

## Κεφάλαιο 6

### ΤΑΞΙΝΟΜΗΣΗ ΜΕ ΔΙΣΚΟΥΣ

- 6.1 Εισαγωγή
- 6.2 Ταξινόμηση με μία συσκευή δίσκων
- 6.3 Ταξινόμηση με δύο συσκευές δίσκων
- 6.4 Συγχώνευση με μία συσκευή δίσκων
- 6.5 Συγχώνευση με δύο συσκευές δίσκων
- 6.6 Άλλες μέθοδοι εξωτερικής συγχώνευσης
- 6.7 Ασκήσεις

## Κεφάλαιο 6

# ΤΑΞΙΝΟΜΗΣΗ ΜΕ ΔΙΣΚΟΥΣ

### 6.1 Εισαγωγή

Όπως είναι γνωστό, αν το αρχείο δεν χωρά στην κύρια μνήμη, τότε η εξωτερική ταξινόμηση γίνεται σε δύο στάδια. Αρχικά το αρχείο υποδιαιρείται σε τμήματα που χωρούν στην κύρια μνήμη και ταξινομούνται με μία μέθοδο εσωτερικής ταξινόμησης, οπότε συγχωνεύονται στη συνέχεια. Σε κάθε στάδιο απαιτείται ανάγνωση και αποθήκευση του αρχείου τουλάχιστο μία φορά. Με τη χρήση μεγάλης κυρίας μνήμης, δύο συσκευών δίσκων και υποστήριξη απομονωτικών μνημών για την επικάλυψη του χρόνου προσπέλασης κάποιας εγγραφής από το δίσκο με το χρόνο αποθήκευσης κάποιας άλλης στο δίσκο, είναι δυνατό να επιτευχθούν πολύ αποτελεσματικές μέθοδοι.

Έστω ότι δίνεται ένα αρχείο 6.000.000 εγγραφών των 400 bytes. Έτσι το αρχείο έχει μέγεθος 2,4 Gb, δηλαδή είναι ένα μεγάλο αρχείο που μπορεί, για παράδειγμα, να συναντηθεί συχνά στην πράξη σε εφαρμογές τραπεζών, δημοσίων οργανισμών ή αεροπορικών εταιρειών. Δεδομένου ότι δεν υπάρχει υπολογιστής με τόσο μεγάλη κύρια μνήμη, θα υποθεθεί ότι η διαθέσιμη κύρια μνήμη είναι 10 Mb.

Αν υπάρχει μόνο μία συσκευή δίσκου, τότε η καλύτερη επιλογή για το πρώτο στάδιο είναι η χρησιμοποίηση του αλγορίθμου ταξινόμησης σωρού (heapsort), που παράγει ταξινομημένα τμήματα μεγέθους ίσου με το μέγεθος της κύριας μνήμης. Ο αλγόριθμος αυτός, σε αντίθεση με όλους τους γνωστούς, είναι ο μόνος που με τη βοήθεια απομονωτικών μνημών επιτρέπει επικάλυψη της εισόδου με την έξοδο. Συνεπώς, σε τελική ανάλυση το

συνολικό χρονικό κόστος ισούται με το κόστος για είσοδο και έξοδο των δεδομένων, ενώ η καθ' εαυτή ταξινόμηση γίνεται παράλληλα και δεν επιβαρύνει χρονικά.

Αν υπάρχουν δύο διαθέσιμες συσκευές δίσκου, τότε χρησιμοποιείται η τεχνική ταξινόμησης που ονομάζεται **επιλογή αντικατάστασης** (replacement selection). Αυτή η μέθοδος παράγει τμήματα με διπλάσιο μέγεθος από το μέγεθος της διαθέσιμης κύριας μνήμης, κατά μέσο όρο. Επιπλέον, με τη βοήθεια απομονωτικών μνημών, η προσπέλαση από τον ένα δίσκο μπορεί να επικαλυφθεί χρονικά με την αποθήκευση στον άλλο. Έτσι τελικά, το πρώτο στάδιο της ταξινόμησης των τμημάτων απαιτεί χρόνο ίσο περίπου με την προσπέλαση του αρχείου.

Το δεύτερο στάδιο της συγχώνευσης είναι παρόμοιο προς τη συγχώνευση δομών κύριας μνήμης και συνεπώς αρκετά απλό. Αν υποθεθεί ότι δίνονται 100 ταξινομημένα τμήματα, τότε από κάθε τμήμα λαμβάνεται μία εγγραφή και εκτελείται μία συγχώνευση 100 δρόμων. Όταν μία εγγραφή κάποιου τμήματος κατευθυνθεί στην έξοδο, τότε αυτή αντικαθίσταται από την επόμενη του ίδιου τμήματος και έτσι η διαδικασία προχωρεί επαναληπτικά μέχρι να εξαντληθούν όλα τα τμήματα.

Η χρήση περισσότερων από δύο συσκευών δίσκου δεν εξυπηρετεί τόσο σημαντικά στη μείωση του χρόνου ταξινόμησης, όσο η χρήση των όλο και μεγαλύτερων και φθηνότερων κύριων μνημών. Επειδή στο παρελθόν δεν υπήρχαν μεγάλες κύριες μνήμες, χρησιμοποιούνταν άλλοι αλγόριθμοι εξωτερικής ταξινόμησης (πχ. η πολυφασική συγχώνευση). Στο τέλος του παρόντος κεφαλαίου οι μέθοδοι ταξινόμησης με ταινίες εξετάζονται και κοινοτομολογούνται για την εφαρμογή τους σε δίσκους. Ωστόσο, σημειώνεται προκαταβολικά ότι μία μέθοδος που θα παρουσιασθεί στη συνέχεια, και η οποία απαιτεί δύο συσκευές δίσκων, ξεπερνά τη μέθοδο της πολυφασικής συγχώνευσης ακόμη και αν αυτή εφαρμοσθεί με 10 ή 20 συσκευές δίσκων.

## 6.2 Ταξινόμηση με μία συσκευή δίσκων

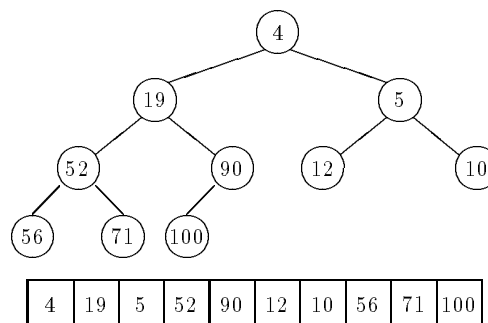
Όταν υπάρχει διαθέσιμη μόνο μία συσκευή δίσκων, τότε στο πρώτο στάδιο χρησιμοποιείται ο αλγόριθμος της ταξινόμησης σωρού με σκοπό τη δημιουργία των ταξινομημένων τμημάτων. Η γνωστή αυτή μέθοδος προτάθηκε από τον Williams το 1964, ενώ μία παραλλαγή γνωστή ως **ταξινόμηση δένδρου** (treesort) προτάθηκε από τον Floyd το 1967. Ο αλγόριθμος αυτός είναι ο

μοναδικός από τους αλγορίθμους εσωτερικής ταξινόμησης που επιτρέπει την επικάλυψη της εισόδου με την έξοδο.

