

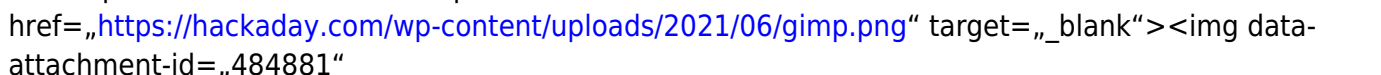
Linux Fu: PDF For Penguins

[Originalartikel](#)

[Backup](#)

PostScript started out as a programming language for printers. While PostScript printers are still a thing, there are many other ways to send data to a printer. But PostScript also spawned the Portable Document Format or PDF and that has been crazy successful. Hardly a day goes by that you don't see some kind of PDF document come across your computer screen. Sure, there are other competing formats but they hold a sliver of market share compared to PDF. Viewing PDFs under Linux is no problem. But what about editing them? Turns out, that's easy, too, if you know how.

GUI Tools

<https://hackaday.com/wp-content/uploads/2021/06/gimp.png>  <https://hackaday.com/2021/07/06/linux-fu-pdf-for-penguins/gimp/> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png?w=400> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png?w=800> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png?resize=250,140> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png?resize=400,225> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png?resize=800,449> <https://hackaday.com/wp-content/uploads/2021/06/gimp.png?resize=1536,862>

You can use lots of tools to edit PDF files, but the trick is how good the results will look. Anything will work for this: LibreOffice Draw, Inkscape, or even GIMP. If all you want to do is remove something with a white box or make an annotation, these tools are usually great, but for more complicated changes, or pixel-perfect output, they may not be the right tool. The biggest problem is that most of these tools deal with the PDF as an image or, at least, a collection of objects. For example, columns of text will probably turn into a collection of discrete lines. Changing something that causes a line to wrap will require you to change all the other lines to match. Sometimes text isn't even text at all, but images. It largely depends on how the creator made the PDF to begin with. If you don't mind using a Web-based tool, <https://www.pdfescape.com> PDFEscape is free and works very well. Other options include Scribus and Okular. Both of these tools can't really edit the file but can import them as images that you can further manipulate. For example, Okular's review mode can add annotations like highlights and freehand lines. Unsurprisingly, emacs can display a PDF file if it is running under X. You can use Control+C Control+C to switch to view a text representation. After all, most of the PDF file format is text and emacs can even handle binary files. So if you don't mind working inside the PDF format — very much like PostScript — you can do your editing in emacs or even another text editor. There are a few dedicated non-free editors out there and at least one <a

href=„<https://sourceforge.net/projects/pdfedit/>“ target=„_blank“>open-source PDF-specific editor. Of course, like most things in Linux, you can also use the command line.</p><h2>Hiding Text</h2><p>The problem with working with PDFs as text — even in emacs — is that they are often compressed and otherwise unreadable. For example, words may appear a character at a time separated by formatting code or other data. So searching for arbitrary text in the PDF may not work.</p><p>You can convert the file to use more uncompressed text, although that’s no panacea. For example, if you open up this segment from an article on ham radio and want to change the word “convention”, it is hard to tell exactly where that text is, but it is somewhere in this general area:</p><pre>3 0 obj << /Length 14770 /Filter /FlateDecode >>
streamH�|Wɒ�8��+p\$gJ,�c��v&
#65533;cS�Ҍc��J�\$���\ZV
5533;����\0��
�CTR�������r�A
533;[�}�7}������[]���
������l5u���`M�>
�/��?l�8�@��gBzq�r!A
533;%�AE���˜
᥉��x!\$��X8^%\$��A�Dᦙ
3;B���(���b�[H
�>����#��{a��&
#65533;e0\$^H&|/����U1\$^��#�
�/�G�Us��/,/,��\
<i�'qC���\$xe�"X�x22��&
#65533;����G��F�Lp]Mnm�\$
]
#TI��G�q�l��'3;!���
!+��{�����bʏ
33;�Qja����Q i�
GRn�\0�g;L����x�Zܿ&#
13107;�n�2�R&
:.,x�r�ky�[JPK��/���Sᦙ
3;�i��������]rA
533;F�p����k��
|���Ql�mx>1�\�1�Q�A
533;y)ХǺ�Z�U.^�](pN��dx�A
533;���֬,��{��cYa�\
�t�s�}�ْ{<\0ZW�:��Ȅ�
;Oɴ��cS�UzluP�֨o}ި��Uq
f��o��V��bT%mj|��tᦙ
3;����v�{s�Rj˺���</pr
e><p>Good luck finding it in that soup. You want to convert it to unpacked text.</p><pre>qpdf -qdf
input.pdf output.txt</pre><p>The resulting file is actually a PDF even though I named it .txt.
However, it has everything unpacked. That’s still not great, though, but at least you could
find the part you need to change:</p><pre>1.2632 -1.1242 TD0.0739 Tc0.1263 Tw(One potentially
confusing Stamp)Tj-1.2632 -1.1368 TD0.026 Tc0.1248 Tw[(con)38.6(v)20.7(ention is that the I/O pin
numbers)]Tj0 -1.1242 TD0.0262 Tc0.0072 Tw[<646f6e90>13.6(t correspond to the IC pin
numbers.))TjT*</pre><p>Again, good luck searching for the word “convention,” for
example. But it is still better than the first example. You can also find metadata even in unprocessed
files using things like /Author and /Title.</p><h2>Command Line Magic</h2><p>The

qpdf

tool can convert a PDF file to another PDF file. It can optimize the output for Web serving, text editing, and it can do simple things like remove pages or merge pieces of multiple files. You can read [the documentation](http://qpdf.sourceforge.net/files/qpdf-manual.html), but here we use the QDF mode to produce a legitimate PDF file with all the objects in numerical order and with normal Unix-style line endings. This allows you to more easily edit the file with a text editor, but as you've seen that doesn't always make it simple. Removing entire objects is a headache, but if you get rid of all the mentions of an object, you can run fix-qdf to recreate the proper QDF file.

