

Ιόνιο Πανεπιστήμιο – Τμήμα Πληροφορικής  
Εισαγωγή στην Επιστήμη των Υπολογιστών  
2022-23

# Αναπαράσταση Μη Αριθμητικών Δεδομένων

(κείμενο, ήχος και εικόνα στον υπολογιστή)

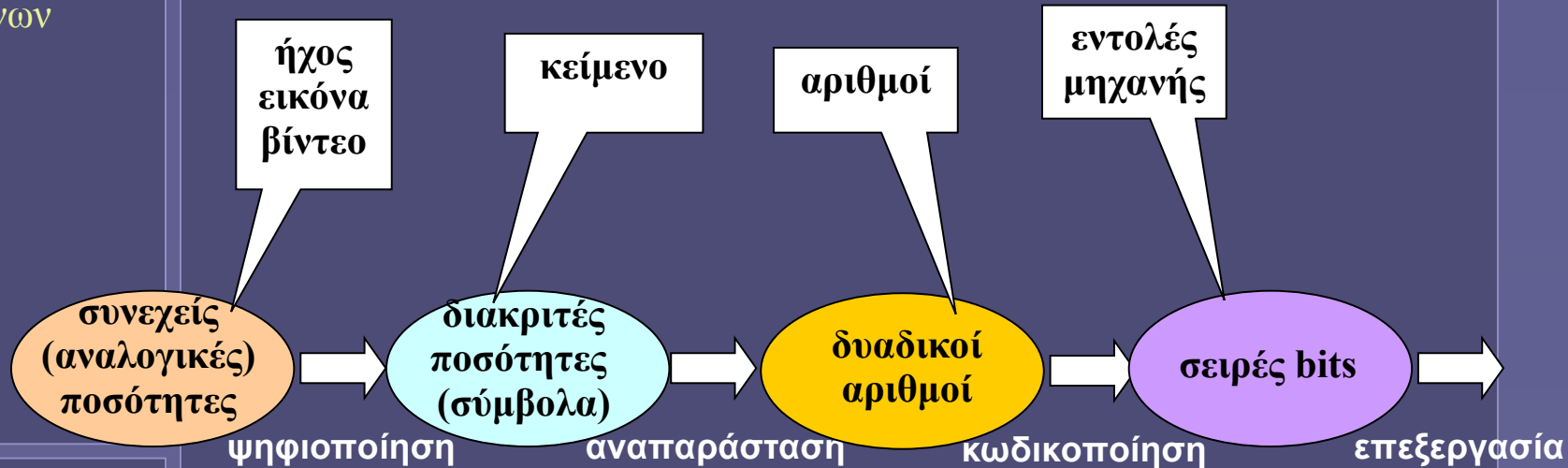
<http://mixstef.github.io/courses/csintro/>

Μ.Στεφανιδάκης



# Αναπαράσταση δεδομένων

- Αναπαράσταση δεδομένων



!

Δεδομένα:  
ανεξάρτητα από  
τύπο και  
προέλευση, στον  
υπολογιστή  
υπάρχουν σε μία  
μορφή: 0 και 1

- Ψηφιοποίηση
  - Διαδικασία μετατροπής συνεχών τιμών σε διακριτά σύμβολα
- Αναπαράσταση
  - Διαδικασία αντιστοίχισης συμβόλων σε δυαδικούς αριθμούς
- Κωδικοποίηση
  - Αποθήκευση δυαδικών αριθμών σε σειρές bits

# Η ερμηνεία της αναπαράστασης

- Αναπαράσταση δεδομένων

!

Στα ερωτήματα αυτά μπορεί να απαντήσει μόνο ο προγραμματιστής της εφαρμογής που χειρίζεται τα δεδομένα!

