

Βασικές έννοιες πληροφορικής

Απόστολος Συρόπουλος

Ξάνθη

Νοέμβριος 2022

Χαρακτήρας Ένα γράμμα, ένα ψηφίο ή οποιοδήποτε άλλο σύμβολο το οποίο χρησιμοποιείται ως μέρος της οργάνωσης, του ελέγχου ή της αναπαράστασης δεδομένων.

Κωδικοποίηση Η διαδικασία μετατροπής πληροφοριών από μία μορφή σε μία άλλη. Η αντίστροφη διαδικασία ονομάζεται αποκωδικοποίηση.

Bit Το bit (μπιτ) (συμβολίζεται με B) είναι ένα δυαδικό ψηφίο, δηλαδή είτε το ψηφίο 0 είτε το ψηφίο 1. Η συντομογραφία bit προέρχεται από τη σύντμηση των λέξεων Binary digit. Η λέξη bit όμως στα αγγλικά σημαίνει και «μικροσκοπικό τμήμα» ή «κομματάκι».

Byte (μπάιτ) Μια ακολουθία από 8 δυαδικά ψηφία ικανή να αναπαραστήσει έναν χαρακτήρα αλφαριθμητικών δεδομένων. Η λέξη byte προέρχεται από την λέξη bite (δαγκωνιά). Τα πολλαπλάσια του byte δίνονται παρακάτω:

Σύμβολο	Όνομα	Τιμή	Σύμβολο	Όνομα	Τιμή
KiB	kibibyte	$1024^1 \text{ B} = 2^{10} \text{ B}$	KB	kilobyte	1000^1 B
MiB	mebibyte	$1024^2 \text{ B} = 2^{20} \text{ B}$	MB	megabyte	1000^2 B
GiB	gibibyte	$1024^3 \text{ B} = 2^{30} \text{ B}$	GB	gigabyte	1000^3 B
TiB	tebibyte	$1024^4 \text{ B} = 2^{40} \text{ B}$	TB	terabyte	1000^4 B
PiB	pebibyte	$1024^5 \text{ B} = 2^{50} \text{ B}$	PB	petabyte	1000^5 B
EiB	exbibyte	$1024^6 \text{ B} = 2^{60} \text{ B}$	EB	exabyte	1000^6 B
ZiB	zebibyte	$1024^7 \text{ B} = 2^{70} \text{ B}$	ZB	zettabyte	1000^7 B
YiB	yobibyte	$1024^8 \text{ B} = 2^{80} \text{ B}$	YB	yottabyte	1000^8 B
—	—	—	RB	ronnabyte	1000^9 B
—	—	—	QB	quettabyte	1000^{10} B

ASCII Ένας κώδικας για την ανταλλαγή πληροφοριών μεταξύ Η/Υ κατασκευασμένων από διαφορετικές εταιρείες ο οποίος καθορίστηκε από το ANSI (Αμερικανικό Εθνικό Ινστιτούτο Τυποποίησης) το 1969. Χονδρικά οι χαρακτήρες του ASCII είναι τα γράμματα της Αγγλικής γλώσσας (πεζά και κεφαλαία), τα ψηφία, τα σημεία στίξης της Αγγλικής και διάφορα σύμβολα όπως το «+», το «/», το κενό αλλά και διάφορα σύμβολα ελέγχου. Στο ASCII κάθε χαρακτήρας αναπαράσσεται με 1 byte.

Επεκτεταμένο ASCII Ένα σύνολο από κώδικες που επεκτείνουν το ASCII. Στο ASCII κάθε χαρακτήρας αναπαρίστανται με 7 bit (δυαδικά ψηφία) κωδικοποιώντας έτσι μόνο 128 μοναδικά σύμβολα. Ένα επεκτεταμένο ASCII χρησιμοποιεί 8 bit για την αναπαράσταση του κάθε χαρακτήρα επιτρέποντας έτσι την κωδικοποίηση μέχρι 256 μοναδικών συμβόλων.

Unicode Αποτελεί βιομηχανικό πρότυπο το οποίο με συνέπεια επιτρέπει την αναπαράσταση και διαχείριση κειμένου γραμμένο στα περισσότερα από τα γνωστά συστήματα γραφής.

UCS-2 και UTF-16 Η κωδικοποίηση UCS-2, η οποία είναι γνωστή και ως κωδικοποίηση ISO-10646-UCS-2, είναι ίσως η πιο φυσική κωδικοποίηση του Unicode έκδοση 2. Σ' αυτή κάθε χαρακτήρας αναπαριστάται με 2 byte και έτσι κάθε χαρακτήρας αντιστοιχεί σε έναν αριθμό από το 0 ως το 65535. Όλοι οι χαρακτήρες που δεν μπορούν να αναπαρασταθούν με την κωδικοποίηση UCS-2, αναπαριστώνονται με 4 byte στην κωδικοποίηση UTF-16.

UCS-4 και UTF-32 Στην κωδικοποίηση UCS-4 κάθε χαρακτήρας αναπαριστάται με 4 byte τα οποία είναι αρκετά για την κωδικοποίηση μέχρι 1.048.575 χαρακτήρων. Το ίδιο συμβαίνει και με την κωδικοποίηση UTF-32, δηλαδή κάθε χαρακτήρας αναπαριστάται με 4 byte. Οι δύο κωδικοποιήσεις έχουν διαφορές που δεν θα περιγραφούν εδώ.

UTF-8 Η πιο κοινή κωδικοποίηση του Unicode η οποία είναι συμβατή με την κωδικοποίηση ASCII. Στο UTF-8 αναπαριστώνονται οι χαρακτήρες του ASCII με 1 byte· οι χαρακτήρες που χρησιμοποιούνται στην Ιαπωνική, την Κινεζική και την Κορεατική γραφή με 3 byte· όλοι οι άλλοι χαρακτήρες που αντιστοιχούν σε σύμβολα όλων των άλλων σύγχρονων γραφών (π.χ. ελληνικής, κυριλλικής, ταϊλανδέζικης, βιετναμέζικης, αραβικής, εβραϊκή κ.λπ.) με 2 byte και, τέλος, με 4 byte όλα τα υπόλοιπα σύμβολα (π.χ. βυζαντινά μουσικά σύμβολα, δυτικά μουσικά σύμβολα, αιγυπτιακά ιερογλυφικά, τα σύμβολα Γραμμικής Β, τα σύμβολα του δίσκου της Φαιστού κ.λπ.).