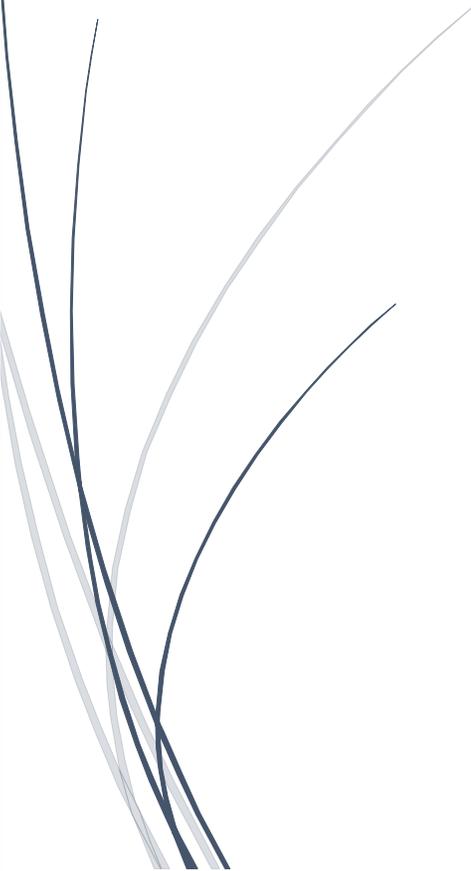




3/20/2023

OCRBot User Guide



For general information or technical issues, submit an incident through ServiceNow.

OCRBot User Instructions

Table of Contents

A.	Introduction.....	2
B.	OCRBot Download & Installation	2
	Minimum System Requirements.....	2
	Steps to Download and Install OCRBot	2
C.	OCRBot Features	4
	Basic.....	4
	Settings Menu	6
	Help Menu.....	7
	OCRBot File Processing.....	7
	Input	7
	Output	7
	Processing Options.....	8
	User Experience.....	8
	Logs.....	9
	Advanced.....	10
	Technical Requirements.....	10
	Steps to Download and Install EverMap.....	10
	ReOCR.....	14
	Layering	15
	OCRBot Remediation and OCRBot Publish Mode	15
D.	Acronyms.....	18
E.	Glossary	18

A. Introduction

This document describes how to download, install and use the OCRBot tool.

B. OCRBot Download & Installation

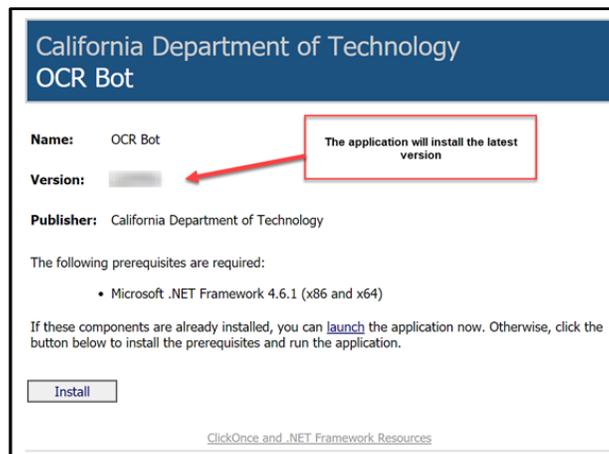
Minimum System Requirements

- Windows 10
- Internet connection
- Ability to download and install software locally
- Microsoft Azure Cognitive Service Endpoint and Key

Steps to Download and Install OCRBot

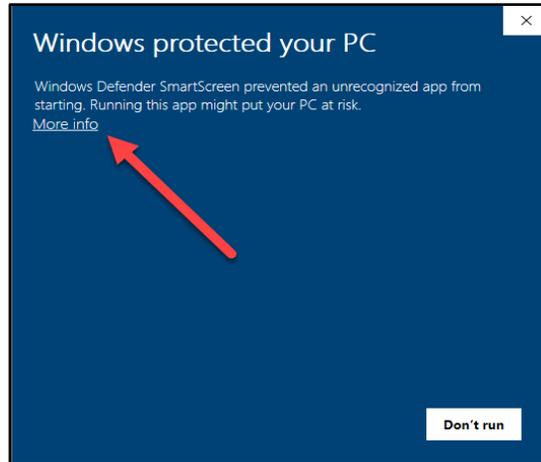
Follow the instructions provided in the encrypted email to download the tool. Once downloaded to your personal computer (PC), follow these steps to install the tool.

