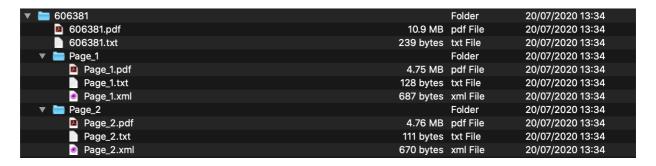
PDF Stripper

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:



- 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