



Κεφ. 3: **Αναπαράσταση** **αλφαριθμητικών** **χαρακτήρων και** **αριθμών**

3.1 Κωδικοποίηση χαρακτήρων

Οι λέξεις και οι φράσεις σε ένα κείμενο, δημιουργούνται από χαρακτήρες.

Οι χαρακτήρες ομαδοποιούνται σε σύνολα χαρακτήρων (character sets ή repertoires). Ένα σύνολο χαρακτήρων μπορεί να περιλαμβάνει αλφαβητικούς χαρακτήρες, αριθμούς και άλλα σύμβολα.

Ένα σύνολο χαρακτήρων λέγεται κωδικοποιημένο σύνολο χαρακτήρων όταν κάθε χαρακτήρας αντιστοιχίζεται σε ένα συγκεκριμένο αριθμό που καλείται code point. Η τιμή του code point, αντιπροσωπεύει τη θέση του χαρακτήρα στο κωδικοποιημένο σύνολο χαρακτήρων



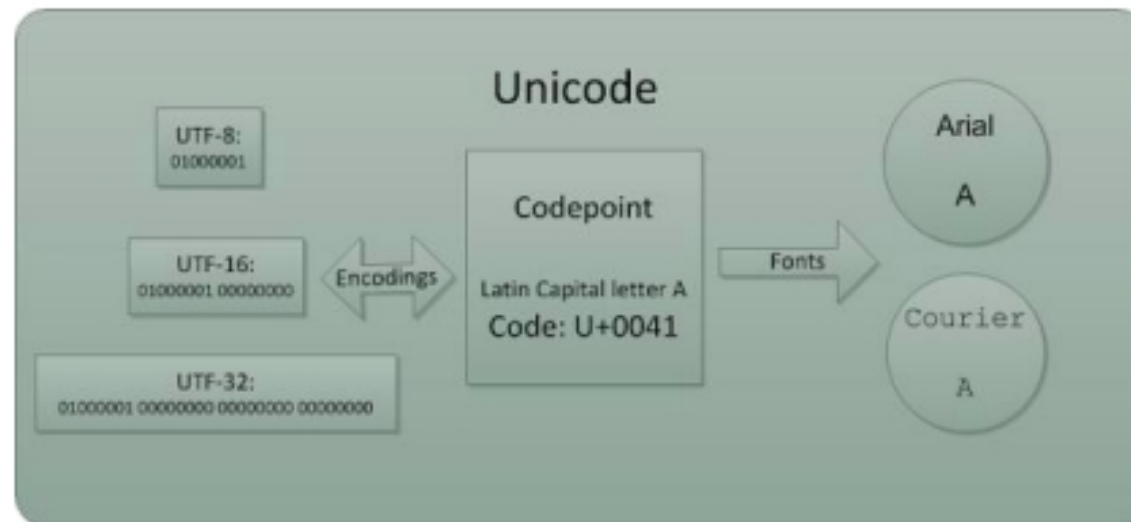
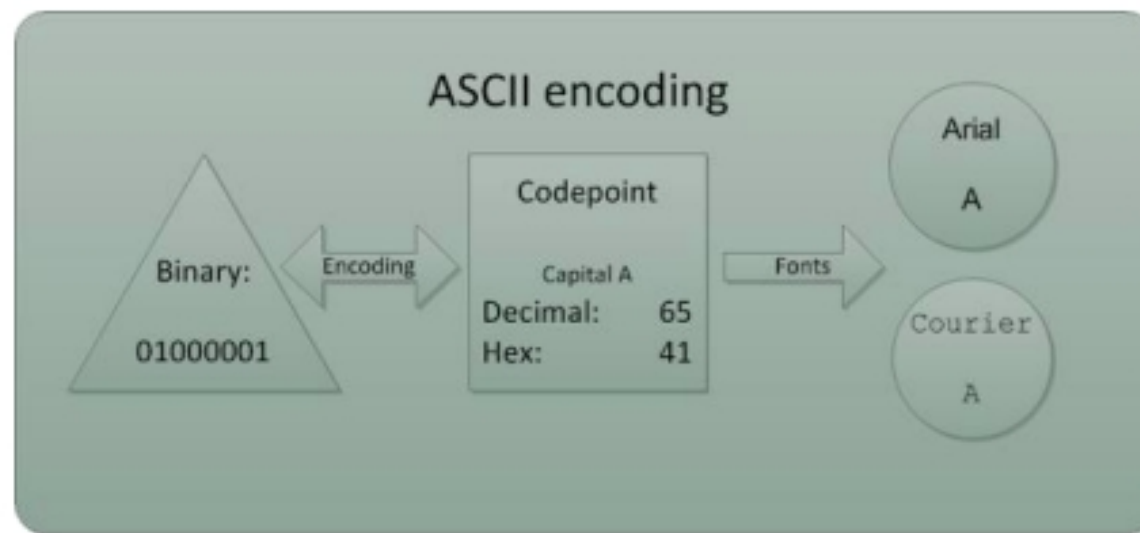
Για παράδειγμα,