Σχεδόν πλήρες δυαδικό δένδρο (almost complete binary tree) είναι εκείνο το δυαδικό δένδρο που:

1. όλα τα φύλλα του βρίσκονται σε δύο διαδοχικά επίπεδα, έστω το  $k$ -οστό και  $(k-1)$ -οστό επίπεδο,
2. όλοι οι κόμβοι του επιπέδου  $(k-2)$  έχουν βαθμό δύο, και
3. τα φύλλα του  $k$ -οστού επιπέδου καταλαμβάνουν τις αριστερότερες θέσεις του επιπέδου.

Σωρός μεγίστων (max heap) είναι το σχεδόν πλήρες δυαδικό δένδρο, όπου η τιμή του κλειδιού κάθε κόμβου είναι μικρότερη από τις τιμές των κλειδιών των παιδιών του συγκεκριμένου κόμβου (δες βιβλίο για Δομές Δεδομένων, Κεφάλαιο 6.2). Έτσι είναι βέβαιο ότι στη ρίζα είναι πάντα αποθηκευμένο το μικρότερο κλειδί. Στο Σχήμα 6.1 παρουσιάζεται ένας σωρός.



Σχήμα 6.1: Σωρός ως δυαδικό δένδρο και ως διάγραμμα.

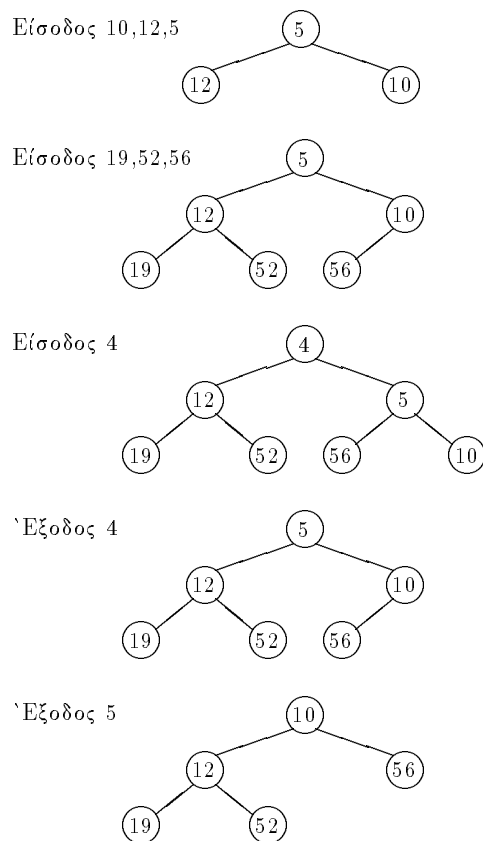
Ένα πλήρες δυαδικό δένδρο πολύ εύκολα μπορεί να παρασταθεί με έναν πίνακα, τοποθετώντας το περιεχόμενο των κόμβων στον πίνακα κατά επίπεδα και από αριστερά προς τα δεξιά. Δηλαδή, στην πρώτη θέση του πίνακα τοποθετείται η ρίζα, στη δεύτερη το αριστερό παιδί της ρίζας, στην τρίτη το δεξί παιδί της ρίζας, κοκ. Άρα, τα παιδιά ενός κόμβου που βρίσκεται στην  $i$ -οστή θέση του πίνακα είναι αποθηκευμένα στις θέσεις  $2i$  και  $2i+1$ , ενώ ο πατέρας του κόμβου αυτού είναι αποθηκευμένος στη θέση  $\lfloor i/2 \rfloor$ .

Η βασική ιδέα του αλγορίθμου ταξινόμησης σωρού είναι η εξής. Η εισαγόμενη εγγραφή καταλαμβάνει την αριστερότερη θέση του τελευταίου επιπέδου του δένδρου (ή με άλλα λόγια την πρώτη κενή θέση του πίνακα). Το κλειδί της νέας εγγραφής συγκρίνεται προς το κλειδί του κόμβου πατέρα. Αν το νέο κλειδί είναι μικρότερο, τότε πρέπει οι δύο εγγραφές να ανταλλάχθούν, ώστε να ισχύει η βασική ιδιότητα των σωρών. Αυτή η διαδικασία σύγκρισης και αναρρίχησης μπορεί να επαναληφθεί, ίσως και μέχρι τη ρίζα. Η διαδοχική ανάγνωση εγγραφών και η τοποθέτησή τους στο σωρό συνεχίζεται μέχρι να γεμίσει ο διαθέσιμος χώρος στην κύρια μνήμη. Έτσι το περιεχόμενο της κύριας μνήμης αποτελεί ένα ταξινομημένο τμήμα.

Κατόπιν η ρίζα του σωρού μεταφέρεται στην έξοδο, ενώ η τελευταία εγγραφή του σωρού καταλαμβάνει τη θέση της ρίζας. Πλέον δεν ισχύει ο ορισμός του σωρού και πρέπει το κλειδί αυτό να συγκριθεί με τα κλειδιά των δύο παιδιών του. Το μικρότερο κλειδί ανταλλάσσεται με την εγγραφή της ρίζας που (δυνητικά) μπορεί να συνεχίσει την καθοδική πορεία μέχρι το τελευταίο επίπεδο. Μετά από όλη αυτή την επαναληπτική διαδικασία, ο ορισμός του σωρού αποκαθίσταται και συνεπώς η νέα ρίζα μπορεί να μεταφερθεί στην έξοδο. Κατ' αυτόν τον τρόπο ο αλγόριθμος συνεχίζει μέχρι τη συνολική μεταφορά όλου του σωρού στην έξοδο. Στη συνέχεια η μνήμη δέχεται νέες εγγραφές από το δίσκο για τη δημιουργία του νέου τμήματος.

Στο Σχήμα 6.2 δίνεται ένα παράδειγμα, υποθέτοντας ότι στη μνήμη χωρούν 7 εγγραφές. Έτσι στη μνήμη εισάγονται 7 εγγραφές με κλειδιά 10, 12, 5, 19, 52, 56 και 4, ενώ στη συνέχεια μεταφέρονται στην έξοδο οι εγγραφές με κλειδιά 4 και 5. Στο σχήμα παρουσιάζονται μερικά βήματα της διαδικασίας που υλοποιεί την ανάπτυξη αυτή. Κάθε δομή του σχήματος αντιστοιχεί στο σωρό μετά από την αποκατάσταση των ιδιοτήτων του (δηλαδή, μετά από τις αναγκαίες συγκρίσεις και ανταλλαγές).

