

通用資料提取平台利用用戶建立的規則集，提取非結構化文件的內容

關鍵詞：

- 金融科技、文件處理、數據提取、工作流自動化、光學字符識別 (OCR)、多語言

解決難題

- 傳統方法需要透過人手將文件內容輸入系統
- 過程耗時並會出人為錯誤，特別是在輸入包含大量條目的表格時
- 服務費和人工成本高

針對上述問題，應科院研發用於非結構化文件的通用資料提取平台，利用光學字符識別 (OCR) 以及計算機視覺技術，將文件中的文字轉換為數字格式，並根據用戶的要求提取有用資訊。該平台可協助金融和政府機構，自動輸入銀行賬單、土地登記、工資證明等文件內容。

創新點

通用資料提取平台容許使用者制定規則集，以提取非結構化文件的內容，並將結果輸出為人類或機器可讀的格式。

創新大綱：

- 由使用者制定規則集，以決定如何提取文件的內容
- 提取特定關鍵字附近欄位的資訊
- 提取包含特定欄目的表格
- 將結果輸出為機器可讀的格式，以供進一步處理
- 在客戶端安全、快速地處理機密文件。
- 質感設計介面允許用戶以簡單的方式驗證和更新資訊

主要影響

- 簡化非結構化文件的處理流程
- 協助金融與政府機構自動輸入資料
- 利用計算機視覺技術提升光學字符識別 (OCR) 準確度

示範圖片



項目完成日期

- 2020 及 2021 年

應用領域

- 非結構化文檔處理
- 自動輸入數據
- 在本地安全地處理文件

專利

- 美國申請號 16/823,398,
中國申請號 202080000398.2 和
香港申請號 62020017194.5

[ASTRI Patent Search](#)

商業合作

- 知識產權授權模式
- 合作開發技術