Working with PDF Files

Cal Esneault, Cajun Clickers Computer Club, LA

The Portable Document Format (PDF) file format is used extensively to pass documents between people and organizations that use different computer hardware and operating systems. Introduced as a proprietary format in 1993 by Adobe Systems, it was released as an open standard in 2008. With rare exception, if you have a PC you can read a PDF file (for example, Adobe Acrobat Reader for Windows OS). If you produce a documentwith almost any word processor, you can export a PDF version so that a recipient doesn't need to have your specific software to access it. As you browse the Internet, you will find many supporting documents are in PDF format for you to print them or for you to read them on an electronic display.

Although PDF files are easy to create and read, editing PDF files is more complex since the default is to produce a read-only file. You can usually copy text or images and paste them as components into other document software. However, many times we want toextract whole PDF pages without having to reformat the results after pasting smaller parts. PDF Shuffler, a small python-gtk open-source program for Linux, is a great method to easily rearrange, split, or merge pages from PDF files.

As an example, I downloaded a 14-page PDF file from federal government resources titled "Glossary of Computer and Internet Terms for Older Adults" and a 1 page PDFFile titled "Basic Computer Technology" from the New York State Library. After startingPDF Shuffler, added and joined both files by hitting the "+" icon.
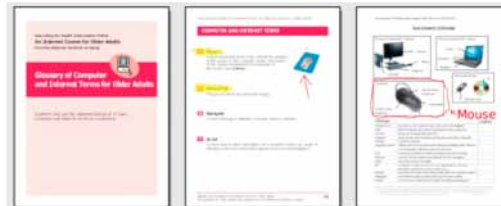
The following image is a screenshot of PDF Shuffler displaying the first 6 pages of the combined file. You can

then select any page and delete it. Pages can be rearranged by a "drag-and-drop" mouse action.

I removed all but pages 1 and 10 from the "Glossary" file and left the single page from the "Technology" file. I next saved the 3-page result as a single PDF file with a new unique name. This shorter customized version can now be sent to a recipient with just the specific information I desired. The simplicity of PDF Shuffler is its greatest attribute.

Although you may not want to change the content on a PDF page, you may want to draw attention to specific points. Xournal is a Linux application for taking notes or sketching with a stylus. It also has an "Annotate PDF" feature. Within Xournal, select a PDF file to annotate. You are then able to add a layer of lines or text above the original page. Below is a view of the saved version of the customized PDF file described previously. Yellow highlighting, arrows, and text were added to accentuate points about a computer mouse.

For filling out PDF forms, flpsed is an even simpler program for Linux systems which only inserts text.

The programs mentioned here are available in the repositories for many Debian-based distributions, such as Ubuntu or Linux Mint. Give them a try next time you are working with a PDF on your Linux box.

Author: Cal Esneault, Former President and leader of many Open Source Workshops & SIGs, Cajun Clickers Computer Club, LA