Generic Information Extraction Platform for Unstructured Documents

Generic Information Extraction Platform extract information in unstructured documents based on user defined ruleset

Keywords:
- FinTech, document processing, data extraction, workflow automation, optical character recognition (OCR), multi-language,

Problems addressed

- Conventional approach requires manual data entry to transfer data from document to system
- It is time consuming and may contain typo, especially when entering a vast amount of data into tables
- Service charges and labour costs are high

To address the issues, ASTRI's Generic Information Extraction Platform utilizes OCR engine and various computer vision techniques to convert text in document to digital format and extract useful information based on user's requirements. The platform enables financial and government institutes to automate data entry of bank statements, land registry documents, salary proof and so on.

Innovations

The Generic Information Extraction Platform allows users to define ruleset to extract information from unstructured documents and export result to human or machine-readable formats.

The innovation outline:

- **Ruleset** is defined by user on how to extract the value.
- **Extract Field Values** near field name or specific keyword.
- **Extract Table** when column names matched the ruleset
- **Auto Workflow** enables users to export results to machine readable format for further processing.
- **Client-side processing** enables secure and fast processing of sensitive documents.
- **Material UI** allows users to validate and update results easily.

Key impact

- Streamline the workflow of processing unstructured documents.
- Assist data entry of financial and government documents.
- Advanced computer vision techniques to improve OCR accuracy

Innovation snapshot

Project completed

- 2020 and 2021

Applications

- Unstructured document processing
- Automate data entry
- Locally and securely process documents

Patent(s)

- US App. No. 16/823,398;
  CN App. No. 202080000398.2 and
  HK App. No. 62020017194.5

ASTRI Patent Search

Commercialisation opportunities

- IP licensing
- Technology co-development

Contact details

Director, Commercialisation
Priscilla Yeung
Email: priscillayeung@astri.org
Telephone: (852) 3406 0280