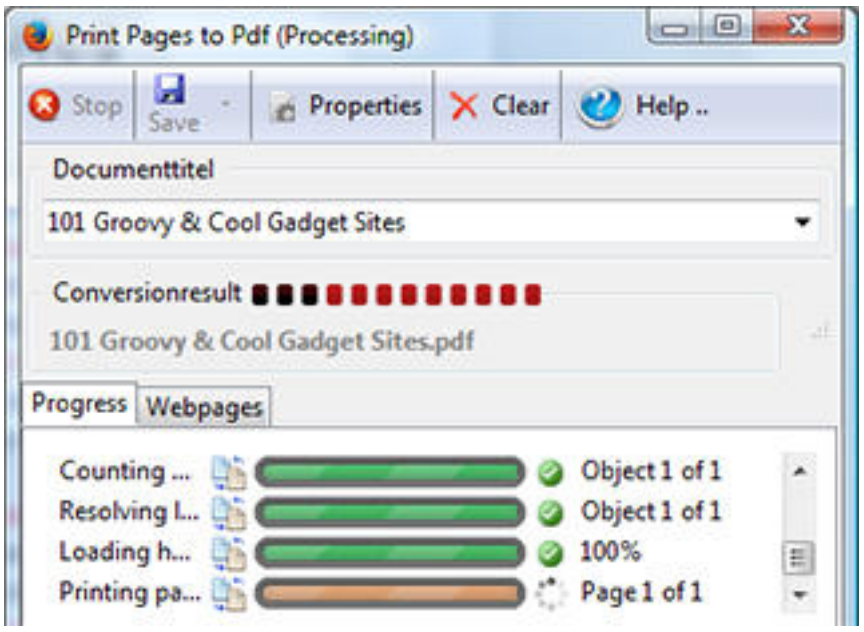
'Driverless' Printing: Generating PDF files from a Web Page

John R. Joyce, Ph.D.

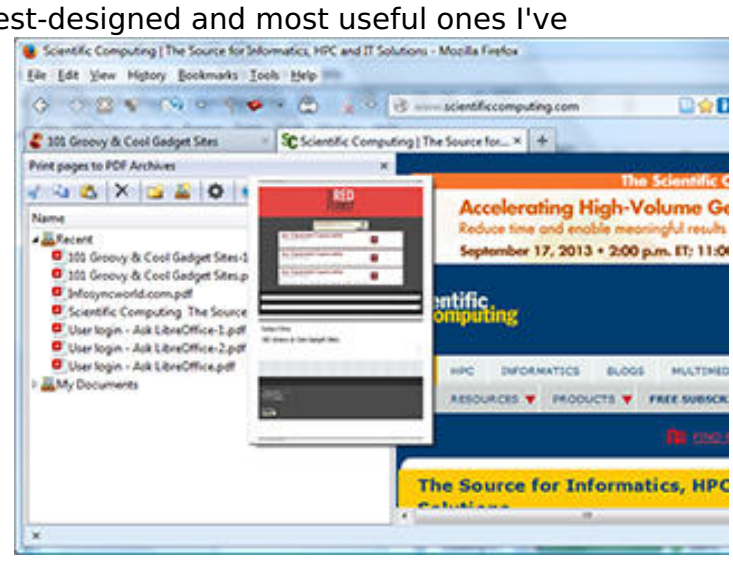


One of the most frustrating things about performing online research in a controlled facility, such as a library, is that frequently the only option for saving a copy of the search results is to use their specified printer at 50 cents to $1 a page, possibly more. Many of these sites purposely do not have PDF, or even Microsoft Silverlight, virtual print drivers installed either. So, what is one to do, acquiesce and pay their price for a hardcopy that you would have to scan and employ *optical character recognition* (OCR) to store electronically? Nonsense, for there are always ways around any problem!

To nip a few issues in the bud, yes, most browsers will allow you to save the Web page, usually in HTML, which you can later load and print out. However, while this is an approach you might have to take on occasion, it is a major productivity hit. Another way around this issue is to use a Web browser, such as Firefox for Android, that allows you to directly save the Web page being viewed as a PDF file. Google Chrome also supports the ability to save a Web page with an option under its Print menu. Unfortunately, not all Web browsers support such a feature. Since a controlled facility is...unlikely...to allow you to install any print drivers or other applications on your own, what other options do you have?

Well, I have found what I believe to be an excellent solution, at least for Mozilla Firefox users, though in a locked down environment you might have to use FirefoxPortable to access it! This wonder is a plug-in called Print pages to PDF[1] by The Ripper. And no, before anyone gets upset about trusting and using software by one of those nasty hackers or pirates, The Ripper is just the user name for Reinhold Ripper, a developer in Germany. He originally developed this plug-in in response to his frustration with performing research and generating a list of links, only to have many of them disappear over time.

I've worked with many different Firefox plug-ins over the years, and this is one of the best-designed and most useful ones I've encountered. As the name implies, this plug-in is designed to generate PDF files from a



Web page, but that is something like stating that the sun is hot! I won't attempt to provide an in-depth examination of it here, as Martin Brinkmann[2] of ghacks.net has already posted a good summary of of it. For additional details, you can also peruse the plug-in's Web site[3], which not only describes the features in depth, but is generously illustrated to help clarify both the features and their use. I'll restrict myself here to providing a high-level overview of the plug-in and caveats regarding potential issues you might encounter.