- Κάπου στη μνήμη του υπολογιστή...
  - Βρίσκεται αποθηκευμένη η σειρά bits  
**0100110111010001**
- Πόσα σύμβολα αναπαριστά;
  - Πόσα bits ανά σύμβολο;
- Ποιος ο τύπος των δεδομένων;
- Ποια συγκεκριμένη ποσότητα συμβολίζει;
- Πώς θα το χειριστεί ο υπολογιστής;

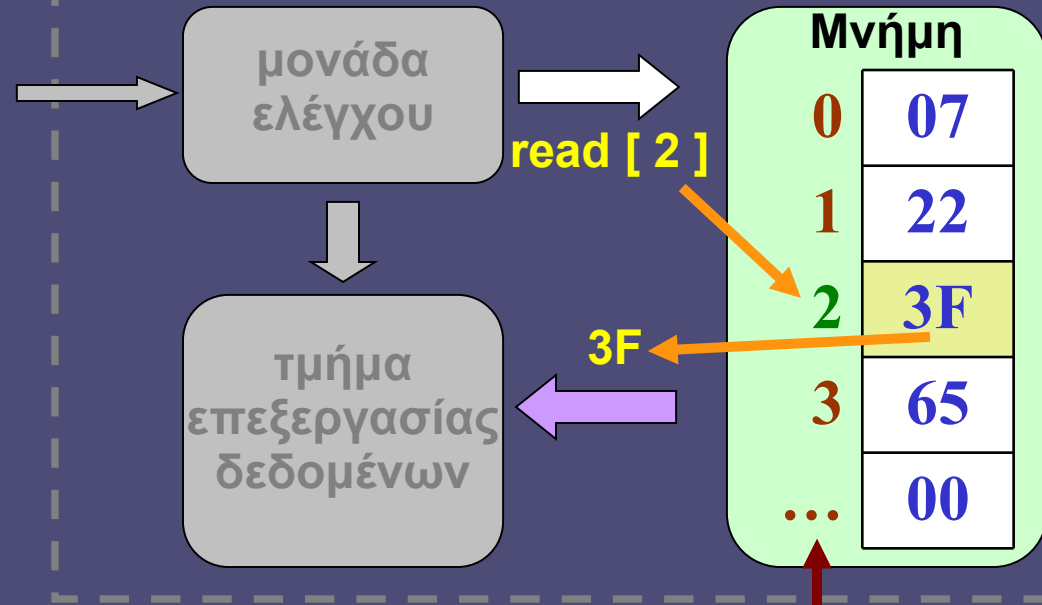
# Αναπαράσταση με δυαδικούς αριθμούς

- Αναπαράσταση δεδομένων

- Σειρά από  $n$  bits
  - Δυαδικός αριθμός με  $n$  bits ( $n \geq 1$ ) μπορεί να αναπαραστήσει  $2^n$  διαφορετικά σύμβολα
- Μη αριθμητικά δεδομένα
  - Κείμενο, εντολές μηχανής, ήχος, εικόνα...
    - Σύνολο διαφορετικών αντικειμένων (συμβόλων)
  - Αντιστοίχιση κάθε συμβόλου σε μοναδικό δυαδικό αριθμό
    - “Αναπαράσταση”
    - Η ακριβής αντιστοίχιση ορίζεται σε ένα πρότυπο (standard)

