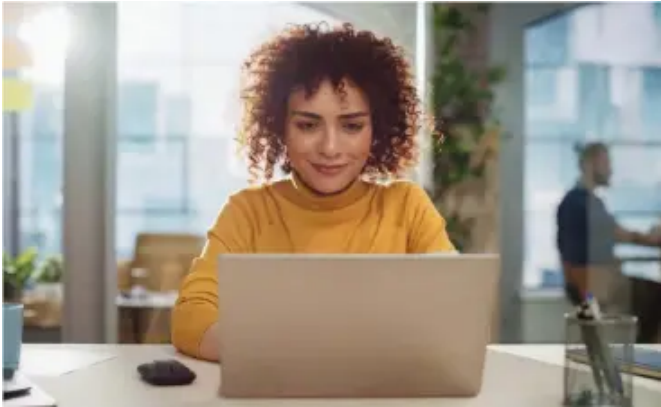- Utilities (Computer software)

Mastering PDFS: How to create, convert, and search

PDFS are very handy, but they have some quirks. Here's how to become an expert at handling them.

PCWorld (USA) · 1 Jan 2024 · 104 · BY ANDERS LUNDBERG

When it comes to preserving any type of document in a format that anyone can later open and be sure it looks right, it's hard to beat the PDF format. PDF, or portable document format, was developed by Adobe in the early 1990s to make it easy to share formatted documents regardless of the sender's and receiver's systems.

The format was based on and is essentially a simplification of postscript, a programming language developed by Adobe 10 years earlier to describe documents for printing in a technologyneutral way. PDF and postscript use code to describe where text, images, and graphic elements should be placed on a page. A rasterizer turns the code into either pixels on a screen or dots in a printout.

PDF files are extremely common and are used for all sorts of things: as an archive format for scanned and digitized documents of various kinds, as a standard format for scientific articles, for digital contracts, and to save something for later printing, just to name a few examples. READ PDF FILES Windows still has no built-in program for viewing and editing PDF files. The default program for opening a PDF in the system is Microsoft Edge, which like most browsers, has a built-in PDF viewer. It works really well for just reading, with the basic features you would expect. For example, you can view the table of contents and click on sections to jump directly to them, and mark up the files with tools like a marker pen and text annotations. You can also save and print.

If you want more functionality than that, you'll have to install a third-party program. If you don't need more-advanced editing features, there are many free alternatives. Adobe, who originally developed the format, has Acrobat Reader DC (fave.co/3gb5fhf). It's safe to stick with the original, and it's a good and relatively fast program that's particularly suitable for those who need to fill in forms, for example, but it uses a lot of resources and can be a bit cumbersome if you don't have a new, fast computer.

A free program that is much more lightweight is Sumatra PDF (fave.co/ 33TMOBT), which also handles ebooks in epub format and various other file types. Sumatra is fast even on older hardware, and unlike some other freeware, it's not full of adverts and doesn't try to push any paid features on you.

Other popular options include Foxit (fave. CO/3RBMYK0), PDF Gear (fave.co/3tknkkb), Pdf-xchange (fave.co/46sfyvc), and the open-source software Okular (fave.co/ 3GEDNMM).

SAVE AS A PDF VIA PRINTING

Windows 10 and 11 have a built-in feature to save almost any document as a PDF via a virtual printer. As long as the program you want to save from can print, you can select the PDF printer and get a PDF file. The printer is called Microsoft Print to PDF.

The problem with this solution for saving as a PDF is that it does not always produce files with text that can be selected, searched, and labeled. This is true, for example, if you save a web page as a PDF. From Word it seems to work better, but Word can already save as a PDF via Save as. Fortunately, Chrome, Edge, and Firefox now have a built-in Save as PDF function that you can also access via the print dialog box. Select the Save as PDF instead of Microsoft Print to PDF and it will save with text that you can select, copy, and search.

For other programs, you can experiment. If it doesn't work well with Microsoft's virtual printer, install another option to see if it works better. Cute PDF Writer (fave.co/3t9kq7p), for instance, works the same way, but you choose Cute PDF in the printer dialog instead of Microsoft.

CONVERT OTHER FORMATS TO PDF, AND VICE VERSA

In addition to saving a PDF via the print dialog, some programs can save directly to PDF via Save as. This works differently in various programs, so the result is sometimes different from printing but sometimes identical. Whichever method you use, you will need to save documents in the program's own format to be able to open and edit in that program again.

Another method of creating a PDF from a file of a different format is with PDF converters. They take files of all possible formats as input and spit out a PDF file. For some file formats, it works almost identically to the print method, but for others it can give different results. Which is better varies, so it's best to try both.

There are also converters that take a PDF file as input and convert the content to another format. The most common are Microsoft Office (Word, Excel, Powerpoint) and images (JPEG, PNG) but there are also others, such as PDF to EPUB for books.

The easiest way to convert single files to or from PDF is with an online tool. There are many similar sites with slightly different sets of features. Adobe also has a number of tools under the name Acrobat online (fave.co/3trahye), including converters to and from Word, Excel, Powerpoint, and JPEG. Other sites include ILOVEPDF (fave. CO/47PEVZD) and PDF Converter (fave.co/3nrbwd5).

