

## ΕΝΟΤΗΤΑ 8

### ΔΙΑΧΕΙΡΙΣΗ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΑΡΧΕΙΩΝ

#### Περιεχόμενα

1. Δομή αρχείων
2. Λειτουργίες επί των αρχείων
3. Οργάνωση αρχείων
4. Ιδιοχαρακτηριστικά αρχείων
5. Κατάλογοι αρχείων
6. Διαχείριση του χώρου του δίσκου
7. Υλοποίηση αρχείων

## 1. Δομή αρχείων

- Ένα αρχείο είναι η μονάδα αποθήκευσης πληροφοριών στο δίσκο. Αποτελείται από ένα αριθμό εγγραφών η κάθε μία των οποίων αποτελείται από ένα αριθμό από πεδία.
- Πολλά αρχεία μαζί αποτελούν μία βάση δεδομένων.

## 2. Λειτουργίες επί των αρχείων

- Οι πιο κοινές κλήσεις σε ένα Λ.Σ. που σχετίζονται με τα αρχεία είναι οι εξής:
  - Δημιουργία, άνοιγμα, κλείσιμο, ενός αρχείου.
  - Αντιγραφή, διαγραφή ενός αρχείου.
  - Αλλαγή ονόματος ενός αρχείου.
  - Προσπέλαση στα περιεχόμενα ενός αρχείου.
- Η προσπέλαση στα περιεχόμενα ενός αρχείου μπορεί να γίνει για τους εξής λόγους:
  - Για διάβασμα ενός αριθμού εγγραφών.
  - Για προσθήκη νέων ή αφαίρεση υπάρχοντων εγγραφών.
  - Για αλλαγή στα περιεχόμενα κάποιων εγγραφών.
- Σκοπός ενός *συστήματος διαχείρισης αρχείων* (file management system) είναι η υποστήριξη για χάρη του χρήστη των ανωτέρω λειτουργιών και η εκτέλεσή τους όσο γίνεται πιο αποδοτικά, η διαφύλαξη της ακεραιότητας των περιεχομένων ενός αρχείου και η παροχή ρουτινών για διασύνδεση με συσκευές Ε/Ε.
- Επίσης, η αντιμετώπιση των προβλημάτων που προκύπτουν από τη μακροχρόνια αποθήκευση πληροφοριών όπως: μεγάλη ποσότητα, διατήρηση των πληροφοριών και μετά τον τερματισμό της διεργασίας που τις δημιούργησε, πρόσβαση σε αυτές από πολλές διεργασίες, κλπ.

### 3. Οργάνωση αρχείων

- Ανάλογα με τον τρόπο προσπέλασης στα περιεχόμενα ενός αρχείου έχουμε τα εξής είδη οργάνωσης των εγγραφών ενός αρχείου:
  - Σειριακό.
  - Σειριακό με δείκτες.
  - Τυχαίας προσπέλασης.

### 4. Ιδιοχαρακτηριστικά αρχείων

- Κάθε αρχείο εκτός του ονόματος και των περιεχομένων του έχει επίσης και μερικές άλλες πληροφορίες που είναι χρήσιμες σε ένα Λ.Σ. για να καθορισθεί ο τρόπος χρήσης του αρχείου από τους χρήστες. Αυτές οι πληροφορίες λέγονται *ιδιοχαρακτηριστικά* (attributes) και περιλαμβάνουν μεταξύ άλλων:
  - Τον δημιουργό του αρχείου.
  - Τον ιδιοκτήτη του αρχείου.
  - Ένδειξη αν το αρχείο μπορεί να διαβασθεί ή/και να τροποποιηθεί.
  - Ένδειξη αν το αρχείο είναι μορφής ASCII ή δυαδικής.
  - Μήκος εγγραφής και μέγεθος του αρχείου.
  - Ώρα δημιουργίας και τελευταίας τροποποίησης του αρχείου.

### 5. Κατάλογοι αρχείων

- Τα αρχεία οργανώνονται σε καταλόγους (directories). Συνήθως οι κατάλογοι οργανώνονται σε δένδροειδή και ιεραρχική δομή. Η εύρεση κάποιου αρχείου γίνεται μέσω του *ονόματος διαδρομής* (path name) που είναι η ακολουθία των ονομάτων των καταλόγων από το σημείο αναφοράς μέχρι τον κατάλογο όπου βρίσκεται το αρχείο. Αν το σημείο αναφοράς είναι ο πρωταρχικός (root) κατάλογος, τότε το όνομα διαδρομής είναι *απόλυτο* (absolute path name) ενώ αν είναι ο τρέχων κατάλογος εργασίας (working directory) τότε το όνομα διαδρομής είναι *σχετικό* (relative path name).

