

## Αναπαράσταση Μη Αριθμητικών Δεδομένων

(κείμενο, ήχος και εικόνα στον υπολογιστή)

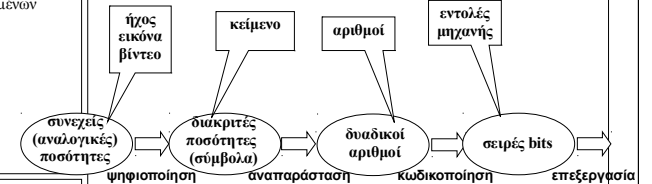
<http://mixstef.github.io/courses/csintro/>



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!

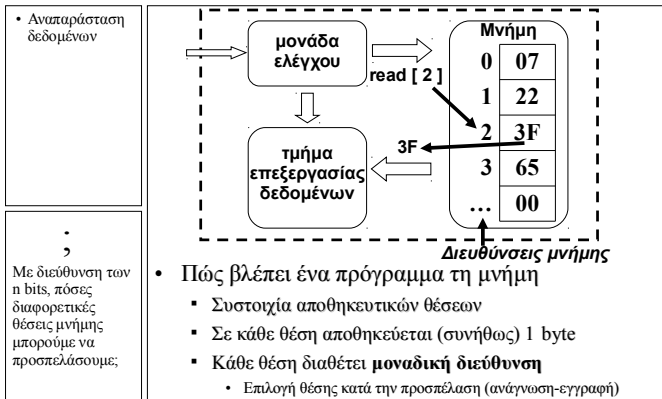
- Κάπου στη μνήμη του υπολογιστή...
  - Βρίσκεται αποθηκευμένη η σειρά bits 0100110111010001
- Πόσα σύμβολα αναπαριστά;
  - Πόσα bits ανά σύμβολο;
- Ποιος ο τύπος των δεδομένων;
- Ποια συγκεκριμένη ποσότητα συμβολίζει;
- Πώς θα το χειριστεί ο υπολογιστής;

## Αναπαράσταση με δυαδικούς αριθμούς

- Αναπαράσταση δεδομένων

- Σειρά από  $n$  bits
  - Δυναμικός 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)

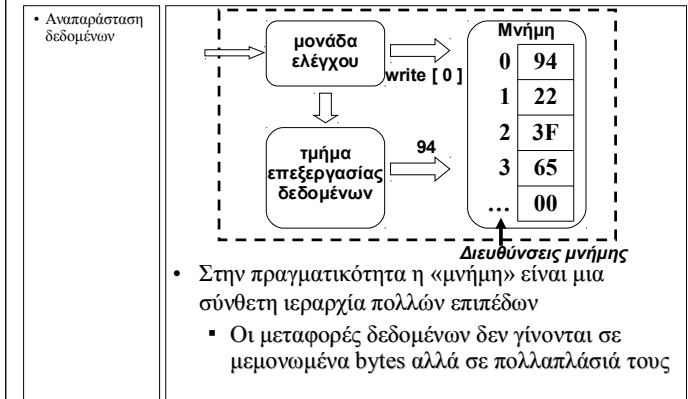
## Το απλουστευμένο μοντέλο μνήμης



Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

5

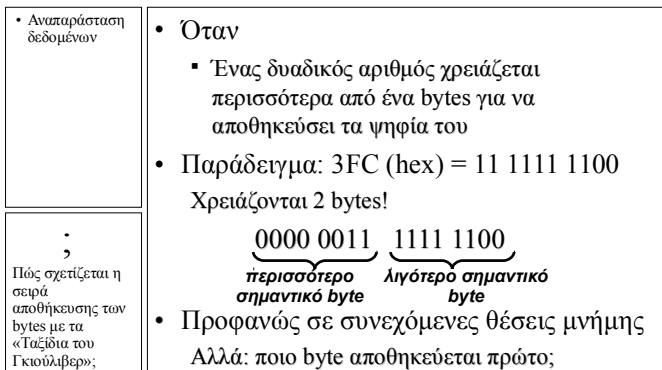
## Το απλουστευμένο μοντέλο μνήμης



Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

6

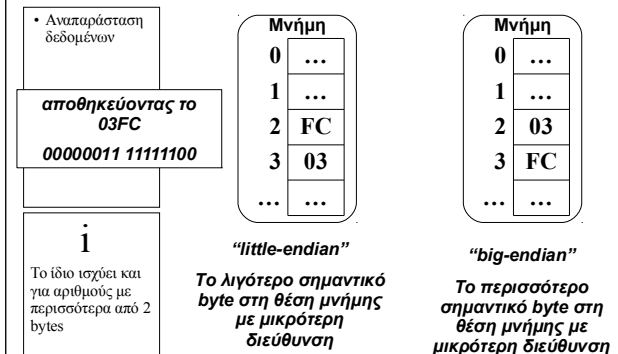
## Θέματα αποθήκευσης δυαδικών αριθμών



Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

7

## Θέματα αποθήκευσης δυαδικών αριθμών



Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

8

## Αρχικές αναπαράστασεις κειμένου

- Αναπαράσταση δεδομένων
- Κείμενο

- Οι πρώτες αναπαράστασεις 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, τηλετυπα...)
    - Νέα γραμμή (LINE FEED – LF)
    - Επιστροφή κεφαλής εκτύπωσης (CARRIAGE RETURN – CR)
    - Καμπανάκι (BELL) κλπ

