

American Standard Code for Information Interchange **ASCII**

Code	..0	..1	..2	..3	..4	..5	..6	..7	..8	..9	..A	..B	..C	..D	..E	..F
0...	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1...	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2...	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3...	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4...	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5...	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6...	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7...	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

- 1963
- 7Bit → 128 Zeichen
- Wenn man kein Amerikaner ist, hat man Probleme beim Zeichenvorrat (und die eigentlich auch)
- „Lösung“: Wir nehmen das ganze Byte. **Erweiterung(en) auf 8Bit.**

American Standard Code for Information Interchange **ASCII**

Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym
0	00	null	32	20	Leerzeichen	64	40	@	96	60	`	128	80		160	A0		192	C0	À	224	E0	à
1	01		33	21	!	65	41	A	97	61	a	129	81		161	A1	í	193	C1	Á	225	E1	á
2	02		34	22	"	66	42	B	98	62	b	130	82		162	A2	ç	194	C2	Â	226	E2	â
3	03		35	23	#	67	43	C	99	63	c	131	83		163	A3	£	195	C3	Ã	227	E3	ã
4	04		36	24	\$	68	44	D	100	64	d	132	84		164	A4	¤	196	C4	Ä	228	E4	ä
5	05		37	25	%	69	45	E	101	65	e	133	85		165	A5	¥	197	C5	Å	229	E5	å
6	06		38	26	&	70	46	F	102	66	f	134	86		166	A6	¦	198	C6	Æ	230	E6	æ
7	07		39	27	'	71	47	G	103	67	g	135	87		167	A7	§	199	C7	Ç	231	E7	ç
8	08	back-space	40	28	(72	48	H	104	68	h	136	88		168	A8	¨	200	C8	È	232	E8	è
9	09	horizontal tab	41	29)	73	49	I	105	69	i	137	89		169	A9	©	201	C9	É	233	E9	é
10	0A	line feed	42	2A	*	74	4A	J	106	6A	j	138	8A		170	AA	ª	202	CA	Ê	234	EA	ê
11	0B		43	2B	+	75	4B	K	107	6B	k	139	8B		171	AB	«	203	CB	Ë	235	EB	ë
12	0C		44	2C	,	76	4C	L	108	6C	l	140	8C		172	AC	¬	204	CC	Ì	236	EC	ì
13	0D	carriage return	45	2D	-	77	4D	M	109	6D	m	141	8D		173	AD	-	205	CD	Í	237	ED	í
14	0E		46	2E	.	78	4E	N	110	6E	n	142	8E		174	AE	®	206	CE	Î	238	EE	î
15	0F		47	2F	/	79	4F	O	111	6F	o	143	8F		175	AF	¯	207	CF	Ï	239	EF	ï
16	10		48	30	0	80	50	P	112	70	p	144	90		176	B0	°	208	D0	Ð	240	F0	ð
17	11		49	31	1	81	51	Q	113	71	q	145	91		177	B1	±	209	D1	Ñ	241	F1	ñ
18	12		50	32	2	82	52	R	114	72	r	146	92		178	B2	²	210	D2	Ò	242	F2	ò
19	13		51	33	3	83	53	S	115	73	s	147	93		179	B3	³	211	D3	Ó	243	F3	ó
20	14		52	34	4	84	54	T	116	74	t	148	94		180	B4	´	212	D4	Ô	244	F4	ô
21	15		53	35	5	85	55	U	117	75	u	149	95		181	B5	µ	213	D5	Õ	245	F5	õ
22	16		54	36	6	86	56	V	118	76	v	150	96		182	B6	¶	214	D6	Ö	246	F6	ö
23	17		55	37	7	87	57	W	119	77	w	151	97		183	B7	·	215	D7	×	247	F7	×
24	18		56	38	8	88	58	X	120	78	x	152	98		184	B8	,	216	D8	Ø	248	F8	ø
25	19		57	39	9	89	59	Y	121	79	y	153	99		185	B9	¡	217	D9	Ù	249	F9	ù
26	1A		58	3A	:	90	5A	Z	122	7A	z	154	9A		186	BA	º	218	DA	Ú	250	FA	ú
27	1B		59	3B	;	91	5B	[123	7B	{	155	9B		187	BB	»	219	DB	Û	251	FB	û
28	1C		60	3C	<	92	5C	\	124	7C		156	9C		188	BC	¼	220	DC	Ü	252	FC	ü
29	1D		61	3D	=	93	5D]	125	7D	}	157	9D		189	BD	½	221	DD	Ý	253	FD	ý
30	1E		62	3E	>	94	5E	^	126	7E	~	158	9E		190	BE	¾	222	DE	þ	254	FE	þ
31	1F		63	3F	?	95	5F		127	7F		159	9F		191	BF	¿	223	DF	ß	255	FF	ÿ

