An OPCUG Product Review

PDF Document Management Revisited
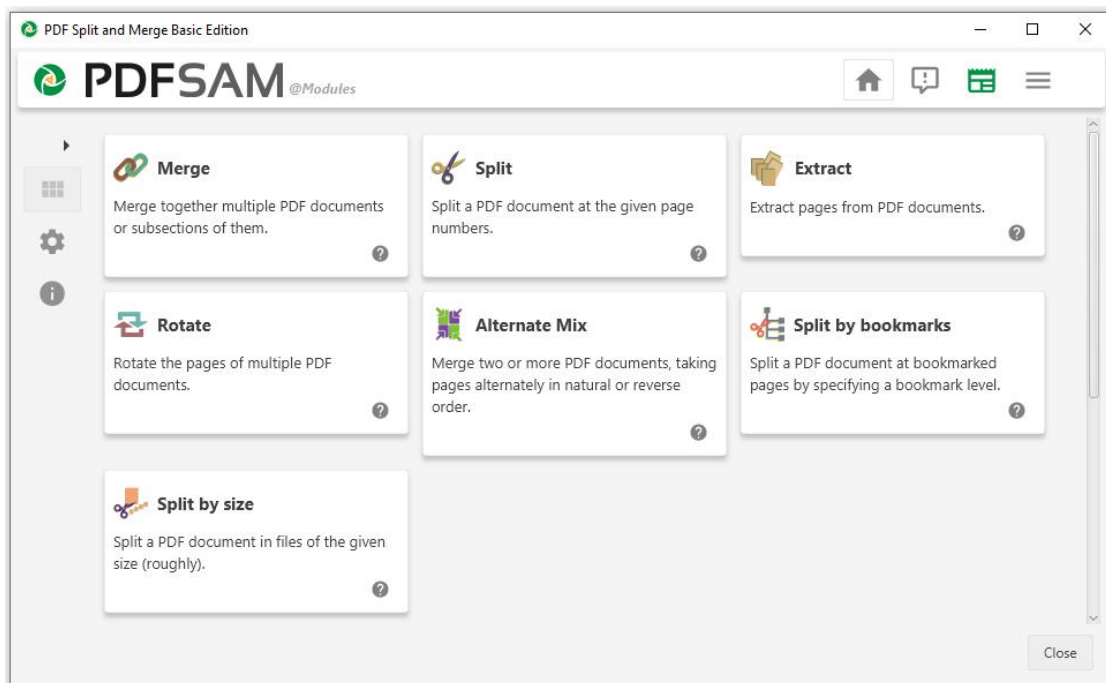### *By Alan German*

Recently, I received a single PDF file that contained 1647 pages as a combination of 30 separate documents. Normally, I would use the PDFTK Builder utility (see Modifying PDF Documents) to split up the original file into its constituent pages, and to then combine relevant sets of pages to form the individual documents as stand-alone PDF files.   However, not so normal was the behaviour of PDFTK Builder when operating on the subject file.   The split function did indeed break out 1647 individual pages, but it also managed to garble the order of the pages such that re-combining the pages for any given document proved to be next to impossible.

For example, the pages were named in sequential order, using the name of the original file and a _00x.pdf suffix (where x is the page number).   The problem was that while pages _001 to _003 were indeed pages 1 to 3 of the first document, page _004 was a page from the middle of a different document altogether!   Life is too short to spend time trying to manually correlate 1647 numbered pages into meaningful sequences, so a different solution was required.