## 6. Διαχείριση του χώρου του δίσκου

- Ένα θέμα είναι το μέγεθος του μπλοκ και το αν αυτό θα πρέπει να είναι σταθερό ή μεταβλητό. Στην περίπτωση σταθερού μεγέθους σπαταλάται χώρος αλλά η χρήση μεταβλητού μεγέθους είναι δύσκολο να υλοποιηθεί.
- Επειδή λοιπόν συνήθως χρησιμοποιείται η πρώτη μέθοδος το επόμενο ερώτημα που πρέπει να απαντηθεί είναι ποιο είναι το μέγεθος του μπλοκ. Αν είναι σχετικά μεγάλο, τότε σε κάθε πράξη Ε/Ε μεταφέρονται περισσότερες εγγραφές που είναι ωφέλιμο όταν η προσπέλαση στο αρχείο είναι σειριακή. Αν είναι σχετικά μικρό, τότε ωφελεί την τυχαία προσπέλαση όπου κατά κανόνα χρειάζεται η μεταφορά λίγων εγγραφών.
- Ένα άλλο θέμα είναι το πότε γίνεται η κατανομή των μπλοκ δίσκου στα αρχεία: στατικά τη στιγμή δημιουργίας του αρχείου ή δυναμικά. Η πρώτη μέθοδος προϋποθέτει ότι γνωρίζουμε εκ των προτέρων το μέγιστο μέγεθος του αρχείου. Η δεύτερη θέτει το επιπλέον ερώτημα του πόσο χώρο δίνουμε αρχικά για την αποθήκευση του αρχείου και αν αυτός είναι ο ίδιος σε μέγεθος κάθε φορά.
- Ένα τελευταίο θέμα είναι η διαχείριση των ελεύθερων μπλοκ, δηλαδή η εύκολη και γρήγορη εύρεση ελεύθερων μπλοκ για να δοθούν σε αρχεία. Εδώ χρησιμοποιούνται οι ακόλουθοι μέθοδοι:
  - *Χάρτες bits* όπου υπάρχει ένα διάνυσμα των  $n$  bits όπου  $n$  είναι ο αριθμός των μπλοκ σε ένα δίσκο. Η τιμή 0 δείχνει ότι το μπλοκ είναι ελεύθερο και η τιμή 1 ότι έχει δοθεί σε κάποιο αρχείο. Είναι εύκολο να υλοποιηθεί και μια και χρειάζεται λίγο χώρο μνήμης μπορεί να βρίσκεται εξ' ολοκλήρου στην κύρια μνήμη.
  - Συνδεδεμένη λίστα των ελεύθερων μπλοκ δίσκου.
  - Αποθήκευση όσων περισσότερων αριθμών ελεύθερων μπλοκ δίσκου χωράνε σε ένα μπλοκ. Δυνατή εδώ και η χρήση ομάδας τέτοιων μπλοκ συνδεδεμένων μεταξύ τους σε μορφή λίστας.
  - Αποθήκευση της διεύθυνσης του πρώτου μπλοκ και ενός αριθμού που υποδεικνύει το πλήθος των ελεύθερων μπλοκ δίσκου που έπονται. Βασίζεται στο ότι η δέσμευση και αποδέσμευση των μπλοκ γίνεται ομαδικά.

## 7. Υλοποίηση αρχείων

- Το βασικό ζήτημα στην υλοποίηση της αποθήκευσης των αρχείων στο δίσκο είναι να γνωρίζουμε κάθε στιγμή ποια μπλοκ του δίσκου ανήκουν σε ποιο αρχείο. Εδώ υπάρχουν οι εξής μέθοδοι:
- Η *συνεχής κατανομή* (contiguous allocation) αποθηκεύει κάθε αρχείο σαν ένα συνεχόμενο μπλοκ δεδομένων. Είναι εύκολη στην υλοποίηση (χρειάζεται να θυμόμαστε μόνο το πρώτο μπλοκ κάθε αρχείου και το μέγεθος του αρχείου) και έχει καλή απόδοση. Όμως δεν είναι ευέλικτη (ιδιαίτερα αν δεν γνωρίζουμε το μέγιστο πιθανό μέγεθος του αρχείου) και οδηγεί σε (εξωτερικό) κατακερματισμό του δίσκου και ανάγκη συμπίεσής του περιοδικά. Τα μπλοκ δίνονται εκ των προτέρων και είναι διαφορετικού μεγέθους.
- Η *κατανομή συνδεδεμένης λίστας* (chained allocation) αποθηκεύει ένα αρχείο κάνοντας χρήση μίας ομάδας μπλοκ, όχι κατ' ανάγκη συνεχόμενων. Η πρώτη λέξη του κάθε μπλοκ είναι ένας δείκτης στο επόμενο. Η μέθοδος αυτή δεν προκαλεί κατακερματισμό και είναι σχετικά εύκολη η υλοποίησή της (χρειάζεται η αποθήκευση της διεύθυνσης του πρώτου μπλοκ κάθε αρχείου). Αλλά η χρήση δεικτών δεσμεύει κάποιο ποσό μνήμης. Επίσης η τυχαία προσπέλαση είναι πιο αργή δημιουργώντας την ανάγκη *συγχώνευσης* (consolidation) των μπλοκ σε συνεχόμενες θέσεις μνήμης. Τα μπλοκ δίνονται δυναμικά ένα-ένα όποτε παρίσταται ανάγκη.
- Η *κατανομή συνδεδεμένης λίστας με δείκτες* (indexed allocation) απαλείφει τα προβλήματα των προηγούμενων μεθόδων με τη χρήση ενός πίνακα για κάθε αρχείο, κάθε εγγραφή του οποίου είναι ένας δείκτης στο αντίστοιχο μπλοκ του αρχείου. Τα μπλοκ μπορεί να δίνονται ένα κάθε φορά που παρίσταται ανάγκη (παραλλαγή της δεύτερης μεθόδου) ή σε ομάδες περιοδικά. Για τη δεύτερη αυτή περίπτωση το μέγεθος της ομάδας μπορεί να είναι σταθερό ή μεταβλητό (παραλλαγές της πρώτης περίπτωσης). Η τυχαία προσπέλαση τώρα είναι πιο εύκολη. Όμως ο πίνακας αυτός πρέπει να βρίσκεται εξ' ολοκλήρου στην κύρια μνήμη. Και εδώ μπορεί να γίνεται περιοδικά συγχώνευση των μπλοκ.