- στο κωδικοποιημένο σύνολο χαρακτήρων ASCII, ο χαρακτήρας που αναπαριστά το κεφαλαίο γράμμα "Α" στο λατινικό αλφάβητο, μπορεί να απεικονιστεί στον αριθμό 65 (δεκαδικό σύστημα αρίθμησης) ή 41 (δεκαεξαδικό σύστημα αρίθμησης).
- στο κωδικοποιημένο σύνολο χαρακτήρων Unicode, το code point για το χαρακτήρα ά είναι 225 (δεκαδικό σύστημα αρίθμησης), ή E1 (δεκαεξαδικό σύστημα αρίθμησης).



Τα κωδικοποιημένα σύνολα χαρακτήρων, μερικές φορές αναφέρονται ως **κωδικοσελίδες** (code pages).

Η **κωδικοποίηση χαρακτήρων** (character encoding) είναι το κλειδί που μετατρέπει τα code points σε δυαδική μορφή για να αποθηκευτούν σε έναν υπολογιστή ή να μεταδοθούν. Επομένως τα μέσα και οι συσκευές εισόδου, εξόδου, αποθήκευσης και επικοινωνίας μεταφέρουν, αποθηκεύουν και χειρίζονται δεδομένα κειμένου υπό τη μορφή ενός κώδικα χαρακτήρων που απεικονίζει κάθε χαρακτήρα σε έναν μοναδικό αριθμό (code point). Η κωδικοποίηση μετατρέπει αυτούς τους αριθμούς σε δυαδική μορφή, ώστε κάθε χαρακτήρας να αντιστοιχίζεται σε μία μοναδική διαφορετική ακολουθία δυαδικών ψηφίων (σειρές από bytes).



Οι πιο διαδεδομένοι κώδικες για την αναπαράσταση χαρακτήρων είναι οι **κώδικες ASCII και UNICODE**.



3.1.1 Κώδικας χαρακτήρων ASCII

Ο κώδικας **ASCII** (American Standard Code for Information Interchange, Αμερικανικός Πρότυπος Κώδικας για Ανταλλαγή Πληροφοριών) είναι ένα κωδικοποιημένο σύνολο χαρακτήρων του λατινικού αλφάβητου. Αναπτύχθηκε στη δεκαετία του 1960 από τον Αμερικάνικο Οργανισμό Προτυποποίησης, ώστε να επιτρέπει σε ψηφιακές συσκευές να επικοινωνούν μεταξύ τους και να επεξεργάζονται, να αποθηκεύουν και να μεταδίδουν πληροφορίες σχετικές με χαρακτήρες



Αρχικά στον κώδικα ASCII χρησιμοποιούνταν 7 bit για την κωδικοποίηση των χαρακτήρων και το 8ο bit για έλεγχο ορθότητας κατά τη μεταφορά στοιχείων. Το bit αυτό ονομάστηκε ψηφίο ισοτιμίας (parity bit). Το αποτέλεσμα ήταν να επιτρέπει την απεικόνιση 128 (2^7) διαφορετικών χαρακτήρων. Σε κάθε χαρακτήρα αντιστοιχίζεται ένας μοναδικός αριθμός από το 0 έως το 127.

0 - 31	χαρακτήρες ελέγχου (μη εκτυπώσιμοι)
32 - 63	αριθμοί, κενό, σημεία στίξης, σύμβολα πράξεων
64 - 95	κεφαλαία λατινικά γράμματα και ειδικά σύμβολα
96 - 127	πεζά λατινικά γράμματα και ειδικά σύμβολα

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

	H	E	L	L	O
ASCII (δεκαδικός)	72	69	76	76	79
Δυαδική απεικόνιση	1001000	1000101	1001100	1001100	1001111



Σημαντικό μειονέκτημα του κώδικα ASCII είναι ότι δεν είναι δυνατή η αναπαράσταση χαρακτήρων άλλου αλφαβήτου εκτός από το λατινικό. Γι' αυτό το λόγο χρησιμοποιήθηκε και το όγδοο bit για την κωδικοποίηση χαρακτήρων και δημιουργήθηκαν **επεκτάσεις του κώδικα**.

Έτσι έγινε δυνατή η αναπαράσταση 128 επιπλέον χαρακτήρων ($2^8=256$), που χρησιμοποιήθηκαν από κάθε χώρα για την αναπαράσταση γραμμάτων του τοπικού αλφαβήτου, καθώς και άλλων χαρακτήρων (π.χ. μαθηματικά σύμβολα).



Σας ευχαριστώ για την προσοχή σας