# Το απλουστευμένο μοντέλο μνήμης

- Αναπαράσταση δεδομένων



Διευθύνσεις μνήμης

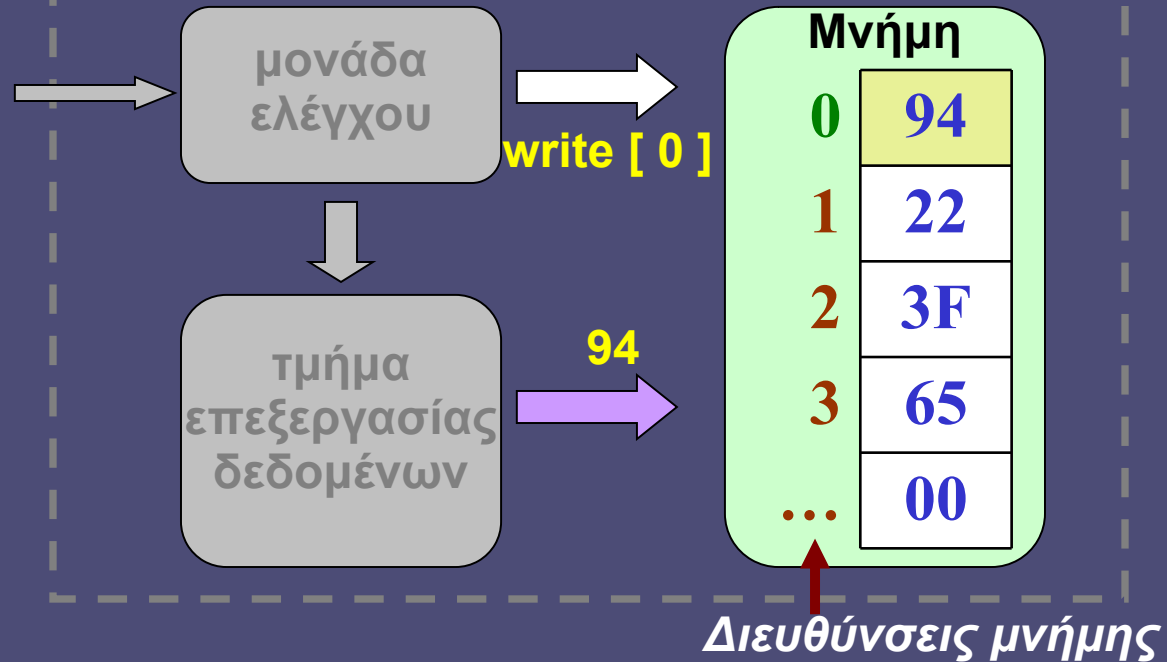
- Πώς βλέπει ένα πρόγραμμα τη μνήμη
  - Συστοιχία αποθηκευτικών θέσεων
  - Σε κάθε θέση αποθηκεύεται (συνήθως) 1 byte
  - Κάθε θέση διαθέτει μοναδική διεύθυνση
    - Επιλογή θέσης κατά την προσπέλαση (ανάγνωση-εγγραφή)

•  
;

Με διεύθυνση των  $n$  bits, πόσες διαφορετικές θέσεις μνήμης μπορούμε να προσπελάσουμε;

# Το απλουστευμένο μοντέλο μνήμης

- Αναπαράσταση δεδομένων



- Στην πραγματικότητα
  - Η «μνήμη» είναι μια σύνθετη ιεραρχία **πολλών επιπέδων**
  - Οι μεταφορές δεδομένων δεν γίνονται σε μεμονωμένα bytes αλλά σε **πολλαπλάσιά τους**

# Θέματα αποθήκευσης δυαδικών αριθμών

- Αναπαράσταση δεδομένων

;

Πώς σχετίζεται η σειρά αποθήκευσης των bytes με τα «Ταξίδια του Γκιούλιβερ»;

- Όταν για έναν δυαδικό αριθμό χρειάζονται **περισσότερα από ένα bytes** για να αποθηκευτούν τα ψηφία του

- Παράδειγμα: 3FC (hex) = 11 1111 1100  
**Χρειάζονται 2 bytes!**

0000 0011 1111 1100  
**περισσότερο σημαντικό byte**      **λιγότερο σημαντικό byte**

- Προφανώς σε συνεχόμενες θέσεις μνήμης  
**Αλλά: ποιο byte αποθηκεύεται πρώτο;**

