

February 1, 2021

PDF Redactor

UiPathTeam PDF Redactor Custom Activity

Version 2.0

February 2021

Bernard Lawes

Bernard.lawes@uipath.com



What is Redaction?

To hide or remove confidential text before publication or distribution...

It's an essential step in guarding personally identifiable information (PII)...



Why makes Redaction so challenging?

- Lack of **Awareness** of redaction protocols
- Manual, Tedious, and **Error Prone**
- **Time-Consuming** even with Desktop Software
- **Batch** Processing is not readily available
- **Wastes** lots of Paper (Environmental Impact)
- Mistakes can lead to **costly compliance risks**

The image shows a redacted invoice. At the top left is the logo for 'EDGE & PLATTEN A JOHN L. LEWIS, INC.'. To the right of the logo, there is a header section with redacted text. Below the logo, there is a section for 'Bank Payment' with redacted details. Further down, there is a section for 'Bill To' with redacted information. A paragraph of text follows, also containing redacted portions. Below this text is a table with columns for 'Invoice #', 'Invoice Date', 'Due Date', 'Amount', 'Term', 'Balance', and 'Days'. The table contains one row of data. Below the table is another section with redacted text. At the bottom of the page, there are three boxes, each containing redacted text and a signature line.

Invoice #	Invoice Date	Due Date	Amount	Term	Balance	Days

PDF Redactor Activity Results

General Contracting Agreement

This agreement ("Agreement") is made this 2/4/2019 between Natasha Penelope, the identified with the EIN of 123-45-0000, henceforth known as "Contractor," and Bernard Lawes, who is a minor, with a License Number of L200-080-51182 henceforth known as "Owner."

Pursuant to the work described in previous agreement dated on 2/4/2019 for case number 237923232 the Contractor and Owner agree to the following:

The Contractor and Owner agree that the following work will be done on 492 North Gate, Miami, FL, 33130:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. On Account Number 21727328232 in repr Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in repr

Timeline

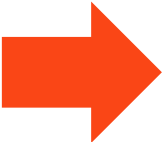
The Contractor and Owner agree that the work detailed above and included in the attached appendings will be completed according to the following timeline:

Work Start: 06/10/2020
Work Frequency: 8 hours / day
Work Completion: 10/06/2020

Payment

Owner agrees to pay the Contractor a total of {amount in dollars}, payable in the following manner:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. \$1,000 Duis aute irure dolor in repr Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim \$2800.00 nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in repr Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis



General Contracting Agreement

This agreement ("Agreement") is made this 2/4/2019 between Natasha Penelope, the identified with the EIN of [REDACTED] henceforth known as "Contractor," and [REDACTED] [REDACTED] who is a minor, with a License Number of [REDACTED] henceforth known as "Owner."

Pursuant to the work described in previous agreement dated on 2/4/2019 for case number [REDACTED] the Contractor and Owner agree to the following:

The Contractor and Owner agree that the following work will be done on 492 North Gate, Miami, FL, 33130:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. On Account Number [REDACTED] in repr Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in repr

Timeline

The Contractor and Owner agree that the work detailed above and included in the attached appendings will be completed according to the following timeline:

Work Start: [REDACTED]
Work Frequency: 8 hours / day
Work Completion: [REDACTED]

Payment

Owner agrees to pay the Contractor a total of {amount in dollars}, payable in the following manner:

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. [REDACTED] Duis aute irure dolor in repr Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim [REDACTED] nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in repr Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis

Value Proposition

- 1 Fast and secure automated redaction without 3rd party software

- 2 Ensure released documents are in compliance with data privacy laws

- 3 Reduce time required to safely publish or deliver requested documents

How does the PDF Redactor Activity work?

Leveraging OCR, a UiPath Robot locates and redacts sensitive data based on given Regex patterns. The activity comes with pre-built regex patterns for items such as social security number, phone number, dates and others. In addition, the user can define their own custom regex patterns as well.