Στη συνέχεια μελετάται η επίδοση του αλγορίθμου. Το ύψος ενός πλήρους δυαδικού δένδρου είναι  $\log n$ , όπου  $n$  είναι το πλήθος των κόμβων του δένδρου. Συνεπώς, ο αριθμός των απαραίτητων συγκρίσεων/ανταλλαγών για τη διατήρηση της βασικής ιδιότητας του σωρού είναι στη χειρότερη περίπτωση  $\log n$ , όπου πλέον  $n$  είναι οι μέχρι τώρα εγγραφές του σωρού. Δηλαδή, αν το κάθε τμήμα έχει μέγεθος 10 Mb και περιέχει 25.000 εγγραφές, τότε το ύψος του δένδρου είναι μεταξύ 14 και 15. Άρα, μία εισαγωγή εγγραφής στο σωρό μπορεί να προκαλέσει το μέγιστο 14 ανταλλαγές. Ακόμη αν υποθεθεί ότι η είσοδος συνίσταται στην προσπέλαση σελίδων μεγέθους 2400 bytes με συντελεστή ομαδοποίησης 6, τότε πρέπει να γίνουν το μέγιστο 84



Σχήμα 6.2: Ταξινόμηση σωρού με επικάλυψη I/O. Η κύρια μνήμη χωρά 7 εγγραφές.

ανταλλαγές μέχρι να εισαχθεί η επόμενη σελίδα. Με την προϋπόθεση ότι μία σύγκριση/ανταλλαγή απαιτεί 10  $\mu$ s έπεται ότι στη χειρότερη περίπτωση απαιτούνται 0.84 ms Αυτός ο χρόνος προσεγγίζει τον απαιτούμενο χρόνο για την προσπέλαση μίας σελίδας των 2400 bytes σε μία συσκευή IBM 3380.

Όταν η ρίζα μεταφερθεί στην απομονωτική μνήμη εξόδου για την αποθήκευση στο δίσκο, η τελευταία εγγραφή μεταφέρεται στη θέση της ρίζας, οπότε απαιτούνται νέες συγκρίσεις/ανταλλαγές. Ο αριθμός των ανταλλαγών αυτών είναι και πάλι της τάξης  $\log n$ , όπου  $n$  είναι οι εγγραφές που έχουν απομείνει στο δένδρο. Έτσι και στις δύο φάσεις μπορεί να υπάρξει πολύ μεγάλη επικάλυψη στους χρόνους εισόδου και εξόδου.

Στη συσκευή IBM 3380 ο μέσος χρόνος εντοπισμού και ο μέσος χρόνος περιστροφής είναι 24,3 ms, ενώ ο χρόνος για τη σειριακή προσπέλαση 10 Mb δεδομένων είναι 3500 ms. Έτσι, κατά τους υπολογισμούς αγνοείται ο χρόνος εντοπισμού, και τελικά ο χρόνος για τη δημιουργία των ταξινομημένων τμημάτων ισούται με:

$$2b \times ebt$$

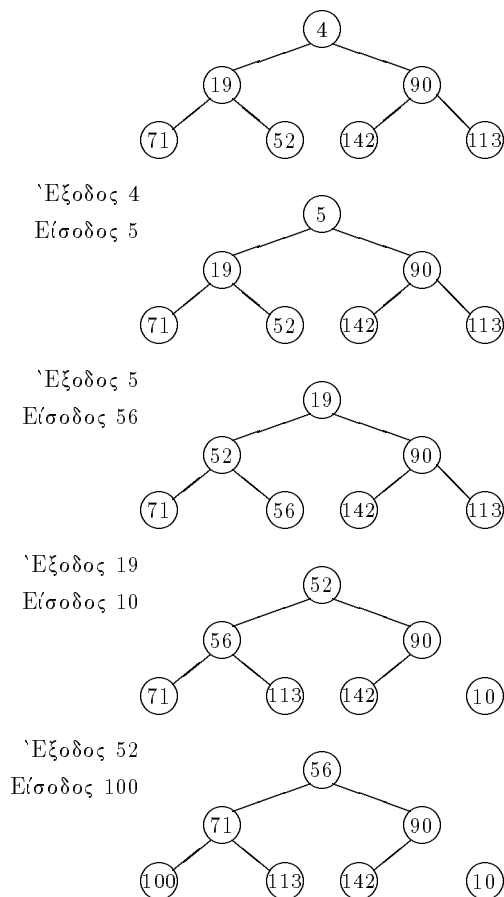
Ο χρόνος αυτός περιλαμβάνει το χρόνο για την προσπέλαση των τμημάτων και το χρόνο αποθήκευσης των ταξινομημένων τμημάτων και πάλι στο δίσκο.

### 6.3 Ταξινόμηση με δύο συσκευές δίσκων

Αν υπάρχουν δύο διαθέσιμες συσκευές δίσκων, τότε είναι δυνατό να προκύψουν ταξινομημένα τμήματα μεγαλύτερα από το μέγεθος της διαθέσιμης κύριας μνήμης. Μάλιστα έχει αποδειχθεί ότι το μέγεθος των τμημάτων είναι διπλάσιο της κύριας μνήμης κατά μέσο όρο. Άλλο σημαντικότερο πλεονέκτημα είναι ότι όλοι οι απαιτούμενοι χρόνοι για την προσπέλαση, την ταξινόμηση και την επαναποθήκευση στο δίσκο μπορεί να επικαλυφθούν.

Στην περίπτωση αυτή, για την εσωτερική ταξινόμηση χρησιμοποιείται ο αλγόριθμος της επιλογής αντικατάστασης, που έχει πολλά κοινά σημεία με τον αλγόριθμο ταξινόμησης σωρού. Όπως και με τον τελευταίο αλγόριθμο, αρχικά δημιουργείται ένας σωρός από εγγραφές που εισάγονται από τη μία συσκευή. Ο σωρός αυτός καταλαμβάνει όλο το χώρο της διαθέσιμης κύριας μνήμης. Στη συνέχεια, οι εγγραφές με το μικρότερο κλειδί που ομαδοποιούνται σε μία σελίδα αποθηκεύονται στον άλλο δίσκο και ταυτόχρονα μία νέα σελίδα μεταφέρεται από την πρώτη συσκευή στην κύρια μνήμη.

Η νέα σελίδα μπορεί να περιέχει μερικές εγγραφές με κλειδιά μεγαλύτερα από τα κλειδιά των εγγραφών, που έχουν ήδη αποθηκευθεί στο δίσκο. Αυτές οι εγγραφές μπορούν να ενσωματωθούν στον αρχικό σωρό. Όμως η νέα σελίδα μπορεί να περιέχει και εγγραφές με κλειδιά μικρότερα από αυτά που έχουν ήδη περάσει στην έξοδο. Αυτές οι εγγραφές δεν εισάγονται στον υπάρχοντα σωρό, αλλά σχηματίζουν ένα νέο σωρό. Έτσι, στην κύρια μνήμη δημιουργούνται δύο σωροί. Στο Σχήμα 6.3 παρουσιάζεται ένα παράδειγμα εφαρμογής της μεθόδου, θεωρώντας ότι η κύρια μνήμη χωρά 7 εγγραφές. Η αρχική δομή παρουσιάζει τον αρχικό σωρό των 7 εγγραφών, αλλά στη συνέχεια η έξοδος και η είσοδος γίνονται ταυτόχρονα ανά μία εγγραφή.



Σχήμα 6.3: Επιλογή αντικατάστασης με επικάλυψη I/O. Η κύρια μνήμη χωρά 7 εγγραφές.