First off, you might encounter issues with simply locating the plug-in to install. For reasons known only to itself, Mozilla has gone to a very aggressive upgrade schedule for Firefox, which has included frequently changing how plug-ins interface and behave. Unfortunately, this has resulted in breaking a number of plug-ins, including Print pages to PDF, requiring them to be updated by their authors. By itself this is not an issue, as it is not uncommon for this to happen to one plug-in or another. Unfortunately, Mozilla's review cycle is currently longer than their update cycle, so they frequently have a new version of Firefox out before the last fixed plug-in update has been reviewed. Fortunately, there is a way around this. While the automatic update feature won't install it, you can go to the plug-in's version page[4] at https://addons.mozilla.org/firefox/addon/print-pages-to-pdf/versions/ and download the latest, but unreviewed, version of the plug-in directly. I've found The Ripper to be very responsive in terms of quickly posting an updated version of his plug-in whenever a fix is needed. When you tell the system to install the plug-in into Firefox,, it will warn you that this version of the plug-in has not been reviewed and ask for confirmation that you really want to install it. Just click on the Yes button to continue and complete the install. If prompted, tell the system to restart Firefox to complete the installation. That's all you need to do to be able to generate PDF files with this application.

On initial installation of this plug-in, the default configuration is to insert a new menu option under Tools to capture the Firefox tabs as a PDF file. If you go into Preferences, which you can do by clicking on the drop-down menu option on the Print page to PDF entry under the Firefox Tools menu, you can readily configure the plug-in to also display capture buttons in the Tools and/or Status bars. At this

point on clicking the newly displayed plug-in button, the default configuration is to generate a PDF for all of the open tabs. If you click on the drop down menu portion of this button, it will provide a list of conversion options, which includes generating a PDF of just the active tab or generating a PDF containing just the test from the Web pages. Again, if you go into Preferences you can alter the default button function to take on any of the capture options.

One of the reasons that I'm so impressed with this plug-in is the degree of configurability built into it. Other behaviors that you can configure include:

- Whether the conversion window remains open or closes upon capture completion
- Whether the PDF is immediately generated or waits for you to configure conversion options
- Whether the plug-in generates a single PDF file when capturing multiple tabs or individual PDF files for each tab
- Whether the PDF is automatically opened in your default viewer after generation
- PDF conversion options such as page size (initial default is A4), whether links are maintained in the PDF, whether to include the Web page background in the PDF, whether to generate an outline of the PDF, as well as default headers and footers to use, et al
- A default storage directory for generated PDF files
- Many more ...

There is one recent enhancement to this plug-in that can potentially trip you up. Previously, Print pages to PDF ignored any Javascript code embedded in the Web page. Currently, it attempts to execute this code to generate a more accurate representation of the selected Web page. Unfortunately, depending on the nature of this code, it may trigger errors preventing the creation of the PDF or significantly extending PDF generation time as it waits for each Javascript error to sequentially time out. Fortunately, if you encounter such a situation you can go back into the plug-ins configuration and un-check the 'Javascript' option on the Javascript tab under 'PDF (Webpage)'.

To give credit where it is due, this plug-in uses the Open Source library wkhmltopdf[5] originally written by Jakob Truelsen to perform much of the actual conversion, so is constrained by the capabilities of this library. In turn, this library is a shell that uses the Webkit widget that was built into the QT[6] cross-platform application and user interface framework after version 4.4.

If you are a Firefox user, you owe it to yourself to check out this plug-in. If you aren't but encounter situations where you need to capture Web pages, but are constrained by the environment in which you are working, perhaps you should give it a try. Its flexibility and extensive conversion options make it more capable than the native PDF conversion options of the other browsers I've seen. The ability to access it using FirefoxPortable is simply a bonus.

References

1. Print pages to Pdf :: Add-ons for Firefox. *Mozilla Corp.* (2013). https://addons.mozilla.org/en-US/firefox/addon/print-pages-to-pdf

2. Brinkmann, M. Print Multiple Tabs or Bookmarks As One PDF Document In Firefox | Ghacks. *gHacks.net* (2012). http://www.ghacks.net/2012/03/25/print-multiple-tabs-or-bookmarks-as-one-pdf-document-in-firefox

3. Overview - Print pages to Pdf. *Print Pages Pdf* (2012). http://printpagestopdf.de.vu/index.php/en/help.html

4. Print pages to Pdf :: Versions :: Add-ons for Firefox. *Mozilla Corp.* (2013). https://addons.mozilla.org/en-US/firefox/addon/print-pages-to-pdf/versions

5. wkhtmltopdf - Convert html to pdf using webkit (qtwebkit) - Google Project Hosting. *wkhtmltopdf* at http://code.google.com/p/wkhtmltopdf

6. Qt Project. *Qt Proj.* http://qt-project.org [1]

John Joyce is a laboratory informatics specialist based in Richmond, VA. He may be reached at editor@ScientificComputing.com [2].

Source URL (retrieved on *09/01/2015 - 10:27pm*):
http://www.scientificcomputing.com/blogs/2013/09/driverless-printing-generating-pdf-files-web-page

Links:
[1] http://qt-project.org
[2] mailto:editor@ScientificComputing.com