

Abby Cloud OCR Script with ApplicationManager

1. Short Introduction

Original Use Case: <https://community.plunet.com/t/83hkkh6/recommendations-for-ocr-conversion-tasks>

Definition: Having an option to create a job that converts an image-based document using OCR in an editable format. Job shall send data to <https://www.abbyy.com/de/cloud-ocr-sdk/>

2. Application Requirements:

2.1. General Requirements

- It shall be possible to configure the specific parameters for the Cloud API:
 - Server, user name, password
 - Profile type
 - Output file type
 - Allowed image formats
 - Language Mapping for Source Language in Plunet and recognized language in Abby Cloud OCR
- Workflow
 - Starting the application
 - application shall be able to receive source language parameter with ApplicationManager call
 - application shall be able to receive storing location of configuration file with ApplicationManager call
 - Preparation
 - Create a log file in Job _IN Folder
 - Retrieve configuration
 - Retrieve parameters passed along by application manager
 - Getting the relevant files
 - retrieve source files from Order Source Level (language folder)
 - identify files with the allowed image format
 - Create a task in Abby Cloud OCR with the parameters set in the configuration file PLUS the passed recognition language from ApplicationManager plus files
 - Check for task completion
 - Upon Completion, download files
 - rename completed file to originalfilename_OCR.OUTPUTFILETYPE
 - store completed file in Order Source Level (language folder)
 - store completed file in _IN Folder of the Job
 - Delete task in Abby
- Additional requirements
 - it shall be possible to set a debugging level in the config file
 - it shall be possible to have multiple source files
- Currently not implemented:
 - CHECK FILE functionality should be possible since log file is created in Job folder - no option to test yet

2.2. Abby Cloud OCR API

API Documentation: <https://support.abbyy.com/hc/en-us/sections/360004931659-API-v2-JSON-version->

Code Samples: <https://support.abbyy.com/hc/en-us/sections/360004895259-Code-samples>

3. Solution

3.1. Executable File and Config File

To be found in dist folder

Application converted into .exe with PYInstaller

3.2. Source

abby_ocr.py

Application written in Python 3.8


3.3. Configuration

3.3.1. File Location

Executable and configuration file shall be stored on server where Plunet runs.

3.3.2. Plunet Job Configuration

Settings for automatic job

Initial job status	<input type="text" value="03 Assigned - waiting"/>	
1.	<input type="text" value="Launch external application"/>	
	Launch application (UNC)	<input type="text" value="C:\Plunet\App\Abby_OCR\abby_ocr.exe [SrcLang"/>
New	<input type="text" value="--- Please select ---"/>	
Final job status	<input type="text" value="13 Without invoice"/>	

Launch external application

File name of result file (IN folder)	<input type="text"/>
File name of test file (IN folder)	<input type="text"/>
Test interval (sec.)	<input type="text" value="50"/>

3.3.3. Parameters

Example: `C:\Plunet\App\Abby_OCR\abby_ocr.exe [SrcLang]`
`"C:\Plunet\App\Abby_OCR"`

Explanation: EXECUTABLE FILE LOCATION || Plunet Flag for Source Language
|| Location of the configuration File

4. Testing the integration

Create a project and put a pdf or tif file into the language source folder of the order.
Create an Abby OCR Job and run it.

Due to the configuration, it would make sense to choose English as the source language.

Sample File can be obtained from Abby OCR Documentation page.