

PDF (autore: Vittorio Albertoni)

Premessa

Il formato PDF (Portable Document Format) è stato sviluppato da Adobe System e consente di rappresentare documenti in modo indipendente dall'hardware e dal software utilizzati per generarli.

Ovviamente, se vogliamo produrre e lavorare con il formato PDF in maniera nativa, il software deputato è di casa Adobe e non è né libero né gratuito.

La Adobe, tuttavia, ha intelligentemente reso di pubblico dominio le specifiche del formato PDF. Da qui la promozione di questo formato a standard internazionale e la intervenuta possibilità di creare e manipolare file PDF utilizzando diversi altri software, anche open source e liberi.

In questo manualetto ne faccio una rassegna.

Indice

1	Visualizzare file PDF	1
2	Creare file PDF	2
3	Proteggere file PDF	2
4	Manipolare file PDF	4
4.1	Interventi sulle pagine	5
4.2	Interventi nelle pagine con aggiunta di contenuti	5
4.3	Interventi nelle pagine con modifica dei contenuti	5

1 Visualizzare file PDF

Qualsiasi sistema operativo contiene, per default, almeno un lettore di file PDF.

Su Windows, Mac OS X e Android troviamo la versione gratuita di Adobe Reader, software proprietario della Adobe, da alcuni anni non più disponibile per Linux.

Ma il mondo Linux si è abbondantemente rifatto con prodotti liberi che troviamo già installati con il sistema operativo o troviamo nei repository delle varie distro. Tra i tanti segnalo qui i due più famosi e ricchi di funzionalità.

Forse il più diffuso è **evince**, disponibile anche per Windows all'indirizzo

<https://wiki.gnome.org/Apps/Evince/Downloads>

e presente in tutti i repository Linux.

Utile anche per visualizzare altri formati, come dvi, postscript, tiff e xps.

Non da meno **okular**, sviluppato dalla comunità KDE, disponibile per Windows e Mac OS X all'indirizzo

<https://okular.kde.org/download.php>

pure presente in tutti i repository Linux e pure utile per visualizzare altri formati, come dvi, postscript, tiff, xps e ebook nel formato epub.

2 Creare file PDF

Un tempo, se si voleva evitare di utilizzare il software proprietario Adobe per creare un file PDF, si ricorreva al metodo detto della stampante virtuale.

Avendo caricato sul computer un certo software, si poteva creare un file PDF come opzione di stampa su file da un qualsiasi editor di testo o grafico con cui se ne era predisposto il contenuto. Esiste tuttora, per esempio, PDF Creator, un software libero, peraltro disponibile solo per Windows, che ci consente di fare questa cosa.

Oggi è inutile tutto ciò in quanto tutte le suite per ufficio e i programmi di elaborazione grafica sono dotati di una funzione di esportazione del lavoro in formato PDF.

Il software libero ci offre tre eccellenze per creare file PDF attraverso l'esportazione di nostre composizioni.

Per testi, anche arricchiti da illustrazioni, formule, grafici e quant'altro abbiamo **LibreOffice** e la meno libera **OpenOffice**.

Per prodotti anche di maggiore complessità e presentati in formati editoriali classici abbiamo **T_EX** e **L^AT_EX**, con il semplificante editor **L_AX**, che ho presentato nell'allegato «pdf_lyx» al mio articolo «Scrivi e pubblica veri libri» del novembre appena scorso.

Per prodotti altamente elaborati sul piano grafico, come giornalini, pieghevoli, ecc. abbiamo **Scribus**.

Tutti software facilmente accessibili e scaricabili attraverso la rete e disponibili per tutti i sistemi operativi Linux, Windows e OS X.

Potremmo essere interessati a produrre un file PDF utilizzando, tale e quale, materiale esistente.

In questi casi, se il materiale esistente è cartaceo, possiamo trarne un file PDF utilizzando lo scanner e il software libero **XSane Image Scanning Program**. Di casa in tutte le distribuzioni Linux, ne esiste una versione un po' antiquata anche per Windows, scaricabile da <http://xsane.org/>. Con questo software collegato allo scanner possiamo produrre direttamente un file PDF contenente la fotocopia - in bianco e nero, in scala di grigio o a colori - di qualsiasi documento già esistente.

Se il materiale esistente è costituito da un file grafico (in formato jpg, png, gif, ecc.) abbiamo l'insuperabile **GIMP GNU Image Manipulation Program**. Una volta importato il file grafico che ci interessa lo possiamo esportare in formato PDF, magari dopo averlo ritoccato a nostro piacimento. GIMP è in genere disponibile sul computer subito con l'installazione di Linux e si trova comunque nel repository ed è quindi installabile con il gestore dei programmi. All'indirizzo <https://www.gimp.org/> troviamo eventualmente tutto ciò che ci può servire, comprese le versioni per Windows e Mac OS X.

