ΚΕΦΑΛΑΙΟ 3 ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΛΓΟΡΙΘΜΟΙ

Τα **δεδομένα (data)** είναι η αφαιρετική αναπαράσταση της πραγματικότητας και συνεπώς μία απλοποιημένη όψη της.

Η συλλογή των ακατέργαστων δεδομένων και ο συσχετισμός τους δίνει ως αποτέλεσμα την ***πληροφορία*** **(information).** Δεν είναι εύκολο να δοθεί επακριβής ορισμός της έννοιας της πληροφορίας, αλλά μπορεί να θεωρηθεί ότι ο αλγόριθμος είναι το μέσο για την παραγωγή πληροφορίας από τα δεδομένα.

Η μέτρηση, η κωδικοποίηση, η μετάδοση της πληροφορίας αποτελεί αντικείμενο μελέτης ενός ιδιαίτερου κλάδου, της ***Θεωρίας Πληροφοριών*** **(Information Theory),** που είναι ένα ιδιαίτερα σημαντικό πεδίο της Πληροφορικής. Πληροφορική θεωρείται η επιστήμη που μελετά τα δεδομένα από τις ακόλουθες σκοπιές:

- **Υλικού.** Το υλικό (hardware), δηλαδή η μηχανή, επιτρέπει στα δεδομένα ενός προγράμματος να αποθηκεύονται στην κύρια μνήμη και στις περιφερειακές συσκευές του υπολογιστή με διάφορες αναπαραστάσεις (representations). Τέτοιες μορφές είναι η δυαδική, ο κώδικας ASCII (βλ. παράρτημα), ο κώδικας EBCDIC, το συμπλήρωμα του 1 ή του 2 κ.λπ.

- **Γλωσσών προγραμματισμού.** Οι γλώσσες προγραμματισμού υψηλού επιπέδου (high level programming languages) επιτρέπουν τη χρήση διάφορων τύπων (types) μεταβλητών (variables) για να περιγράψουν ένα δεδομένο. Ο μεταφραστής κάθε γλώσσας φροντίζει για την αποδοτικότερη μορφή αποθήκευσης, από πλευράς υλικού, κάθε μεταβλητής στον υπολογιστή.

- **Δομών Δεδομένων.** Δομή δεδομένων (data structure) είναι ένα σύνολο δεδομένων μαζί με ένα σύνολο επιτρεπτών λειτουργιών επί αυτών. Για παράδειγμα, μία τέτοια δομή είναι η **εγγραφή (record),** που μπορεί να περιγράφει ένα είδος, ένα πρόσωπο κλπ. Η εγγραφή αποτελείται από τα **πεδία (fields)** που αποθηκεύουν **χαρακτηριστικά (attributes)** διαφορετικού τύπου, όπως για παράδειγμα ο κωδικός, η περιγραφή κλπ. Άλλη μορφή δομής δεδομένων είναι **το αρχείο** που αποτελείται από ένα σύνολο εγγραφών. Μία επιτρεπτή λειτουργία σε ένα αρχείο είναι η σειριακή προσπέλαση όλων των εγγραφών του.

- **Ανάλυσης Δεδομένων.** Τρόποι καταγραφής και αλληλοσυσχέτισης των δεδομένων μελετώνται έτσι ώστε να αναπαρασταθεί η γνώση για πραγματικά γεγονότα. Οι τεχνολογίες των Βάσεων Δεδομένων (Databases), της Μοντελοποίησης Δεδομένων (Data Modelling) και της Αναπαράστασης Γνώσης (Knowledge Representation) ανήκουν σε αυτή τη σκοπιά μελέτης των δεδομένων.

Αλγόριθμοι + Δομές Δεδομένων = Προγράμματα

Δομή Δεδομένων είναι ένα σύνολο αποθηκευμένων δεδομένων που υφίστανται επεξεργασία από ένα σύνολο λειτουργιών.

