

Extract PDF Ranges

Description

Extract PDF Ranges splits PDF files into multiple parts... one part for each page range specified. By default, each range file is sent to the outgoing connection.

Output to a folder can also be specified. The resulting folder of files can then be merged with Enfocuss' **Merge PDF** tool.

A custom private data value can also be attached to the output.

A common use is for extraction of *n* number of variable data samples, like the first 5 and last 5 records of a data populated, final output PDF.

Compatibility

Switch 2020 Spring and higher. Windows and Mac OSX.

Incoming connections

Extract PDF Ranges only accepts PDF files from one or many incoming connections (all other file types and folders are logged and sent to the Error connection).

Outgoing connections

Extract PDF Ranges uses outgoing data Traffic Light connections. Flows should be set up with both a Success and an Error outgoing Data connection.

Note that flows without an outgoing Data Error connection will log and send Error jobs to the flow's "Problem jobs folder". This is not suggested.

Flow elements properties

- **Page ranges** (Inline value, or single-line text with variables)
 - Comma delimited list of page ranges to be extracted. The page range for each extraction will be added to the file name (see the test flows below).
- **Labels** (drop-down list)
 - The type of labels for each output file: "Absolute numeric from source", "Original from source" or "New absolute numeric".
- **Output to folder?** (Yes/No)
 - Output all files to a folder before sending to outgoing connection.
- **Custom private data value** (Inline value, or single-line text with variables)
 - Private data to attach to each output (the default value is "xRanges"). The key to access the value is "DevMint.xRanges.customPrivateData".

Properties detailed info

Page ranges: "Reverse" page ranges (eg: 100-50) and "Out-of-range" ranges (eg:50-150 for a 100 page document) will be ignored.

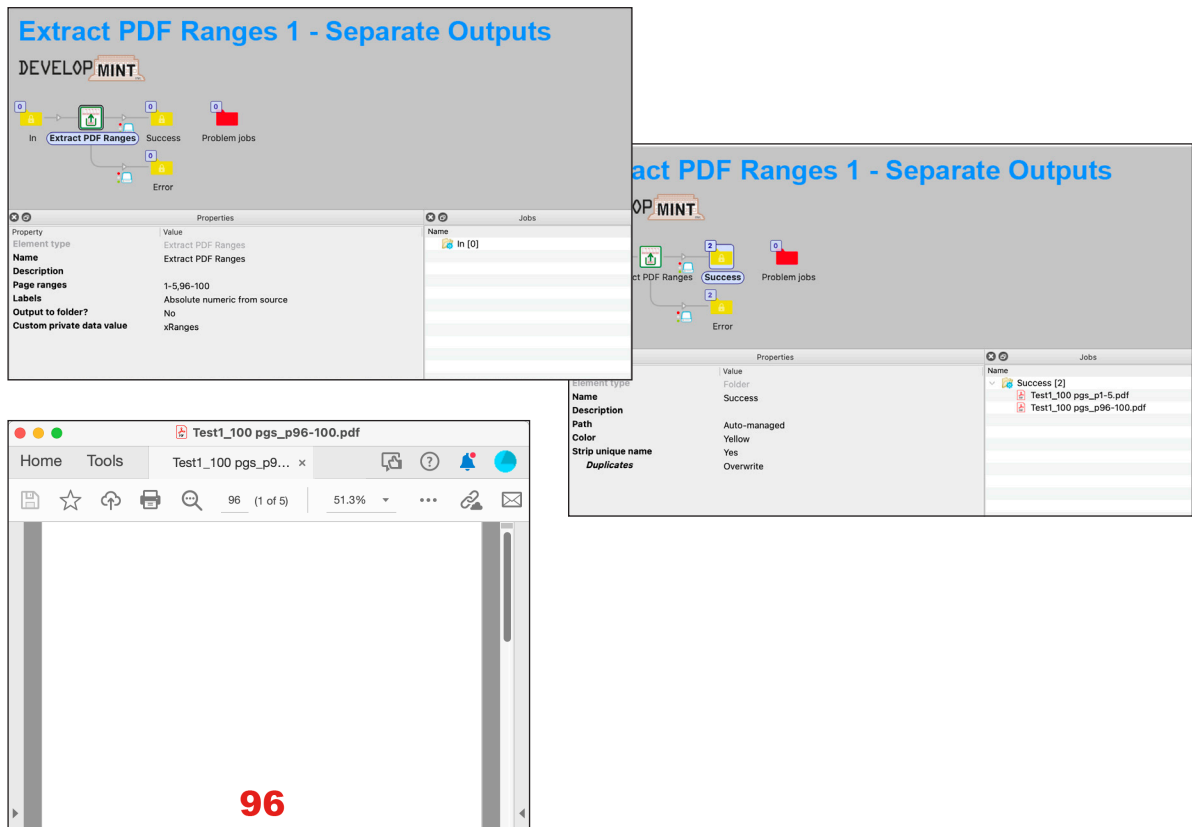
Output to folder: Note that sort identifiers ("_rngA", "_rngB"...) will be added to the names of files output to a folder. The reason for this is, if the files are merged, some ranges would be merged out of order (eg: a range of "1,77,153,229" would be merged in the order of "1,153,229,77").

Test files and flows

Sample test files and flows illustrated below are available for download as a zip archive.

Test flow 1:

This flow illustrates the simplest setup. It is set up to export two ranges as two separate files, using the “Absolute numeric from source” page labels setting.



Test flows continued on next page...

Test flow 2:

This flow demonstrates the folder output of ranges. Similar to Flow 1, the ranges are extracted, but output to a folder. Note that a sort identifier has been added to each file. Finally, the files are merged using Enfocus' **Merge PDF** tool.

