

Μέσω email μπορούν να διακινηθούν μόνο εκτυπώσιμοι χαρακτήρες ASCII (και πιθανώς unicode). Επομένως, μπορούμε να επισυνάπτουμε απευθείας μόνο αρχεία κειμένου. Έχετε αναρωτηθεί, πώς είναι δυνατόν να επισυνάπτουμε και αρχεία τα οποία είναι από τη φύση τους δυαδικά και άρα αποτελούνται και από bytes που δεν μπορούν να εμπίπτουν στους παραπάνω χαρακτήρες (π.χ. έγγραφα pdf, εικόνες, κλπ);

Η απάντηση είναι ότι τα αρχεία αυτά, πρώτα μετασχηματίζονται ή κωδικοποιούνται σε αρχεία κειμένου, που αποτελούνται μόνο από εκτυπώσιμους χαρακτήρες ASCII, και στη συνέχεια επισυνάπτονται στο μήνυμά μας. Ο παραλήπτης, με τη σειρά του, πρέπει να τα αποκωδικοποιήσει για να αποκαταστήσει το αρχικό περιεχόμενό τους. Όλα αυτά, βέβαια, τα αναλαμβάνει αυτόματα η εφαρμογή που χειρίζεται το email μας.

Ένας (ίσως ο πιο συνηθισμένος) τρόπος για την κωδικοποίηση δυαδικών αρχείων και τον μετασχηματισμό τους σε αρχεία κειμένου είναι η λεγόμενη κωδικοποίηση κατά base64. Υπάρχει μάλιστα και εφαρμογή που το κάνει από το τερματικό. Για παράδειγμα, αν εκτελέσετε:

```
$ base64 picture.jpg
```

θα δείτε πώς κωδικοποιείται η εικόνα σε κείμενο που αποτελείται μόνο από χαρακτήρες ASCII.

I. ΚΩΔΙΚΟΠΟΙΗΣΗ BASE64

Στο εργαστήριο αυτό ζητείται να υλοποιήσετε μία δική σας “base64”, δηλαδή ένα πρόγραμμα το οποίο:

1. Παίρνει μέσω ορισμάτων στη main() το όνομα ενός αρχείου.
2. Ανοίγει και διαβάζει το αρχείο ως δυαδικό.
3. Κωδικοποιεί τα bytes του κατά base64.
4. Αποθηκεύει το αποτέλεσμα σε ένα αρχείο κειμένου το οποίο έχει ίδιο όνομα με το αρχικό, μόνο που του έχει προστεθεί η κατάληξη “.enc64”.

Η διαδικασία κωδικοποίησης είναι σχετικά απλή (δείτε και το αντίστοιχο άρθρο στη wikipedia). Στην ουσία:

1. 3 bytes του αρχείου τοποθετούνται στα 3 πρώτα bytes ενός ακεραίου, δηλαδή, αν τα bytes είναι, με τη σειρά, a , b και c , σχηματίζεται ο αριθμός $(a \ll 16) | (b \ll 8) | c$.
2. Τα 24 bits χωρίζονται σε τέσσερις δάδες.
3. Κάθε εξάδα από bits μετασχηματίζεται σε χαρακτήρα ASCII. Συγκεκριμένα, ο αριθμός 0 δίνει το ‘A’, ο 1 το ‘B’, κλπ. και ο 63 δίνει το ‘/’. Ο πλήρης πίνακας με την αντιστοίχιση υπάρχει στο άρθρο της wikipedia.

Αν το αρχικό αρχείο δεν έχει μέγεθος πολλαπλάσιο του 3, η κωδικοποίηση γίνεται σαν να υπάρχουν τα απαιτούμενα 1 ή 2 bytes και θεωρείται ότι έχουν τιμή 0. Αν έλλειπε 1 byte, κωδικοποιούνται οι τρεις πρώτες δάδες και παράγεται και ένας χαρακτήρας ‘=’. Αν έλλειπαν 2 bytes, κωδικοποιούνται οι δύο πρώτες εξάδες και παράγονται επίσης 2 χαρακτήρες ‘=’.

II. ΕΠΙΠΛΕΟΝ ΕΞΑΣΚΗΣΗ: ΑΠΟΚΩΔΙΚΟΠΟΙΗΣΗ BASE64

Στο πρόγραμμά σας προσθέστε και τη λειτουργία της αποκωδικοποίησης κατά base64. Το πρώτο όρισμα στην main() θα πρέπει να καθορίζει, πλέον, αν θα κάνετε κωδικοποίηση ή αποκωδικοποίηση. Συγκεκριμένα, με -e θα κάνετε κωδικοποίηση ενώ με -d αποκωδικοποίηση, ως εξής:

```
$ ./a.out -e file
$ ./a.out -d file
```