RPA Design and Development

v4.0

UiPath

Lesson 13 | PDF Automation

PDF Automation – Exam Topics

1. Extract data from native and scanned PDF.

2. Extract a single piece of data from a single and multiple native PDFs.

PDF Extraction

Process of extracting the raw data from PDF documents which can contain text and images. PDFs can be of two types:

Native PDF

Scanned PDF

Activities for PDF Extraction

There are two activities for extracting text from PDFs:

Read PDF Text

- Reads all characters from a specified PDF file and stores them in a string variable.
- It extracts text from a Native PDF

Read PDF with OCR

- Reads all characters from a specified PDF file and stores it in a string variable by using OCR technology.
- It extracts text from a Scanned PDF

Other PDF Activities

Some other activities related to PDFs in Studio are:

Get PDF Page Count

- Provides the total number of pages in a PDF file

Extract PDF Page Range

- Extracts text from a specified range of pages from a PDF document

Export PDF Page As Image

- Creates an image from a page in a specified PDF file

Join PDF Files

- Joins multiple PDF files stored in an array of strings into a single PDF file

Extract Images From PDF

- Extracts images from a specified PDF file and saves them in a folder

Manage PDF Password

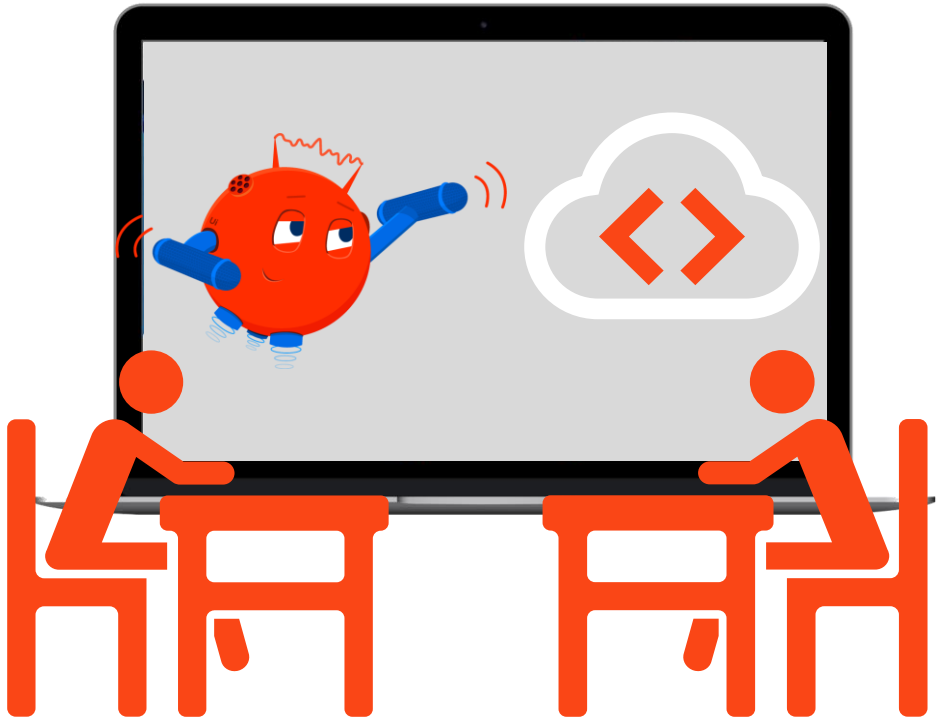
- Manages the password of a specified PDF file if current password is known

Classroom Exercise

Demonstrate the use of **Read PDF Text** activity by extracting text from a PDF file and storing it in a Notepad file.

- Install the dependency UiPath.PDF.Activities
- Go to https://www.uipath.com/resources/automation-whitepapers and download a PDF whitepaper
- Open the downloaded PDF file
- Scrape the text from the file using the Read PDF Text activity
- Save the scraped text directly in a .txt file

Practice Exercise



Build a workflow using the **Read PDF Text** activity and extract only Email IDs and Phone Numbers from a PDF file and store in a Notepad.

- Download the practice excel file available on www.rpachallenge.com
- Convert the file to PDF to use in this exercise
- Read data from the PDF file using a Read PDF Text activity
- Extract only Phone Numbers and email IDs from the PDF and store it in a Notepad file