IBM PC (MS-DOS): Codepage 437 / Codepage 850

	..0	..1	..2	..3	..4	..5	..6	..7	..8	..9	..A	..B	..C	..D	..E	..F
0..	NULL 0	⊙ 263A	⊙ 263B	♥ 2665	♦ 2666	♣ 2663	♠ 2660	• 2022	◼ 2508	○ 25CB	◻ 2509	♂ 2642	♀ 2640	♪ 266A	♪ 266B	* 263C
1..	▶ 25BA	◀ 25C4	‡ 2195	!! 203C	¶ B6	§ A7	■ 25AC	‡ 21A8	‡ 2191	‡ 2193	‡ 2192	‡ 2190	‡ 221F	‡ 2194	‡ 25B2	‡ 25B3
2..		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3..	0 30	1 31	2 32	3 33	4 34	5 35	6 36	7 37	8 38	9 39	:	;	<	=	>	?
4..	@ 40	A 41	B 42	C 43	D 44	E 45	F 46	G 47	H 48	I 49	J 4A	K 4B	L 4C	M 4D	N 4E	O 4F
5..	P 50	Q 51	R 52	S 53	T 54	U 55	V 56	W 57	X 58	Y 59	Z 5A	[5B	\ 5C] 5D	^ 5E	_ 5F
6..	` 60	a 61	b 62	c 63	d 64	e 65	f 66	g 67	h 68	i 69	j 6A	k 6B	l 6C	m 6D	n 6E	o 6F
7..	p 70	q 71	r 72	s 73	t 74	u 75	v 76	w 77	x 78	y 79	z 7A	{ 7B	 7C	} 7D	~ 7E	△ 2302
8..	Ç C7	ü FC	é E9	â E2	ä E4	à E0	ã E5	ç E7	ê EA	ë EB	è E8	ï EF	î EE	í EC	Ä C4	Å C5
9..	É C9	æ E6	Æ C6	ô F4	ö F6	ò F2	û FB	ù F9	ÿ FF	ÿ D6	ÿ DC	€ A2	£ A3	¥ A5	ℳ 28A7	f 192
A..	á E1	í ED	ó F3	ú FA	ñ F1	ñ D1	º AA	º BA	¿ BF	¿ 2310	¿ AC	¿ BD	¿ BC	¿ A1	« AB	» BB
B..	☼ 2591	☼ 2592	☼ 2593	 2502	 2524	 2561	 2562	 2556	 2555	 2563	 2551	 2557	 2550	 255C	 2558	 2510
C..	L 2514	L 2534	T 252C	 251C	— 2500	† 253C	† 255E	† 255F	† 255A	† 2554	† 2569	† 2566	† 2560	† 2550	† 256C	† 2567
D..	⌚ 2568	⌚ 2564	⌚ 2565	⌚ 2559	⌚ 2558	⌚ 2552	⌚ 2553	⌚ 256B	⌚ 256A	⌚ 2518	⌚ 250C	⌚ 2588	⌚ 2584	⌚ 258C	⌚ 2590	⌚ 2580
E..	α 3B1	β 0F	Γ 393	π 3C0	Σ 3A3	σ 3C3	μ B5	τ 3C4	φ 3A6	θ 3A9	Ω 384	δ 221E	∞ 3C6	φ 3C6	ε 3B5	η 2229
F..	≡ 2261	± B1	≥ 2265	≤ 2264	∫ 2320	∫ 2321	÷ F7	≈ 2248	• B0	• 2219	• B7	√ 221A	÷ 207F	² B2	■ 25A0	NBSP A0
	..0	..1	..2	..3	..4	..5	..6	..7	..8	..9	..A	..B	..C	..D	..E	..F