Since PDFTK Builder is freeware, nothing is really lost by abandoning this program and seeking a similar utility. This is always my preferred course of action when things go south with a piece of software.   In the present instance, the new program that I identified was named PDFsam Basic, where "sam" is short for split-and-merge, and "Basic" denotes the free (and open-source) version of the software package.
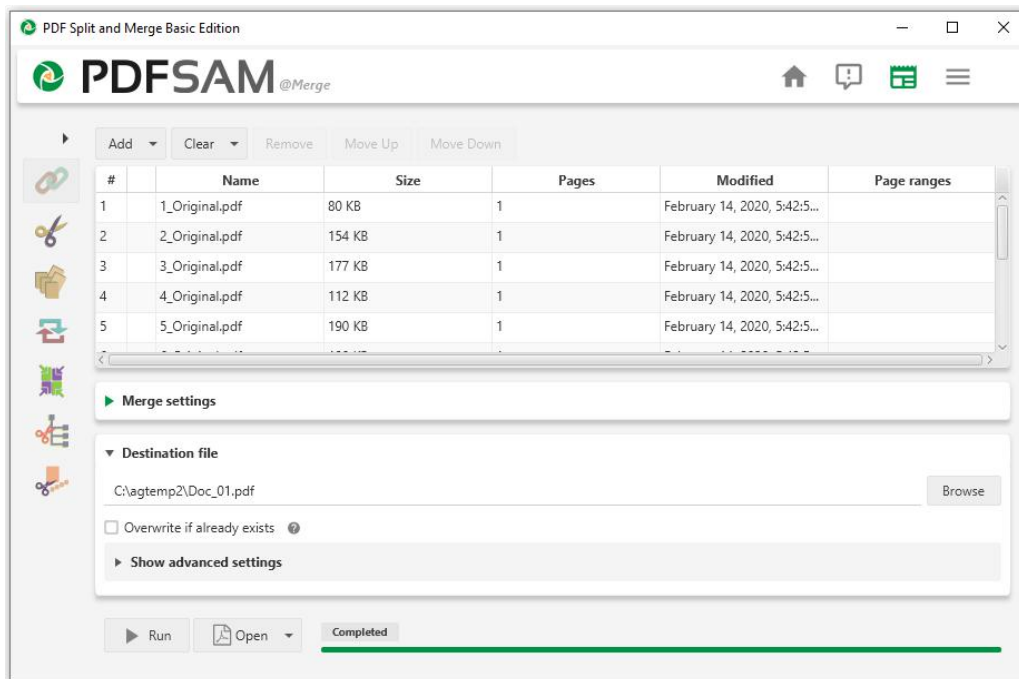
PDFsam Basic contains individual modules to perform merge, split, extract, and rotate operations on the pages of PDF files.   Furthermore, the split function can be invoked by page numbers, size, and bookmarks.   Another feature, named mix, is applied to two or more PDF files, taking pages alternately from the input files, a process designed to accommodate single-page scans of double-sided documents.

The first two of the available modules, Split and Merge, were eminently suitable for my purposes, these being the same procedures available previously with PDFTK Builder. PDFsam's split module extracted the 1647 pages as individual PDF files. By default the file names are in the form x_PDFsam_ and the original file name (where x is the page number). Some customization of the file names is allowed so, even though I was only using these on a temporary basis, I opted to modify the output files using an x_Original.pdf name format.

Once the progress bar had reached its end, and *Completed* was displayed, I checked that the numbered pages were those from the original file – and that they were in the correct sequence!

Selecting the *Home* icon then took me back to the main menu, from which I could select the Merge module. Now, it was a simple matter of using *Add* to select the pages I required for the first document, changing the name of the output file from PDFsam_merge.pdf to something more meaningful, and pressing *Run*. This process could then be repeated for subsequent documents by using *Clear*, and then *Add*, to create a new list of pages to be merged into other documents.

While split and merge can be used as indicated, PDFsam has another tool in its arsenal – the extract module – which is applicable to the current task. In effect, extract allows specific pages to be selected, as comma-separated numbers and/or page ranges, which are then copied from the original PDF file and written out to a new PDF file. The end result is the same as using split and merge, but extract is a more efficient means of performing this particular task.

As noted, PDFsam Basic has a number of other features. In addition, the program can be further customized through a variety of settings. The availability of the extract module in this utility, the fact that the program appears to work correctly with large PDF files, and its genesis as open-source makes this one a keeper!

Bottom Line

PDFsam Basic (Open-Source)
Andrea Vacondio
Version 4.1.1
https://pdfsam.org

Copyright © Ottawa PC Users' Group (OPCUG) https://opcug.ca/copyright-and-usage/