Teniamo presente che i file PDF a colori creati da scanner o per esportazione di file grafici da GIMP riproducono in maniera molto fedele gli originali ma possono risultare alquanto pesanti. Un modo per creare un file PDF di qualità accettabile da un file grafico senza occupare spazio superiore a quello occupato da quest'ultimo è quello di importare il file grafico in LibreOffice Draw e da lì esportarlo in formato PDF.

3 Proteggere file PDF

Il sistema di protezione del file PDF prevede due tipi di password:

- la PASSWORD DI APERTURA (in inglese chiamata USER PASSWORD) necessaria per aprire il documento e vederlo;
- la PASSWORD DEI PERMESSI (in inglese chiamata OWNER PASSWORD) necessaria per poter fare certe cose sul documento (ad esempio modificarlo).

Se attiviamo la protezione solo prevedendo la password di apertura, una volta aperto il documento con questa password sarà possibile fare su di esso tutto ciò che è consentito dalla tecnologia.

Se attiviamo la protezione solo prevedendo la password dei permessi, il documento si aprirà e sarà visibile senza utilizzare la password ma per fare la cosa che abbiamo protetto sarà necessaria questa password.

Se attiviamo la protezione prevedendo entrambe le password, ci verrà chiesta una prima password per aprire il documento e renderlo visibile (e qui dovremo inserire la password di apertura) e, come ci accingeremo a fare quella cosa che avevamo assoggettato alla password dei permessi, dovremo inserire anche questa.

Il software libero ci regala un potente programma a riga di comando che, tra le tante cose che fa con i file PDF, si occupa anche di crittografarli e decrittografarli: si chiama **qpdf**.

Se usiamo il sistema operativo Linux molto probabilmente lo troviamo installato per default, comunque lo troviamo sicuramente nel repository e lo possiamo installare con il gestore dei programmi.

Se usiamo Windows possiamo scaricarlo da <https://sourceforge.net/projects/qpdf/>. Ci troveremo un file zip che estrarremo in C:\ e poi inseriremo nel path l'indirizzo fino alla sottodirectory bin, in modo da poter lanciare il comando qpdf a terminale, in qualsiasi posizione ci troviamo. Se usiamo Mac OS X possiamo installare qpdf attraverso Homebrew.

In rete (<http://qpdf.sourceforge.net/files/qpdf-manual.pdf>) si trova il manuale, purtroppo in lingua inglese, che ci dice tutto su questo prezioso software (chi usa Linux, a programma installato, trova il manuale in /usr/share/doc/qpdf). Qui richiamo solo le funzioni di crittografia.

La sintassi per crittografare un file .pdf già prodotto è la seguente:

```
qpdf --encrypt <1> <2> <3> <4> -- <file.pdf> <filecifrato.pdf>
```

Al posto di <1> indichiamo la password di apertura oppure, se non vogliamo usarla, la stringa vuota "" (due doppi apici senza spazio in mezzo). Più numerosi i caratteri, più bassa la probabilità di riuscire a violare la password.

Al posto di <2> indichiamo la password dei permessi oppure, se non vogliamo usarla, la stringa vuota "" (due doppi apici senza spazio in mezzo). Più numerosi i caratteri, più bassa la probabilità di riuscire a violare la password.

Al posto di <3> indichiamo i bit per la crittografia, potendo scegliere tra 40, 128 e 256. Più alto il numero di bit più bassa la probabilità di riuscire a violare la password.

Al posto di <4> indichiamo la o le restrizioni, per la varietà delle quali, correlata anche alla scelta indicata al punto <3>, rimando al manuale. Qui cito l'esempio relativo alla inibizione della modifica del file. In questo caso al posto di <4> indichiamo --modify=no se al punto <3> abbiamo indicato 40 oppure indichiamo --modify=none se al punto <3> abbiamo indicato 128 o 256. Se non si indicano restrizioni potremo fare ciò che vogliamo con il file e tanto valeva non usare la password dei permessi.

Al posto di <file.pdf>, dopo avere chiuso la stesura dei parametri della funzione encrypt con la doppia lineetta, indichiamo il file da crittografare.

Al posto di <filecifrato.pdf> indichiamo il nome che intendiamo dare al file crittografato.