Οι διαδικασίες εισόδου/εξόδου μπορούν να συντονισθούν. Σε αντίθεση με τη μέθοδο ταξινόμησης σωρού, στη μέθοδο αυτή όταν μία εγγραφή περνά στην απομονωτική μνήμη εξόδου, τότε τη θέση της στη ρίζα του σωρού δεν καταλαμβάνει η τελευταία εγγραφή του πίνακα αλλά μία νέα εγγραφή. Έτσι η είσοδος και η έξοδος γίνονται ταυτόχρονα σε ένα βήμα με επικάλυψη των σχετικών χρόνων. Συνεπώς είναι βέβαιο ότι ο νέος σωρός θα γίνει μεγαλύτερος από τον αρχικό, επειδή οι περισσότερες από τις εισαγόμενες εγγραφές θα έχουν μικρότερα κλειδιά από τις ήδη περασμένες στο δίσκο.



Συνήθως στο σημείο αυτό δεν υπάρχει μεγάλη επικάλυψη μεταξύ εισόδου και εξόδου, επειδή οι περισσότερες νέες εγγραφές κατευθύνονται στο νέο σωρό. Σε κάποια χρονική στιγμή ο ένας σωρός θα αδειάσει, οπότε στο σημείο αυτό τελειώνει η δημιουργία ενός ταξινομημένου τμήματος. Έτσι η διαδικασία συνεχίζεται με το νέο σωρό και τη δημιουργία του επόμενου ταξινομημένου τμήματος. Η χρήση μεγαλύτερων σελίδων και απομονωτικών μνημών μετριάξει τη δυσμενή επίπτωση του φαινομένου αυτού.

Αν οι εγγραφές έρχονται περίπου ταξινομημένες, τότε τα προκύπτοντα τμήματα θα είναι πολύ μεγάλα. Αντίθετα, αν οι εγγραφές που έρχονται στην είσοδο είναι σχεδόν διατεταγμένες κατά φθίνουσα τάξη, τότε τα τμήματα θα είναι μικρά. Έχει αποδειχθεί ότι για τιμές κλειδιών που υπακούουν σε τυχαία κατανομή, το μέγεθος των ταξινομημένων τμημάτων είναι περίπου διπλάσιο από το μέγεθος της διαθέσιμης κύριας μνήμης.

Η επιλογή αντικατάστασης είναι μία μέθοδος εξαιρετικά αποτελεσματική όταν υπάρχουν δύο συσκευές δίσκων. Έτσι αμέσως μετά την αρχική φόρτωση της κύριας μνήμης με τις πρώτες σελίδες, η είσοδος με την έξοδο μπορούν να γίνουν παράλληλα στις δύο συσκευές. Αν υπάρχει μόνο μία συσκευή, τότε η μέθοδος δεν είναι αποτελεσματική, γιατί ο βραχίονας του δίσκου θα μετακινείται συνεχώς εμπρός και πίσω για να εκτελέσει τις διαδοχικές αναγνώσεις και αποθηκεύσεις. Ο απαιτούμενος χρόνος από τη μέθοδο αυτή είναι:

$$b \times ebt$$

χωρίς να λαμβάνεται υπ' όψη ο χρόνος για την προσπέλαση των πρώτων σελίδων που εκτελείται χωρίς παραλληλισμό στην έξοδο, επειδή θεωρείται μικρός σε σχέση με την προσπέλαση όλου του αρχείου.

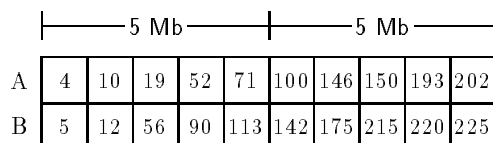
Σε τελική ανάλυση, η μέθοδος επιλογής αντικατάστασης σε σχέση με τη μέθοδο ταξινόμησης σωρού υπερτερεί σε δύο σημεία.

- τα παραγόμενα τμήματα έχουν κατά μέσο όρο διπλάσιο μέγεθος, και
- το χρονικό κόστος υποδιπλασιάζεται.

Όμως από την άλλη πλευρά υπάρχει επιπρόσθετο κόστος επειδή απαιτούνται δύο συσκευές δίσκων.

## 6.4 Συγχώνευση με μία συσκευή δίσκων

Στο σημείο αυτό πλέον αρχίζει η δεύτερη φάση, όπου τα ταξινομημένα τμήματα πρέπει να συγχωνευθούν ώστε να προκύψει η τελική ταξινομημένη μορφή του αρχείου. Στη συνέχεια υποτίθεται ότι διαθέσιμη υπάρχει μία συσκευή δίσκων με χωρητικότητα τουλάχιστο διπλάσια από το μέγεθος του αρχείου των 2,4 Gb. Αρχικά, λοιπόν, χρησιμοποιείται ο αλγόριθμος ταξινόμησης σωρού, οπότε παράγονται 240 ταξινομημένα τμήματα μεγέθους 10 Mb. Ο απαιτούμενος χρόνος για τη σειριακή προσπέλαση αρχείου αυτού του μεγέθους είναι περίπου 14 λεπτά. Συνεπώς, ο απαιτούμενος χρόνος για τη δημιουργία των τμημάτων είναι 28 λεπτά.



Μνήμη A					Μνήμη B					Έξοδος
4	10	19	52	71	5	12	56	90	113	
	10	19	52	71	5	12	56	90	113	4
	10	19	52	71		12	56	90	113	5
		19	52	71		12	56	90	113	10
		19	52	71			56	90	113	12
			52	71			56	90	113	19
				71			56	90	113	52
								90	113	56
								90	113	71
100	146	150	193	202				90	113	
100	146	150	193	202					113	90
	146	150	193	202					113	100
	146	150	193	202						113
	146	150	193	202	142	175	215	220	225	
	146	150	193	202		175	215	220	225	142
		150	193	202		175	215	220	225	146
		⋮					⋮			⋮

Σχήμα 6.4: Συγχώνευση δύο ταξινομημένων τμημάτων των 10 Mb με κύρια μνήμη των 10 Mb.

Η πρώτη μέθοδος συγχώνευσης που θα μελετηθεί είναι η συγχώνευση δύο δρόμων. Η μέθοδος αυτή κάθε φορά συγχωνεύει δύο μόνο ταξινομημένα τμήματα. Επειδή η διαθέσιμη μνήμη είναι 10 Mb, δεν είναι δυνατό να μεταφερθούν στη μνήμη δύο τμήματα την ίδια στιγμή. Για το λόγο αυτό υποτίθεται ότι υπάρχουν απομονωτικές μνήμες των 5 Mb, όπου φορτώνονται τα πρώτα μισά των δύο τμημάτων. Ακόμη, επειδή διατίθεται μία μόνο συσκευή δίσκου δεν είναι δυνατόν τα τμήματα να προσπελασθούν με επικάλυψη των χρόνων. Μόνο κατά τη φάση της εξόδου είναι δυνατό να εφαρμοσθεί η τεχνική της χρήσης διπλής απομονωτικής μνήμης και να επιτευχθεί παραλληλισμός. Το Σχήμα 6.4 παρουσιάζει ένα παράδειγμα συγχώνευσης, όπου φαίνεται πως γίνεται διαδοχικά η είσοδος και η έξοδος.