# Θέματα αποθήκευσης δυαδικών αριθμών

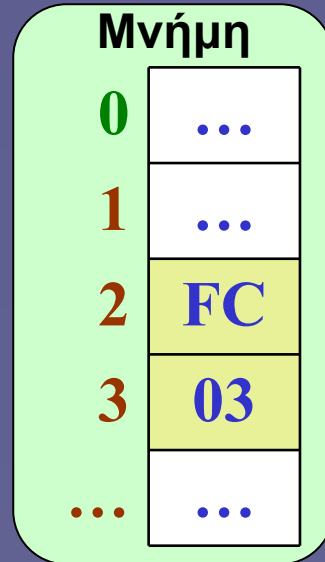
- Αναπαράσταση δεδομένων

αποθηκεύοντας το  
03FC

00000011 11111100

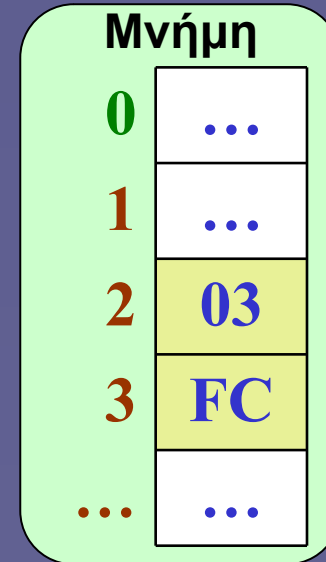
**i**

Το ίδιο ισχύει και  
για αριθμούς με  
περισσότερα από 2  
bytes



*“little-endian”*

Το λιγότερο  
σημαντικό byte στη  
θέση μνήμης με  
μικρότερη διεύθυνση



*“big-endian”*

Το περισσότερο  
σημαντικό byte στη  
θέση μνήμης με  
μικρότερη διεύθυνση



# Αρχικές αναπαραστάσεις κειμένου

- Αναπαράσταση δεδομένων
- Κείμενο

- **Οι πρώτες αναπαραστάσεις κειμένου**
  - Στον υπολογιστή
  - 6-7 bits ανά χαρακτήρα
    - Πόσοι διαφορετικοί χαρακτήρες;
- **Μη εκτυπώσιμοι χαρακτήρες**
  - Χαρακτήρες ελέγχου
    - Ιδιαίτερα χρήσιμοι για τις συσκευές εξόδου της εποχής (εκτυπωτές, τηλέτυπα...)
    - Νέα γραμμή (LINE FEED – LF)
    - Επιστροφή κεφαλής εκτύπωσης (CARRIAGE RETURN – CR)
    - Καμπανάκι (BELL) κλπ

# Κώδικας ASCII

- Αναπαράσταση δεδομένων
- Κείμενο

- Βασικό αρχικό πρότυπο αναπαράστασης κειμένου
  - 7 bits ανά χαρακτήρα

STANDARD ASCII ΚΩΔΙΚΑΣ

hex	char	hex	char	hex	char
20		40	@	60	`
21	!	41	A	61	a
22	"	42	B	62	b
23	#	43	C	63	c
24	\$	44	D	64	d
25	%	45	E	65	e
26	&	46	F	66	f
27	'	47	G	67	g
28	(	48	H	68	h
29	)	49	I	69	i
2A	*	4A	J	6A	j
2B	+	4B	K	6B	k
2C	,	4C	L	6C	l
2D	-	4D	M	6D	m
2E	.	4E	N	6E	n
2F	/				

Δεν φαίνεται όλος ο πίνακας

•  
1

ASCII: American  
Standard Code for  
Information  
Interchange

# Κείμενο σε κώδικα ASCII

- Αναπαράσταση δεδομένων
- Κείμενο

;

Με 7 bits ανά χαρακτήρα και χρήση bytes, 1 bit μένει αχρησιμοποίητο. Πόσοι επιπλέον χαρακτήρες αν χρησιμοποιηθεί και το bit αυτό;

