

Σαρώνοντας με ελεύθερο λογισμικό

Αριστείδης Κοντογιώργης και Αντώνης Τσολομούτης

Πανεπιστήμιο Αιγαίου
Τμήμα Μαθηματικών
832 00 Καρλόβασι
Σάμος
kontogar@aegean.gr και
atsol@aegean.gr

Περίληψη

In this article we show how to scan many pages of a document in one PDF file with free (GPL) tools. The simple method leads to better results than commercial programs (like Acrobat Distiller).

Στο άρθρο αυτό περιγράφουμε πως μπορεί να σαρώσει κανείς (scan) ένα οποιοδήποτε έγγραφο και να παράγει ένα αρχείο PDF χρησιμοποιώντας εξ' ολοκλήρου ελεύθερο GPL λογισμικό¹. Υπάρχουν κάποιοι περιορισμοί που θα ήθελε ο καθένας να θέσει ώστε να θεωρεί την διαδικασία χρήσιμη. Θέλουμε να ικανοποιήσουμε τις εξής συνθήκες:

1. Το παραγόμενο αρχείο να είναι μικρό
2. Να μπορεί να εκτυπωθεί και το τυπωμένο να είναι αναγνώσιμο (καλή ποιότητα)
3. Να μπορώ να έχω πολλές σελίδες στο ίδιο αρχείο

Όσοι έχουν ασχοληθεί με σάρωση ίσως γνωρίζουν ότι το μόνο πρόγραμμα που τα κάνει όλα αυτά είναι το Acrobat Distiller της Adobe, ένα ακριβό πρόγραμμα με το επιπλέον μειονέκτημα ότι υπάρχει μόνο για MS-Windows και Macintosh. Συνεπώς όποιος δεν έχει στην διάθεση του αυτό το εργαλείο ή εργάζεται σε άλλο λειτουργικό σύστημα δεν μπορεί να το χρησιμοποιήσει.

Ξεκινάμε λοιπόν με ένα GPL πρόγραμμα για σαρώσεις. Αναφερόμαστε στο πρόγραμμα sane και το γραφικό περιβάλλον του xsane. Και τα δύο προγράμματα

¹ General Public License

μπορούν να λειτουργήσουν σε Linux, Unix, MS-Windows, Macintosh χρησιμοποιώντας την κατάλληλη έκδοση. Ένα πλεονέκτημα που έχει το sane είναι ότι μπορεί να σαρώσει και να σώσει το αποτέλεσμα σε αρχείο PostScript. Βέβαια το αρχείο αυτό είναι συνήθως πολύ μεγάλο. Για παράδειγμα, σελίδες A4 με 8bit χρώμα μπορεί να δώσουν ακόμα και 200Mb ανά σελίδα σε ανάλυση 600dpi. Αλλά τα αρχεία αυτά είναι «ενδιάμεσα» και θα σβηστούν όταν τελειώσουμε τη διαδικασία. Μεγάλα αρχεία σώζει και ο Distiller κατά την διάρκεια της σάρωσης στον tmp κατάλογο που χρησιμοποιεί. Τα ενδιάμεσα αυτά αρχεία είναι πολύ μικρότερα όταν σαρώνουμε χωρίς χρώμα (ασπρόμαυρα) που είναι το επιθυμητό για σάρωμα κειμένου. Για παράδειγμα στα 600dpi μια σελίδα κειμένου δίνει περίπου 8Mb Postscript αρχείο το οποίο θεωρείται μικρό για τους σημερινούς δίσκους.

Αφού σαρώσουμε όλες τις σελίδες με το sane και τις σώσουμε σε PostScript (το xsane μας βοηθάει να το κάνουμε αυτό εύκολα αφού σώζει αυτόματα τα αρχεία με αύξοντες αριθμούς) θα πρέπει να τα ενώσουμε σε ένα αρχείο και να παράγουμε PDF.

Η μετατροπή PostScript σε PDF είναι απλή δουλειά και μπορεί να γίνει πολύ εύκολα με το πρόγραμμα ps2pdf που υπάρχει σε κάθε διανομή του GhostScript. Το μόνο πρόβλημα που μένει είναι πως θα ενωθούν όλα τα PostScript αρχεία.

Το πακέτο psutils παρέχει το πρόγραμμα psmerge το οποίο υπόσχεται ακριβώς αυτό το οποίο ζητάμε. Δυστυχώς όμως μετά από πολλές προσπάθειες δεν καταφέραμε να το κάνουμε να δουλέψει με τα PostScript αρχεία που παίρναμε από την σάρωση.

Τη λύση την δίνει (ίσως ανέλπιστα) η γνωστή από τον χώρο του Unix εντολή cat! (υπάρχει και για MS-Windows πχ στο πακέτο cygwin). Αν υποθέσουμε ότι έχουμε τέσσερις σελίδες σαρωμένες στα αρχεία 001.ps, 002.ps, 003.ps, 004.ps δίνουμε στη γραμμή εντολών (bash σε Unix ή MS-Windows (cygwin)):

```
for i in *.ps; do cat $i >> final.ps; done
```

Αυτή θα παράγει το τελικό PostScript αρχείο final.ps το οποίο περιέχει και τις 4 σελίδες. Ολοκληρώνουμε τη διαδικασία με μιά μετατροπή σε PDF:

```
ps2pdf final.ps
```

και έχουμε ένα σωστό PDF αρχείο μεγέθους συνήθως μικρότερου από το μέγεθος που παράγει ο Distiller. Τώρα μπορούμε να σβήσουμε όλα τα αρχεία PostScript.

Αν έχουμε πρόβλημα χώρου μπορούμε να παράγουμε κατευθείαν το τελικό PDF αρχείο χωρίς να παραχθεί το (συνήθως μεγάλο) final.ps. Αυτό γίνεται χρησιμοποιώντας ένα pipe προς την ps2pdf ως εξής:

```
for i in *.ps; do cat $i; done | ps2pdf - final.pdf
```

Αν κανείς έχει σοβαρό πρόβλημα χώρου μπορεί να κάνει αρκετή οικονομία μετατρέποντας το PostScript αρχείο από τη σάρωση κάθε σελίδας σε PDF. Έτσι θα έχει (για το παραπάνω παράδειγμά μας) τέσσερα PDF αρχεία (τα 001.pdf, 002.pdf, 003.pdf, 004.pdf) τα οποία θα θέλει να ενώσει σε ένα. Αυτό μπορεί να το κάνει πάλι χρησιμοποιώντας pipes ως εξής:

```
for i in *.pdf; do pdf2ps $i - ; done | ps2pdf - final.pdf
```

Η παραπάνω εντολή θα πρέπει να δουλέψει και να δώσει το επιθυμητό αρχείο final.pdf. Στην τρέχουσα έκδοση όμως του GhostScript (έκδοση 7.04) αποτυγχάνει. Αν έχετε τέτοιο πρόβλημα μπορείτε να εγκαταστήσετε το GPL πρόγραμμα XPDF το οποίο παρέχει την εντολή pdftops. Τώρα η παραπάνω εντολή γίνεται:

```
for i in *.pdf; do pdftops $i - ; done | ps2pdf - final.pdf
```

η οποία δουλεύει σωστά.

Όσο αφορά στα παραγόμενα μεγέθη σαρώσαμε μία σελίδα (με κείμενο και φωτογραφίες). Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα:

Ανάλυση/Χρώμα	Distiller	Ελεύθερο Λογισμικό
300dpi Ασπρόμαυρο	231 Kb	189 Kb
600dpi Ασπρόμαυρο	691 Kb	459 Kb
600dpi Έγχρωμο (8bit)	16 Mb	12 Mb