Έστω, λοιπόν, ότι κατά τη διαδικασία της συγχώνευσης εξαντλείται το μισό ενός τμήματος. Τότε το δεύτερο μισό του ίδιου τμήματος έρχεται στη μνήμη και η συγχώνευση συνεχίζεται με την παράλληλη έξοδο. Τελικά, ο απαιτούμενος χρόνος για τη συγχώνευση δύο ταξινομημένων τμημάτων είναι:

$$(4 + 3) \times (r + s) + 2 \times 2 \times seg \times ebt$$

όπου *seg* είναι ο αριθμός των σελίδων κάθε ταξινομημένου τμήματος. Το πρώτο γινόμενο, προφανώς, δίνει το χρόνο εντοπισμού και περιστροφής. Ο συντελεστής 4 αντιστοιχεί στις τέσσερις προσπελάσεις στο δίσκο για τα ημίσεα των τμημάτων, ενώ ο συντελεστής 3 αντιστοιχεί στις φορές που γίνεται αποθήκευση στο δίσκο (οι δύο πρώτες αναγνώσεις γίνονται πριν την πρώτη αποθήκευση).

Αν στις παραμέτρους δοθούν οι γνωστές τιμές, τότε το πρώτο γινόμενο ισούται με 170 ms, ενώ το δεύτερο ισούται με 14000 ms. Αν και το κόστος που προκύπτει από το πρώτο ημιάθροισμα δεν είναι τόσο μεγάλο όσο το κόστος από το δεύτερο, εντούτοις διατηρείται στον τύπο. Ο λόγος είναι ότι, στη συνέχεια θα δοθούν μέθοδοι συγχώνευσης περισσότερων από δύο δρόμους εκεί θα υπάρξει διαφοροποίηση.

Ο απαιτούμενος χρόνος για το πρώτο πέρασμα συγχώνευσης των 120 ζευγών τμημάτων των 10 Mb είναι περίπου 28.4 λεπτά. Στα επόμενα περάσματα θα πρέπει να συγχωνευθούν 60 ζεύγη τμημάτων των 20 Mb, 30 ζεύγη των 40 Mb, κοκ. Δηλαδή, θα χρειασθούν συνολικά  $\lceil \log_2 240 \rceil$  περάσματα. Στον Πίνακα 6.1 φαίνεται για κάθε πέρασμα πόσα τμήματα υπάρχουν και ποιό είναι το μέγεθός τους.

Μετά το πρώτο πέρασμα ο αριθμός των προσπελάσεων στο δίσκο για αποθήκευση προσεγγίζει σχετικά περισσότερο προς τον αριθμό των προσπε-

Πέρασμα	1	2	3	4	5	6	7	8
Μέγεθος τμήματος (Mb)	10	20	40	80	160	7×320 + 160	3×640 + 480	1280 1120
Αριθμός	240	120	60	30	15	8	4	2

Πίνακας 6.1: Συγχώνευση δύο δρόμων για 240 τμήματα.

λάσεων για ανάγνωση. Έτσι, όταν κατά το τέταρτο πέρασμα το μέγεθος των τμημάτων είναι 80 Mb χρειάζονται 32 αναγνώσεις και 31 αποθηκεύσεις για τη συγχώνευση δύο τμημάτων. Για το λόγο αυτό, τελικά γίνεται δεκτό ότι για κάθε κομμάτι τμήματος που έρχεται στην απομονωτική μνήμη εισόδου αντιστοιχεί ακριβώς μία προσπέλαση στο δίσκο. Αν, λοιπόν, το αρχείο αποτελείται από  $nsg$  τμήματα, τότε κάθε πέρασμα απαιτεί χρόνο:

$$2 \times 2 \times nsg \times (r + s) + 2 \times b \times ebt$$

Το κόστος αυτό πρέπει να πολλαπλασιασθεί επί τον αριθμό των περασμάτων, για να δώσει το συνολικό χρονικό κόστος των περασμάτων. Επίσης στο ποσό αυτό πρέπει να προστεθεί και ο αρχικός χρόνος για την ταξινόμηση των τμημάτων. Για το συγκεκριμένο παράδειγμα, λοιπόν, προκύπτει ότι συνολικά απαιτούνται:

$$8 \times 28,4 + 28 = 255 \text{ λεπτά ή } 4,25 \text{ ώρες}$$

Στη συνέχεια θα μελετηθεί η συγχώνευση τεσσάρων δρόμων. Επειδή θα εξετασθεί μόνο η δεύτερη φάση της συγχώνευσης, υποτίθεται ότι τα τμήματα είναι ήδη ταξινομημένα. Αρχικά, λοιπόν, από τέσσερα ταξινομημένα τμήματα μεταφέρονται στην κύρια μνήμη τα πρώτα τέταρτά τους που συγχωνεύονται κατά το γνωστό τρόπο. Μετά το πρώτο πέρασμα θα προκύψουν 60 τμήματα των 40 Mb. Το δεύτερο πέρασμα θα δώσει 15 τμήματα των 160 Mb, ενώ το τρίτο πέρασμα θα δώσει 4 τμήματα. Συνεπώς, χρειάζονται συνολικά 4 περάσματα. Ωστόσο, επειδή κάθε φορά έρχονται στην κύρια μνήμη μικρά κομμάτια (τέταρτα) των τμημάτων, το κόστος του χρόνου εντοπισμού και περιστροφής είναι σχετικά μεγαλύτερο. Ο απαιτούμενος χρόνος για ένα πέρασμα της μεθόδου αυτής είναι:

$$2 \times 4nsg \times (r + s) + 2 \times b \times ebt$$

που για το συγκεκριμένο παράδειγμα δίνει 28,8 λεπτά. Άρα για τα τέσσερα περάσματα μαζί με το κόστος της αρχικής ταξινόμησης των τμημάτων

απαιτούνται:

$$4 \times 28,8 + 28 = 143 \text{ λεπτά}$$

δηλαδή 2,4 ώρες, που είναι σημαντικά μικρότερος χρόνος από τον απαιτούμενο χρόνο για τη συγχώνευση δύο δρόμων.

