自动化、光学字符识别 (OCR)、多语言

### 解决难题

- 传统方法需要手动将数据从文档传输到系统
- 过程耗时并且可能出现拼写错误，尤其是在输入包含大量数据的表格时
- 服务费和人工成本高

为了解决这些问题，应科院研发了用于非结构化文件的通用信息提取平台。应用光学字符识别（OCR）引擎和各种计算机视觉技术，将文档中的文本转换为数字格式，并根据用户需求提取有用信息。该平台可帮助金融和政府机构自动输入银行对账单、土地登记、工资证明等文件。

### 创新点

通用信息提取平台允许用户定义规则集，以提取非结构化文档中的信息，并将结果导出为人类或机器可读的格式。

#### 创新要点：

- 由用户制定规则集，定义如何提取文件的内容
- 提取特定关键字附近的资讯
- 当列名与规则集匹配时提取表格
- 自动工作流程，将结果导出为机器可读格式，以供进一步处理
- 可在客户端安全、快速地处理机密文件
- Material UI 使用户轻松地验证和更新信息

### 主要影响

- 简化的非结构化文档处理流程
- 协助金融和政府机构自动录入文件
- 应用计算机视觉技术提升光学字符识别（OCR）的准确性

### 示例图片



### 项目完成日期

- 2020 及 2021 年

### 应用

- 非结构化文档处理
- 自动输入数据
- 在本地安全地处理文件

### 专利

- 美国申请号 16/823,398,  
中国申请号 202080000398.2 和  
香港申请号 62020017194.5

[ASTRI Patent Search](#)

### 商业合作

- 知识产权许可
- 技术合作开发

[<应科院授权及研发项目检索>](#)