

# PDF Redactor

UiPathTeam PDF Redactor Custom Activity

November 2021

Bernard Lawes  
[info@uipath.com](mailto:info@uipath.com)



# What is Redaction?

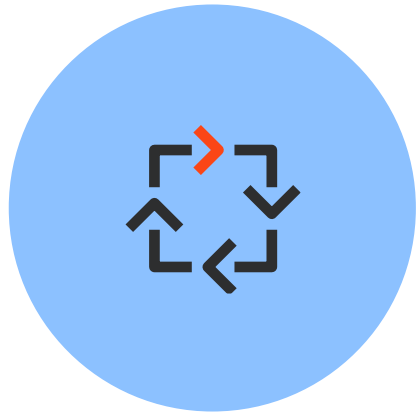
To hide or remove confidential text before publication or distribution...

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



# Redaction Plugin for Document Understanding

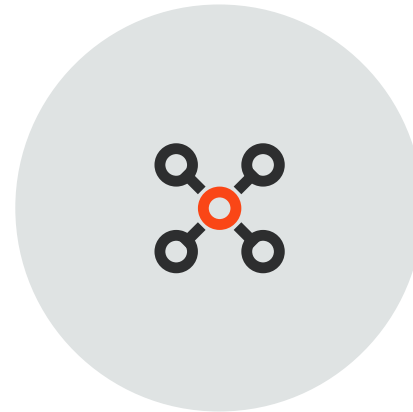
## Digitize



Digitize OCR Activities

Leverage Studio's available OCR Libraries

## Extract



Extraction Activities

Identify PII data using Regex, Form Extractor, or Machine Learning

## Redact



Redaction Activity

Irreversibly redact any text on the document

# Why is 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 sample of a redacted document, likely a medical form. It features a header with a logo and contact information. The main body contains several sections with redacted text, including a patient information section, a medical history section, and a table with redacted data. The table has columns for 'Date', 'Time', 'Location', and 'Status'. The bottom of the form includes a section for 'Patient Signature' and 'Physician Signature'.

Date	Time	Location	Status
1/1/2020	10:00 AM	Room 101	Admitted
1/2/2020	11:00 AM	Room 101	Discharged
1/3/2020	12:00 PM	Room 101	Admitted
1/4/2020	1:00 PM	Room 101	Discharged

# What does it do?

## 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

# Why Employ PDF Redaction



- 1 Fast and secure automated redaction without 3<sup>rd</sup> 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 PDF Redactor Work?

- Leveraging OCR, a UiPath Robot locates and redacts sensitive data based on given Regex patterns and extracted data from UiPath Document Understanding.
- 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.
- Enabled to ingest ExtractionResults and DocumentObjectModels from Document Understanding, the activity can redact any text value in the document. Moreover, the activity benefits from Document Understanding's ever-growing extraction capabilities using Form Extraction, Machine Learning, and others

# UiPath™ Marketplace



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>



# UiPathTeam.PDFRedaction.Activities

## Available Activities

### PDF Redaction

This activity redacts PDF Documents without the use of any other third party desktop software.

### DU Redaction Plugin

Plugin to Redact Documents using output from UiPath Document Understanding, namely the Document Object Model and the Extraction Result



# UiPathTeam.PDFRedaction.Activities

## Functional Arguments

### Activity: PDF Redaction

Argument	Description	Available Options						Type	Direction
FileInput	Full path to the Input PDF File							String	In
FileOutput	Full Path the redacted Output PDF File							String	In/Out
FormulaAuto	Pre-Built regular expression	Ssn	Phone	Email	Ein	Dates	currency	String[]	In
Formula	User-defined custom regular expressions							String	In
Keywords	String array of keywords to be redacted							String[]	In
HighlightOnly	If set to true, outlines words instead of redacting them							Boolean	In
RedactColor	Color of the redaction							Color	In
Silent	If set to true, the process will not output the console							Boolean	In
OCREngine	Sets the OCR Engine. Defaults to OmniPage	OmniPage		Google		UiPath		String	In
OCRAPIKey	API key for selected OCR Engine							String	In
OCREndpoint	Endpoint URL for the selected OCR Engine							String	In