COMPRESS PDFS

If you save a document as a PDF via printing or a built-in function, the resulting file is often very large. If you don't need high-resolution color images even though you will print the file, you can compress it considerably using one of a wide range of tools for this purpose. When a test article on Pcworld was saved via Chrome's built-in print function, the file went from 1.2 to 0.17 megabytes. On really large files, the reduction can be tens of megabytes, and if you have many PDF files, it will make a noticeable difference.

Here it may be appropriate to use one of the many online PDF tools available. Adobe has it online, although compression is otherwise only included in the paid version of Acrobat. It is also available at ILOVEPDF.

PASSWORDPROTECTED FILES

The PDF format supports protecting files with encryption and a password. It uses the AES encryption, and it is secure provided that the password is long. The vast majority of PDF programs can open and save encrypted files, so it is rarely a big problem that a file is password protected.

If you have a password-protected file but you don't want to have to fill in the password every time you want to open it, you can either open it and print it to PDF, making a new password-free copy of

it, or use a third-party program that supports it. Adobe Acrobat Pro works, of course, but you can also try PDF Gear (fave.co/3tknkkb), for example.

FAST, SMART PREVIEW WITH POWERTOYS

Microsoft Powertoys is a program that brings together a number of interesting and sometimes very practical features that are a bit experimental or too much in the niche for advanced users to be included in Windows. One of the most useful, Kika, is basically a copy of Apple's Overview feature in Mac OS.

Select a file in Explorer and use the keyboard shortcut Ctrl+space to preview the contents. Use the keyboard to move the selection to other files and view those instead. If you prefer a different key-board shortcut, you can change it in the Powertoy settings. The preview of PDFS in Kika supports tables of contents with links to chapters or sections. You can highlight and copy text and even print directly from here, so in many cases you can get by without opening the file in any program.

A third-party tool that works with more file types is Quicklook, available for free on Github (fave.co/41hup0w).

MANAGE PDFS WITH SCANNED TEXT

A common use of the PDF format is for scanned multi-page documents, books, and other originally handwritten or printed materials. Because they have been scanned, they are images and not digital text, so it is not possible to highlight text and copy or search them.

The solution if you want to be able to do something with the text in such files is software with OCR functionality. Several free programs have this built in, such as PDF Gear (fave.co/3tknkkb) and Pdf-xchange (fave. co/46sfyvc). The best is probably the open source program OCRMYPDF (fave.co/ 46NBCAR), but it is a command-line program and the installation procedure is somewhat complic-ated. For those who frequently handle scanned documents and books, it may be worth the effort, especially since it's easy to bake into automated workflows to process hundreds or thousands of files. For everyone else, it's easier to use an online service like ilovepdf.com. SEARCH IN PDFS

Since PDFS are often used for large documents, such as long manuals and course literature, it is common to have to search them. Unfortunately, this is not always as easy as in Word documents, for example. The reason lies in the way the PDF format works. Unlike in text document formats, the text does not lie as a block of unformatted text to which formatting is applied, but as a stream of charac-ters, together with information about where they are placed on the page, what font they are in, and so on.

For PDFS saved directly from Word or other programs, this is rarely a problem, but for scanned and Ocr-read files, it can be.

This is especially true when text is in columns, tables, lists, or text boxes. Try selecting two lines of text and see what happens—more often than not the selection continues in the wrong place. It can also cause problems for the search function—for example, when a word has been hyphenated and no longer has all the characters next to each other. The word "text document" hyphenated to "text-" and "document" will not appear in the search results for the full word.

So if you don't get a hit with a word you're looking for, you can try entering just part of the word, preferably a bit that normally comes before or after any line breaks and is as unique as possible. In the above example, you can try both text and document, or just "docu" as it is not a letter combination that is included in many other words other than document and words formed from the same root.

FILL IN PDF FORMS

One thing that sets PDFS apart from most other file formats is that they can be made interactive. Many organizations send forms as PDFS to be filled in. These can include input fields for text such as

name and address, tick boxes, and sometimes even multiple choice menus.

How well such forms work with PDF programs other than Adobe Acrobat (including Reader DC) varies. Some work well in most programs, while others do not work at all. Therefore it may be a good idea to have Acrobat Reader DC installed on your computer even if you do not have it set as the default program for opening PDF files.

If the preset fields don't work well, one tip is to use the markup tool to add text yourself. This works even on forms that have no text fields.

SIGN PDF DOCUMENTS

A special class of PDF forms are documents that need to be digitally signed. Here it is important to distinguish between two different types of signatures: images or vector graphic elements that look as much as possible like your handwritten signature (also known as an electronic signature), and fully digital signatures.

The former is used for many kinds of documents both between individuals and between organizations and individuals: for example, on a rental contract where there is also plenty of documented communication to prove that an agreement has been made. Acrobat Reader DC has a built-in function for this, and you can sign with a scanned image of your signature, a signature you draw directly on the computer, or just your name. Right-click in an open PDF file and select "Sign yourself."

The latter has a higher level of security, and the person signing must prove their identity. The EU has developed common rules for valid digital signatures, and in Sweden the most common way to authenticate yourself is with Bank ID. This is used, for example, to sign contracts for property sales and other important matters.