

## MINICOMO EN LINUX METADATOS EN EL ARCHIVO PDF

Por: Milton Marte Feliú

Palabras clave: metadatos, archivo, pdf, web, semántica, consola, Linux

La importancia de los metadatos de los archivos en la [web semántica](#), radica en la facilidad que ofrece a las aplicaciones de búsqueda para que el usuario termine encontrando la información de manera fácil en el menor tiempo. Es definido por Agudelo Benjumea, M. (2009) como "...un registro de metadatos consiste en un conjunto de atributos o elementos necesarios para describir un recurso determinado, que funciona como identificador de los materiales digitales diseñados", otra definición se puede leer en wikipedia (2011) "La **Web semántica o Web 3.0** (del [inglés semantic web](#)) es la "Web de los datos"...[\[1\]](#) Se basa en la idea de añadir [metadatos semánticos](#) y [ontológicos](#) a la [World Wide Web](#)", queda claro que al elaborar un documento no importa el fin, esa parte no podrá ser olvidada.

Unir varios pdf con la herramienta de consola pdftk.  
Para instalar por consola:

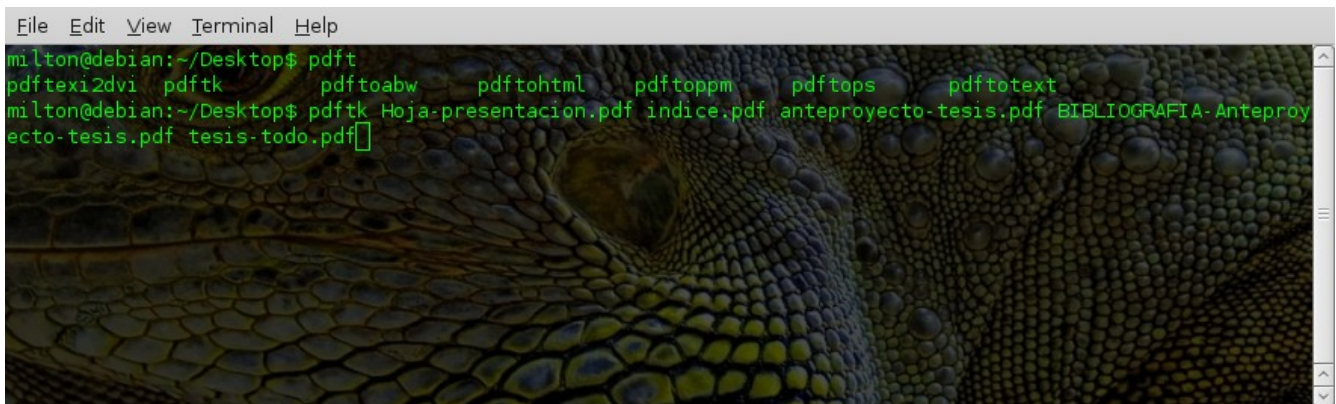
`sudo aptitude pdftk` (Ubuntu)  
`apt-get install pdftk` (Debian)

Comando:

`pdftk Hoja-presentacion.pdf anteproyecto.pdf indice.pdf bibliografia.pdf cat  
output tesis-completa.pdf`