Ovviamente qpdf prevede anche un comando per eliminare la protezione. La sua sintassi è la seguente:

```
qpdf --password=<1> --decrypt <filecifrato.pdf> <file.pdf>
```

Al posto di <1> indichiamo la password; se il file era cifrato con entrambe le password basta indicare la prima, cioè la password di apertura, per liberare il file da entrambe; se il file aveva solo la password dei permessi indichiamo quella.

Al posto di <filecifrato.pdf> indichiamo il nome del file che vogliamo liberare dalla protezione.

Al posto di <file.pdf> indichiamo il nome che intendiamo dare al file non più protetto.

La crittografia dei file pdf è una cosa seria, nel senso che il file criptato non contiene alcuna traccia leggibile della password, per cui non è assolutamente possibile ricostruire la password stessa se non con metodi di cracking brute-force.

Pertanto le password vanno rammentate e ben custodite, pena l'inutilizzabilità dei file protetti: è pertanto opportuno che, al momento della crittografia, al file criptato sia dato un nome diverso da quello del file non criptato, in modo che quest'ultimo sia sempre disponibile a chi lo ha creato.

A proposito di cracking brute-force segnalo l'esistenza di un software che si trova facilmente nei repository Linux e che si chiama **pdfcrack** ed è disponibile solo per Linux.

Lo si può comunque scaricare da <https://sourceforge.net/projects/pdfcrack/>.

Con il comando a terminale

```
pdfcrack <filecifrato.pdf>
```

si avvia un processo destinato a scoprire la password di apertura del file cifrato indicato.

Se il file era cifrato a 40 bit con password da 4 o 5 caratteri, in pochi secondi scopriremo la password e il tempo salirà, in modo esponenziale, per password composte da un maggior numero di caratteri. Teniamo presente che password composte da meno di 4 caratteri si scoprono istantaneamente.

Se il file era cifrato a 128 bit i pochi secondi possono diventare pochi minuti per password brevi ma possono diventare giorni per password lunghe.

Se il file era cifrato a 256 bit il nostro software non ci proverà nemmeno.

Quanto sopra valga non tanto perché ci mettiamo illecitamente a violentare file protetti ma per renderci conto di come sia importante scegliere adeguatamente parametri e password per proteggere i nostri file.

Crittografia al momento della creazione del file

Se creiamo un file PDF con LibreOffice o con Scribus, dopo aver scelto da menu ESPORTA NEL FORMATO PDF... si apre la finestra nella quale possiamo scegliere alcune caratteristiche per il file PDF che andiamo a creare.

Se apriamo la scheda SICUREZZA di questa finestra abbiamo la possibilità di indicare le password per criptare il file e per scegliere le azioni eventualmente da abbinare alla password dei permessi.

Entrambi i software criptano a 128 bit.

4 Manipolare file PDF

La finalità principale della tecnologia PDF è quella di produrre un file di modesta pesantezza che si presti alla diffusione in rete o attraverso la rete e che renda visibile un documento indipendentemente dal software utilizzato per produrlo. Il software che abbiamo visto nei precedenti capitoli è al servizio di questa finalità principale.

Col tempo, proprio anche grazie alla facilità con cui il file PDF può essere trasferito per email, si sono aggiunte altre funzionalità per le quali il software che abbiamo visto non basta più.

Pensiamo all'esigenza che può verificarsi di eliminare una parte di un documento contenuto in un file PDF o di mettere insieme in un unico documento più pagine allocate in diversi file PDF; all'esigenza di raccogliere su un file PDF prodotto come bozza di un documento le osservazioni di altre persone che collaborano alla sua stesura; all'esigenza di raccogliere su un file PDF che riproduce un contratto la firma del contraente; all'esigenza di eliminare un fastidioso errore di stampa, ecc. ecc.

Tutte occasioni in cui è necessario manipolare uno o più file PDF per generarne uno diverso.

Dagli esempi fatti possiamo distinguere tre categorie di interventi:

- interventi sulle pagine che non modificano i contenuti,
- interventi nelle pagine che aggiungono contenuti,
- interventi nelle pagine che modificano i contenuti.

Coloro che usano il sistema operativo Windows o il sistema operativo Mac OS X trovano caricato sul loro computer il software Adobe Reader che, adeguatamente arricchito con funzionalità acquistabili con un abbonamento mensile, consente di fare parecchie di queste cose. Per chi voglia risparmiare e per chi usa il sistema Linux vediamo cosa possiamo trovare nel mondo del software libero.

4.1 Interventi sulle pagine

Per dividere e unire pagine lavorando su diversi file PDF, che è una delle prestazioni che Adobe Reader non fa gratuitamente, abbiamo **PDFsam Basic**.

