

Κεφάλαιο 1.

Δεδομένα

1.1 Πληροφορίες και δεδομένα

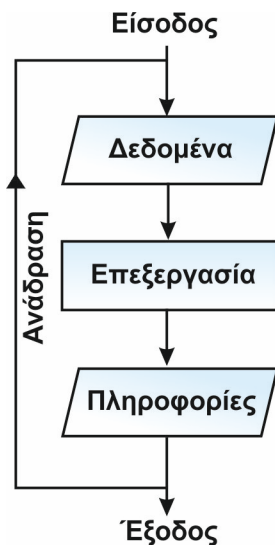
Οι όροι πληροφορία και δεδομένο χρησιμοποιούνται από πολλούς αδιάκριτα. Έχουν όμως σημαντικές διαφορές, οπωσδήποτε λεπτές και δυσδιάκριτες από πρώτη άποψη. Το πρότυπο ελληνικής ορολογίας που εκπονεί η Τεχνική Επιτροπή 48 του ΕΛΟΤ προτείνει τον ακόλουθο ορισμό της λέξης δεδομένο (αγγλικά datum-πληθ. data):



“Μια παράσταση γεγονότων, εννοιών ή εντολών σε τυποποιημένη μορφή που είναι κατάλληλη για επικοινωνία, ερμηνεία ή επεξεργασία από τον άνθρωπο ή από αυτόματα μέσα”.

Η **επεξεργασία δεδομένων** (data processing) αναφέρεται στην εκτέλεση με τον υπολογιστή διαφόρων πράξεων/λειτουργιών πάνω στα δεδομένα σε μορφή κατανοητή από τον άνθρωπο ή από άλλη μηχανή. Το αποτέλεσμα της επεξεργασίας δεδομένων είναι η **πληροφορία**. Για παράδειγμα ένας τηλεφωνικός αριθμός και ένα ονοματεπώνυμο αποτελούν δεδομένα. Δεν παρέχουν όμως καμία πληροφορία. Πληροφορία παράγεται μόνο αν σχετισθούν μεταξύ τους, ότι δηλαδή το τηλέφωνο του τάδε είναι το συγκεκριμένο. Με βάση τις πληροφορίες λαμβάνονται **αποφάσεις** και γίνονται διάφορες ενέργειες. Στη συνέχεια οι ενέργειες αυτές παράγουν νέα δεδομένα, αυτές νέες πληροφορίες, οι τελευταίες νέες αποφάσεις και ενέργειες κ.ο.κ. (βλ. σχήμα 1.1).

Η διεργασία αυτή μπορεί να ανακυκλώνεται, με τις πληροφορίες που παράχθηκαν, να επανατροφοδοτούν (feedback) την είσοδο για επανάληψη του κύκλου.



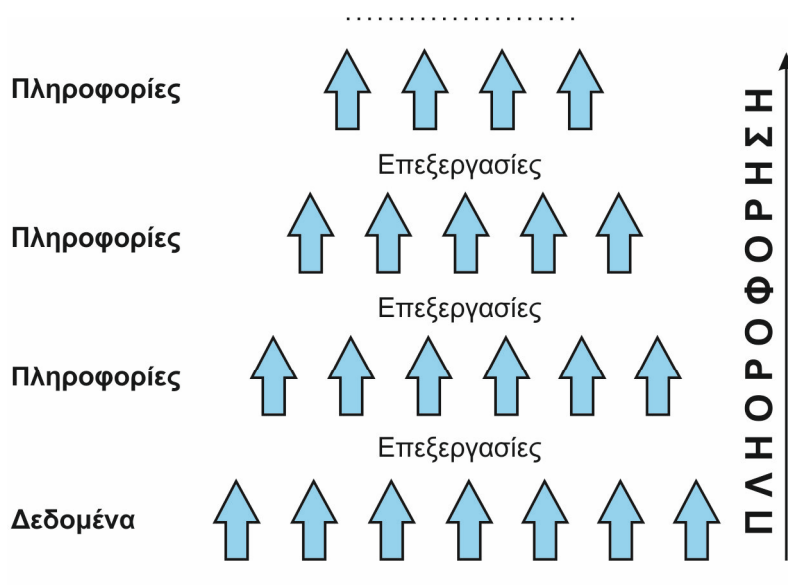
Σχήμα 1.1. Σχέση δεδομένων και πληροφοριών

