

Computer systems

Dataweergave:
Codes



DEEL 1: Talstelsels

- Talstelsels - Verschillende notaties *(week 01)*
- Negatieve binaire getallen *(week 02)*
- Niet-gehele binaire getallen *(week 03)*
- BCD-getallen *(week 03)*

DEEL 2: Codes

- Ascii *(week 03)*
- Unicode *(week 03)*

- American Standard Code for Information Interchange.
- ASCII = 7-bits code
 - 128 bitconfiguraties (2^7)
 - 96 voor het weergeven van tekst (cijfers, letters en leestekens)
 - 32 voor het realiseren van stuurtekens



- De ASCII tabel
 - Kenmerken van ASCII-code
 - “Extended ASCII”
- ISO-8859
- UNICODE (ISO 10460)
- EBCDIC code



ASCII	000	001	010	011	100	101	110	111
0000	NULL	DLE		0	@	P	`	p
0001	SOM	DC1	!	1	A	Q	a	q
0010	STX	DC2	"	2	B	R	b	r
0011	ETX	DC3	#	3	C	S	c	s
0100	EOT	DC4	\$	4	D	T	d	t
0101	ENQ	NAK	%	5	E	U	e	u
0110	ACK	SYN	&	6	F	V	f	v
0111	BELL	ETB	'	7	G	W	g	w
1000	BS	CAN	(8	H	X	h	x
1001	HT	EM)	9	I	Y	i	y
1010	LF	SUB	*	:	J	Z	j	z
1011	VT	ESC	+	;	K	[k	{
1100	FF	FS	,	<	L	\	l	
1101	CR	GS	-	=	M]	m	}
1110	SO	RS	.	>	N	^	n	ESC
1111	SI	US	/	?	O	_	o	DEL

- **Transmissie-controle-tekens**
 - STX (begin van tekst), ACK (bevestiging van ontvangst), EOT (einde van de transmissie), enz...
- **Informatie-opmaaktekens**
 - CR (carriage return), LF (line feed), Backspace, enz.....
- **Controletekens (device control)**
 - voor randapparaten zoals DC1 (inschakeling van een hulpapparaat : "XON") en DC3 (uitschakeling van een hulpapparaat : "XOFF").
- **Informatie-scheidingstekens** of separator codes.



Hex	Code	Betekenis	Hex	Code	Betekenis
0	NULL	Null	10	DLE	Data Link Escape
1	SOH	Start of Heading	11	DC1	Device Control 1
2	STX	Start of text	12	DC2	Device Control 2
3	ETX	End of text	13	DC3	Device Control 3
4	EOT	End of transmission	14	DC4	Device Control 4
5	ENQ	Enquiry	15	NAK	Negative Acknowledge
6	ACK	Acknowledge	16	SYN	Synchronous Idle
7	BEL	Bell	17	ETB	End of Transmission block
8	BS	Backspace	18	CAN	Cancel
9	HT	Horizontal Tab	19	EM	End of Medium
A	LF	Linefeed	1A	SUB	Substitute
B	VT	Vertical Tab	1B	ESC	Escape
C	FF	Form Feed	1C	FS	File Separator
D	CR	Carriage Return	1D	GS	Group separator
E	SO	Shift Out	1E	RS	Record separator
F	SI	Shift In	1F	US	Unit separator



- voorbeeld: de cijfers van 0 - 9 in ASCII

30h 0

31h 1

32h 2

33h 3

....

39h 9

Ofwel de binaire waarde voorafgegaan door 011:

0 = 0110000

- voorbeeld: Hoofdletters en kleine letters

Hoofdletter A = 100 0001

Kleine letter a = 110 0001



- Uitgebreide tekenset
- = 8-bits code (bit7 “1”)
 - 256 bitconfiguraties (2^8)
 - 128 gestandaardiseerd (ASCII)
 - 96 voor tekst (cijfers, letters en leestekens)
 - 32 voor stuurtekens
 - 128 niet gestandaardiseerd (landafhankelijk)