1. Click on the “Install” button and the “setup.exe” file will download to your local PC.



Your experience may differ depending on your security settings

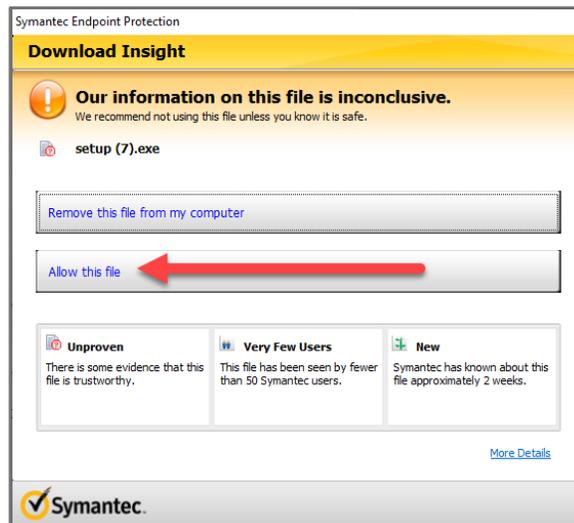
2. Run the “setup.exe” you downloaded.
 - A Windows Defender security prompt may popup stating it is protecting your PC, select “More Info”.



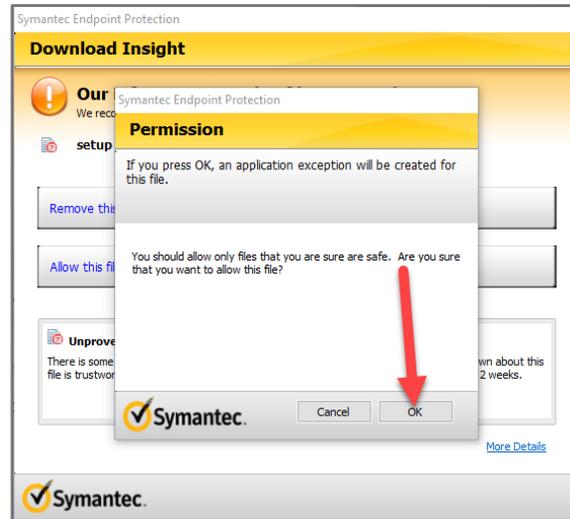
3. Select "Run Anyway".



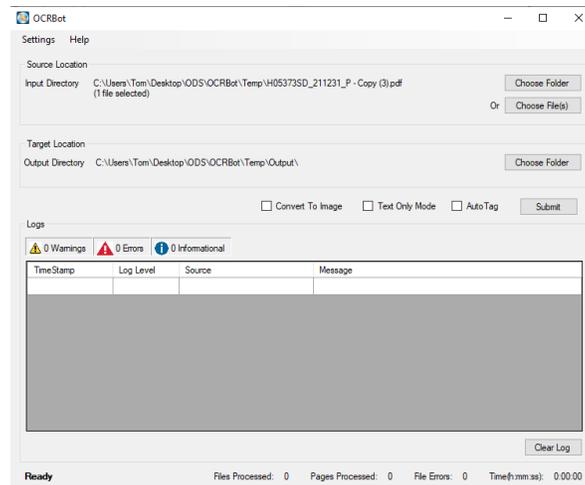
4. You may receive a prompt for "Symantec Endpoint Protection", if present select "Allow this file".



5. A secondary prompt will display, select “OK” to continue.
 - a. Symantec will prompt a second time, please repeat the previous two steps.



6. Once the installation is complete the OCRBot will start.
 - a. Your OCRBot view may be different depending on external dependencies, for example Adobe Acrobat Pro DC and EverMap AutoBatch Plug-in.

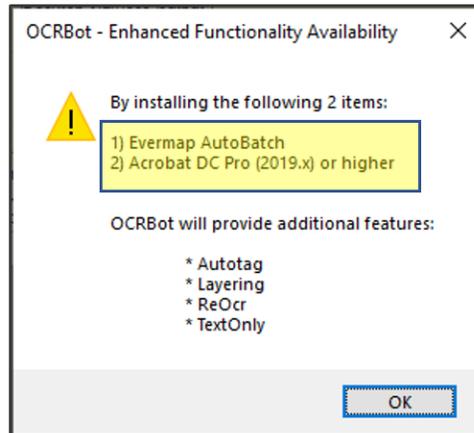


C. OCRBot Features

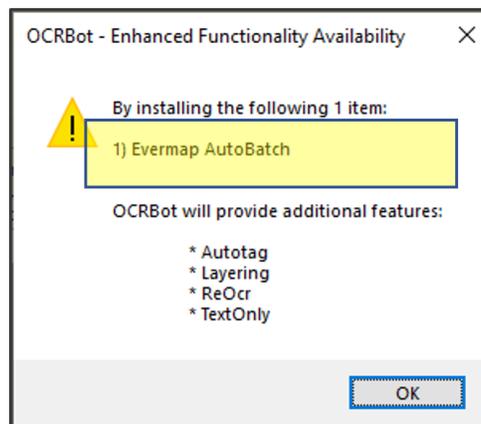
Basic

Without the support tools (Adobe Acrobat Pro DC and EverMap AutoBatch Plug-in), OCRBot will run in basic mode. In basic mode, Auto Tag, Text Only Mode, ReOcr, and layering are not available.