- 7 bits ανά χαρακτήρα
  - 128 χαρακτήρες
  - Αναπαράσταση με τους αριθμούς 0...127
- Κανονικοί χαρακτήρες (εκτυπώσιμοι)
  - 32...47, 58...64, 91...96, 123...126 = σημεία στίξης κ.ά. (32 = SPACE)
  - 48...57 = ψηφία 0...9
  - 65...90 = κεφαλαία λατινικά (A-Z)
  - 97...122 = πεζά λατινικά (a-z)
- Χαρακτήρες ελέγχου (μη εκτυπώσιμοι)
  - 0...31, 127 – πιο γνωστά: 9 (TAB), 13/10 (CR/LF, σήμανση “νέας γραμμής”)

# Κείμενο σε κώδικα ASCII

- Αναπαράσταση δεδομένων
- Κείμενο

!

Εφόσον η κωδικοποίηση είναι με 1 byte ανά χαρακτήρα, δεν τίθεται θέμα “little-” ή “big-endian”

- Παράδειγμα

<b>H</b>	<b>a</b>	<b>v</b>	<b>e</b>		<b>a</b>		<b>n</b>	<b>i</b>	<b>c</b>	<b>e</b>		<b>d</b>	<b>a</b>	<b>y</b>	<b>!</b>
72	97	118	101	32	97	32	110	105	99	101	32	100	97	121	33

- Στις γλώσσες προγραμματισμού
  - “string” (συμβολοσειρά)
  - Σε γλώσσες όπως η C, το 0 (αριθμητικό) συμβολίζει το τέλος του string
  - Ο υπολογιστής μπορεί να κάνει πράξεις (π.χ. σύγκριση) με τα strings

# Επεκτάσεις κώδικα ASCII

- Αναπαράσταση δεδομένων
- Κείμενο



Χρησιμοποιώντας τον ISO-8859-1 δεν είναι δυνατή η αναπαράσταση των ελληνικών!

- Χρήση του 1 επιπλέον bit του byte
  - 128 + 128 χαρακτήρες, αριθμοί 0...255
  - 0...127 αντιστοιχούν στον αρχικό ASCII
  - 127...255: επεκταμένα αλφάβητα
- Επέκταση αλφαβήτων (πρότυπα)
  - Χαρακτήρες που δεν υπάρχουν στον ASCII
  - Διαφορετικά ανά γλώσσα! Π.χ.:
    - ISO-8859-1: Δυτική Ευρώπη (Å, Ñ, Æ, ä, ø κλπ)
    - ISO-8859-7: Νέα Ελληνικά
    - ...και πολλά άλλα πρότυπα για τις υπόλοιπες γλώσσες
  - Επίσης: μη πρότυπες λύσεις
    - Για Windows, Mac ..

# Κώδικας ISO-8859-7

- Αναπαράσταση δεδομένων
- Κείμενο

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	<i>unused</i>															
1x																
2x	SP	!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3x	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4x	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5x	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6x	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7x	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8x	<i>unused</i>															
9x																
Ax	NBSP	·	´	£	€	Ɔp		§	¨	©	,	*	¬	SHY		—
Bx	°	±	²	³	´	ˆ	À	Á	Â	Ã	Ä	Å	Ö	½	Υ	Ω
Cx	ΐ	Α	Β	Γ	Δ	Ε	Ζ	Η	Θ	Ι	Κ	Λ	Μ	Ν	Ξ	Ο
Dx	Π	Ρ		Σ	Τ	Υ	Φ	Χ	Ψ	Ω	Ϊ	Ϋ	ά	έ	ή	ί
Ex	ύ	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο
Fx	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω	ϊ	ϋ	ό	ύ	ώ	

[Wikipedia]

# Κείμενο σε κώδικα ISO-8859-7

- Αναπαράσταση δεδομένων
- Κείμενο

!