0		25	˘	50	2	75	K	100	d	125	}	150	û	175	»	200	Ł	225	ß	250	×
1	☺	26	®	51	3	76	L	101	e	126	~	151	ù	176	░	201	Œ	226	G	251	Ö
2	☹	27	¬	52	4	77	M	102	f	127		152	ij	177	▒	202	ƒ	227	p	252	n
3	♥	28	Ł	53	5	78	N	103	g	128	Ç	153	Ö	178	▓	203	Ŧ	228	S	253	²
4	♦	29	«	54	6	79	O	104	h	129	ü	154	Ü	179	⋈	204	Ŧ	229	s	254	n
5	♣	30	Ù	55	7	80	P	105	i	130	é	155	ç	180	┆	205	=	230	m	255	
6	♠	31	Ú	56	8	81	Q	106	j	131	â	156	£	181	┆	206	≡	231	t		
7	⌘	32		57	9	82	R	107	k	132	ä	157	¥	182	┆	207	≡	232	F		
8	■	33	!	58	:	83	S	108	l	133	à	158	□	183	┆	208	≡	233	Q		
9	○	34	"	59	;	84	T	109	m	134	å	159	ı	184	┆	209	≡	234	W		
10	◼	35	#	60	<	85	U	110	n	135	ç	160	á	185	┆	210	≡	235	d		
11	♂	36	\$	61	=	86	V	111	o	136	ê	161	í	186	┆	211	≡	236	¥		
12	♀	37	%	62	>	87	W	112	p	137	ë	162	ó	187	┆	212	≡	237	f		
13	♪	38	&	63	?	88	X	113	q	138	è	163	ú	188	┆	213	≡	238	e		
14	🎵	39	'	64	@	89	Y	114	r	139	ï	164	ñ	189	┆	214	≡	239	Ç		
15	⚙	40	(65	A	90	Z	115	s	140	î	165	Ñ	190	┆	215	≡	240	º		
16	▶	41)	66	B	91	[116	t	141	ì	166	ª	191	┆	216	≡	241	±		
17	◀	42	*	67	C	92	\	117	u	142	Ë	167	º	192	┆	217	≡	242	³		
18	↕	43	+	68	D	93]	118	v	143	Å	168	¿	193	┆	218	≡	243	£		
19	!!	44	,	69	E	94	^	119	w	144	É	169	¬	194	┆	219	≡	244	ó		
20	¶	45	-	70	F	95	_	120	x	145	æ	170	Ø	195	┆	220	≡	245	õ		
21	§	46	.	71	G	96	`	121	y	146	Æ	171	½	196	┆	221	≡	246	,		
22	n	47	/	72	H	97	a	122	z	147	ô	172	¼	197	┆	222	≡	247	»		
23	↑	48	0	73	I	98	b	123	{	148	ö	173	ı	198	┆	223	≡	248	°		
24	↑	49	1	74	J	99	c	124		149	ò	174	«	199	┆	224	≡	249	×		

Verschillende versies:

- ISO-8859-1 of Latin-1 = West-Europese talen (Frans, Spaans, Catalaans, Baskisch, Portugees, Italiaans, Albanees, Nederlands, Duits, Deens, Zweeds, Noors, Fins, Faroese, IJslands, Iers, Schots en Engels).
- ISO-8859-2 of Latin-2 = Centraal en Oost-Europese talen (Tsjechisch, Hongaars, Roemeens, Croatisch, Slovaaks, Sloveens, Servisch)
- ISO-8859-3 of Latin-3 = Zuid-Europese talen (Esperanto, Maltees, ..)
- ISO-8859-4 of Latin-4 = Baltische staten (Estlands, Litouws, Lets, Groenlands en Laplands).
- ISO-8859-5 of Cyrillic = talen met cyrillisch schrift (Bulgaars, Macedonisch, Russisch).
- ISO-8859- of Arabic = Arabische talen
- ISO-8859-7 of Greek = Grieks
- ISO-8859-8 of Hebrew = Hebreeuws en Yiddisch
- ISO-8859-9 of Latin-5 = Turks (aanpassing van Latin-1 door de IJslandse tekens te vervangen door Turkse)
- ISO-8859-11 of Thai = Thaise taal
- ISO-8859-15 of Latin-9 = Update van Latin-1 (vervanging van de minst gebruikte symbolen `|` `~` `'` `, ¼ ½ ¾` met vergeten Franse en Finse letters en plaatsing van het € symbool in cel A4 in plaats van het voormalige internationale currency symbool ₧).