	..0	..1	..2	..3	..4	..5	..6	..7	..8	..9	..A	..B	..C	..D	..E	..F
0..	NULL 0	⊙ 263A	⊙ 263B	♥ 2665	♦ 2666	♣ 2663	♠ 2660	• 2022	◼ 2508	○ 25CB	◻ 2509	♂ 2642	♀ 2640	♪ 266A	♪ 266B	* 263C
1..	▶ 25BA	◀ 25C4	‡ 2195	!! 203C	¶ B6	§ A7	■ 25AC	‡ 21A8	‡ 2191	‡ 2193	‡ 2192	‡ 2190	‡ 221F	‡ 2194	‡ 25B2	‡ 25B3
2..		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3..	0 30	1 31	2 32	3 33	4 34	5 35	6 36	7 37	8 38	9 39	:	;	<	=	>	?
4..	@ 40	A 41	B 42	C 43	D 44	E 45	F 46	G 47	H 48	I 49	J 4A	K 4B	L 4C	M 4D	N 4E	O 4F
5..	P 50	Q 51	R 52	S 53	T 54	U 55	V 56	W 57	X 58	Y 59	Z 5A	[5B	\ 5C] 5D	^ 5E	_ 5F
6..	` 60	a 61	b 62	c 63	d 64	e 65	f 66	g 67	h 68	i 69	j 6A	k 6B	l 6C	m 6D	n 6E	o 6F
7..	p 70	q 71	r 72	s 73	t 74	u 75	v 76	w 77	x 78	y 79	z 7A	{ 7B	 7C	} 7D	~ 7E	△ 2302
8..	Ç C7	ü FC	é E9	â E2	ä E4	à E0	ã E5	ç E7	ê EA	ë EB	è E8	ï EF	î EE	í EC	Ä C4	Å C5
9..	É C9	æ E6	Æ C6	ô F4	ö F6	ò F2	û FB	ù F9	ÿ FF	ÿ D6	ÿ DC	ø F8	£ A3	Ø D8	× D7	f 192
A..	á E1	í ED	ó F3	ú FA	ñ F1	ñ D1	º AA	º BA	¿ BF	¿ 2310	¿ AC	¿ BD	¿ BC	¿ A1	« AB	» BB
B..	☼ 2591	☼ 2592	☼ 2593	 2502	 2524	 2561	 2562	 2556	 2555	 2563	 2551	 2557	 2550	 255C	 2558	 2510
C..	L 2514	L 2534	T 252C	 251C	— 2500	† 253C	† 255E	† 255F	† 255A	† 2554	† 2569	† 2566	† 2560	† 2550	† 256C	† 2567
D..	ø F0	ø D0	È CA	È CB	È C8	ı 131	Í CD	İ CE	İ CF	ı 2518	ı 250C	ı 2588	ı 2584	ı A6	ı CC	ı 2580
E..	Ó D3	ß 0F	Ô D4	Ô D2	ö F5	Ö D5	μ B5	þ FE	Þ DE	Ú DA	Ú DB	Ú D9	ý FD	Ý D0	- AF	' B4
F..	SHY AD	± B1	— 2017	¾ BE	§ B6	÷ A7	• B8	° B0	° AB	• B7	• B9	• B3	• B2	■ 25A0	NBSP A0	
	..0	..1	..2	..3	..4	..5	..6	..7	..8	..9	..A	..B	..C	..D	..E	..F

MS-DOS-Codepages

437	Englisch
708	Arabisch (ASMO)
720	Arabisch (Microsoft)
737	Griechisch
775	Baltisch
850	Westeuropäisch
852	Mitteleuropäisch
855	Kyrillisch
857	Türkisch
858	Westeuropäisch mit Euro
860	Portugiesisch
861	Isländisch
862	Hebräisch
863	Kanadisches Französisch
864	Arabisch (IBM)
865	Nordisch
866	Russisch
869	Griechisch

Binäre Codierung: Zeichen

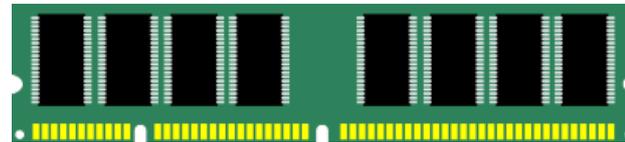
ISO 8859-1/ISO 8859-N

Code	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
0...	nicht belegt															
1...	nicht belegt															
2...	SP	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3...	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4...	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5...	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6...	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7...	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
8...	nicht belegt															
9...	nicht belegt															
A...	NBSP	ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	SHY	®	¯
B...	°	±	²	³	´	µ	¶	·	,	¹	º	»	¼	½	¾	¿
C...	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D...	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E...	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F...	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