1. A popup will display at startup:

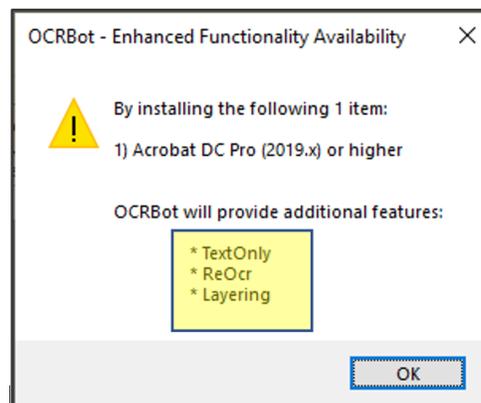


2. If Adobe Acrobat Pro DC (2019.021.20058 or higher) is installed and EverMap is not installed, a popup displays at startup:

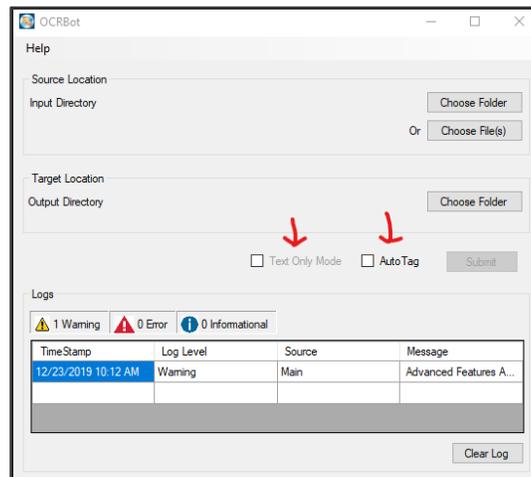


3. If Adobe Acrobat Pro DC (version less than 2019.021.20058) and EverMap AutoBatch Plug-in are installed, Auto Tag will be available.

- A popup will display at startup:



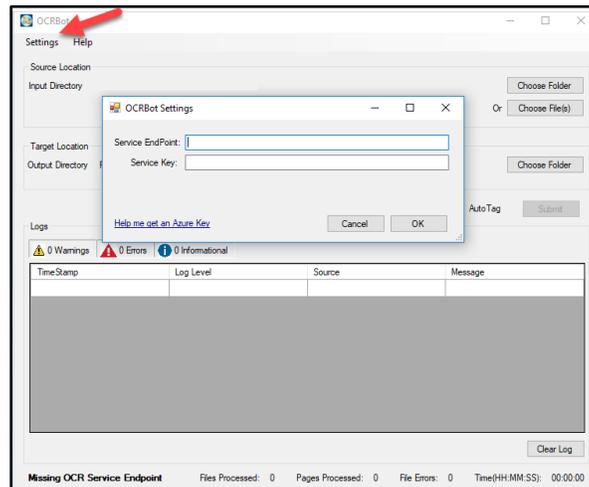
- OCRBot will have “Text Only Mode” disabled and “Auto Tag” enabled:



Settings Menu

If you do not have a Microsoft Azure Cognitive Service Endpoint and Service Key, submit an [OCRBot Service Request](#).

- Click the “Settings”, then click the “OCR Service Endpoint”
- On the “OCRBot Settings” window
 - Enter the Service EndPoint URL on the “Service EndPoint” text field.
 - Enter the CDT Cognitive Service Key or your own Microsoft Cognitive Service Key value on the “Service Key” text field.



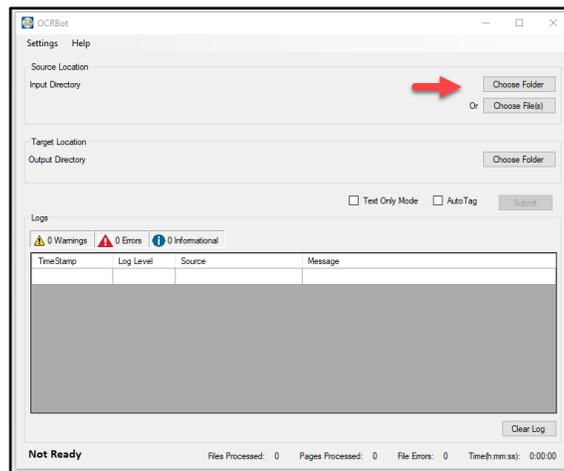
Help Menu

The OCRBot Help Menu will provide access to the latest User Guide, Data Flow, Release Notes, Privacy Policy, FAQs and About. The source code is available for download at the bottom left side under the “About”.