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

9

## 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	/	4F	O	6F	o

**i**  
ASCII: American  
Standard Code for  
Information  
Interchange

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

10

## Κείμενο σε κώδικα ASCII

- Αναπαράσταση δεδομένων
- Κείμενο

- 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”)

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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

11

## Κείμενο σε κώδικα ASCII

- Αναπαράσταση δεδομένων
- Κείμενο

- Παράδειγμα

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

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

- 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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

12

## 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 ..

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

13

## Κώδικας ISO-8859-7

- Αναπαράσταση δεδομένων
- Κείμενο

	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	xA	xB	xC	xD	xE	xF
0x	unused															
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	unused															
9x																
Ax	NBSP	·	´	ε	€	Δ	!	§	¨	©	.	+	~	SHY		—
Bx	·	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¸
Cx	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î
Dx	Ï	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ
Ex	ß	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î
Fx	ï	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ

[Wikipedia]

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

14

## 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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

15

## 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 (~100.000 χαρακτήρες)
  - Θα μπορούσε να καλύψει πάνω από 1 εκ. χαρακτήρες!
  - Κάθε χαρακτήρας αναπαρίσταται με περισσότερα από ένα bytes
- Το πρότυπο Unicode περιέχει επίσης
  - Πληροφορία ισοδύναμων ή παρόμοιων χαρακτήρων
  - Οδηγίες συνδυασμών τόνων/διακριτικών και γραμμάτων

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

16

## Ελληνικά και Unicode

- Αναπαράσταση δεδομένων
- Κείμενο

Greek and Coptic

03FF

	037	038	039	03A	03B	03C	03D	03E	03F
0			ι	Π	ϖ	π	β	η	ζ
1			Α	Ρ	α	ρ	θ	λ	Ϛ
2			Β		β	ς	Υ	Ϙ	ϙ
3			Γ	Σ	γ	σ	Υ	Ϙ	ϙ
4			Δ	Τ	δ	τ	Ϊ	Ϛ	Θ
5			Ε	Υ	ε	υ	φ	ϙ	Ε
6			Α	Ζ	Φ	Ζ	Φ	Ψ	Ξ

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

17

## Κείμενο σε Unicode

- Αναπαράσταση δεδομένων
- Κείμενο

### Παράδειγμα

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

δεκαεξαδικό

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
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

18

## Unicode σε κωδικοποίηση UTF-8

- Αναπαράσταση δεδομένων
- Κείμενο

### 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

!

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

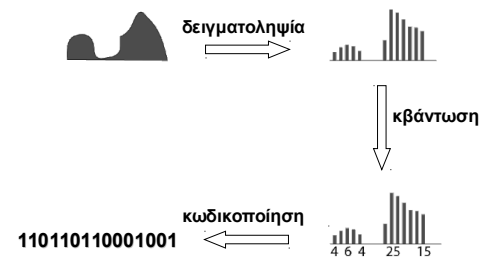
- Το βασικό λατινικό αλφάβητο (ASCII) χρησιμοποιεί 1 byte ανά χαρακτήρα
  - Προς τα πίσω συμβατότητα
- Τα ελληνικά, 2 bytes
- Αλφάβητα Άπω Ανατολής, 3+ bytes

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

19

## Ήχος: Ψηφιοποίηση και Αποθήκευση

- Αναπαράσταση δεδομένων
- Κείμενο
- Ήχος

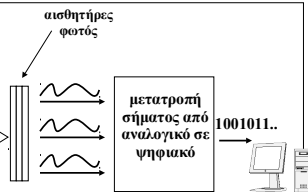


Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

20

## 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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

21

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



R:144 G:128 B:118	R:193 G:164 B:179	...
R:201 G:174 B:134	...	...
...	...	...

- 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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

22

## 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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

23

## 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

Εισαγωγή στην Επιστήμη των Υπολογιστών – “Αναπαράσταση Μη Αριθμητικών Δεδομένων”

24

## Κωδικοποίηση εντολών μηχανής

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

opcode	operand1	operand2	..	operandN
--------	----------	----------	----	----------

Περιγράφει το είδος της πράξης που θα εκτελεστεί

Περιγράφουν την προέλευση των δεδομένων εισόδου (αριθμό καταχωρητή, διεύθυνση μνήμης κλπ) και τον προορισμό των δεδομένων εξόδου (αποτέλεσμα πράξης)

*Το είδος της πράξης προσδιορίζει τον τύπο, την προέλευση και τον αριθμό των δεδομένων που συμμετέχουν στην πράξη !*