Input / Output Arguments

Str	Description	Examples or Available Options		Type	In / Out	Status
FileInput	Input PDF File			String	In	
FileOutput	Full path of output redacted PDF File			String	Out	
Formula	User-defined Custom Regular expression			String	In	
FormulaAuto	Pre-Built and auto-generated regular expressions for PII data	- ssn - ein - phone - email	- dates - currency	String[]	In	
Keywords	String array of keywords to be redacted			Boolean	In	
Method	Available options are to Redact (black box), Highlight (red outline), or to Obfuscate (reveals only last 4 characters)	- redact - highlight - obfuscate		String	In	New
Obfuscate	Deprecated parameter. Please use the Method argument to obfuscate data.			Boolean	In	Deprecated
OCRAPIKey	API Key for the selected OCR Engine			String	In	New
OCREndpoint	Endpoint URL for the selected OCR Engine			String	In	New
OCREngine	OCR Engine used to digitize the document. OmniPage is the default OCR Engine	- OmniPage - Google Cloud Vision - UiPath		String	In	New
RedactColor	Color used for the redaction. Default color is black	- black - blue - gray - white	- green - red - orange	String	In	
Silent	Set to true to Hide Status Bar			Boolean	In	

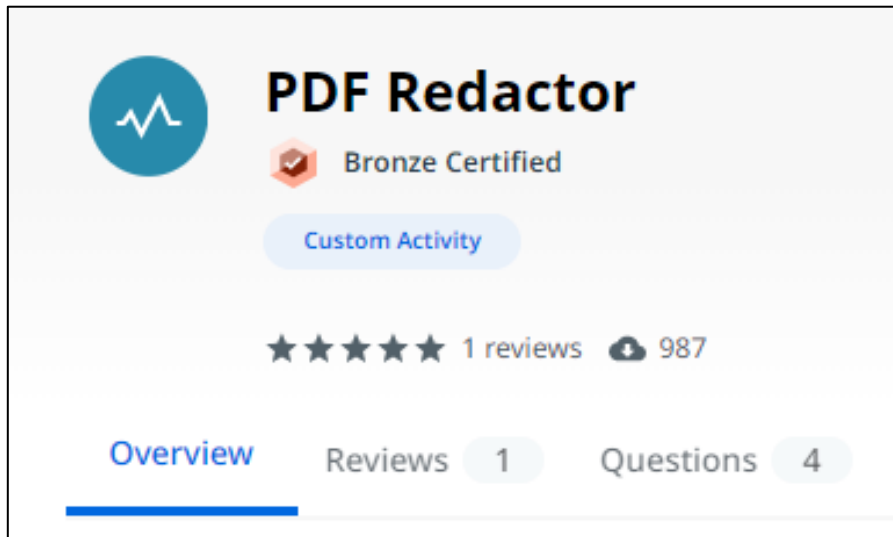
Limitations

Limitation	Description
Contiguous Words Only	The activity does not redact phrases, sentences, or any group of words separated by a space, tab, or carriage return. Regex patterns will match only contiguous words.
OCR Engines	Available OCR Engines • Google Vision OCR • OMNIPage • UiPath Document OCR

Dependencies

Library	Description	License Information	Version
PDFSharp.Extensions	Extension methods for PDFSharp to support and simplify some common operations including image extraction.	https://github.com/gherees/PDFSharp.Extensions/blob/master/LICENSE	0.1.2.2
PDFSharp.Standard		https://github.com/CLMSUK/PDFsharp/blob/master/LICENSE	1.51.12
UiPath.OCR.Activities	The UiPath.OCR.Activities package contains the "UiPath Screen OCR" and "UiPath Document OCR" activities, that use UiPath's OCR engines.	https://www.uipath.com/hubfs/legalspot/UiPath_MSSA.pdf	2.41
UiPath.OmniPage.Activities	The UiPath OmniPage Activities pack contains the OmniPage OCR activity, using the Nuance OmniPage OCR Engine.		1.50
UiPath.PDF.Activities			3.2.2
UiPath.System.Activities	Package contains core activities which enable the automation of desktop applications, browsers, and virtual machines. Please check the documentation for more details.		20.4.0
UiPath.UIAutomation.Activities	Core activities which enable the robots to manipulate data tables and collections, work with files and folders, communicate with Orchestrator. Package also contains workflow operators, dialog forms, debugging and invoking methods.		20.4.2

Where to find help for PDF Redactor?



Download Custom Activity Here

<https://connect.uipath.com/marketplace/components/pdf-redactor>

Forum Discussion and Examples

<https://forum.uipath.com/t/pdf-redaction-custom-activity/236846>



Bernard Lawes

Pre-Sales Engineer

Bernard.lawes@uipath.com

www.uipath.com