Στη γενική μορφή, μία συγχώνευση  $P$  δρόμων για κάθε πέρασμα απαιτεί (σε ms) χρόνο ίσο με:

$$2P \times nsg \times (r + s) + 2 \times b \times ebt$$

Ο αριθμός των περασμάτων είναι  $\lceil \log_P nsg \rceil$ , οπότε ο συνολικός χρόνος είναι:

$$2 \times b \times ebt + \lceil \log_P nsg \rceil \times (2 \times P \times nsg \times (r + s) + 2 \times b \times ebt)$$

Αν  $P = nsg$ , τότε απαιτείται ένα πέρασμα και ο σχετικός χρόνος είναι:

$$4b \times ebt + 2 \times nsg^2 \times (r + s)$$

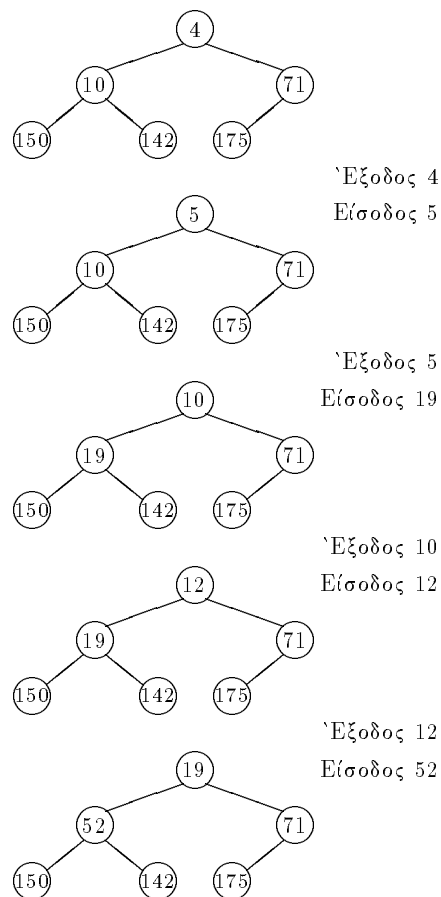
Αν χρησιμοποιηθεί ταξινόμηση 16 δρόμων, τότε το συγκεκριμένο αρχείο απαιτεί δύο περάσματα. Λαμβάνοντας υπ' όψη και τον αυξημένο χρόνο εντοπισμού και περιστροφής, τελικά προκύπτει ότι ο συνολικός χρόνος ταξινόμησης και συγχώνευσης είναι 62 λεπτά + 28 λεπτά = 1,5 ώρες.

Μέχρι στιγμής δεν αναφέρθηκε πως γίνεται η συγχώνευση. Η χρησιμοποιούμενη δομή είναι ένας σωρός, που αποτελείται από τα μικρότερα κλειδιά των τμημάτων. Ένα παράδειγμα της διαδικασίας αυτής παρουσιάζεται στο Σχήμα 6.5, όπου δίνονται 6 ταξινομημένα τμήματα των 4 εγγραφών, για να ενοποιηθούν με συγχώνευσης 6 δρόμων. Η διαδικασία στηρίζεται στη μέθοδο της επιλογής αντικατάστασης. Δηλαδή, όταν η ρίζα του σωρού μεταφέρεται στην έξοδο, τότε η εγγραφή με το μικρότερο κλειδί από το ίδιο τμήμα ενσωματώνεται στο σωρό και παίρνει τη θέση της ρίζας. Βέβαια στη συνέχεια γίνονται οι απαραίτητες συγκρίσεις/ανταλλαγές ώστε να αποκατασταθεί ο ορισμός του σωρού. Σημειώνεται ότι στο σχήμα κάθε φορά παρουσιάζεται η τελική μορφή του σωρού μετά τις απαραίτητες διεργασίες.

## 6.5 Συγχώνευση με δύο συσκευές δίσκων

Αν υπάρχουν διαθέσιμες δύο συσκευές δίσκων, τότε ελαττώνεται ο απαιτούμενος χρόνος, τόσο στη φάση της δημιουργίας των ταξινομημένων τμημάτων, όσο και στη φάση της συγχώνευσης. Στη δεύτερη, λοιπόν, αυτή φάση

A	4	5	19	90
B	71	100	113	225
Γ	10	12	52	56
Δ	150	193	215	232
E	142	146	250	278
ΣΤ	175	202	220	255



Σχήμα 6.5: Συγχώνευση έξι δρόμων.

χρησιμοποιείται η τεχνική της διπλής απομονωτικής μνήμης για τον παραλληλισμό της εισόδου με την έξοδο. Ωστόσο, ο χρόνος της φάσης αυτής δεν υποδιπλασιάζεται σε σχέση με το χρόνο συγχώνευσης όταν χρησιμοποιείται μία μόνο συσκευή δίσκου. Το επόμενο παράδειγμα είναι κατατοπιστικό.

Ας υποθεθεί ότι χρησιμοποιείται η συγχώνευση δύο δρόμων. Η διαθέσιμη κύρια μνήμη είναι 10 Mb, ενώ τα δύο τμήματα που θα συγχωνευθούν έχουν μέγεθος 20 Mb. Συνεπώς, η κύρια μνήμη μπορεί να χωρέσει τα πρώτα τέταρτα (5 Mb) των τμημάτων. Αν η στατιστική κατανομή των τιμών των κλειδιών στα δύο τμήματα είναι ίδια, τότε είναι πολύ πιθανό ότι τα δύο

τιμήματα θα εξαντληθούν περίπου ταυτόχρονα. Έτσι, θα χρειασθούν δύο συνεχόμενες προσπελάσεις στο δίσκο για την πλήρωση των απομονωτικών μνημών με τα επόμενα τέταρτα των τμημάτων και θα δημιουργηθούν νεκροί χρόνοι αναμονής. Αυτό φαίνεται στο Σχήμα 6.6.

	— 5 Mb —				— 5 Mb —				— 5 Mb —				— 5 Mb —			
A	4	10	52	71	113	142	175	215	220	232	250	291	316	346	351	382
B	5	12	19	56	90	109	146	150	193	202	225	255	278	302	356	401

Είσοδος	Μνήμη A			Μνήμη B				Έξοδος
	10	52	71	5	12	19	56	4
	10	52	71	12	19	56		5
		52	71	12	19	56		10
		52	71		19	56		12
		52	71			56		19
			71			56		52
90			71					56
100							90	71
113						9	100	
142						9	10	113
⋮								
		⋮						
						⋮		
								⋮

Σχήμα 6.6: Διπλή απομονωτική μνήμη των 5 Mb για κάθε ταξινομημένο τμήμα.

Καλύτερος συντονισμός της εισόδου με την έξοδο μπορεί να επιτευχθεί με διπλασιασμό του αριθμού των απομονωτικών μνημών για κάθε τμήμα και με ταυτόχρονο υποδιπλασιασμό του μεγέθους τους. Δηλαδή, πρέπει να χρησιμοποιηθούν μνήμες των 2,5 Mb. Έτσι, όταν έχουν συγχωνευθεί οι μισές εγγραφές από κάθε τμήμα, τότε ακολουθεί νέα είσοδος από εγγραφές του ίδιου τμήματος. Με τη μέθοδο αυτή δεν μηδενίζονται οι χρόνοι αναμονής, ωστόσο επιτυγχάνεται καλύτερη επικάλυψη. Από την άλλη πλευρά υπάρχει επιπλέον κόστος εξαιτίας της αύξησης των χρόνων εντοπισμού και περιστροφής. Η χρήση απομονωτικών μνημών των 2,5 Mb φαίνεται στο Σχήμα 6.7.