OCRBot File Processing

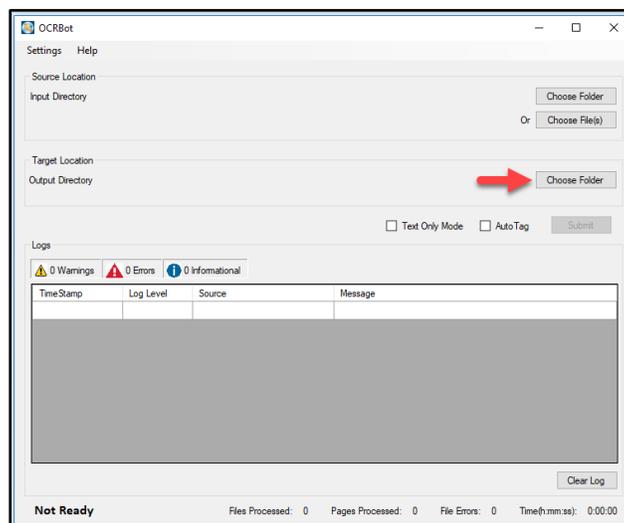
Input

1. Click either “Choose Folder” or “Choose File(s)” to select the input directory location to process the file(s).



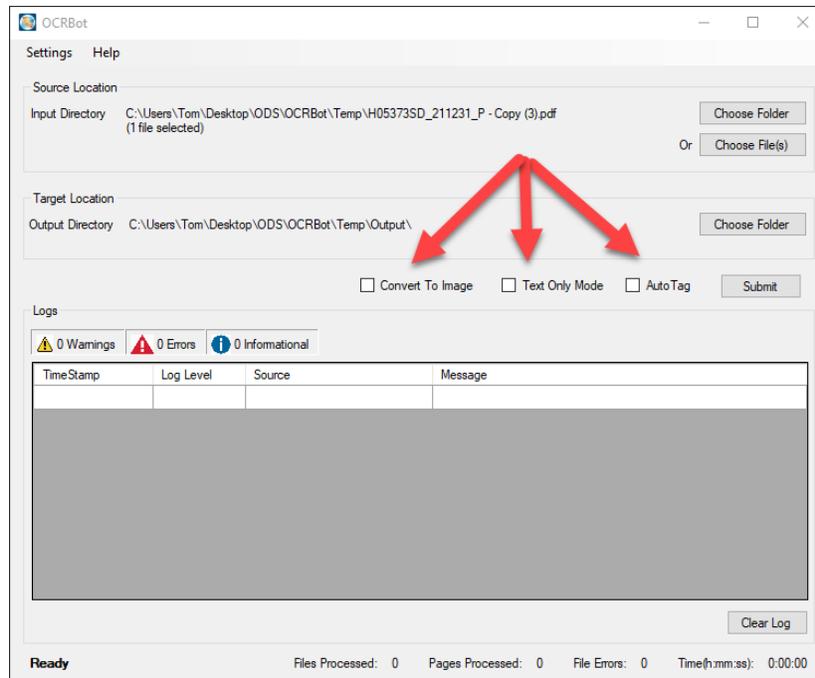
Output

2. Click “Choose Folder” to select an output directory location for the file to be stored after they have been processed. The output folder should be a separate folder from the input folder. Any duplicate file names in the output folder will cause OCRBot to skip processing of that file.



Processing Options

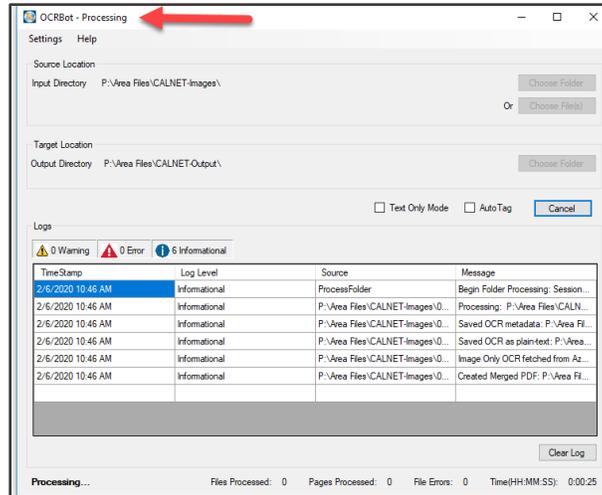
3. Choose processing options then click Submit.
 - a. **Convert To Image** – Convert the PDF to an image to prepare for OCR processing (useful if the document contains unnecessary information like tags that are causing problems)
 - b. **Text Only Mode** – Output PDF will have all text visible and all images will be hidden.
 - c. **Auto Tag** – Apply the Adobe Acrobat Pro DC Auto Tag to the output PDF.



