PDF Data extraction with uipath

|  |  |
| --- | --- |
| Raised hand with solid fill | This report on UiPath PDF Data Extraction will cover all the concepts on how to extract data from PDFs using UiPath.  Activities that can read every character in a document can be used to read and extract the content of.pdf files separately.   * The first one explains how to read the .pdf file while using a specific activities. * The second one explains how to read the .pdf file while using the [Read PDF With OCR](https://docs.uipath.com/activities/other/latest/document-understanding/read-pdf-with-ocr) activity. |

# Overview

Several robots will be built, First idea will talk about how to extract information from several invoices, including (invoice date, Two robots will be built. The first company name, invoice number, etc.)

## OBJECTIVE

|  |  |
| --- | --- |
| Raised hand with solid fill | First Idea  In the first project there will be 3 invoices, which will be automatically opened and information extracted from them. Invoice Number, Order Number, Invoice Date, Due Date, Total Due. |

.

## **TECHNICAL REQUIREMENTS**

|  |  |
| --- | --- |
| Raised hand with solid fill | This robot contains three basic parts: Data Table, PDF, Excel. |

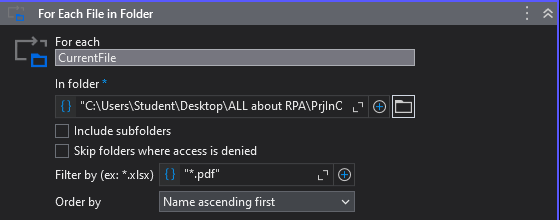
# PROCEDURES

* To store information and add it to Excel, we need a way that is easy for Excel to deal with and fill the columns and rows quickly. Therefore, we use [Build Data Table](https://docs.uipath.com/activities/other/latest/workflow/build-data-table), which falls under the Data Table commands.

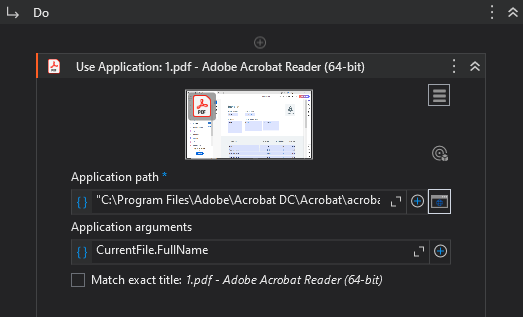
We title each header that must be inserted into Excel appropriately to what must be saved You can specify the type of variables that can be added within the columns. We chose for all columns to be of type string.

* In order to loop each item in the folder containing the invoices, we use [For Each File in Folder](https://docs.uipath.com/activities/other/latest/workflow/for-each-file-x)

inside in folder box we install the path of the folder, If it contains subfolders, we check the box, and if it contains files that require link permission, we check the box. Inside the Filter by box, we use the idea of ​​[Wildcards](https://docs.uipath.com/studio/standalone/2023.4/user-guide/selectors-with-wildcards), where we put Asterisk to include all files. It is possible to concatenate more than one extension within the filter by box



* Inside the for each scope at the do we put [Use Application/Browser](https://docs.uipath.com/activities/other/latest/ui-automation%22/n-application-card) To select the PDF page and access it



Within the Application arrgument you must enter a command CurrentFile.FullName to make sure that all files have been accessed

* Inside the [Use Application/Browser](https://docs.uipath.com/activities/other/latest/ui-automation%22/n-application-card) scope use the [Get Text](https://docs.uipath.com/activities/other/latest/ui-automation%22/n-get-text) activity to store all information inside variables that fit the columns included in the [Build Data Table](https://docs.uipath.com/activities/other/latest/workflow/build-data-table) later



After that use [Add Data Row](https://docs.uipath.com/activities/other/latest/workflow/add-data-row) activity to add information to the headers that were created within the [Build Data Table](https://docs.uipath.com/activities/other/latest/workflow/build-data-table) to add the variables that were created using the [Get Text](https://docs.uipath.com/activities/other/latest/ui-automation%22/n-get-text) activity to the headers that were created inside the build data table, taking into considration the order when entering. The ArrayRow box is filled because it receives variables that are all of the same type.

A screenshot of a computer program

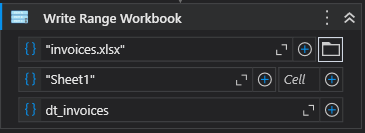
Description automatically generated

A screenshot of a computer program

Description automatically generated

In the Data table field, a variable of type [Data table](https://docs.uipath.com/studio/standalone/2021.10/user-guide/data-table-variables) is created that contains the desired data to be transferred in the next step to the Excel sheet.

* A new Excel file is created using the [Write Range Workbook](https://docs.uipath.com/activities/other/latest/productivity/write-range) activity, where the file is initially named, then the name of the page is specified, and from what cell it begins, after which the variable of type Data table is called.



# Recommendations

* To verify whether the robot has opened all files, we put a [message box](https://docs.uipath.com/activities/other/latest/workflow/message-box) to send the name of the file that was opened before the use [Use Application/Browser](https://docs.uipath.com/activities/other/latest/ui-automation%22/n-application-card).

A screenshot of a computer

Description automatically generated

* If there is data in a specific sheet and you want to add more to the same sheet, you can use [Append Range Workbook](https://docs.uipath.com/activities/other/latest/productivity%22/append-range) instead of [Write Range Workbook](https://docs.uipath.com/activities/other/latest/productivity/write-range).

A screenshot of a computer

Description automatically generated

* You can use [Clear Data Table](https://docs.uipath.com/activities/other/latest/workflow/clear-data-table) if there is a problem with data duplication every time.

A black rectangular object with white text

Description automatically generated

* In case you want to stop after each loop, you can use Break activity.



* To avoid problems navigating within files, choose Always at the close property in the properties panel.

A screenshot of a computer

Description automatically generated



# Results & COCLUSION

You can access the results and watch the full robot run by [clicking here](https://drive.google.com/file/d/11Ln-BNy58WRMblJwkgD2t4LxFVApXKRm/view?usp=drive_link)