Ας θεωρηθεί και πάλι το αρχείο των 6.000.000 εγγραφών προς 400 bytes, που σημαίνει ότι το αρχείο καταλαμβάνει 2,4 Gb. Αν χρησιμοποιηθεί ο αλγόριθμος επιλογής αντικατάστασης, τότε προκύπτει ότι δημιουργούνται 120

Είσοδος	Μνήμη Α			Μνήμη Β				Έξοδος
	10	52	71	5	12	19	56	4
	10	52	71		12	19	56	5
		52	71		12	19	56	10
113		52	71			19	56	12
142	113		71				56	19
90	113	142					56	52
100	113	142		90				56
146	113	142		90	100			71
150	113	142		100		113		90
175	113	142				113	150	100
⋮		⋮				⋮		⋮

Σχήμα 6.7: Διπλή απομονωτική μνήμη των 2.5 Mb για κάθε ταξινομημένο τμήμα.

ταξινομημένα τμήματα των 20 Mb. Ο απαιτούμενος χρόνος για την παραγωγή των τμημάτων αυτών είναι 14 λεπτά περίπου.

Αν η συγχώνευση είναι 120 δρόμων, τότε θα γίνει μόνο ένα πέρασμα και κάθε τμήμα θα συνεισφέρει με το 1/120 του στο διαθέσιμο χώρο των 10 Mb. Αυτή η ποσότητα αντιστοιχεί στο 1/240 των 20 Mb των τμημάτων. Οι απομονωτικές μνήμες έχουν το μισό μέγεθος από την τελευταία ποσότητα, άρα κάθε ταξινομημένο τμήμα υποδιαιρείται σε 480 ανεξάρτητα κομμάτια που πρέπει να εισαχθούν στην κύρια μνήμη.

Ακόμη υποτίθεται ότι όλα τα τμήματα είναι αποθηκευμένα στη μία συσκευή, ενώ η αποθήκευση γίνεται στην άλλη συσκευή. Κάθε φορά που ένα κομμάτι κάποιου τμήματος εξαντλείται, τότε ένα άλλο κομμάτι του ίδιου τμήματος έρχεται από την πρώτη συσκευή στην κύρια μνήμη. Ο συνολικός αριθμός κομματιών είναι  $120 \times 480 = 57.600$ . Ο χρόνος εντοπισμού και περιστροφής κυριαρχεί σε σχέση με το χρόνο μεταφοράς δεδομένων. Προκύπτει ότι ο συνολικός απαιτούμενος χρόνος (σε ms) για τη συγχώνευση είναι:

$$120 \times 480 \times (s + r) + b \times ebt = 120 \times 480 \times 24,3 + 1000.000 \times 0,84$$

δηλαδή 37,3 λεπτά. Σε σχέση, λοιπόν, με τα 62 λεπτά της συγχώνευσης με μία συσκευή έχει επιτευχθεί σημαντική βελτίωση.



Για να ολοκληρωθεί η συγχώνευση στο συγκεκριμένο αρχείο με δύο περάσματα, πρέπει η συγχώνευση να είναι 11 δρόμων, ενώ κάθε ταξινομημένο τμήμα να αποτελείται από 44 κομμάτια. Με τη μέθοδο αυτή απαιτείται η παρουσία δύο κομματιών από κάθε τμήμα στην κύρια μνήμη, άρα οι απομονωτικές μνήμες πρέπει να έχουν μέγεθος  $1/22$  της διαθέσιμης μνήμης. Γενικά, για μία συγχώνευση  $P$  δρόμων απαιτούνται  $4P$  κομμάτια από κάθε τμήμα και συνεπώς ο συνολικός αριθμός προσπελάσεων στο δίσκο είναι  $4P \times nsg$ . Ο απαιτούμενος χρόνος για συγχώνευση 11 δρόμων είναι:

$$44 \times 120 \times 24,3 + 840.000 \text{ ms}$$

που ισούται με 968 sec. Επομένως, για δύο περάσματα απαιτούνται 1936 sec ή 32,3 λεπτά. Προσθέτοντας και το χρόνο των 14 λεπτών που απαιτείται για την αρχική ταξινόμηση των τμημάτων προκύπτει ένα συνολικό χρονικό κόστος της τάξης των 46 περίπου λεπτών. Αυτός ο χρόνος είναι περίπου μισός του χρόνου των 90 λεπτών που απαιτούνται για την ταξινόμηση και συγχώνευση δύο περασμάτων με μία συσκευή δίσκων.

Το χρονικό κόστος για την αρχική ταξινόμηση με επιλογή αντικατάστασης και στη συνέχεια τη συγχώνευση  $P$  δρόμων δίνεται από τη σχέση:

$$b \times ebt + [\log_P nsg] \times (4P \times nsg \times (r + s) + b \times ebt)$$

Σε κάθε πέρασμα αντιστοιχούν  $4P \times nsg$  προσπελάσεις στο δίσκο για ανάγνωση και ισάριθμες για αποθήκευση. Οι προσπελάσεις για αποθήκευση στο δίσκο γίνονται παράλληλα, γι' αυτό το λόγο και δεν υπεισέρχεται στον υπολογισμό του κόστους. Αν η συγχώνευση γίνεται με ένα πέρασμα, δηλαδή  $P=nsg$ , τότε ο αντίστοιχος χρόνος είναι:

$$2b \times ebt + 4nsg^2 \times (r + s)$$

Αν υπάρχουν διαθέσιμες περισσότερες από δύο συσκευές δίσκων επιτυγχάνεται καλύτερος παραλληλισμός. Για παράδειγμα, χρησιμοποιώντας μία συσκευή για έξοδο και τις υπόλοιπες για την είσοδο δεδομένων μπορεί να γίνει καλύτερη επικάλυψη των χρόνων εντοπισμού και περιστροφής. Ακόμη και όταν αυτή η επικάλυψη γίνει κατά το θεωρητικά βέλτιστο τρόπο, τότε εύκολα προκύπτει ότι ο απαιτούμενος χρόνος για τη συγχώνευση ισούται με το χρόνο για μία σειριακή ανάγνωση ή αποθήκευση του αρχείου, δηλαδή 14 λεπτά. Προσθέτοντας και τον αρχικό χρόνο ταξινόμησης των τμημάτων προκύπτει συνολικός χρόνος 28 λεπτών. Αυτός ο χρόνος αποτελεί το θεωρητικό κάτω όριο και δεν μπορεί να βελτιωθεί περισσότερο.