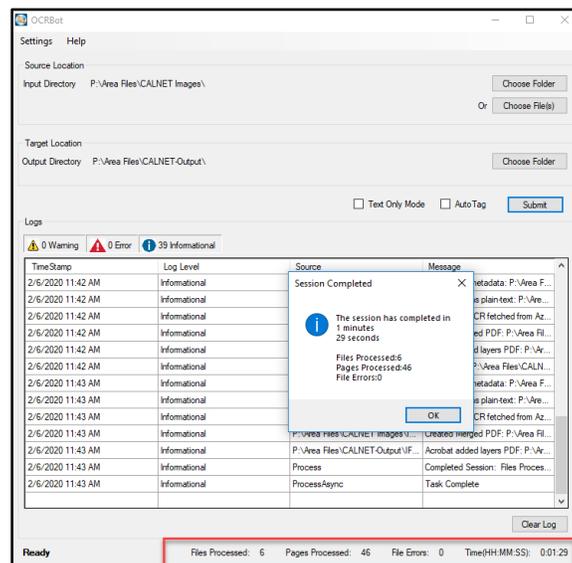
The application will preserve all settings (Source Location, Target Location, Service Endpoint and Service Key, Text Only Mode and Auto-Tag check boxes).

User Experience

1. When the OCRBot is processing, the word “Processing” is displayed.
2. The “Submit” button changes to “Cancel”, and the file and folder selection buttons are disabled.
 - a. At any time a batch can be cancelled and restarted.

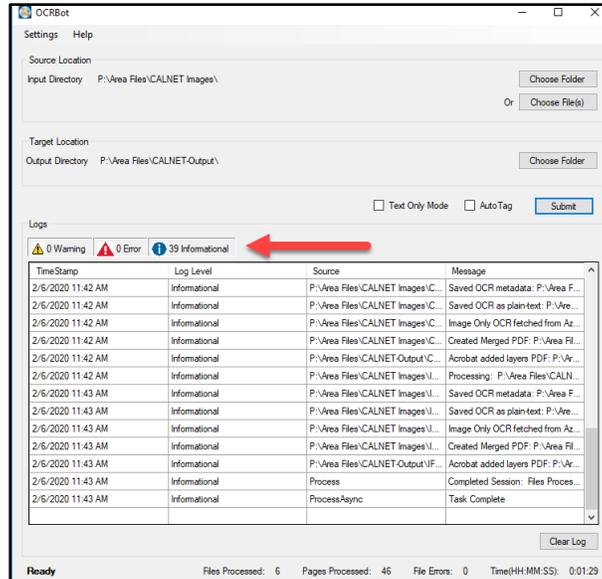


3. Once processing is complete, a “Session Completed” popup window displays the process information.
 - The last entry in the log will be “Task Complete”.
 - At the bottom of the screen the number of Files processed, number of Pages Processed, File Errors, and the Time (HH:MM:SS) to process the files is displayed.



Logs

1. **Warning** – Generated when a document is not fully processed.
2. **Error** – Generated when a processing failure occurs.
3. **Informational** – General process steps.
4. **Clear Log** – Clears all log entries from the Log Window and starts a new log output file.



Advanced

Technical Requirements

With Adobe Acrobat Pro DC (version \geq 2019.021.20058) and

EverMap Autobatch Plug-in, OCRBot will provide additional options and features:

- Autotag
- TextOnly
- Convert to Image
- ReOCR
- Layering
- OCRBot Remediation Mode
- OCRBot Publish Mode
- CDT will provide the EverMap AutoBatch link via encrypted email to download the Plug-in for advanced feature functionality.

Steps to Download and Install EverMap

Follow the instructions below after clicking the link to download the “SetupAutoBatchsl.exe” file.

Contact your department technical support staff if you are unable to install software with your login credentials on your local PC.

1. Click “Allow this file” when prompted.

Symantec Endpoint Protection

Download Insight

 **Our information on this file is inconclusive.**
We recommend not using this file unless you know it is safe.

 **setupautobatchsl (2).exe**

[Remove this file from my computer](#)

[Allow this file](#)

 Unproven There is some evidence that this file is trustworthy.	 Very Few Users This file has been seen by fewer than 5 Symantec users.	 Very New Symantec has known about this file approximately 5 days.
------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------

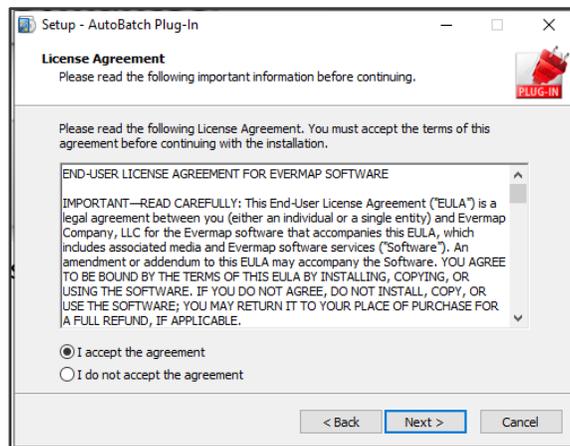
[More Details](#)

 Symantec.

