



por Guido Socher ([homepage](#))

Sobre el autor:

Hace algún tiempo comentamos que linuxfocus deseaba tener los artículos disponibles como archivos pdf. Como respuesta recibimos una serie de sugerencias que están resumidas en este tip. Muchas gracias por todas esas sugerencias.

Traducido al español por:
guadapeich ([homepage](#))

Introducción

Probablemente habrás notado que ahora tenemos archivos PDF para todos los artículos en idiomas que usan los códigos de caracteres iso8859-1. No fue fácil conseguirlo ya que deseábamos que se generaran automáticamente para evitar que las páginas html y los documentos PDF difirieran.

Aquí está nuestra experiencia con una lista de posibilidades para generar PDF.

La idea

Todos los sistemas linux tienen la utilidad de ghostscript ps2pdf. ps2pdf trabaja muy bien y la calidad del PDF generado es buena. En otras palabras, siempre podemos generar archivos PDF si manejamos el documento como un archivo postscript.

El sistema de impresión de linux está basado completamente en postscript así que ¿sería fácil?. Realmente el problema es hacerlo desde la consola con un script. No quieres utilizar el ratón cuando necesitas convertir unos cuantos miles de artículos.

LF Tip: generando PDF desde páginas html



Resumen:

Desde ahora en LinuxFocus incluiremos al menos un consejo cada mes. Si tienes alguno, envíalo a guido@linuxfocus.org

Si no te preocupan los colores, las tablas y las imágenes entonces una combinación de "lynx -dump | nenscript" y ps2pdf hará el trabajo. Sin embargo si necesitas tablas e imágenes sigue leyendo.

Los candidatos

html2ps

Es un perl script y la versión probada aquí fue html2ps 1.0 beta3. La página del proyecto es <http://user.it.uu.se/~jan/html2ps.html>

El programa trabaja bastante bien. Sin embargo depende de una serie de módulos de perl y tiene problemas con páginas que usen tablas para dar formato al html. Es una buena solución si tienes una composición simple.

latex

Existe un conversor de latex a PDF. Utilizando xslt puedes convertir html a Latex. Un requisito para esto es tener un documento html con sintaxis correcta. Puedes conseguirlo con la utilidad tidy:

```
HTML --(tidy)--> XHTML --(XSLT)--> Latex --(pdflatex)--> PDF
```

No investigué esta posibilidad porque xslt y latex me parecieron muy pesados y complejos.

control remoto del navegador

Si de alguna forma fuera posible controlar remotamente un navegador entonces tendríamos la ventaja de que el PDF generado sería idéntico a lo que normalmente ves en tu navegador. El problema es que es necesario un display X11. No es entonces posible correrlo desde una tarea de cron.

El proyecto mozilla ha mejorado la impresión y presentación pero sin embargo suprimió algunas de las opciones de control remoto que nestcape communicator tenía. La solución siguiente sólo trabajará con communicator 4.X

```
netscape -noraise -remote "openurl(http://somepage) "  
sleep(10) # there is no way to know if the page is completely loaded  
# so we just wait a bit  
netscape -noraise -remote saveas(somepage.ps,PostScript)  
sleep(10)  
ps2pdf somepage.ps
```

Algunos lectores pensaban que la impresión remota también sería posible con konqueror pero nadie consiguió dar una solución que funcionase.

htmldoc

Htmldoc es una utilidad muy bien escrita que se encuentra en <http://www.htmldoc.org/>. El siguiente comando hace exactamente lo que queríamos:

```
htmldoc -t pdf --webpage -f file.pdf file.html
```

Usamos la versión 1.8.24 y trabajó perfectamente. El único problema es que el archivo pdf resultante es en comparación 10 veces mayor que cualquiera de los PDF generados por las otras soluciones sin importar que opción de compresión se use. Un gran problema si tienes miles de documentos.

Conclusión

Finalmente usamos una combinación de control remoto de netscape y htldoc. No confiamos sólo en htldoc debido al tamaño de los archivos generados. Si tú tienes nuevas sugerencias e ideas con respecto a este asunto entonces escríbenos.

<p><u>Contactar con el equipo de LinuFocus</u> © <u>Guido Socher</u> "some rights reserved" see linuxfocus.org/license/ http://www.LinuxFocus.org</p>	<p>Información sobre la traducción: en --> -- : Guido Socher (homepage) en --> es: guadapeich (homepage)</p>
--	--

2005-04-11, generated by lfparsr_pdf version 2.51