- Uitbreiding van ASCII code tot 8 bit code (ISO-8859) is nog steeds onvoldoende
- Principe: een unieke 16-bits code voor elk teken of symbool (een codepunt).
 - mogelijkheid tot 65536 unieke codepunten
 - nog steeds onvoldoende: ca. 200.000 symbolen
- Eerste 255 karakters uit de UNICODE komen overeen met die van ISO-8859-1
 - eenvoudige conversie tussen ASCII en UNICODE.



- Resterend gebied: elk alfabet een eigen zone
 - Latijns alfabet = 336 codepunten
 - Grieks alfabet = 144 codepunten.
 - de Han karakters uit Chinees en Japans = 20992 codepunten.
 - Koreaans alfabet = 11156 codepunten.

- Unicode wordt ondersteund door verschillende besturingssystemen (o.a. Windows NT/2000/XP/Vista/7 en Unix), verschillende programmeertalen (o.a. JAVA) en een groot aantal toepassingspakketten.



- Principe: hou de omzetting zo simpel mogelijk
 - Voorbeeld: hoofdletter A (hexadecimale waarde 41)

ASCII (7-bit)															
										1	0	0	0	0	1
ISO-8859 (8-bit)															
									0	1	0	0	0	0	1
UNICODE 8-bit (UCS-1, UTF-8)															
									0	1	0	0	0	0	1
UNICODE 16-bit (UCS-2, UTF-16)															
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1

- Extended Binary Coded Decimal InterChange.
 - Een zuivere 8-bit code en bevat dus 256 bitrijen, die echter niet allen benut worden.
 - Het aantal afdrukbare tekens blijft hetzelfde als bij de ASCII-code. Het aantal besturingstekens neemt echter toe. Voor de rest zijn er meer vrije plaatsen, hetgeen de flexibiliteit van de code vergroot.
 - Vooral gebruikt in minicomputers en mainframe-omgeving (vooral IBM).



0	NUL	25	EM	50	SYN	75		100		125	'	150	o	175		200	H	225		250	.
1	SOH	26	CC	51		76	<	101		126	=	151	p	176		201	I	226	S	251	
2	STX	27	CU1	52	PN	77	(102		127	"	152	q	177		202		227	T	252	
3	ETX	28	IFS	53	RS	78	+	103		128		153	r	178		203		228	U	253	
4	PF	29	IGS	54	UC	79		104		129	a	154		179		204		229	V	254	
5	HT	30	IRS	55	EOT	80	&	105		130	b	155		180		205		230	W	255	EO
6	LC	31	IUS	56		81		106	!	131	c	156		181		206		231	X		
7	GE	32	DS	57		82		107	,	132	d	157		182		207		232	Y		
8	DEL	33	SOS	58		83		108	%	133	e	158		183		208	}	233	Z		
9	GE	34	FS	59	CU3	84		109	-	134	f	159		184		209	J	234			
10	SMM	35		60	DC4	85	U	110	>	135	g	160		185		210	K	235			
11	VT	36	BYP	61	NAK	86		111	?	136	h	161		186		211	L	236			
12	FF	37	LF	62		87		112		137	i	162	s	187		212	M	237			
13	CR	38	ETB	63	SUB	88		113		138		163	t	188		213	N	238			
14	SO	39	ESC	64	Sp	89		114		139		164	u	189		214	O	239			
15	SI	40		65		90	!	115		140		165	v	190		215	P	240	□		
16	DLE	41		66		91	\$	116		141		166	w	191		216	Q	241	1		
17	DC1	42	SM	67		92	*	117		142		167	x	192	{	217	R	242	2		
18	DC2	43	CU2	68		93)	118		143		168	y	193	A	218		243	3		
19	TM	44		69		94	;	119		144		169	z	194	B	219		244	4		
20	RES	45	ENQ	70		95	¬	120		145	j	170		195	C	220		245	5		
21	NL	46	ACK	71		96	-	121		146	k	171		196	D	221		246	6		
22	BS	47	BEL	72		97	/	122	:	147	l	172		197	E	222		247	7		
23	IL	48		73		98		123	#	148	m	173		198	F	223		248	8		
24	CAN	49		74	c	99		124	@	149	n	174		199	G	224		249	9		