Οι αναπαραστάσεις αλφαβήτων με 1 byte ανά χαρακτήρα έχουν (σχεδόν) καταργηθεί

- Παράδειγμα

Γ	ε	ι	α		σ	ο	υ	!
195	229	233	225	32	243	239	245	33

- Επέκταση κώδικα ASCII
  - 0...127 όπως στον ASCII
  - 128...159 πρόσθετοι χαρακτήρες ελέγχου
  - 160...255 ελληνικά και σχετικά σύμβολα

# Πρότυπο Unicode

- Αναπαράσταση δεδομένων
- Κείμενο

!

Με περισσότερα από 1 bytes ανά χαρακτήρα τίθεται θέμα **σειράς αποθήκευσης** των bytes!

- Για την αναπαράσταση όλων των αλφαβήτων
  - Καλύπτει ιδεογράμματα, φωνητικές αναπαραστάσεις και διάφορα σύμβολα (~100.000 χαρακτήρες έχουν οριστεί)
  - Θεωρητικά μπορεί να καλύψει πάνω από 1 εκ. χαρακτήρες
- Κάθε χαρακτήρας αναπαρίσταται με έναν δυαδικό αριθμό (codepoint)
  - 0 έως 10FFFF
  - Χρειάζονται **περισσότερα από ένα bytes** για την αποθήκευση ενός τέτοιου αριθμού



# Πρότυπο Unicode

- Αναπαράσταση δεδομένων
- Κείμενο

- Το πρότυπο Unicode περιέχει επίσης
  - Πληροφορία ισοδύναμων ή παρόμοιων χαρακτήρων
  - Συνδυασμούς τόνων/διακριτικών και γραμμάτων
  - Οδηγίες για την ταξινόμηση των γραμμάτων ανά γλώσσα

# Ελληνικά και Unicode

- Αναπαράσταση δεδομένων
- Κείμενο

Greek and Coptic									03FF
	037	038	039	03A	03B	03C	03D	03E	03F
0			ι̇ 0390	Π 03A0	Ϝ 03B0	π 03C0	β 03D0	Ϡ 03E0	Ϻ 03F0
1			Α 0391	Ρ 03A1	α 03B1	ρ 03C1	θ 03D1	ϡ 03E1	ϙ 03F1
2			Β 0392		β 03B2	ς 03C2	Υ 03D2	Ϸ 03E2	Ϙ 03F2
3			Γ 0393	Σ 03A3	γ 03B3	σ 03C3	Ύ 03D3	ϸ 03E3	ι 03F3
4	΄ 0374	΄ 0384	Δ 0394	Τ 03A4	δ 03B4	τ 03C4	Ύ̇ 03D4	Ϣ 03E4	Θ 03F4
5	΄ 0375	Ⲁ 0385	Ε 0395	Υ 03A5	ε 03B5	υ 03C5	φ 03D5	ϣ 03E5	€ 03F5
6		Ά 0386	Ζ 0396	Φ 03A6	ζ 03B6	φ 03C6	ω 03D6	ħ 03E6	ε 03F6

δεν φαίνεται όλος ο πίνακας

# Κείμενο σε Unicode

- Αναπαράσταση δεδομένων
- Κείμενο

## Παράδειγμα

δεκαδικό →  
δεκαεξαδικό →

Γ	ε	ι	α		σ	ο	υ	!
915	949	953	945	32	963	959	965	33
0393	03B5	03B9	03B1	0020	03C3	03BF	03C5	0021

Κωδικοποίηση big-endian

03	93	03	B5	03	B9	03	B1	00	20	03	C3	03	BF	03	C5	00	21
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Κωδικοποίηση little-endian

93	03	B5	03	B9	03	B1	03	20	00	C3	03	BF	03	C5	03	21	00
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

# Unicode σε κωδικοποίηση UTF-8

- Αναπαράσταση δεδομένων
- Κείμενο



Η κωδικοποίηση UTF-8 έχει επικρατήσει σε όλα τα προγράμματα που χειρίζονται κείμενα Unicode

