Tiff to Text III Manual

Revised 1/30/2012

Tiff to Text III

Tiff to Text III is a utility that utilizes Microsoft Office Document Imaging's OCR engine to batch process TIFF images into text files that match TIFF images. It is ideally suited for law firms that need to process tens of thousands of TIFF images into matching text files.

It will also transform TIFF images into Searchable Tiff Images and PDF files. When creating a PDF a hidden layer of text will be placed on the page of the pdf; however the hidden text will not be aligned with the words on the pdf. Therefore, when searching the PDF the user will be taken to the correct page, but not directly to the word in the PDF.

The program can be used to process a batch of images or it can be used to monitor a folder and subfolders and process the images as they arrive.

The program will always produce a mirror copy of the input folder structure and the files must be moved to a new location.

MODI and MSPView

Tiff to Text III uses the OCR engine found in Microsoft Office Document Imaging (MODI). This is one of the best OCR engines on the market. It was released as part of Office 2003 and Office 2007. In Office 2003 it will be installed on the first use and in Office 2007 the user must do a custom install to add it to their PC.
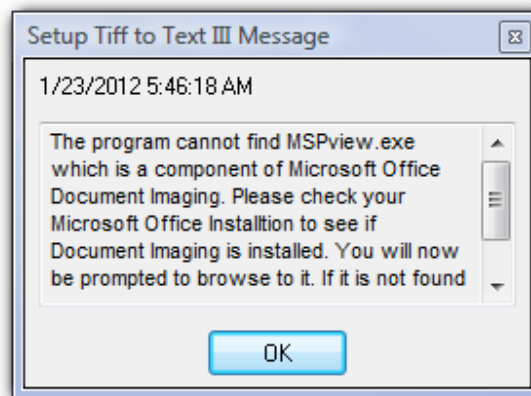
If the message above appears MSPView.exe cannot be found. It will be necessary to install it from the Office Installation or optionally it can be installed from Sharepoint Designer which is free. When installing from Sharepoint Designer do a custom install to add it. Sharepoint Designer can be downloaded here.

Setup and Operation

Setting up the program

From the Main Menu click on "Setup".

1. This is the root input folder that contains the images to be processed
2. This is the root output folder where the processed files will be placed
3. Enter a path if saving the original Images
4. If this box is checked all sub folders of the input folder will be processed
5. If this box is checked the processed files will be placed in a subfolder that matches the input folder. If left unchecked all processed images will be placed in the root output folder. The program will not allow duplicate files to be created. Therefore if processing subfolders of the input folder the user should check this box if there is a possibility of duplicate file names
6. If this box is checked the input folder and subfolders will be monitored for images, if unchecked it will run once and stop. If placing text files in the same folder as the input folder this must be unchecked.
7. Place a check mark in this box to display the name of the file as it is being processed
8. Place a check mark here to activate the Time Out Option*
9. Enter a time out in minutes, this allows the program to continue if it encounters a corrupt image file. Set the time to the longest it could possibly take to OCR a file.
10. Check this box if a tiff image is to be placed in the output folder, the tiff image will be searchable. When using this option the input and output folder must not be the same.
11. Check this box if a text file is to be placed in the output folder it will have a matching name to the input file
12. Check this box if a PDF file is to be placed in the output folder it will have a matching name to the input file
13. Place a check mark in this box to save a copy of the original file
14. Check here to create a blank text file if no text exists

15. Create a log of Blank text files for manual review
16. Saves the settings and prompts the user to run the program
17. Opens this Help File
18. Exits without saving the file.

* The Time Out option is designed to continue processing if a corrupt file is found. It assumes a file has to be OCR'd in a certain amount of time, (8) if it is not finished, it assumes there is a problem with the image and it aborts the processing of the file and moves on to the next one. The advantage is that if there are thousands of single page files to be processed the user does not need to continuously check to see if the program encountered a corrupt image file. If a failure does occur it will be written in a log file in the output folder.