Η **επεξεργασία πληροφοριών** δημιουργεί πληροφορίες ανώτερου επιπέδου (δευτερογενείς, τριτογενείς κ.λπ.). Για παράδειγμα οι πληροφορίες, αν ένας συγκεκριμένος υπάλληλος: είναι παντρεμένος, έχει δύο παιδιά, παράγουν την πληροφορία της οικογενειακής του κατάστασης.

Η **Πληροφόρηση** αποτελεί τη συνολική εικόνα, που παρέχει ένα σύνολο πληροφοριών. Για παράδειγμα ο διευθυντής προσωπικού μιας επιχείρησης γνωρίζει (έχει πληροφόρηση) για τις ηλικίες των υπαλλήλων, όταν ξέρει τις ημερομηνίες γέννησης αυτών.

Στο σχήμα 1.2 παρουσιάζεται η οικοδόμηση της πληροφόρησης μέσα σε έναν οργανισμό ξεκινώντας από τα δεδομένα της βάσης και προχωρώντας προς τα ανώτερα διοικητικά επίπεδα. Όσο ανεβαίνουμε την κλίμακα, τόσο η πληροφόρηση γίνεται πιο γενικευμένη και τελικά μαζί με την αποκτώμενη εμπειρία καταλήγει σε **γνώση**.

Παλιότερα χρησιμοποιήθηκε ο όρος Μηχανογράφηση για την επεξεργασία δεδομένων. Όμως έχει σχεδόν εκλείψει πλέον αντικαθιστάμενος από τον όρο πληροφορική και τα παράγωγά του.



Δεδομένα: Πρωτογενή Ακατέργαστα Μεγέθη

Πληροφορία: Αποτέλεσμα Επεξεργασίας Δεδομένων

Πληροφόρηση: Αποτέλεσμα Επεξεργασίας Πληροφοριών

Σχήμα 1.2. Σχέση δεδομένων, πληροφοριών και πληροφόρησης στην επιχείρηση.

1.2 Μελέτη των δεδομένων

Η Πληροφορική ορίζεται ως η επιστήμη που μελετά τα δεδομένα από τις ακόλουθες σκοπιές:

α) Υλικού. Το υλικό επιτρέπει τα δεδομένα ενός προγράμματος να αποθηκεύονται στην κύρια μνήμη και στις περιφερειακές συσκευές ενός υπολογιστή με διάφορες μορφές. Είναι γνωστό ότι οι αριθμοί αποθηκεύονται στο δυαδικό σύστημα. Για την παράσταση αρνητικών ακέραιων χρησιμοποιείται το συμπλήρωμα ως προς 2 και σπανιότερα το συμπλήρωμα ως προς 1. Οι αλφαριθμητικές πληροφορίες χρειάζονται κάποιον κώδικα όπως τον Ascii ή τον EBCDIC.

β) Γλωσσών προγραμματισμού. Οι γλώσσες προγραμματισμού υποστηρίζουν τη χρήση διαφόρων τύπων δεδομένων (data types). Οι πιο συνήθεις τύποι δεδομένων είναι (βλ. και πίνακες 1.1, 1.2):