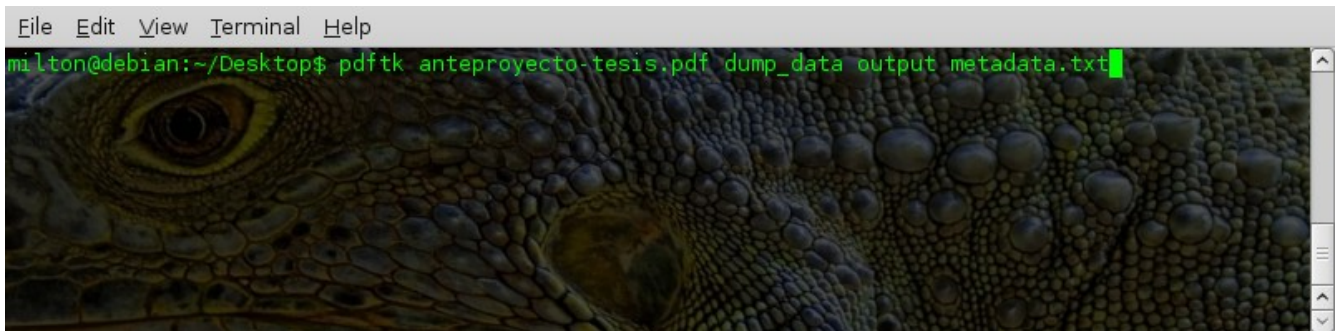
Ayuda

`man pdftk`



```
File Edit View Terminal Help
milton@debian:~/Desktop$ pdft
pdftexi2dvi pdftk pdftoabw pdftohtml pdftoppm pdftops pdftotext
milton@debian:~/Desktop$ pdftk Hoja-presentacion.pdf indice.pdf anteproyecto-tesis.pdf BIBLIOGRAFIA-Antepro
yecto-tesis.pdf tesis-todo.pdf
```

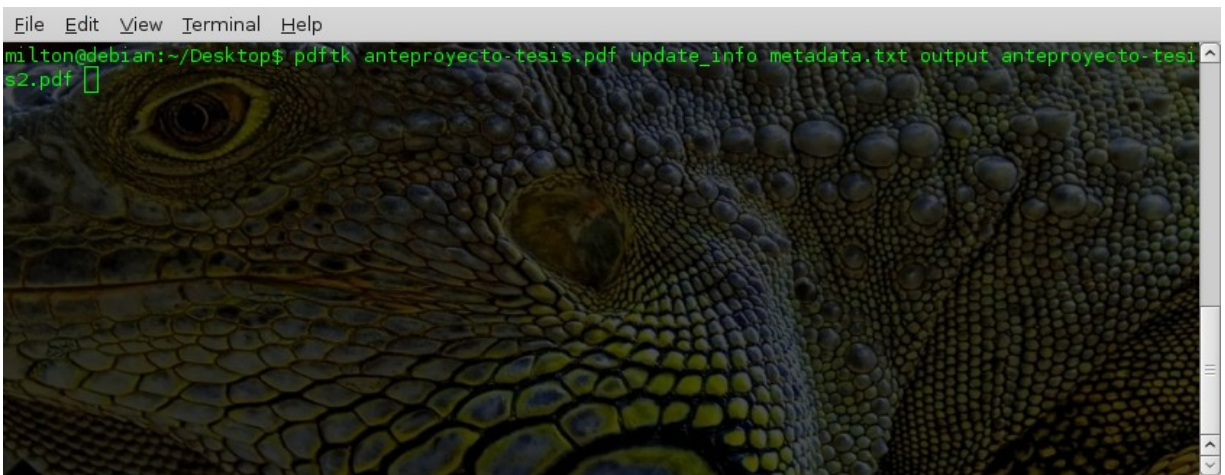
Para extraer los metadatos de un pdf este comando generará un archo de texto.  
`pdftk documento.pdf dump_data output metadata.txt`



```
File Edit View Terminal Help
milton@debian:~/Desktop$ pdftk anteproyecto-tesis.pdf dump_data output metadata.txt
```