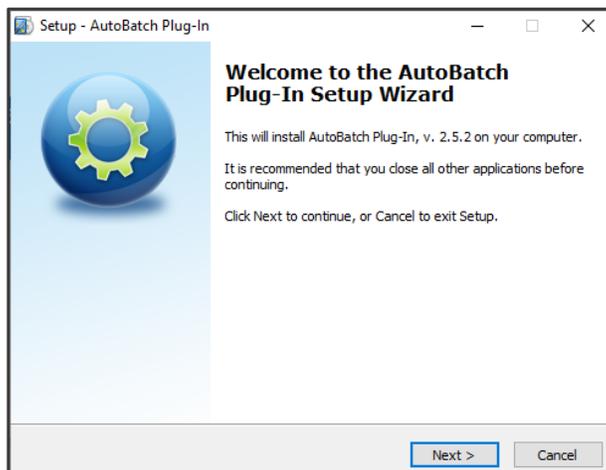
2. Click "OK" when prompted.
3. Click "Next" at the introduction screen to the setup wizard.



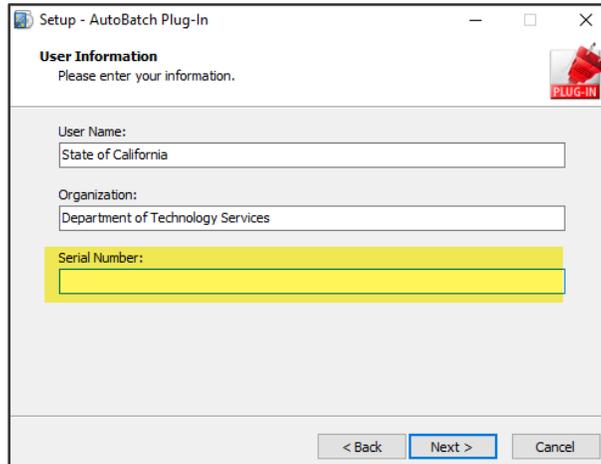
4. Click on the radio box "I accept the agreement", then click "Next" when prompted.



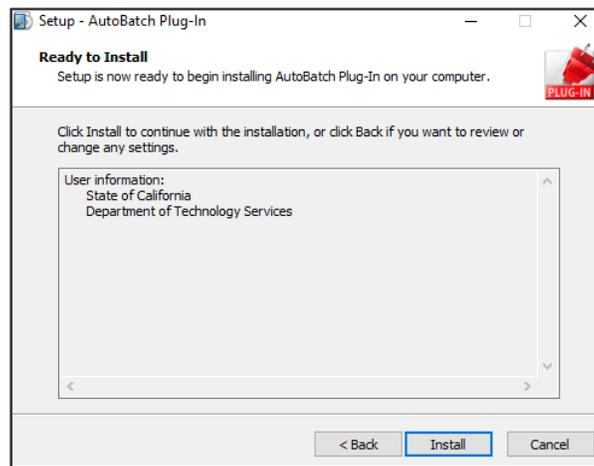
5. Click "Next" again.



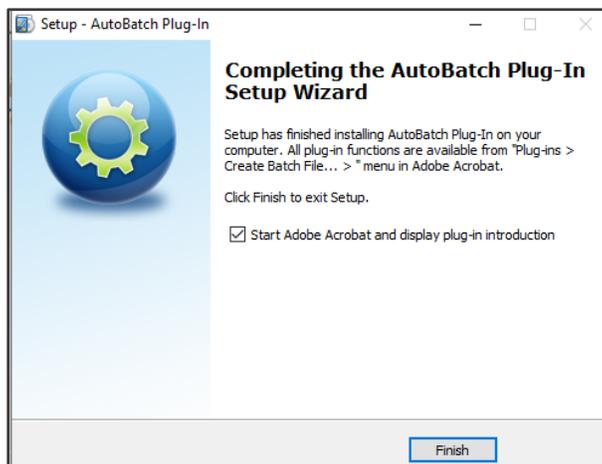
6. Enter the Serial Number provided by CDT via encrypted email. Click “Next” when prompted.