## 6.6 Άλλες μέθοδοι εξωτερικής συγχώνευσης

Στο προηγούμενο κεφάλαιο παρουσιάστηκαν μέθοδοι ειδικά σχεδιασμένες για εξωτερική ταξινόμηση με ταινίες. Οι μέθοδοι αυτές μπορούν να εφαρμοστούν και σε δίσκους. Κατ' αναλογία προς την ανάλυση του προηγούμενου κεφαλαίου προκύπτει ότι ο αντίστοιχος συνολικός χρόνος για είσοδο και έξοδο των δεδομένων με τη μέθοδο της φυσικής συγχώνευσης είναι:

$$4 \lceil \log_P nsg \rceil \times \frac{n \times R}{\frac{Buf}{P+1}} \times \left( s + r + \frac{Buf}{(P+1) \times t'} \right)$$

για την ισοζυγισμένη συγχώνευση προκύπτει:

$$2 \lceil \log_P nsg \rceil \times \frac{n \times R}{\frac{Buf}{2P}} \times \left( s + r + \frac{Buf}{2P \times t'} \right)$$

ενώ για την πολυφασική συγχώνευση προκύπτει:

$$2 PN(P) \times \frac{n \times R}{\frac{Buf}{P+1}} \times \left( s + r + \frac{Buf}{(P+1) \times t'} \right)$$

Συγκρίνοντας τα ζεύγη των τύπων για κάθε μέθοδο και για διάφορες τιμές των παραμέτρων εύκολα φαίνεται η βελτίωση της χρονικής απόκρισης από τη χρήση των δίσκων σε σχέση με τη χρήση ταινιών. Μάλιστα στην πράξη η πολυφασική συγχώνευση έχει χρησιμοποιηθεί και με συσκευές μαγνητικών δίσκων. Ωστόσο, στη συνέχεια αποδείχθηκε η ακαταλληλότητα της μεθόδου για το περιβάλλον των μαγνητικών δίσκων. Για παράδειγμα, η ταξινόμηση του γνωστού αρχείου των 2,4 Gb με 120 τμήματα των 20 Mb απαιτεί 32 περίπου λεπτά όταν είναι διαθέσιμοι δύο δίσκοι. Σε αντίθεση, ο Knuth αναφέρει ότι όταν η πολυφασική συγχώνευση χρησιμοποιεί 10 συσκευές για να ταξινομήσει το ίδιο αρχείο, τότε ο απαιτούμενος χρόνος είναι τετραπλάσιος του χρόνου μίας σειριακής ανάγνωσης, δηλαδή 56 περίπου λεπτά. Δηλαδή, ακόμη και με 10 συσκευές η επίδοση της πολυφασικής συγχώνευσης υστερεί σε σχέση με την επίδοση της προηγούμενης μεθόδου.

Πολλές φορές παρατηρείται το φαινόμενο ότι πολλά εμπορικά πακέτα εξωτερικής ταξινόμησης είναι πολύ αργά. Ένας λόγος είναι ότι χρησιμοποιούν για την αρχική φάση της δημιουργίας των ταξινομημένων τμημάτων κάποιον αλγόριθμο τάξης  $O(n^2)$ . Ακόμη, η αρχιτεκτονική του συστήματος μπορεί να μην είναι η βέλτιστη για το συγκεκριμένο πακέτο. Επίσης, σε

ένα υπολογιστικό περιβάλλον πολλών χρηστών δημιουργούνται προβλήματα επάρκειας κύριας μνήμης, χώρου στο δίσκο, διάρκειας του χρόνου εξυπηρέτησης κάθε χρήστη κλπ. Ωστόσο, είναι βέβαιο ότι υπάρχουν και αξιόλογα διαθέσιμα εμπορικά πακέτα. Για παράδειγμα, αν ένα πακέτο χρειασθεί διπλάσιο ή τριπλάσιο χρόνο από αυτόν που έχει υπολογισθεί αναλυτικά, τότε θεωρείται ότι πρόκειται για μία καλή λύση.

## 6.7 Ασκήσεις

<1> Να εφαρμοσθεί ο αλγόριθμος επιλογής αντικατάστασης σε κύρια μνήμη χωρητικότητας 7 εγγραφών. Οι διαδοχικά εισαγόμενες εγγραφές έχουν κλειδιά: 12, 18, 25, 8, 20, 10, 7, 30, 32, 24, 9, 4, 28, 6, 29.

<2> Με τα δεδομένα της προηγούμενης άσκησης να εφαρμοσθεί ο αλγόριθμος ταξινόμησης σωρού.

<3> Σε μία συσκευή δίσκου αποθηκεύεται ένα αρχείο ασθενών νοσοκομείου με 100.000 εγγραφές των 400 bytes. Ποιά διαδικασία είναι ταχύτερη για την αναζήτηση μίας λίστας ονομάτων 500 ασθενών.

- να ταξινομηθεί το αρχείο και η λίστα, και να σαρωθεί το αρχείο μία φορά; ή
- να γίνουν 500 ανεξάρτητες αναζητήσεις σε αρχείο σωρού;

Αν η λίστα αποτελούνταν από 50 ονόματα τι θα συνέφερε;

<4> Δίνονται δύο αρχεία σωρού με 10.000.000 και 2.000.000 εγγραφές των 100 bytes αντίστοιχα. Τα δύο αρχεία έχουν 1.000.000 κοινές εγγραφές. Να προταθεί μία τουλάχιστον τεχνική για τη δημιουργία ενός μόνο αρχείου χωρίς διπλοεγγραφές και να κοστολογηθεί. Υποτίθεται ότι ο δίσκος έχει τα χαρακτηριστικά του IBM 3380 αλλά πολύ μεγαλύτερη χωρητικότητα.

<5> Δίνονται δύο ταξινομημένα αρχεία μεγέθους 500 και 200 Mb αντίστοιχα. Το περιεχόμενο των 100 Mb είναι κοινό. Πόσος χρόνος χρειάζεται για τη δημιουργία ενός μόνο αρχείου χωρίς διπλοεγγραφές. Υποτίθεται ότι ο δίσκος έχει τα χαρακτηριστικά του IBM 3380 αλλά πολύ μεγαλύτερη χωρητικότητα.

<6> Δίνεται αρχείο των 2000 Mb και σύστημα δύο συσκευών δίσκων και κύριας μνήμης των 20 Mb. Είναι προτιμότερη η συγχώνευση 100 δρόμων

ή η συγχώνευση 10 δρόμων; Είναι προτιμότερη η συγχώνευση 81 δρόμων ή η συγχώνευση 9 δρόμων; Για δεδομένες παραμέτρους του συστήματος να βρεθεί η βέλτιστη τιμή του  $P$ ;

<7> Ένα αρχείο αποτελείται από 2028 ταξινομημένα τμήματα των 4 Mb. Το μέγεθος της εγγραφής είναι 200 bytes, ενώ το μέγεθος της απομονωτικής μνήμης είναι 16 Mb. Να υπολογισθεί ο απαιτούμενος χρόνος για είσοδο και έξοδο των δεδομένων κατά την εκτέλεση μίας ισοζυγισμένης συγχώνευσης  $P$  δρόμων (όπου  $P = 4, 8, 12, 16, 20, 24$ ). Δίνεται επίσης ότι: ο αριθμός των κυλίνδρων είναι 555, αριθμός ατράκτων/κύλινδρο 39, πυκνότητα εγγραφής 19,2 Mb/άτρακτο, μέσος χρόνος αναζήτησης 25 ms, μέσος λανθάνων χρόνος 9,3 ms, μέγεθος κενού αμελητέο.

<8> Να θεωρηθούν τα δεδομένα της Άσκησης 7 και να υπολογισθεί ο απαιτούμενος χρόνος για πολυφασική συγχώνευση.

<9> Να μελετηθούν οι τύποι του Κεφαλαίου 6.6 με σκοπό την εύρεση του βέλτιστου αριθμού δρόμων  $P$  για κάθε περίπτωση.