Οι βασικές λειτουργίες (ή αλλιώς πράξεις) επί των δομών δεδομένων είναι οι ακόλουθες:

- ***Προσπέλαση*** **(access),** πρόσβαση σε ένα κόμβο με σκοπό να εξετασθεί ή να τροποποιηθεί το περιεχόμενό του.
- ***Εισαγωγή*** **(insertion),** δηλαδή η προσθήκη νέων κόμβων σε μία υπάρχουσα δομή.
- ***Διαγραφή*** **(deletion),** που αποτελεί το αντίστροφο της εισαγωγής, δηλαδή ένας κόμβος αφαιρείται από μία δομή.
- ***Αναζήτηση*** **(searching),** κατά την οποία προσπελαύνονται οι κόμβοι μιας δομής, προκειμένου να εντοπιστούν ένας ή περισσότεροι που έχουν μια δεδομένη ιδιότητα.
- ***Ταξινόμηση*** **(sorting),** όπου οι κόμβοι μιας δομής διατάσσονται κατά αύξουσα ή φθίνουσα σειρά.
- ***Αντιγραφή*** **(copying),** κατά την οποία όλοι οι κόμβοι ή μερικοί από τους κόμβους μίας δομής αντιγράφονται σε μία άλλη δομή.
- ***Συγχώνευση*** **(merging),** κατά την οποία δύο ή περισσότερες δομές συνενώνονται σε μία ενιαία δομή.
- ***Διαχωρισμός*** **(separation),** που αποτελεί την αντίστροφη πράξη της συγχώνευσης.

Δομές δεδομένων

Οι δομές δεδομένων διακρίνονται σε δύο μεγάλες κατηγορίες:

- **τις στατικές (static)** και τις
- **δυναμικές (dynamic)**

Οι **δυναμικές δομές** δεν αποθηκεύονται σε συνεχόμενες θέσεις μνήμης αλλά στηρίζονται στην τεχνική της λεγόμενης δυναμικής παραχώρησης μνήμης (dynamic memory allocation).

Με άλλα λόγια, οι δομές αυτές δεν έχουν σταθερό μέγεθος, αλλά ο αριθμός των κόμβων τους μεγαλώνει και μικραίνει καθώς στη δομή εισάγονται νέα δεδομένα ή διαγράφονται κάποια δεδομένα αντίστοιχα. Όλες οι σύγχρονες γλώσσες προγραμματισμού προσφέρουν τη δυνατότητα δυναμικής παραχώρησης μνήμης.

Με τον όρο **στατική δομή δεδομένων** εννοείται ότι το ακριβές μέγεθος της απαιτούμενης κύριας μνήμης καθορίζεται κατά τη στιγμή του προγραμματισμού τους, και κατά συνέπεια κατά τη στιγμή της μετάφρασής τους και όχι κατά τη στιγμή της εκτέλεσης τους προγράμματος. Μία άλλη **σημαντική διαφορά** σε σχέση με τις δυναμικές δομές είναι ότι τα στοιχεία των στατικών δομών αποθηκεύονται σε συνεχόμενες θέσεις μνήμης.

Πίνακες

Μπορούμε να ορίσουμε τον **πίνακα** ως μια δομή που περιέχει στοιχεία του ίδιου τύπου (δηλαδή ακέραιους, πραγματικούς κ.λπ).

Γενικά η αναφορά στα στοιχεία ενός πίνακα γίνεται με τη χρήση του συμβολικού ονόματος του πίνακα ακολουθούμενου από την τιμή ενός ή περισσότερων *δεικτών* (indexes) σε παρένθεση ή αγκύλη.

Ένας πίνακας μπορεί να είναι μονοδιάστατος, αλλά στη γενικότερη περίπτωση μπορεί να είναι δισδιάστατος, τρισδιάστος και γενικά ν-διάστατος πίνακας. Όσον αφορά στους δισδιάστους πίνακες σημειώνεται ότι αν το μέγεθος των δύο διαστάσεων είναι ίσο, τότε ο πίνακας λέγεται *τετραγωνικός* (square) και γενικά συμβολίζεται ως πίνακας n x n.