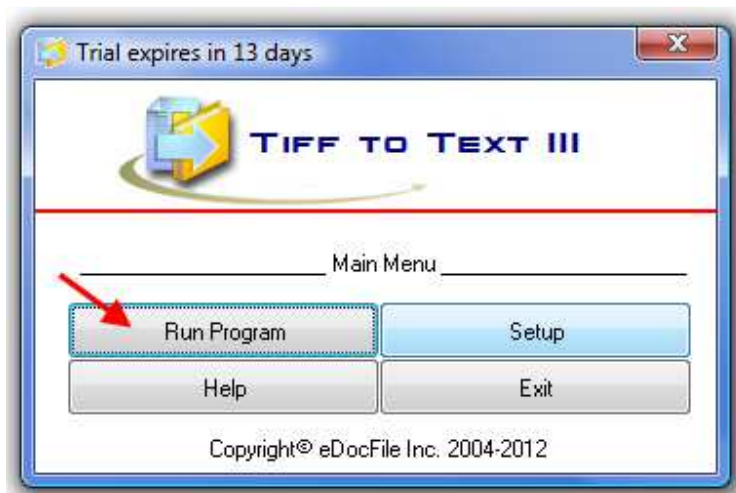Enter the desired options and click on continue.

Click on Yes to Start Processing Files

An Icon will appear on the task bar showing that the program is running.

To shut down or restart the program

To Start Tiff to Text III

Click on "Run Program" to start the program

Click on "Shut Down" to end the program

Tiff Due Diligence

Tiff Due Diligence is a utility that confirms matching files exist for processed TIFF images. It will look at one file folder hierarchy and check another with a different root folder to see if there are matching file names with a different extension.

Please note that Tiff Due Diligence is independent of Tiff to Text III and it stores its settings in the Windows Temp Folder.
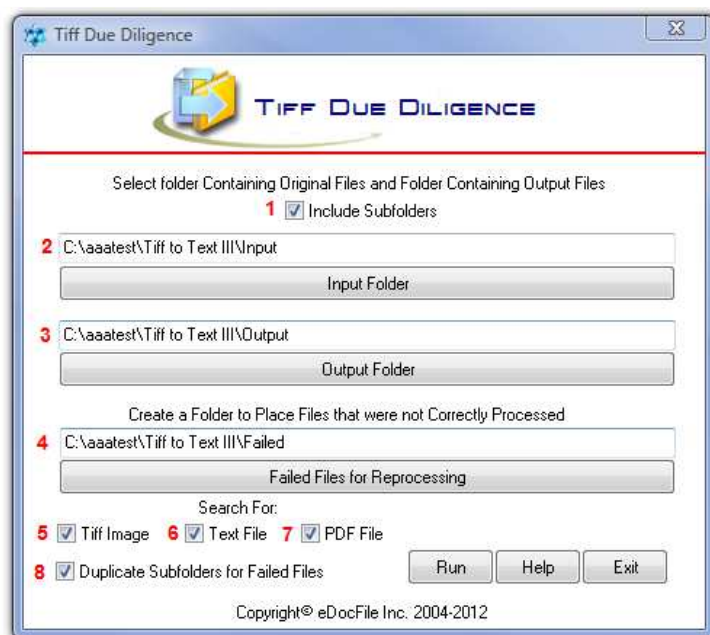
1. Place a check mark in this box to include Sub folders when making the comparison
2. The root input folder to be matched
3. The root output folder
4. A folder for failed files - tiff files that failed processing will be placed in this folder
5. Check for a matching tif image
6. Check for a matching text file
7. Check for a matching PDF file
8. Place a check mark in this box to create a matching file folder hierarchy for failed files. If it is not checked all failed files will be placed in root failed file folder.

Trouble Shooting and Support

The only know issue with Tiff to Text relates to MODI. When Microsoft Office is updated or in some cases removed, MODI is left behind. Since the new version of MODI is installed in a different path, both versions reside on the same PC. The only fix know at this time is to uninstall MODI, check to see if MSPView.exe still exists after the uninstall, if it does delete it and the folder that contains it. Run a simple registry cleaner, (CCleaner.exe works and it is Freeware) and reinstall MODI.

Both the licensed and Trial Version of Tiff to Text are supported.

Please contact us with any questions you have or anytime you need help with Tiff to Text III

eDocfile Inc.
2709 Willow Oaks Drive
Valrico, FL 33594
Phone 813-413-5599
info@edocfile.com <mailto:info@edocfile.com>
www.edocfile.com