**PDF Data Extraction using python**

***Zappy.ZappyActions.ProgrammingScripts***

Extract the pdf file content into an excel file.

# INPUT

# InputFilePath – PDF file path for extraction.

# OPTIONAL

# Passwd – Password, if the PDF file is protected.

* **File Type –** Pdf file type like lattice or stream.

**Stream –** Stream can be used to parse tables that have whitespaces between cells to simulate a table structure.

**Lattice** **–** Lattice is more deterministic in nature, and it does not rely on guesses. It can be used to parse tables that have demarcated lines between cells, and it can automatically parse multiple tables present on a page.

* **Pages –** Page number of PDF file with comma separated like (1,2, 3, …).

**MISC**

* **PythonExePath –** Python exe path.

**OUTPUT**

* **OutputFilePath** **–** Created excel (.xlsx) file path.
* **CommandText –** Output text.