**Read PDF Text activity to read content in PDF**

PDF Activities:

* **Read PDF Text**- Read all Characters from Specific PDF file and Stores in a string
* **Read PDF with OCR**- Read all Characters from Specific PDF file using OCR Technologies
* **Get Text**- Extract text value from specified UI Element

Open uipath->select **FirstDemo1Process** project->select project->right click on **FirstDemo1Process**

->select add->select Flowchart->Name: **InvoiceExtractPDF**->select create

Select Activity->search sequence->select **Sequence**->drag and drop **Sequence** to the flowchart->double click on **Sequence**->Rename: Read PDF Text Activity Demo

Select Activity->search pdf->if the activity isn’t there then have to install the package

To install a package->select Manage Packages->select All Packages->search pdf

->**select UiPath.PDF.Activities**->select install->select save

Select Activity->search **pdf**->select **Read PDF Text**->drag and drop **Read PDF Text** to the + drop activity

Select **Read PDF Text**->in the Properties->File, **FileName**: "give location of PDF file"

->Output, **Text**: press ctrl + k-> Output **Text**: pdfTextExtract->then give enter

Select Activity->search **write text file**->select **Write Text File**->drag and drop **Write Text File** to the + drop activity

Select **Write Text File**->Text: pdfTextExtract ->Write to filename: "extractedReadPDFActivity.txt"

To run the workflow->right click on the particular file which will run->select **Set as Start Node**->select Debug file->select Run File

Refresh project->a file name with “extractedReadPDFActivity.txt” is created

|  |
| --- |
| C:\Users\mesad\Desktop\Capture.JPG |

|  |
| --- |
| C:\Users\mesad\Desktop\Capture2.JPG |

**String functions to extract content from pdf based on the logic applied**

Select Activity->search **assign**->select **Assign**->drag and drop **Assign** to the + drop activity

Select **Assign**->in the Properties->Value:pdfTextExtract.Split(Environment.NewLine.ToArray,StringSplitOptions.RemoveEmptyEntries)

Select Variables->in the Created Variable: **arrayText**->Variable Type: select Array of [T]->select String->select ok

Select Assign->in the Properties->To: arrayText

Select Activity->search **message box**->select **Message Box**->drag and drop **Message Box** to the + drop activity

Select **Message Box**->Text:arrayText(3) ->->to extract the value from index

To run the workflow->right click on the particular file which will run->select **Set as Start Node**->select Debug file->select Run File

|  |
| --- |
| C:\Users\mesad\Desktop\Capture.JPG |

|  |
| --- |
| C:\Users\mesad\Desktop\Capture.JPG |

**Read PDF with OCR Technology to extract item with image**

Using **Read PDF Text** activity can extract any value from PDF

To get the invoice no from pdf->Select **Message Box**->Text: arrayText(3).Split("#".ToCharArray)(1)

To get the phone no from pdf->Select **Message Box**->Text: arrayText(3).Split("#".ToCharArray)(0)

To extract data from image copy which is converted to PDF use: **Read PDF with OCR**

Select Activity->search sequence->select **Sequence**->drag and drop **Sequence** to the flowchart->double click on **Sequence**->Rename: Read PDF Demo 2

Select Activity->search **read pdf with ocr**->select **Read PDF With OCR**

->drag and drop **Read PDF With OCR** to the + drop activity

Select Activity->search **read pdf with ocr**->select **Read PDF With OCR**

->drag and drop **Read PDF With OCR** to the + drop activity

Select activity->search **Microsoft OCR**->select **Microsoft OCR**-> drag and drop **Microsoft OCR** to the Drop OCR Engine

Select **Read PDF With OCR** ->in the Properties->File, **FileName**: "give location of PDF file"

->Output, **Text**: press ctrl + k-> Output **Text**: pdfTextImage->then give enter

Select Activity->search **write text file**->select **Write Text File**->drag and drop **Write Text File** to the + drop activity

Select **Write Text File**->Text: pdfTextExtract ->Write to filename: "extractedReadPDFOCR.txt"

To run the workflow->right click on the particular file which will run->select **Set as Start Node**->select Debug file->select Run File

Refresh project->a file name with “"extractedReadPDFOCR.txt"” is created

Select Activity->search **assign**->select **Assign**->drag and drop **Assign** to the + drop activity

Select **Assign**->in the Properties->Value: pdfTextImage.Split(Environment.NewLine.ToArray,StringSplitOptions.RemoveEmptyEntries)

Select Variables->in the Created Variable: **arrayTextOCR**->Variable Type: select Array of [T]->select String->select ok

Select Assign->in the Properties->To: arrayTextOCR

Select Activity->search **message box**->select **Message Box**->drag and drop **Message Box** to the + drop activity

Select **Message Box**->Text:arrayTextOCR(1) ->to extract the value from index

To run the workflow->right click on the particular file which will run->select **Set as Start Node**->select Debug file->select Run File

|  |
| --- |
| C:\Users\mesad\Desktop\Capture.JPG |

**Get Text Activity to extract content from PDF**

Select Activity->search sequence->select **Sequence**->drag and drop **Sequence** to the flowchart->double click on **Sequence**->Rename: Read PDF getText

Select Activity->search **Start Process**->select **Start Process**->drag and drop **Start Process** to the + drop activity

Select **Start Process**->in the properties->Input, **FileName**: "give location of PDF file"

Select Activity->search **get text**->select **Get Text**->drag and drop **Get Text** to the + drop activity

At first manually open the pdf file then Select **Get Text** ->select **indicate on screen**->select the element on the pdf which will extract

To confirm the unique locator which will be extracted from pdf->select **Get Text**->select **click to open options menu**->select **edit selector**->select **Highlight**

Select **Get Text**->in the Properties ->Output, Value: press ctrl + k-> Value: **invoices** ->then give enter

Select Activity->search **message box**->select **Message Box** ->drag and drop **Message Box** to the + drop activity

Select **Message Box**->Text: **invoices**

To run the workflow->right click on the particular file which will run->select **Set as Start Node**->select Debug file->select Run File

**Note**:

* **Read PDF Text** and **Read PDF with OCR**  activity : use this when there are multiple validation
* **Get Text** activity: use this when processing hundreds of pdf or when is prone to failure for invalid format

|  |
| --- |
| C:\Users\mesad\Desktop\Capture.JPG |