Another way to make common edits to PDF files is to use [PDFtk server](https://www.pdfabs.com/tools/pdftk-server/) (PDFtk without the server moniker is a GUI toolkit for Windows). Using PDFtk you can merge or split documents, rotate pages, and do many other common tasks. For example, to join two files in order:

```
pdftk in1.pdf in2.pdf cat output output.pdf
```

You can omit, say, page 9:

```
pdftk in1.pdf in2.pdf cat 1-8 10-end output output.pdf
```

You can also shuffle merged pages in different orders:

```
pdftk A=in1.pdf B=in2.pdf shuffle A B output output.pdf
```

Text to PDF and Back

If you want to convert text into PDF from the command line you have several options. [Pandoc](https://pandoc.org/MANUAL.html) is an amazing tool that converts markdown to almost anything. It will not only convert markdown to PDF but just about anything else.

You can also use various combinations of [ps2pdf](https://web.mit.edu/ghostscript/www/Ps2pdf.htm) (along with a tool to generate PostScript), pdf2text (part of [poppler-utils](https://poppler.freedesktop.org)), or [Ghostscript](https://www.ghostscript.com) to create PDFs or strip text out of them. Ghostscript can do a lot, including convert a PDF to a number of image formats if you want to, say, display them on a Web page as an image.

Special Printing and Other Tools

[!\[\]\(cbe80b694ebd74fcfe136a095b608235_img.jpg\)](https://hackaday.com/wp-content/uploads/2021/06/printer.png) [data-permalink="https://hackaday.com/2021/07/06/linux-fu-pdf-for-penguins/printer-18/" data-bbox="115 545 475 562" data-orig-file="https://hackaday.com/wp-content/uploads/2021/06/printer.png" data-orig-size="962,734" data-comments-opened="1" data-image-meta="{"aperture":"0","credit":"","camera":"","caption":"","created_timestamp":"0","copyright":"","focal_length":"0","iso":"0","shutter_speed":"0","title":"","orientation":"0"}" data-bbox="115 562 475 595" data-image-title="printer" data-image-description="" data-image-caption="" data-medium-file="https://hackaday.com/wp-content/uploads/2021/06/printer.png?w=400" data-large-file="https://hackaday.com/wp-content/uploads/2021/06/printer.png?w=800" class="alignright size-thumbnail wp-image-484914" data-bbox="115 595 475 628" src="https://hackaday.com/wp-content/uploads/2021/06/printer.png?w=250" alt="" data-bbox="115 628 475 651" width="250" height="191" srcset="https://hackaday.com/wp-content/uploads/2021/06/printer.png 962w, https://hackaday.com/wp-content/uploads/2021/06/printer.png?resize=250,191 250w, https://hackaday.com/wp-content/uploads/2021/06/printer.png?resize=400,305 400w, https://hackaday.com/wp-content/uploads/2021/06/printer.png?resize=800,610 800w" data-bbox="115 651 475 684" referrerpolicy="no-referrer" data-bbox="115 684 475 717"/> Sometimes you want to modify a PDF file so it will print a certain way. We've already talked about how to merge odd and even pages, for example, but there are a few other commands you might want for this purpose:](https://hackaday.com/2021/07/06/linux-fu-pdf-for-penguins/printer-18/)

- [pdfxup](http://manpages.ubuntu.com/manpages/bionic/man1/pdfxup.1.html) – Uses pdflatex and Ghostscript to put multiple pages on one printed page (e.g., 2-up)

pdfjam

– Uses LaTeX to put documents on different size pages or produce multiple pages on one printed page

pdfposter

– Create giant output on multiple pages from a single page<p>If you prefer a GUI you might check out <a href=„<https://pdfsam.org>“ target=„_blank“>PDFsam basic. If you are interested in Java software, there is <a href=„<http://multivalent.sourceforge.net>“ target=„_blank“>Multivalent.</p><h2>Wrap Up</h2><p>As usual, there are many ways to do daily tasks in Linux. Sometimes the challenge isn’t doing the work, but rather finding the tool that best fits your style of working.</p><p>Oddly enough pandoc <a href=„<https://hackaday.com/2017/05/23/stupid-git-tricks/>“>keeps coming up for different reasons. If you prefer your documents on paper, you need a printer and <a href=„<https://hackaday.com/2012/07/16/dealing-with-the-horrors-of-pdfs-by-binding-your-own-books/>“>bookbinding clamp.</p> </html>

From:

<https://schnipsl.qgelm.de/> - Qgelm

Permanent link:

https://schnipsl.qgelm.de/doku.php?id=wallabag:wb2linux-fu_-pdf-for-penguins

Last update: 2025/06/27 11:17