- **Ακέραιοι αριθμοί** που καταλαμβάνουν ένα, δύο, τέσσερα ή οκτώ διαδοχικά bytes μνήμης. Οι ακέραιοι αριθμοί μπορούν είναι προσημασμένοι ή μη και να δίδονται στο δεκαδικό, οκταδικό ή δεκαεξαδικό σύστημα.
- **Πραγματικοί αριθμοί.** Αυτοί διακρίνονται σε *σταθερής υποδιαστολής* (fixed point) και *κινητής υποδιαστολής* (floating point). Στους πρώτους διατίθεται ένα σταθερό πλήθος bytes για το ακέραιο μέρος και για το κλασματικό μέρος. Οι δεύτεροι αφορούν την παράσταση ενός αριθμού με την εκθετική μορφή του. Κατ' αυτή κρατείται χωριστά το κλασματικό μέρος (μαντίσα) και ο εκθέτης της βάσης. Π.χ. στο δεκαδικό σύστημα ο αριθμός 1500000 μπορεί να γραφεί ως $0,15 \times 10^7$.

Για την αποθήκευση αυτού του αριθμού αρκεί να κρατηθούν το κλασματικό μέρος 15 και το 7 που είναι ο εκθέτης του 10. Κάθε ένα μέρος κατακρατείται στο δυαδικό σύστημα σε έναν αριθμό bytes, οπότε ποικίλει το εύρος των δυνατών τιμών και η ακρίβεια.

- **Λογικά δεδομένα** που έχουν μόνο δύο τιμές, αληθής (true) ή ψευδής (false).
- **Συμβολοσειρές** (strings) για την παράσταση αλφαριθμητικών δεδομένων.

Πίνακας 1.1. Αριθμητικοί τύποι δεδομένων

Τύπος	bytes	εύρος τιμών	ακρίβεια
Ακέραιος μη προσημασμένος	1	0 ... 255	
Ακέραιος προσημασμένος	1	-128 ... 127	
Ακέραιος μη προσημασμένος	2	0 ... 65.535	
Ακέραιος προσημασμένος	2	-32.768 ... 32.767	
Ακέραιος μη προσημασμένος	4	0 ... 4.294.967.295	
Ακέραιος προσημασμένος	4	-2.147.483.648 ... +2.147.483.647	
Ακέραιος προσημασμένος	8	$-2^{63} \dots 2^{63} - 1$	
Πραγματικός σταθερής υποδιαστολής	8	$-10^{15} \dots +10^{15}$	15
Πραγματικός κινητής υποδιαστολής	4	$-10^{38} \dots +10^{38}$	7
Πραγματικός κινητής υποδιαστολής	8	$-10^{308} \dots +10^{308}$	15-16
Πραγματικός κινητής υποδιαστολής	10	$-10^{4932} \dots +10^{4932}$	
Δεκαδικός	14	$-10^{28} \dots +10^{28}$	28



Στους αρνητικούς αριθμούς χρησιμοποιείται το συμπλήρωμα ως προς 2. Στο εύρος τιμών των παραστάσεων κινητής υποδιαστολής δίδεται η τάξη μεγέθους. Η ακρίβεια στους ακέραιους είναι ίση με το πλήθος των ψηφίων του αριθμού.

Πίνακας 1.1. Οι αριθμητικοί τύποι δεδομένων που υποστηρίζονται από τις σπουδαιότερες γλώσσες προγραμματισμού

Τύπος	bytes	C / C++	Delphi	Python	Visual Basic
Ακέραιος μη προς.	1	unsigned char	byte		Byte
Ακέραιος προσημ.	1	signed char	shortint		SByte
Ακέραιος μη προς.	2	unsigned short	word		UShort
Ακέραιος προσημ.	2	signed short	smallint		Short
Ακέραιος μη προς.	4	unsigned [long] int	cardinal		UInteger
Ακέραιος προσημ.	4	[signed] [long] int	integer		Integer
Ακέραιος προσημ.	8	long long int	int64	int	Long
Πραγμ. σταθ. υποδ.	8		currency		
Πραγμ. κινητής υποδ.	4	float	single		Single
Πραγμ. κινητής υποδ.	8	double	real, double		Double
Πραγμ. κινητής υποδ.	10		extended	float	
Δεκαδικός	16				Decimal

γ) Δομών Δεδομένων. Ο όρος δομή δεδομένων (data structure) αναφέρεται σε ένα σύνολο δεδομένων μαζί με ένα σύνολο λειτουργιών επιτρεπτών για τα δεδομένα αυτά. Υπάρχουν πολλές μορφές δομών δεδομένων. Άλλες είναι πιο κατάλληλες για χρήση στην κύρια μνήμη και άλλες στις βοηθητικές.