# UiPathTeam.PDFRedaction.Activities

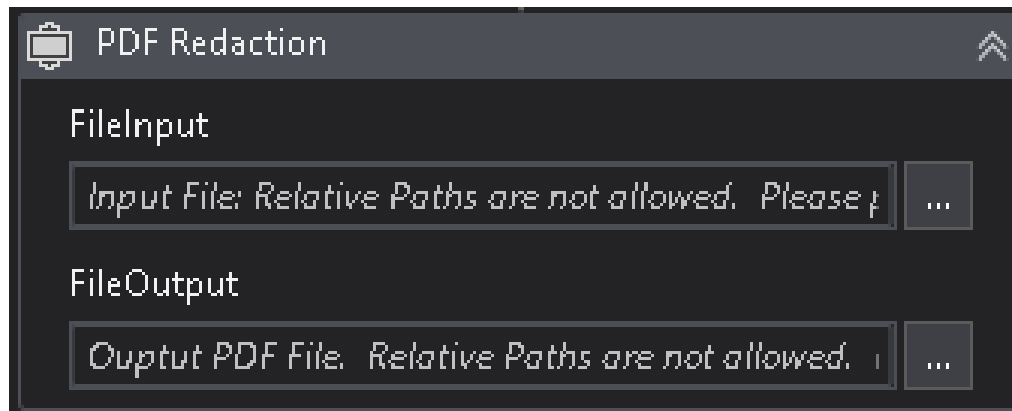
## Functional Arguments

### Activity: DU Redaction Plugin

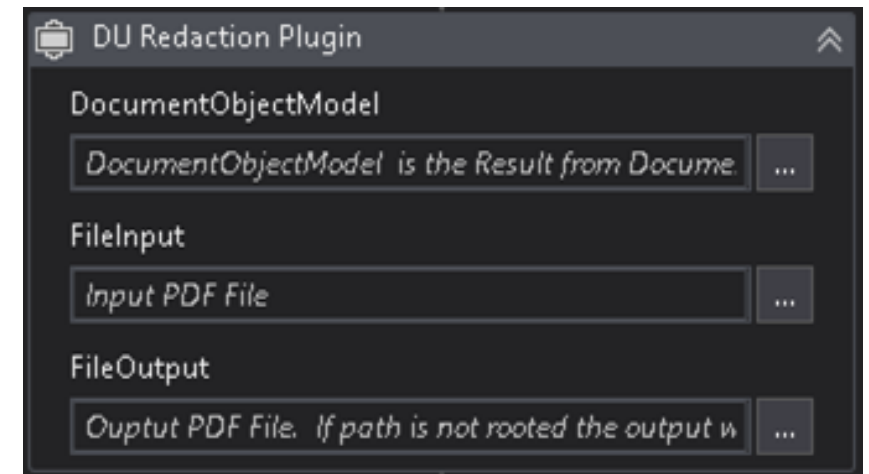
Argument	Description	Available Options						Type	Direction
FileInput	Full path to the Input PDF File							String	In
FileOutput	Full Path the redacted Output PDF File							String	In/Out
FormulaAuto	Pre-Built regular expression	Ssn	Phone	Email	Ein	Dates	currency	String[]	In
Formula	User-defined custom regular expressions							String	In
Keywords	String array of keywords to be redacted							String[]	In
HighlightOnly	If set to true, outlines words instead of redacting them							Boolean	In
RedactColor	Color of the redaction							Color	In
Silent	If set to true, the process will not output the console							Boolean	In
WatermarkFile	Full path to the image file to be used as a watermark							String	In
WatermarkLocation	Origin point for the watermark to appear. Default is (0,0)							Point	In
RedactFields	Array of fields from existing taxonomy to redact							String[]	In
DocumentObjectModel	Output of the UiPath Document Understanding Digitize Activity							DocumentObjectModel	In
ExtractionResult	Output of the UiPath Document Understanding Extraction Activity							ExtractionResult	In

# Screenshots

## PDF Redaction



## DU Redaction Plugin



# Dependencies

Library	Description	License Type	License Information	Version
PDFShap.Standard	.NET Standard class library for using PGP	MIT License	<a href="https://github.com/gheeres/PDFSharp.Extensions/blob/master/LICENSE">https://github.com/gheeres/PDFSharp.Extensions/blob/master/LICENSE</a>	4.0.1
PDFSharp.Extension	Extensions methods for PDFSharp to simplify common operations, including image extraction.	MIT License	<a href="https://github.com/CLMSUK/PDFsharp/blob/master/LICENSE">https://github.com/CLMSUK/PDFsharp/blob/master/LICENSE</a>	20.10.6
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.			20.10.1
UiPath.OmniPage.Activities				
UiPath.OmniPageBundle.Extended				
UiPath.PDF.Activities				
UiPath.System.Activities	Core activities which enable the robots to manipulate data tables and collections, work with files and folders. Package also contains workflow operators, dialog forms, debugging and invoking methods.			
UiPath.IntelligentOCR.Activities	Core activities that enable the usage of a complete document processing framework, from taxonomy definition, digitization, document classification, data extraction, data validation and classifier / extractor training.			
UiPath.OCR.Activities	The UiPath.OCR.Activities package contains the "UiPath Screen OCR" and "UiPath Document OCR" activities, that use UiPath's OCR engines.			

# Limitations

Limitation	Description
Single Contiguous Words	<p>The PDF Redaction activity does not redact phrases or sentences, or any groups of words separated by a space.</p> <p>To redact groups of words, use the DU Redaction Plugin activity</p>
OCR Engines	<p>The PDF Redaction activity is limited to the following OCR engines, Google Vision OCR, OmniPage, and UiPath Document OCR.</p> <p>Use the DU Redaction Plugin activity to make use other OCR Engines.</p>

# Upcoming Features

Features	Description

# Where can you find out more?



UiPath™

- [How AI Can Continuously Improve and Scale Automations](#) (webinar with customers sharing case studies)
- [Guide on Document Understanding](#) (white paper)
- [Academy training](#)
- Documentation on [Document Understanding framework & extractors](#) and [ML model training via AI Fabric](#)
- [AI & RPA](#) webpage





**Bernard Lawes**

Pre-Sales Engineer

[info@uipath.com](mailto:info@uipath.com)

---

[www.uipath.com](http://www.uipath.com)