Actualizar los metadatos

`pdftk documento.pdf update_info metadata.txt output nueva-copia.pdf`



Copiar en texto.txt

InfoKey: Title

InfoValue: Metadatos

InfoKey: Author

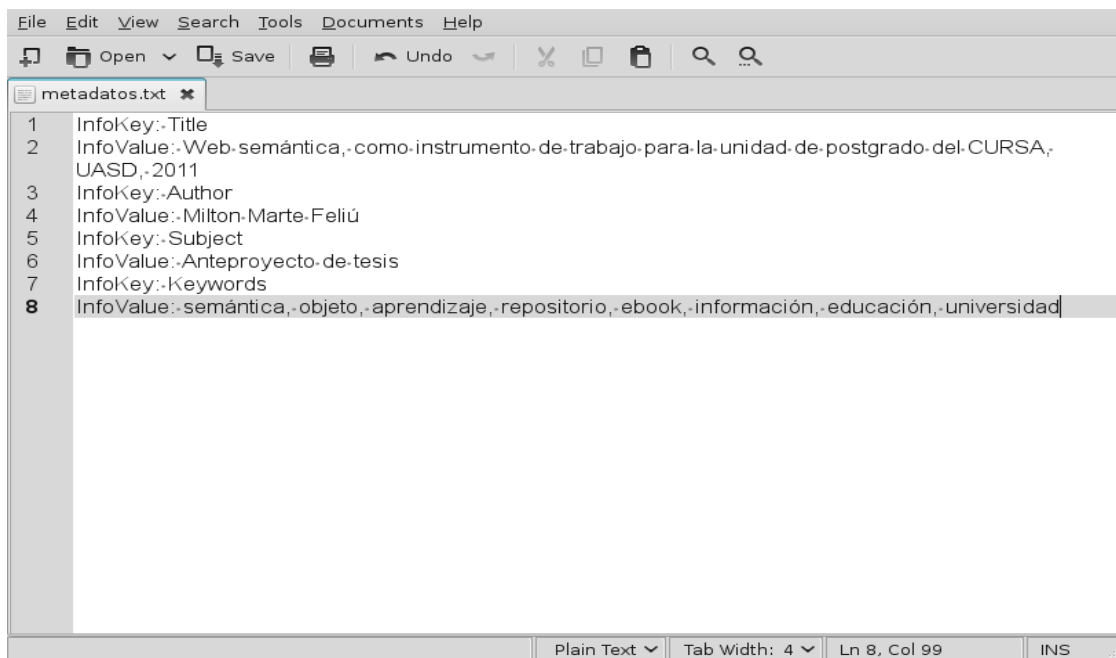
InfoValue: Milton Marte Feliú

InfoKey: Subject

InfoValue: Minicomo en Linux

InfoKey: Keywords

InfoValue: semántica, web, pdf, metadatos



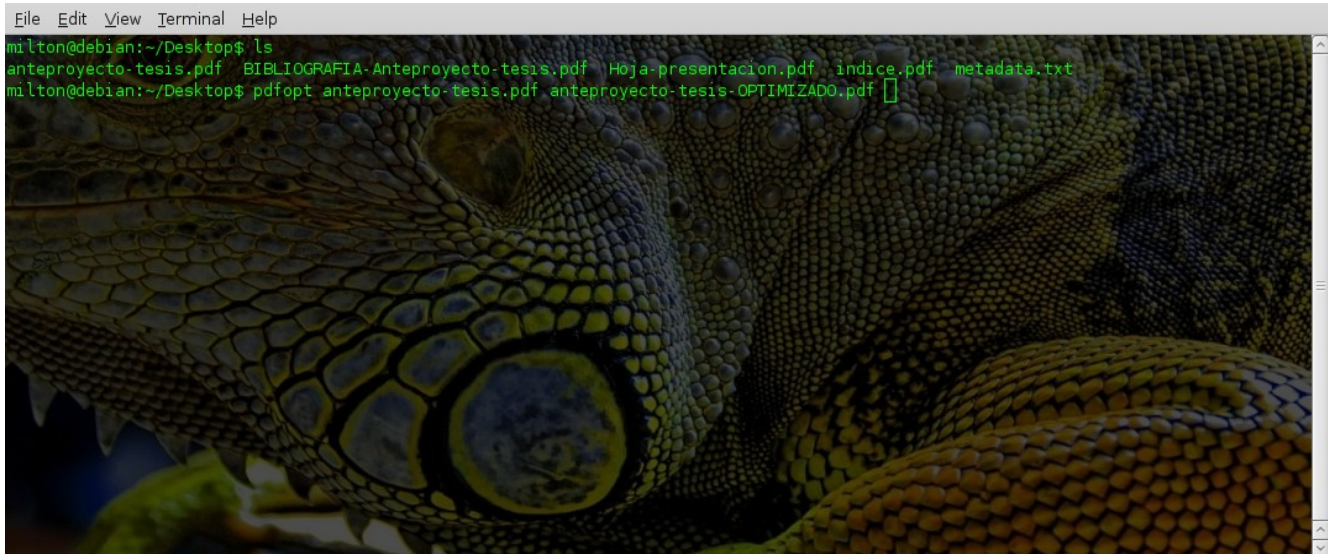
Optimizar el pdf

Instalar a Ghostscript PDF Optimizer

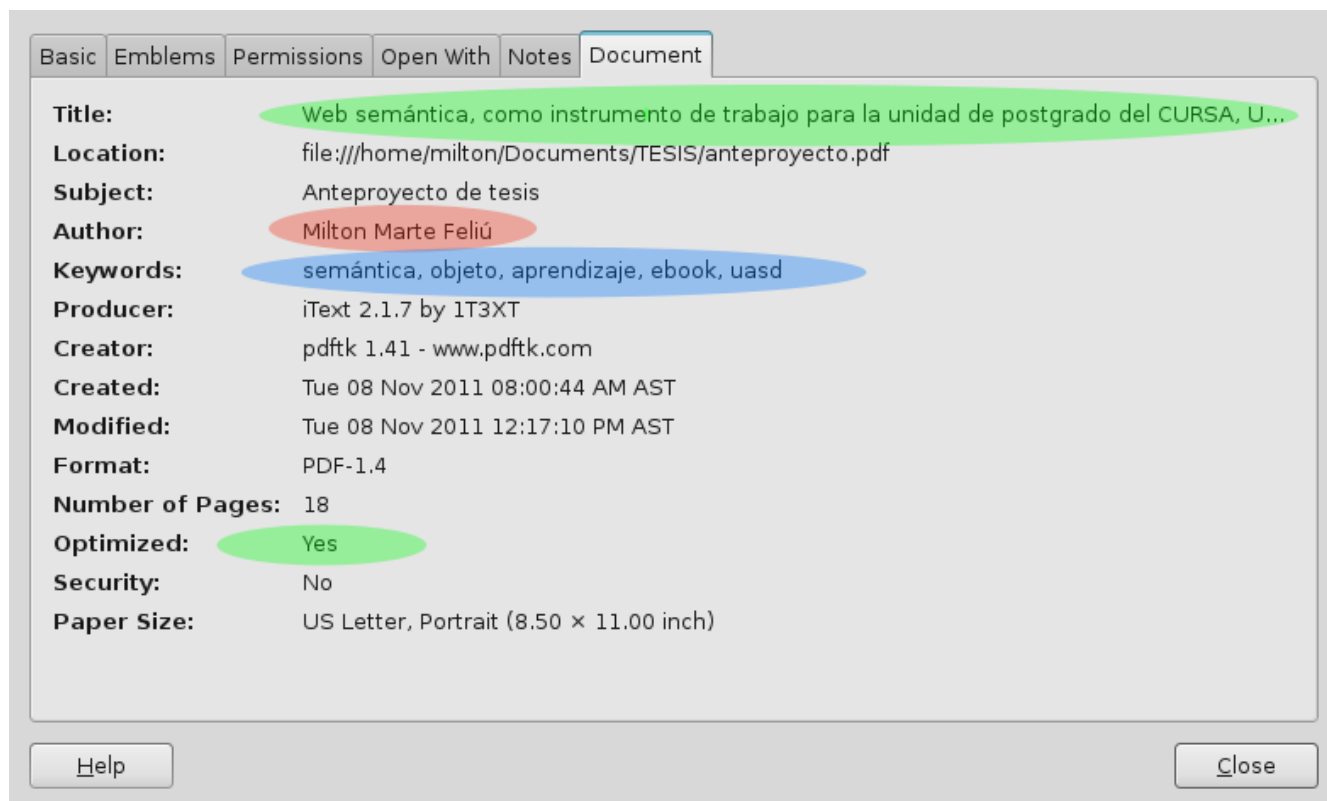
`apt-get install ghostscript`

`pdftop documentos.pdf documento-optimizado.pdf`

```
File Edit View Terminal Help
milton@debian:~/Desktop$ ls
anteproyecto-tesis.pdf  BIBLIOGRAFIA-Anteproyecto-tesis.pdf  Hoja-presentacion.pdf  indice.pdf  metadata.txt
milton@debian:~/Desktop$ pdftop anteproyecto-tesis.pdf anteproyecto-tesis-OPTIMIZADO.pdf
```



En la propiedad del documento se mostrará la información del archivo, como se destaca en la imagen.



## Referencias:

W3C oficina España (1994-2005). *Guía Breve de Web Semántica*

<http://www.w3c.es/divulgacion/guiasbreves/websemantica>

wikipedia (2011). *Web semántica*.

[http://es.wikipedia.org/wiki/Web\\_sem%C3%A1ntica](http://es.wikipedia.org/wiki/Web_sem%C3%A1ntica)

Agudelo Benjumea, Mónica M. (2009). Metadatos.

Ministerio de educación de Colombia. 5 páginas.

pdf Labs (2004–2011). *Tools, Services and Code for PDF Users and Programmers*.

<http://www.pdfabs.com/tools/pdftk-the-pdf-toolkit/>

linuxcommand (2003). *pdfopt*

[http://linuxcommand.org/man\\_pages/pdfopt1.html](http://linuxcommand.org/man_pages/pdfopt1.html)