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

1.3 Οι Δομές Δεδομένων

Τα δεδομένα ενός προβλήματος αποθηκεύονται στον υπολογιστή είτε στην κύρια μνήμη του είτε στη βοηθητική. Η αποθήκευση αυτή δεν γίνεται με κάποιον τυχαίο τρόπο αλλά συστηματικά, δηλαδή χρησιμοποιώντας μια δομή. Η έννοια της δομής δεδομένων (data structure) είναι ιδιαίτερα σημαντική για την Πληροφορική και ορίζεται με το επόμενο τυπικό ορισμό.



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

Λειτουργίες

Κάθε μορφής δομή δεδομένων αποτελείται από ένα σύνολο στοιχείων ή κόμβων. Οι βασικές λειτουργίες ή πράξεις επί των δομών δεδομένων είναι οι ακόλουθες:

- **Προσπέλαση** (access), πρόσβαση σε έναν κόμβο με σκοπό να εξεταστεί ή να τροποποιηθεί το περιεχόμενό του.
- **Ανάκτηση** (retrieval), η με οποιονδήποτε τρόπο λήψη (ανάγνωση) του περιεχομένου ενός κόμβου.
- **Αναζήτηση** (searching), η εξέταση ενός συνόλου στοιχείων δεδομένων προκειμένου να εντοπιστούν ένα ή περισσότερα στοιχεία, που έχουν μια δεδομένη ιδιότητα.
- **Εισαγωγή** (insertion), η προσθήκη ή δημιουργία νέων κόμβων σε μια υπάρχουσα δομή.
- **Διαγραφή** (deletion) ή **ακύρωση** (cancelation), που συνιστά το αντίθετο της εισαγωγής.
- **Μεταβολή ή τροποποίηση** (modification) του περιεχομένου ενός κόμβου μιας δομής.
- **Ταξινόμηση** (sorting), όπου τα στοιχεία μιας δομής διατάσσονται κατά αύξουσα ή φθίνουσα τάξη.
- **Συγχώνευση** (merging), κατά την οποία δύο ή περισσότερες δομές συνενώνονται σε μια ενιαία δομή.
- **Διαχωρισμός** (separation), που αποτελεί την αντίστροφη πράξη της συγχώνευσης.
- **Προσάρτηση** (append), κατά την οποία μία δομή επικολλάται στο τέλος μιας άλλης.
- **Αντιγραφή** (copying), όπου όλοι ή μερικοί κόμβοι μιας δομής αντιγράφονται σε μια άλλη.

Στην πράξη σπάνια υπάρχουν όλες οι προηγούμενες λειτουργίες σε μια δομή. Συνήθως παρατηρείται το φαινόμενο μια δομή δεδομένων να είναι αποδοτικότερη από μια άλλη με κριτήριο κάποια λειτουργία π.χ. την αναζήτηση, αλλά λιγότερο αποδοτική για κάποια άλλη λειτουργία, όπως π.χ. την εισαγωγή. Οι παρατηρήσεις αυτές εξηγούν από τη μια μεριά την ύπαρξη διαφορετικών δομών και από την άλλη τη σπουδαιότητα της επιλογής της κατάλληλης δομής κάθε φορά.

