

PDF HACKS™

100 Industrial-Strength Tips & Tools



O'REILLY®

Sid Steward

HACK
#51

Split and Merge PDF Documents (Even Without Acrobat)

You can create new documents from existing PDF files by breaking the PDFs into smaller pieces or combining them with information from other PDFs.

As a document proceeds through its lifecycle, it can undergo many changes. It might be assembled from individual sections and then compiled into a larger report. Individual pages might be copied into a personal reference document. Sections might be replaced as new information becomes available. Some documents are agglomerations of smaller pieces, like an expense report with all of its lovely and easily lost receipts.

While it's easy to manipulate paper pages by hand, you must use a program to manipulate PDF pages. Adobe Acrobat can do this for you, but it is expensive. Other commercial products, such as *pdfmeld* from FyTek (<http://www.fytek.com>), also provide this basic functionality. The *pdftk* PDF toolkit [Hack #79] is a free software alternative.

Quickly Combine Pages in Acrobat

In Acrobat 6, select File → Create PDF → From Multiple Files.... Click the Browse... button (Choose... on a Macintosh) to open a file selector. You can select multiple files at once. On Windows, you can select a variety of file types, including Microsoft Office documents. Arrange the files into the desired order and click OK.

To quickly combine two PDF documents using Acrobat 5, begin by opening the first PDF in Acrobat. In the Windows File Explorer, select the PDF you want to append, drag it over the PDF open in Acrobat, and then drop it. A dialog will open, asking where you want to insert the PDF. Select After Last Page and it will be appended to the first PDF.

If you have a folder of PDF files to combine and their order in the Windows File Explorer is the order you want in the final document, begin by opening the first PDF in Acrobat 5. Next, in the File Explorer, select the remaining PDFs to merge. Finally, click the first PDF in this selection, as shown in [Figure 5-1](#), drag the selection over the PDF currently open in Acrobat, and then drop it. A dialog will open, asking where you want to insert these PDFs. Select After Last Page and they will be appended to the first PDF. Review the document to ensure the PDFs were appended in the correct order.



Acrobat also allows you to arrange, move, and copy PDF pages using its [Thumbnails view](#) [Hack #14].

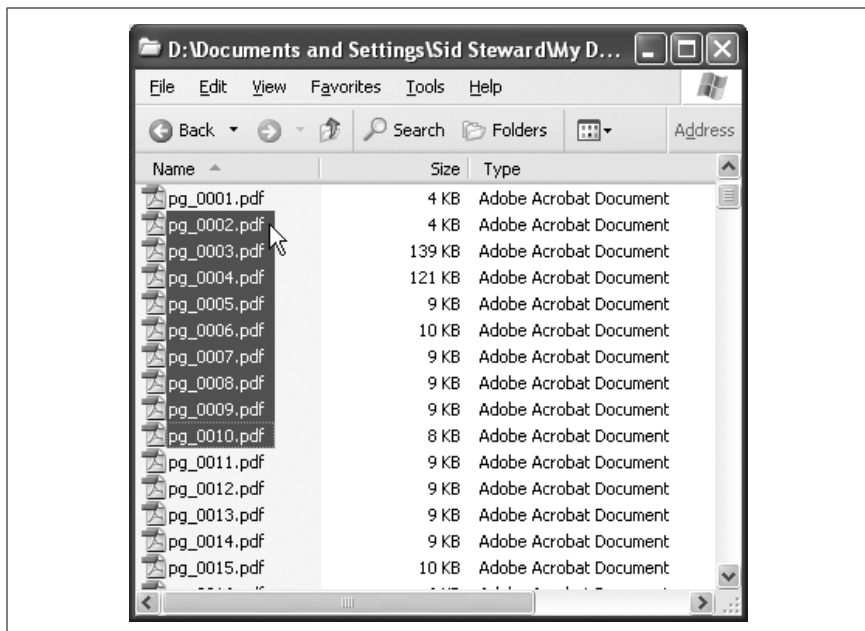


Figure 5-1. Clicking the first document in your selection when you drag-and-drop into Acrobat 5

Manipulate Pages with pdftk, the PDF Toolkit

pdftk is a command-line tool for doing everyday things with PDF documents. It can combine PDF documents into a single document or split individual pages out into a new PDF document. Read “[Tool Up with pdftk](#)” [Hack #79] to install pdftk and our handy command-line shortcut. pdftk is free software.

Open a command prompt and then change the working directory to the folder that holds the input PDF files. Or, you can open a handy command line by right-clicking the folder that holds your input PDF files and selecting Command from the context menu.



Instead of typing the input PDF filename, drag-and-drop the PDF file from the Windows File Explorer into the command prompt. Its full filename will appear at the cursor.

To combine pages into one document, invoke pdftk like so:

```
pdftk <input PDF files> cat [<input PDF pages>] output <output PDF filename>
```

A couple of quick examples give you the flavor of it. Here is an example of combining the first page of *in2.pdf*, the even pages in *in1.pdf*, and then the odd pages of *in1.pdf* to create a new PDF named *out.pdf*:

Split and Merge PDF Documents (Even Without Acrobat)

```
pdftk A=in1.pdf B=in2.pdf cat B1 A1-endeven A1-endodd output out.pdf
```

Here is an example of combining a folder of documents to create a new PDF named *combined.pdf*. The documents will be ordered alphabetically:

```
pdftk *.pdf cat output combined.pdf
```

Now, let's dig into the parameters:

<input PDF files>

Input PDF filenames are associated with handles like so:

```
<input PDF handle>=<input PDF filename>
```

where a handle is a single, uppercase letter. For example, A=in1.pdf associates the handle A with the file *in1.pdf*.

Specify multiple input PDF files like so:

```
A=in1.pdf B=in2.pdf C=in3.pdf
```

A file handle is necessary only when combining specific pages or when the input file requires a password.

[<input PDF pages>]

Describe input PDF page ranges like so:

```
<input PDF handle>[<begin page number>[-<end page number>[<qualifier>]]]
```

where the handle identifies one of the input PDF files, and the beginning and ending page numbers are one-based references to pages in that PDF file. The qualifier can be *even* or *odd*. A few examples make this clearer. If A=in1.pdf:

A1-12

Means the first 12 pages of *in1.pdf*

A1-12even

Means pages 2, 4, 6, 8, 10, and 12

A12-1even

Means pages 12, 10, 8, 6, 4, and 2

A1-end

Means all the pages from *in1.pdf*

A

Means the same thing as A1-end

A10

Means page 10 from *in1.pdf*

You can see from these examples that page ranges also specify the output page order. Notice the keyword *end*, which refers to the final page in a PDF.

Specify a sequence of page ranges like so:

A1 B1-end C5

When combining all the input PDF documents in their given order, you can omit the <input PDF pages> section.

<output PDF filename>

The output PDF filename must be different from any of the input filenames.

If any of the input files are encrypted, you will need to supply their owner passwords [\[Hack #52\]](#).