Στοίβα και Ουρά

Οι πίνακες χρησιμεύουν για την αποθήκευση και διαχείριση δύο σπουδαίων δομών, της **στοίβας (stack)** και της **ουράς (queue)**

Στοίβα

Τα δεδομένα που βρίσκονται στην κορυφή της στοίβας λαμβάνονται πρώτα, ενώ αυτά που βρίσκονται στο βάθος της στοίβας λαμβάνονται τελευταία. Αυτή η μέθοδος επεξεργασίας ονομάζεται Τελευταίο μέσα, πρώτο έξω ή απλούστερα με την αγγλική συντομογραφία **LIFO(Last-In-First-Out).**

Δύο είναι οι κύριες λειτουργίες σε μία στοίβα:

- **η ώθηση (push)** στοιχείου στην κορυφή της στοίβας, και
- **η απώθηση (pop)** στοιχείου από τη στοίβα.

Η διαδικασία της ώθησης πρέπει οπωσδήποτε να ελέγχει, αν η στοίβα είναι γεμάτη, οπότε λέγεται ότι συμβαίνει **υπερχείλιση (overflow)** της στοίβας.

Αντίστοιχα, η διαδικασία απώθησης ελέγχει, αν υπάρχει ένα τουλάχιστον στοιχείο στη στοίβα, δηλαδή ελέγχει αν γίνεται **υποχείλιση (underflow)** της στοίβας.

Μια **βοηθητική μεταβλητή** (με όνομα συνήθως **top**) χρησιμοποιείται για να δείχνει το στοιχείο που τοποθετήθηκε τελευταίο στην κορυφή της στοίβας.

Ουρά

Σε μία ουρά αναμονής με ανθρώπους, συμβαίνει να εξυπηρετείται εκείνος που στάθηκε στην ουρά πρώτος από όλους τους άλλους (αν και υπάρχουν εξαιρέσεις που όμως δεν θα εξετασθούν στο βιβλίο αυτό). Η μέθοδος αυτή επεξεργασίας ονομάζεται Πρώτο μέσα, πρώτο έξω ή απλούστερα ακολουθώντας την αγγλική συντομογραφία **FIFO (First-In-First-Out).**

Δύο είναι οι κύριες λειτουργίες που εκτελούνται σε μία ουρά:

- **η εισαγωγή (enqueue)** στοιχείου στο πίσω άκρο της ουράς, και
- **η εξαγωγή (dequeue)** στοιχείου από το εμπρός άκρο της ουράς

Σε αντίθεση με τη δομή της στοίβας, στην περίπτωση της ουράς απαιτούνται δύο δείκτες: **ο εμπρός (front)** και **ο πίσω (rear)** δείκτης, που μας δίνουν τη θέση του στοιχείου που σε πρώτη ευκαιρία θα εξαχθεί και τη θέση του στοιχείου που μόλις εισήλθε.

Αναζήτηση

Το πρόβλημα της **αναζήτησης (searching)** ενός στοιχείου σε πίνακα είναι ιδιαίτερα ενδιαφέρον λόγω της χρησιμότητας του σε πλήθος εφαρμογών. Η πιο απλή μορφή αναζήτησης στοιχείου σε πίνακα είναι **η σειριακή (sequential) ή γραμμική (linear)** μέθοδος.

Η σειριακή μέθοδος αναζήτησης είναι η πιο απλή, αλλά και η λιγότερη αποτελεσματική μέθοδος αναζήτησης.

Έτσι, δικαιολογείται η χρήση της μόνο σε περιπτώσεις όπου:

- ο πίνακας είναι μη ταξινομημένος
- ο πίνακας είναι μικρού μεγέθους (για παράδειγμα, n < 20)
- η αναζήτηση σε ένα συγκεκριμένο πίνακα γίνεται σπάνια

Ταξινόμηση

Η τακτοποίηση των κόμβων μίας δομής με μία ιδιαίτερη σειρά είναι μία πολύ σημαντική λειτουργία που ονομάζεται ταξινόμηση (sorting) ή διάταξη (ordering). Συνήθως η σειρά αυτή είναι η αύξουσα τάξη (ascending sequence) της τιμής των μεγεθών προς ταξινόμηση.

Δοθέντων των στοιχείων *a1,a2,...,an* η ταξινόμηση συνίσταται στη *μετάθεση* (permutation) της θέσης των στοιχείων, ώστε να τοποθετηθούν σε μία σειρά ak1,ak2,...,akn έτσι ώστε, δοθείσης μίας *συνάρτησης διάταξης* (ordering function), f, να ισχύει: f(ak1)<f(ak2)<...<f(akn)