Διάκριση των δομών δεδομένων

Κατ' αρχήν οι δομές δεδομένων διακρίνονται σε *στατικές* (static) και *δυναμικές* (dynamic). Οι στατικές δομές έχουν σταθερό μέγεθος και μπορούν να κατακρατήσουν συγκεκριμένο πλήθος στοιχείων. Αντίθετα οι δυναμικές δομές δεν έχουν σταθερό μέγεθος και το πλήθος των στοιχείων τους μπορεί να μεγαλώνει ή να μικραίνει, καθώς στη δομή εισάγονται νέα δεδομένα ή διαγράφονται άλλα.

Οι δομές δεδομένων διακρίνονται επίσης σε *γραμμικές* και *μη γραμμικές*. Στις γραμμικές δομές μπορεί να ορισθεί κάποια σχέση διάταξης για δύο οποιαδήποτε διαδοχικά στοιχεία τους. Αυτό σημαίνει ότι κάποιο στοιχείο θα είναι πρώτο και κάποιο τελευταίο. Οποιοδήποτε από τα υπόλοιπα θα έπεται από το προηγούμενό του και θα προηγείται από το επόμενο του. Η σχέση αυτή μπορεί να είναι φυσική και λογική ή μόνο λογική.

Στην πρώτη περίπτωση τα στοιχεία καταλαμβάνουν διαδοχικές θέσεις μνήμης. Για την εύρεση ενός στοιχείου χρησιμοποιείται η *υπολογιζόμενη τεχνική*, κατά την οποία με την εφαρμογή ενός μαθηματικού τύπου βρίσκεται η διεύθυνση του στοιχείου. Η δομή του πίνακα είναι το πιο χαρακτηριστικό παράδειγμα διαδοχικής καταχώρησης (βλ. κεφ. 3 έως 6).

Στη δεύτερη περίπτωση τα στοιχεία δεν καταλαμβάνουν υποχρεωτικά διαδοχικές θέσεις μνήμης. Εδώ για την εύρεση ενός στοιχείου χρησιμοποιείται η λεγόμενη *δεσμική τεχνική*. Με την τεχνική αυτή κάθε στοιχείο διαθέτει επί πλέον και ένα δείκτη (pointer), ο οποίος παραπέμπει στο επόμενο λογικά στοιχείο. Η συνδεσμική λίστα είναι παράδειγμα δομής που χρησιμοποιεί δείκτες (βλ. κεφ. 9).

Στις μη γραμμικές δομές δεν μπορεί να οριστεί μια σχέση διάταξης όπως η παραπάνω. Τέτοιες δομές είναι τα δένδρα και οι γράφοι. Στα δένδρα ένας κόμβος έχει έναν προηγούμενο αλλά πιθανόν πολλούς επόμενους (βλ. κεφ. 10). Στους γράφους κάθε κόμβος μπορεί να έχει πολλούς προηγούμενους και πολλούς επόμενους (βλ. κεφ. 11).

Τέλος διάκριση των δομών μπορεί να γίνει και ανάλογα με το είδος της χρησιμοποιούμενης μνήμης (κύρια μνήμη, ταινία, δίσκος). Με αυτή τη διάκριση δεν εννοούμε, ότι υπάρχουν άλλες δομές για την κύρια μνήμη και άλλες για τη βοηθητική, αλλά ότι κάποιες δομές είναι καταλληλότερες για την κύρια μνήμη, ενώ κάποιες άλλες βρίσκουν μεγαλύτερη χρήση στη βοηθητική. Ο λόγος γι' αυτό είναι ότι οι χρόνοι πρόσβασης στη βοηθητική μνήμη είναι σημαντικά μεγαλύτεροι από την κύρια. Ακόμη υπάρχουν και μερικές δομές, που είναι ακατάλληλες για κάποιο είδος βοηθητικής μνήμης, γιατί αυτή η μνήμη δεν επιτρέπει το συγκεκριμένο είδος πρόσβασης (π.χ. ταινία).

Οι δομές δεδομένων κύριας μνήμης παρουσιάζονται στα κεφάλαια 3 έως 11, ενώ αυτές για τη βοηθητική στα κεφάλαια 12 έως 21.