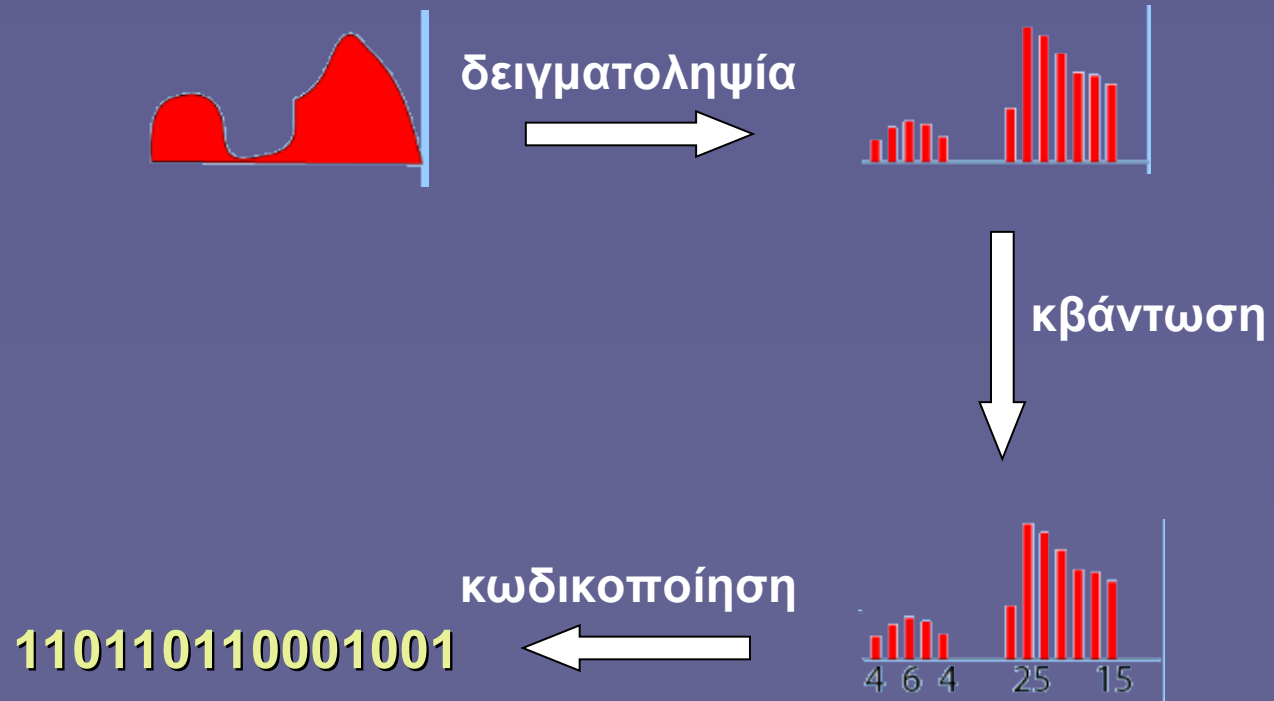
- Αναπαράσταση μεταβλητού μήκους

Unicode	Κωδικοποίηση UTF-8
00...7F	0xxxxxxx
80...7FF	110xxxxx 10xxxxxx
800...FFFF	1110xxxx 10xxxxxx 10xxxxxx
10000...10FFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

- Το βασικό λατινικό αλφάβητο (ASCII) χρησιμοποιεί 1 byte ανά χαρακτήρα
  - Προς τα πίσω συμβατότητα
- Τα ελληνικά, 2 bytes
- Αλφάβητα Άπω Ανατολής, 3+ bytes

