

linux.org

Just a Wonderin'

9-11 minutes



-

- [#1](#)

Does any one know of an application that can convert full web pages into PDF format?

As it would make reading easier for this little black duck and also keep my files tidier.



-

- [#2](#)

Does any one know of an application that can convert full web

pages into PDF format?

As it would make reading easier for this little black duck and also keep my files tidier.

No, I wish I did, and I hope someone can offer some suggestions. You can tell your browser to "print" a web page, and then you can direct the print to a file which will create a PDF. Sometimes this looks okay, but many times it looks terrible and scrambled all over the place. It would look bad like that if you actually print it to paper too. You can do a "print preview" on any web page and get an idea how the PDF will look.

The web page "code" (HTML, CSS, and other things) create the proper look you see in the browser, but printers do not understand it. One simple problem is that the web page you are viewing is usually much WIDER than your standard piece of paper (in portrait view). If you could scale it to fit, it would be too small to read.

Firefox, and probably other browsers, will let you "Save Page As" (either Web Page Complete with images, or as HTML Only) but I guess that is what you are doing now. Saving the complete web page does keep it intact so you can view it properly later.

The only way that I've found is to capture a screenshot of a web page (usually just part of a web page) which saves an image file. That image can be imported into LibreOffice Writer and then exported as a PDF. Or the GIMP image program will

also export to PDF, I think. There are probably many tools to do this, but it is not the solution you are looking for.

Cheers



-

- [#3](#)

Saving the complete web page does keep it intact so you can view it properly later.

Yep that it does!

Reading a web page is like trying to read a newspaper that is written as one column from one edge to the other. That is why I would like to be able to convert it in to a PDF file, as reading it then is just reading an A4 page of writing a lot easier and less distracting.



-

- [#4](#)

Would something like this be suitable?

Save the file to say Downloads, then in Firefox (did we ask what Browser?)

File - open - (navigate to file) and click

Edit - added BTW

BTW I'll play with this, might have to tweak between portrait and landscape to get all print, but shows promise.

- 157.3 KB Views: 15



-

- [#5](#)

That looks nice, but it is still scrambled. The right-side column (Staff Online, Members Online, ads, etc) are still shoved to the end of the document. Sometimes that may not matter, but sometimes it will (to me). Are there settings to overcome this?

The [monthly subscription price](#) is a little steep for me too. I guess that would remove the advertising on every page though. Ouch!

Last edited: Yesterday at 2:40 AM



-

- [#6](#)

Then how about this one? Free version and Pro version.

It saves to PDF and displays it onscreen - too small but 250% was about right for me.

Then save download, open that in Firefox and Automatic Zoom makes it fine for me, addresses those issues that Stan has above, too.

- 220.4 KB Views: 15

Last edited: Yesterday at 6:37 AM



-

- [#7](#)

Then how about this one? Free version and Pro version.

I appreciate your efforts, my friend! But it is still scrambled when I view it. Is it just me? Or my PDF viewer? Maybe [@Nik-Ken-Bah](#) will have better luck when he awakes.

I see the dark blue linux.org "header graphic" (or part of it) across the middle of page 3. And no header graphic is displayed at the top of the document. Online people and Latest Posts are still moved to the bottom of the document. But the advertising is gone now. I see the same thing whether viewed in Firefox or saved to my desktop and opened without Firefox. I love the clear crisp quality of the PDF to show the

text, but I've never found a solution for layout issues. Of course I have not looked for a solution in a long time either, so maybe this is possible now. Or maybe not.

And with that, It's bedtime for me. Breakfast with friends in the morning.

Cheers



-

- [#8](#)

OK, guess we are whistling in the wind until the OP can tell us what Browser he is using.

Naaahh, I'm joshing you

If this is not the definitive answer, then it's my best for now, and I have to go fry more fish elsewhere on the forum (& kill some spammers).

I started thinking of wget solutions (for The Viewers who do not know, wget is a command you can issue at Terminal to download files from the Internet without the use of a browser).

Along the way, I came across a gem called

wkhtmltopdf

which you can install from tar or deb or rpm, and then issue

the command followed by the website page you want to use (don't have to save it) and generate a file with your own choice of name.pdf

You can either use File - Open on your Browser and view it in that (online or offline) or else get a PDF Viewer such as evince

and use that

I am writing this from Arcolinux, which has Evince already installed, but it is in many Distros' Repositories.

So without further ado, here is

- 174.3 KB Views: 15



-

- [#9](#)

So without further ado, here is

Thank you Wiz appreciated.

I'll have to check it out on the morrow in the morn as I have no PDF reader for Vindows and Vindows is the drama queen.

Use Minty for reading PDF's through doc viewer , no dramas with it.





-
- [#10](#)

Installed evince... looks same as Xreader for me, in Firefox and standalone. Page header graphic is back at the top in your WizardRocks.pdf, but your avatar and username is partially cut in two at the page break between pg1 and pg2. Right-column info still pushed to the bottom of the document. I adjusted evince settings many different ways, but it did not change the layout at all.

I'll try to locate that sometime today and see if I can make any progress with it instead.

[EDIT]

Sadly, not much difference here either. Right-column info still pushed to the bottom of the document. The header graphic looks different than the others, but none actually look like the web page header anyway. This one strips the avatars from all of us, but it does show images of "likes" and "emojis." I looked briefly through the man page for wkhtmltopdf, but I didn't spot anything that would cure my complaints. I did try with --background and --images options, but the avatars still did not appear.

[/EDIT]

Again, thanks Wizard! I'm glad [@Nik-Ken-Bah](#) asked this question, and I do hope a solution can be found, especially a

free solution. It is a more difficult problem than most people realize. If it were simple, then my Firefox print function would be satisfactory. Sometimes it is, but not often.

Cheers

Last edited: Today at 1:44 AM



-

- [#11](#)

Well I checked it out and as Stan said about the first, about the header ending up in the middle of the article, it was!

The second one was better and more readable.

What I had to do to create the PDF was download as a HTML file.

Then open in Libre Office which does a fairly decent job at displaying it as a text file including all the Pics.

Highlighted the document and cleared the formatting to ease editing of the doc, pain in the derriere otherwise.

I had to then edit it so that it was more presentable page wise and text was where it should end and start on the page.

Had to adjust the size of the pictures to fit in the frame of an A4 page.

With the pics had to display it with KItty or something like that so I could open it with Pix so that I could crop the picture so it

just displayed the relevant detail more clearly and reduce the size of the picture so that I can rearrange them more neatly.

When opening it in Libre Office reformat the page to Landscape and resize it to A3 size.

move the right width bar so that it sits on the 18 cm mark or equivalent in other rule markings.

When everything is within the 18 cm borders reformat the page to portrait and back to A4 size.

Then when you have it all edited then export it as a Direct PDF file.

A lot of flamin' messing about to get there. A reason that I asked whether was an application that could do it directly.

Input HTML file, Output PDF file.

I have done this before creating a PDF file but that was with pure text in ODT format that I personally created.

Libre Office handles HTML files slightly better than Open Office does. But there are items in AOO that are lacking in Libre Office and one in particular is being able to reduce or increase the line spacing for a number of lines without effecting the paragraph or the rest of the page. It hides in the right hand tool menu.