

# OCR File Splitter

Revised 10-30-2008

## Purpose

OCR File Splitter is a program that is designed to monitor “watch” a file folder for the arrival of a multi-page Tiff image. When one arrives it will be split into smaller multi-page Tiff images based upon the text content of the file or by a fixed number of pages. This makes it ideal for separating incoming facsimiles for further processing or separating a group of files that have been batch scanned with a copier. The program will detect a cover page if it is present and remove it, afterwards it will create a separate file for each of the documents contained within the multi-page tiff image. For instance, if someone were to fax in a group of sales orders, super bills, credit applications etc. (something that has a varying amount of pages) a separate file would be created for each transaction that needs to be processed.

## How it works:

To determine the beginning of the document the program utilizes the OCR engine in Microsoft Office Document Imaging (MODI) a required component. Once the file is OCR'd the text is extracted and compared to three lists of text. Contents from one or both lists must be present while no text can be present from a third list in order for the page to be classified as the first page in a document. Each consecutive page is added to the first page until another first page is determined. This process repeats itself until all pages in the multi-page tiff file are processed.

## System Requirements:

Microsoft Windows

Microsoft Office Document Imaging (MODI) a component of the Office Suite

## Setup:

*If any assistance is needed setting up this program don't hesitate to contact eDocfile Inc. at 813-413-5599 or [sales@edocfile.com](mailto:sales@edocfile.com)*

Begin by clicking on All Programs \ OCR File Splitter \ Setup OCR File Splitter



Enter or Browse to a folder to store settings in and Click on Continue



Click on New from the Job Manager



The Setup Menu will Open

**Setup Image Separation**

**SETUP OCR FILE SPLITTER**

1 Enter Job Name 2 PDF Output Type

3 Source Folder

4 Output Folder

5 Failed Files

Fixed Page Processing

To use Rules Processing Leave Fixed Length of Document Blank

6 Cover Page Length 7 Fixed Length of Document

Rules

8 ☐ Detect Cover Page

The First Page must contain a Word or Phrase from this list

The First Page must also contain a Word or Phrase from this list

The First Page must not contain a Word or Phrase from this list

9

10

11

12

Continue Exit

Copyright© eDocFile Inc. 2004-2008

1. Enter a name for this Job – it will later appear in a drop down list when it is time to run the program
2. Select either PDF or Tiff for Output File Type
3. Enter the Folder that contains the files to be processed
4. Enter the Folder that is to be used for Output
5. Enter the Folder for Files that Failed Processing – This pertains to Fixed Page Processing Only
6. Enter the number of pages contained in the Cover Page if using Fixed Page Processing
7. Enter the number of pages in the document if using Fixed Page Processing – Leave Blank for automatic determination with OCR
8. Check for a Cover Page if Processing with OCR. If checked and there is a cover page the file will have Cover Page in its suffix. If unchecked, it will be named with a counter in its suffix just as the other files
9. Enter Text that is contained on the First Page of the Document only – It is case sensitive,
10. Enter Additional Text that is contained on the First Page of the Document only - It is case sensitive

- The Software will look for a match from 9 and 10 and if any line from 9 is present and any line from 10 is present it will look at 11
11. Enter Text that cannot be on the first page of the Document.
  12. Click on Continue to Save or Exit

If using a fixed page length no rules will be processed. The software looks for a fixed page length and if none is specified it assumes that the rules are to be used.

When entering text to determine the start of a document the software looks for a word or phrase in list one, if any is present (as many as needed can be entered allowing it to do different document types) it checks for a word or phrase in list two if any are matched it looks in list three. If it finds a match in list three it assumes it is not the first page.

When finished entering settings click on Continue.

## Running the Program

Begin by clicking on All Programs \ OCR File Splitter \ OCR File Splitter



The Main Menu will appear



Select a Job to Run from the Drop Down List and Click on “Run Job” to process the files in the folder. If “Monitor File Folder” is checked the files will be processed in the folder and the software will keep checking for new files. If “Show Processing Progress” is checked a small window will appear when files are actually being processed.

Please note that when files are being processed the user's computer's resources will be maxed out while the OCR process is being performed. If the program is to be used to monitor a folder it is best to just let it run on a PC that no one uses for anything mission critical.

## Shutting Down the Program

When the Program finishes processing a batch it will display a window informing the user that the processing is completed.



If running in a Monitor Mode the user will have an Icon in the Sys Tray that is flashing. To exit the monitor function, Click on All Programs \ OCR File Splitter \ Shutdown Splitter



## Support

Contact:  
eDocfile Inc.  
2709 Willow Oaks Drive  
Valrico, FL 33594  
Phone 813-413-5599  
[sales@edocfile.com](mailto:sales@edocfile.com)