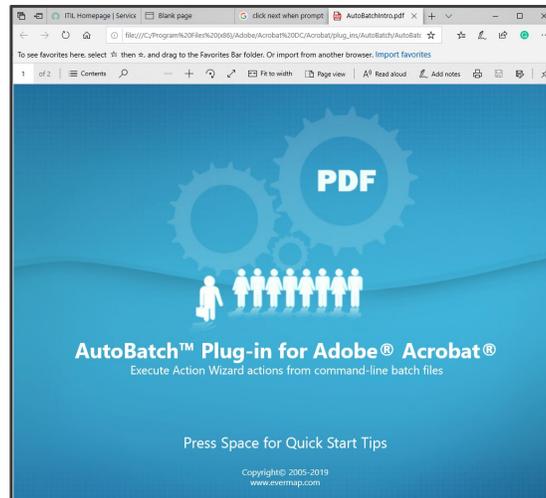
7. Click “Install” when prompted.



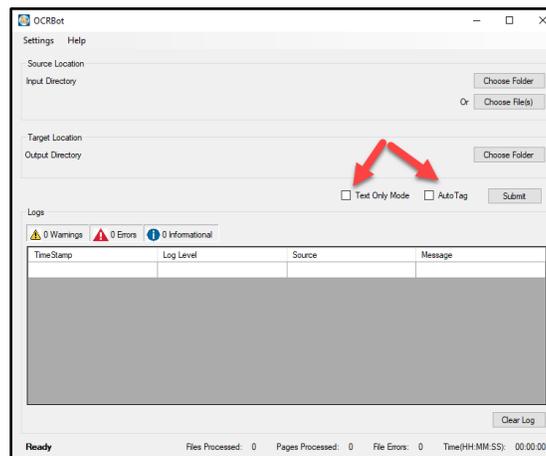
8. Click “Finish” when prompted.



9. Once the installation is complete, the program will present the following Quick Start Tips web page:



10. Once the OCRBot support tools (Acrobat Pro DC with EverMap AutoBatch Plug-in) are installed, the Advanced Options will be available; no further configuration necessary.



ReOCR

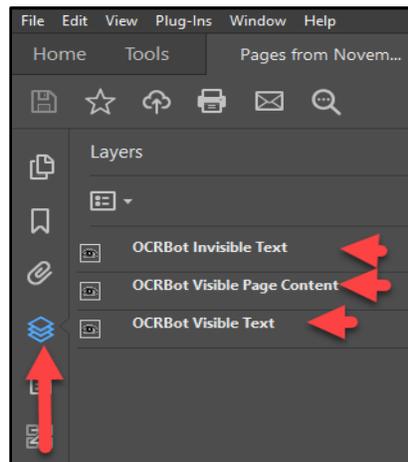
Previous text processed through another OCR will be replaced with the new text from Azure Cognitive Services.

Layering

The visibility of the three (3) layers can be toggled in Adobe Acrobat Pro DC:

- **OCRBot Invisible Text** – Text that has been inserted into the PDF by OCRBot or other similar tools.
- **OCRBot Visible Text** – Other document text.
- **OCRBot Visible Page Content** – Visible content like Images.

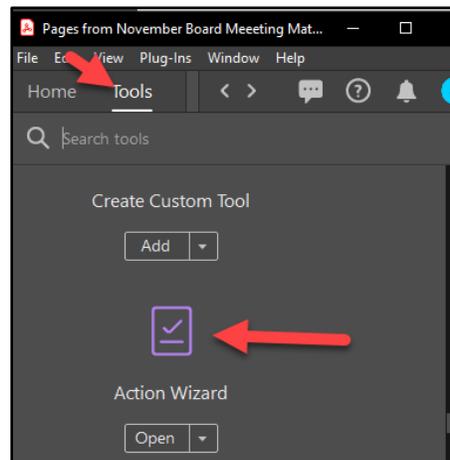
1. Layering feature of Adobe Acrobat used by OCRBot.



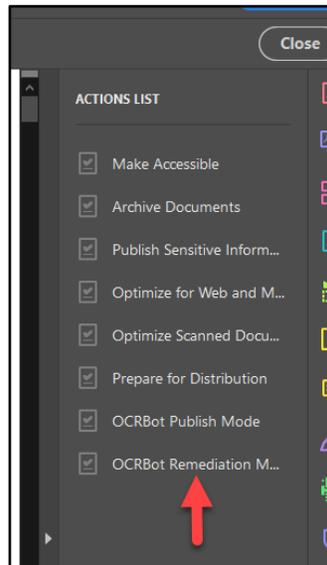
OCRBot Remediation and OCRBot Publish Mode

Requirements: Adobe Acrobat Pro DC >= 2019

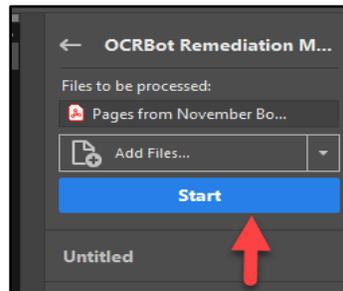
1. Select Tools tab, and Action Wizard:



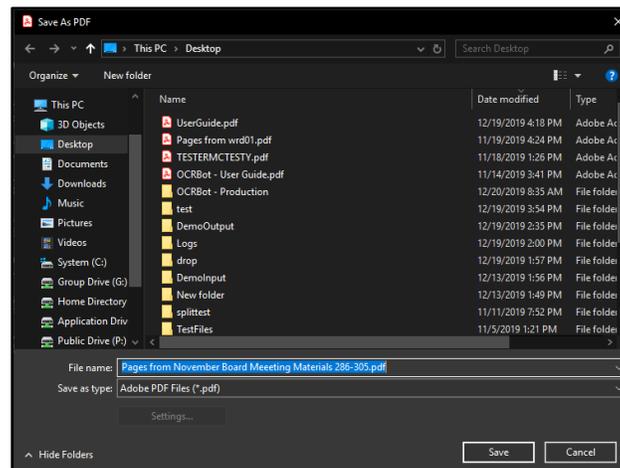
2. Select “OCRBot Remediation Mode” from the Action Lists.



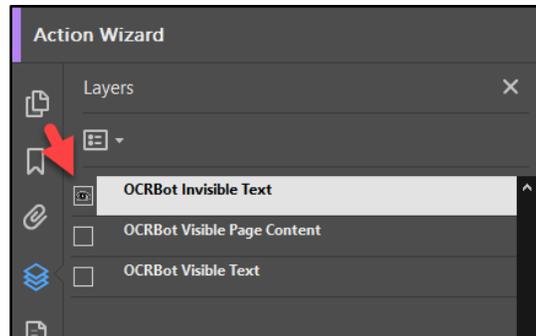
3. Click Start to begin document remediation.



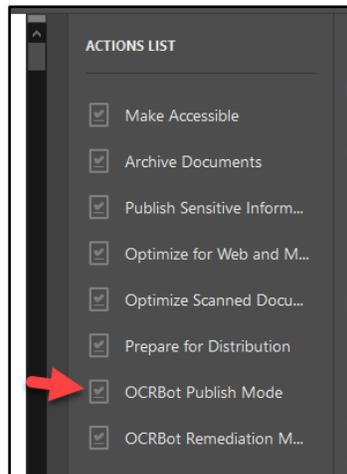
4. At the Save As Dialog, click Save to save the file.



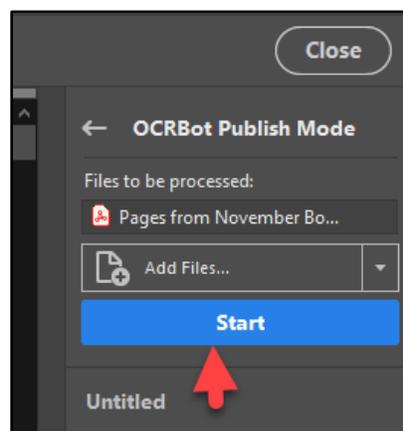
5. To navigate to the Layers choose View → Hide/Show → Navigation Panes → Layers
 - a. Click on the “eye” icon to toggle each layer.



6. When remediation is complete, click the “OCRBot Publish Mode”.



7. Click Start to save your updated file.



D. Acronyms

ACCRONYM	DEFINITION
ADA	Americans with Disabilities Act
API	Application Programming Interface
CDT	California Department of Technology
OCR	Optical Character Recognition
PDF	Portable Document Format

E. Glossary

TERM	DEFINITION
Autotag	A feature in Adobe Acrobat DC that attempts to automatically tag a PDF for accessibility.
Bot	An automated program.
Cognitive Service Endpoint & Key	The Cognitive Service Endpoint is the location of the service. The Cognitive Service Key is the credential to access the service.
EverMap Auto Batch Plug-in	A plug-in that can automatically run Adobe Acrobat features such as Autotag.
EverMap Serial Number	The serial number required to unlock and use EverMap AutoBatch Plug-in.
Layering	Information is stored on different layers in a PDF document to control what is visible and/or printable.
Microsoft Azure Cognitive Service	Azure suite of artificial intelligence tools such as character recognition, speech to text, etc.
OCRBot	A software tool that uses Azure Cognitive Services to perform optical character recognition and Autotag on a PDF document.
OCRBot Publish Mode	A special action wizard in Adobe Acrobat that makes the invisible layer of scanned OCR text invisible, which is the default output of OCRBot. You would use this to revert the changes caused by remediation mode.
OCRBot Remediation Mode	A special action wizard included in Adobe Acrobat that allows you to make the invisible layer of text visible.
PDF Tag	A PDF accessibility markup.
ReOCR	Any existing OCR text will be replaced with new OCR text.
Text Only	OCRBot will make the text data visible and hide images.