Chi usa Linux lo trova probabilmente già installato con il sistema operativo o lo può caricare con il gestore dei programmi. Dal sito

<https://pdfsam.org/it/>

sono scaricabili gli installer per Windows e OS X, oltre, ovviamente, a quello per Linux se non lo si è trovato diversamente.

Sempre su questo sito è acquistabile PDFsam Enhanced, che, ad un costo pari a quattro mesi di abbonamento per avere Adobe Reader rafforzato, ci consente di fare le stesse cose per tutta la vita.

Solo su Linux possiamo trovare **PDFShuffler** e **PDFMod**, che fanno le stesse cose che fa PDFsam Basic.

4.2 Interventi nelle pagine con aggiunta di contenuti

Sostanzialmente le occasioni per questo tipo di interventi sono l'apposizione di annotazioni di commento a un testo, l'indicazione di correzioni ritenute opportune, l'evidenziazione di zone di testo, la compilazione di formulari, l'apposizione di firme attraverso file grafico.

Questi interventi sono possibili con Adobe Reader non arricchito per chi usa Windows o OS X.

Chi usa Linux si può consolare con **Xournal**, che consente di fare tutte queste cose e altre ancora. E' un software che si può trovare nei repository delle distro, comunque all'indirizzo *<http://xournal.sourceforge.net/>*.

Apposizione di annotazioni, compilazione di formulari e apposizione di firme si possono fare anche aprendo il file PDF su cui intervenire con **LibreOffice Draw**. Purtroppo, però, questo software alcune volte non è utilizzabile in quanto deforma il documento originale, importandone i caratteri senza rispettare la loro dimensione e produce un PDF modificato con contenuto grafico molto diverso da quello di prima (ciò accade con LibreOffice 5 nel momento in cui scrivo: speriamo si tratti di un bug destinato ad essere eliminato).

4.3 Interventi nelle pagine con modifica dei contenuti

Esiste sempre il rischio che un file PDF modificato, soprattutto quando prima della modifica aveva un contenuto complesso sul piano grafico, ben definito e ben inquadrato, dopo una modifica ci proponga un contenuto che non è più bello come lo era prima: basti pensare al file contenente un libro illustrato, ben impaginato, con le illustrazioni numerate, i numeri di pagina, i richiami nel testo a pagine e figure, ecc. ed a cosa succederebbe al contenuto di questo file dopo l'aggiunta di un paragrafo a metà di un qualsiasi capitolo modificando il file PDF.

Molto spesso, tuttavia, può essere utile correggere un refuso, fare una piccola aggiunta o intervenire anche più pesantemente su file una cui modifica non vada a creare lo scompiglio esemplificato prima.

Il problema è trovare il software per farlo.

L'inventore del PDF ci propone il costoso Adobe Acrobat Pro DC, che, oltre a tutte le cose viste nei due paragrafi precedenti, consente di modificare i contenuti preesistenti nel file PDF.

Altri costosi software commerciali lo fanno.

Il mondo del software libero ci propose fino a qualche tempo fa PDFEdit, che utilizzava le vecchie librerie QT3 e che dal 2009 non ha più avuto nessuno a curarne l'aggiornamento.

Ciò che rimane è **LibreOffice Draw**, purtroppo con il difetto cui ho accennato alla fine del precedente paragrafo: però quando funziona funziona egregiamente. Unica limitazione: non consente di modificare il testo nei file PDF prodotti con lo scanner.

A questo punto devo segnalare e ricambiare una gentilezza.

All'indirizzo

<https://code-industry.net/free-pdf-editor/>

la Code Industry offre a chi usa il sistema operativo Linux una versione gratuita di **masterpdfeditor4**, un ottimo software di editing PDF disponibile per Windows, OS X e Linux, la cui versione completa a pagamento si può acquistare per 51 euro (enormemente meno di ciò che per fare le stesse cose ci fa pagare la Adobe con i suoi abbonamenti).

La versione gratuita, non utilizzabile per fini commerciali, ha qualche limitazione (ad esempio non fa crittografia) ma se la colleghiamo al software che abbiamo visto nel corso di questo manualetto, le limitazioni praticamente scompaiono.

Anche chi usa Windows e OS X può disporre di una versione non registrata gratuita del programma che, oltre alle limitazioni, quando produce il file PDF inserisce una sia pur leggera filigrana con la scritta «Created in Master PDF Editor». Ciò che non accade con la versione gratuita per Linux. C'è ancora chi ci vuole bene.