

# PDF Stripper

## Description

Strips a PDF file into single pages, placed in subfolders, then extracting text in a text file and metadata in an xml file from each page.

The result:

606381	Folder	20/07/2020 13:34
606381.pdf	10.9 MB pdf File	20/07/2020 13:34
606381.txt	239 bytes txt File	20/07/2020 13:34
Page_1	Folder	20/07/2020 13:34
Page_1.pdf	4.75 MB pdf File	20/07/2020 13:34
Page_1.txt	128 bytes txt File	20/07/2020 13:34
Page_1.xml	687 bytes xml File	20/07/2020 13:34
Page_2	Folder	20/07/2020 13:34
Page_2.pdf	4.76 MB pdf File	20/07/2020 13:34
Page_2.txt	111 bytes txt File	20/07/2020 13:34
Page_2.xml	670 bytes xml File	20/07/2020 13:34

- Input pdf file
- Txt file containing text from the entire pdf file
- Page\_X (X = page number) subfolder containing the single page pdf, a text file containing the page's text and an xml file with page size, mediabox, bleedbox and trimbox dimensions.

## Compatibility

Enfocus Switch 2020 and newer

*There is no need for any 3th party application!*

## Connections

There can be multiple incoming connections. There need to be two outgoing connections:

- Data Success
- Data Error

## Properties detailed info

There are no properties, the newly created folder has the same name as the incoming PDF filename.

## Outgoing connections properties

- Data Success
  - Contains the stripped pdf with subfolders, text and metadata
- Data Error
  - Contains any non-pdf file