

Pstotext: estrae il testo da un file PostScript o PDF

(pstotext - Extract text from PostScript and PDF files)

Pstotext consente in pochi secondi di estrarre il testo contenuto in un file PDF o Postscript.

Se usate debian:

```
# apt-get install pstotext
```

altrimenti cercatelo con google o nella vostra distribuzione.

La sintassi per la conversione è:

```
$ pstotext [opzioni] nome del file pdf
```

così facendo avrete nel vostro terminale l'output del vostro file pdf, che potrete copiare in un documento di testo. Volendo potete automaticamente ridirigere l'output usando ">", esempio:

```
$ pstotext linux.le.it.pdf > linux.le.it.txt
```

così avremo estratto il testo dal file linux.le.it.pdf e lo avremo nel file linux.le.it.txt

NB: VIENE ESTRATTO SOLO IL TESTO NON LE IMMAGINI (per quelle suggerisco di fare lo screenshot e di lavorarle con gimp – www.gimp.org)

Alcune opzioni:

-cork : Specifica che il file PostScript utilizza la codifica «cork», ovvero ciò che viene generato da Dvips quando questo converte file DVI generati da TeX con la codifica T1.

-landscape e -landscapeOther: due opzioni che indicano la rotazione del testo, se è ruotato a 90 gradi in un senso, oppure nell'altro.

-portrait: In questo caso si intende che il testo scorre nel modo consueto, su un foglio orientato in modo verticale.

-output file: Consente di indicare il file di testo da generare, senza bisogno di ridirigere lo standard output.

\$ man pstotext per maggiori informazioni.

--

Vito Russo - Project Manager russo.le.it
<http://russo.le.it>
<http://linux.le.it>
<http://windows.le.it>
<http://coding.le.it>
<http://bsd.le.it>