The screenshots illustrate the workflow steps and their outputs:

- Top-left screenshot:** Shows the workflow diagram with the 'Merge PDF' step highlighted. The 'Properties' panel shows:

Property	Value
Name	Extract PDF Ranges 2 - Merge Outputs
Description	Created at 5/25/22 8:35 PM
Background image	None
Header image	J:\Users\Aserver\Desktop\Enfocus App Dev\App Store Project...
Marker image	None
Show advanced element prope...	No
Time-of-day window	No
Day-of-week window	No
Day-of-month window	No
- Top-right screenshot:** Shows the 'Jobs' panel with a list of extracted PDF files:

Name	Success [1]
	Test2_300 pgs_177,153,229
	Test2_300 pgs_rmgA_p1.pdf
	Test2_300 pgs_rmgB_p77.pdf
	Test2_300 pgs_rmgC_p153.pdf
	Test2_300 pgs_rmgD_p229.pdf
- Bottom-left screenshot:** Shows the 'Properties' panel for the 'Merged' folder:

Property	Value
Element type	Folder
Name	Merged
Description	
Path	Auto-managed
Color	Yellow
Strip unique name	Yes
Duplicates	Overwrite

Test flows continued on next page...

Test flow 3:

This flow shows the use of an embedded range (in the file name).

First, the range has been inserted in the file name, after a separator (double underscore). I used the Luno Tools free Data Mapper app to extract the range from the file name, and attach it to the source file as private data.

Second, I use the Enfocus **Rename job** tool to remove the range from the source file.

Then, **Extract PDF Ranges** uses the attached private data for the page range.

Finally, the folder of files are merged using Enfocus' **Merge PDF** tool.

The image displays three screenshots of the Enfocus workflow 'Extract PDF Ranges 3 - Embedded Range' at different stages:

- Top Left Screenshot:** Shows the 'Data Mapper' tool configuration. The 'Map to' field is set to 'Private Data'. The 'Extra map name' is 'pageRange'. The 'Extra map value' is 'None'. The 'Extra map comment' is 'Value of the extra data map to be added.' The 'Extra map name' is 'None'. The 'Jobs' panel shows 'Error [0]'.
- Top Right Screenshot:** Shows the 'Extract PDF Ranges' tool configuration. The 'Map to' field is set to 'Private Data'. The 'Extra map name' is 'pageRange'. The 'Extra map value' is 'None'. The 'Extra map comment' is 'Value of the extra data map to be added.' The 'Extra map name' is 'None'. The 'Jobs' panel shows 'Error [0]'. The 'Properties' panel shows 'Value: Extract PDF Ranges', 'Element type: Extract PDF Ranges', and 'Description: Single-line text with variables defined. Absolute numeric from Enter one or more page ranges, using absolute numeric values (eg. 8,20-24,50,3). Invalid ranges will be ignored. [Job.PrivateData-Key="pageRange"]'.
- Bottom Screenshot:** Shows the 'Merge PDF' tool configuration. The 'Map to' field is set to 'Private Data'. The 'Extra map name' is 'pageRange'. The 'Extra map value' is 'None'. The 'Extra map comment' is 'Value of the extra data map to be added.' The 'Extra map name' is 'None'. The 'Jobs' panel shows 'Merged [1]' and 'Test3_300 pgs_177;153,229.pdf'. The 'Properties' panel shows 'Value: Merged', 'Element type: Folder', 'Name: Merged', 'Description: Auto-managed', 'Path: Auto-managed', 'Color: Yellow', 'Strip unique name: Yes', and 'Duplicates: Overwrite'.

Test flows continued on next page...

Test flow 4:

This flow shows the use of an “end range”.

With **Extract PDF Ranges** you can specify an end range by prefixing the number of pages from the end of the document with an “R” (or “r”).

For example, to extract the last 5 pages from a document (regardless of the page count), the range would be “R5-R1”. For a 100 page document, as in the example below, “R5-R1” would yield an extracted document with pages 96-100 from the source document. For a 500 page document, “R5-R1” would yield an extracted document with pages 496-500 from the source document.

Note that, as in regular ranges, “Reverse” page ranges and “out-of-range” ranges are still ignored (ie: For a 100 page document, reverse range “R1-R5” and Out-of-range “R105-R101” would both be invalid ranges).

The image displays two screenshots of the Develop Mint interface for a test flow titled "Extract PDF Ranges 4 - End Range".

The left screenshot shows the flow diagram with an "In" input, an "Extract PDF Ranges" element, and "Success" and "Error" outputs. The Properties panel for the "Extract PDF Ranges" element is visible, showing the following details:

Property	Value
Element type	Extract PDF Ranges
Name	Extract PDF Ranges
Description	
Page ranges	1-5,R5-R1
Labels	Absolute numeric from source
Output to folder?	No
Custom private data value	xRanges

The Jobs panel shows a single job named "In [0]".

The right screenshot shows the same flow diagram, but the Jobs panel now displays a "Success [2]" folder containing two PDF files:

- Test4_100 pgs_p1-5.pdf
- Test4_100 pgs_p96-100.pdf

The Properties panel for the "Success" folder is also visible, showing the following details:

Property	Value
Element type	Folder
Name	Success
Description	
Path	Auto-managed
Color	Yellow
Strip unique name	Yes
Duplicates	Overwrite

Test flows continued on next page...

Test flow 5:

This flow shows the “Original from source” page labeling option.

The test source file in this example is a 10pg book with a common page labeling scheme of: CV_1, CV_2, i, ii, 1-4, CV_3, CV_4.

If we want to extract the first 8 pages of the book, and want to keep the original labels, we would expect the extracted document to have the labels: CV_1, CV_2, i, ii, 1-4.