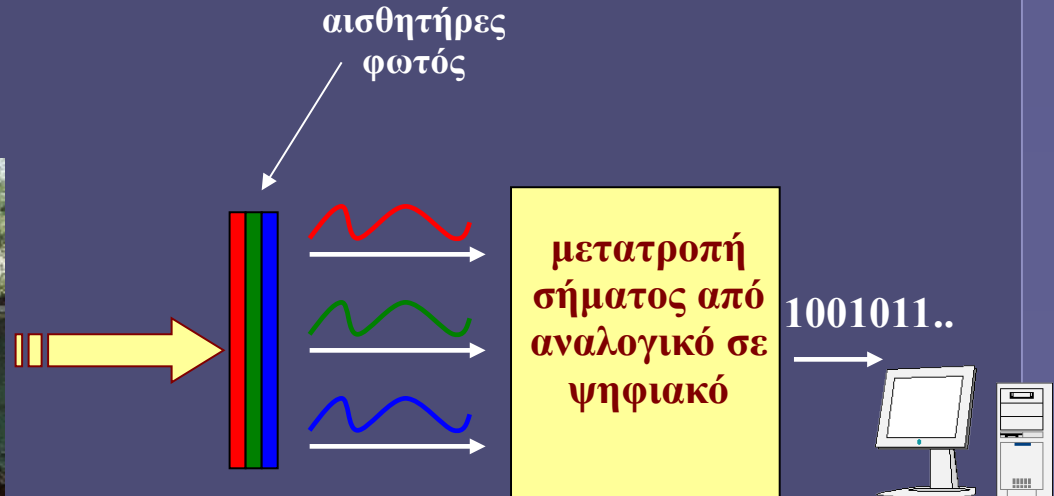
# Ήχος: Ψηφιοποίηση και Αποθήκευση

- Αναπαράσταση δεδομένων
- Κείμενο
- Ήχος



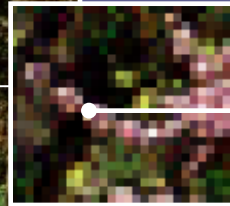
# Εικόνα: από τον αναλογικό στον ψηφιακό κόσμο

- Αναπαράσταση δεδομένων
- Κείμενο
- Ήχος
- Εικόνα



- Φωτοευαίσθητα κύτταρα
  - για τρία χρώματα (κόκκινο-πράσινο-μπλε)
- Μετατροπή σήματος σε ψηφιακή πληροφορία

# Παράδειγμα: απλή αναπαράσταση pixels με 16,7 εκ. χρώματα



1 pixel

<b>R:144</b> <b>G:128</b> <b>B:118</b>	<b>R:193</b> <b>G:164</b> <b>B:179</b>	...
<b>R:201</b> <b>G:174</b> <b>B:134</b>	...	...
...	...	...

- 3 bytes/pixel (24bits): **R**(ed) **G**(reen) **B**(lue)
  - 256 στάθμες ανά συνιστώσα χρώματος
    - $256 \times 256 \times 256 = 16.777.216$  χρώματα
  - εικόνες με μεγαλύτερο βάθος χρώματος
    - 32 έως 48 bits

# Εναλλακτικά: διανυσματικά γραφικά

- Αναπαράσταση δεδομένων
- Κείμενο
- Ήχος
- Εικόνα

- Περιγραφή σχημάτων
  - Ως σύνολο ευθύγραμμων και καμπύλων τμημάτων
  - Με συντεταγμένες
  - Εύρεση σημείων μέσω μαθηματικού τύπου
- Εύκολη αλλαγή μεγέθους γραφικών
  - Χωρίς παραμόρφωση των σχημάτων



# Αναπαράσταση βίντεο

- Αναπαράσταση δεδομένων
- Κείμενο
- Ήχος
- Εικόνα
- Βίντεο

- “Κινούμενη εικόνα” (καρέ)
  - όπως αναπαριστούμε τις απλές εικόνες
  - αλλά: με χρήση συμπίεσης
    - Για μείωση όγκου δεδομένων
    - Γειτονικά καρέ έχουν πολλές ομοιότητες

# Κωδικοποίηση εντολών μηχανής

- Αναπαράσταση δεδομένων
- Κείμενο
- Ήχος
- Εικόνα
- Βίντεο
- Εντολές Μηχανής