ISO 8859

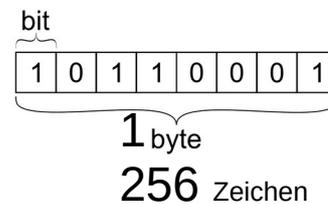
-1	Latin-1, Westeuropäisch
-2	Latin-2, Mitteleuropäisch
-3	Latin-3, Südeuropäisch
-4	Latin-4, Nordeuropäisch
-5	Kyrillisch
-6	Arabisch
-7	Griechisch
-8	Hebräisch
-9	Latin-5, Türkisch
-10	Latin-6, Nordisch
-11	Thai
-12	(existiert nicht)
-13	Latin-7, Baltisch
-14	Latin-8, Keltisch
-15	Latin-9, Westeuropäisch
-16	Latin-10, Südosteuropäisch

Binäre Codierung: Zeichen



(Sprach-)Spezifischer
Zeichenvorrat,
grafische Darstellung

Codierung



Decodierung

(Sprach-)Spezifischer
Zeichenvorrat,
grafische Darstellung

Unicode Transformation Formate (UTF) – Implementieren den Unicode Zeichensatz. **UTF-8**, **UTF-16**, UTF-32

Derzeit: 1.111.998 elementare Zeichen („Codepunkte“) verfügbar
Darstellung: U+00DF (Mindestens 4x4Bit, bis zu U+10FFFF)



Unicode ist ein internationaler Standard, in dem langfristig für jedes sinnvolle Schriftzeichen oder Textelement aller bekannten Schriftkulturen und Zeichensysteme ein digitaler Code festgelegt wird. Ziel ist es, die Verwendung unterschiedlicher und inkompatibler Kodierungen in verschiedenen Ländern oder Kulturkreisen zu beseitigen. Unicode wird ständig um Zeichen weiterer Schriftsysteme durch das Unicode-Konsortium ergänzt.

(Wikipedia, <https://de.wikipedia.org/wiki/Unicode>)

Damit geht auch: <https://unicode.org/emoji/charts/full-emoji-list.html>

No	Code	Browser	App!	Goog	FB	Wind	Twtr	Joy	Sams	GMail	SB	DCM	KDDI	CLDR Short Name
1	U+1F600										–	–	–	grinning face

UTF-8 Codierung mit bis zu 4 Bytes

Regeln:

- Ist das erste Bit eine 0, wird das Zeichen mit 1 Byte codiert
→ 7 Bit zur Codierung übrig, diese entspricht ASCII
- Sonst beginnt das erste Byte mit sovielen 1en, wie zur Codierung Bytes verwendet werden, gefolgt von einer 0.
Alle weiteren Bytes beginnen mit „10“ – **pro Byte** können dann also **6Bit** „Nutzdaten“ untergebracht werden.

Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym	Dez	Hex	Sym
0	00		32	20		64	40	@	96	60	'	128	80		160	A0		192	C0	À	224	E0	à
1	01		33	21	!	65	41	A	97	61	a	129	81		161	A1	í	193	C1	Á	225	E1	á
2	02		34	22	"	66	42	B	98	62	b	130	82		162	A2	ê	194	C2	Â	226	E2	â
3	03		35	23	#	67	43	C	99	63	c	131	83		163	A3	ë	195	C3	Ã	227	E3	ã
4	04		36	24	\$	68	44	D	100	64	d	132	84		164	A4	¼	196	C4	Ä	228	E4	ä
5	05		37	25	%	69	45	E	101	65	e	133	85		165	A5	½	197	C5	Å	229	E5	å
6	06		38	26	&	70	46	F	102	66	f	134	86		166	A6	¾	198	C6	Æ	230	E6	æ
7	07		39	27	'	71	47	G	103	67	g	135	87		167	A7	¸	199	C7	Ç	231	E7	ç
8	08		40	28	(72	48	H	104	68	h	136	88		168	A8	¸	200	C8	È	232	E8	è
9	09	back-space	41	29)	73	49	I	105	69	i	137	89		169	A9	©	201	C9	É	233	E9	é
10	0A	horizontal tab	42	2A	*	74	4A	J	106	70	j	138	90		170	AA	¸	202	CA	Ê	234	EA	ê
11	0B	line feed	43	2B	+	75	4B	K	107	71	k	139	91		171	AB	«	203	CB	Ë	235	EB	ë
12	0C		44	2C	,	76	4C	L	108	72	l	140	92		172	AC	¬	204	CC	Ì	236	EC	ì
13	0D	carriage return	45	2D	-	77	4D	M	109	73	m	141	93		173	AD	¸	205	CD	Í	237	ED	í
14	0E		46	2E	.	78	4E	N	110	74	n	142	94		174	AE	®	206	CE	Î	238	EE	î
15	0F		47	2F	/	79	4F	O	111	75	o	143	95		175	AF	¸	207	CF	Ï	239	EF	ï
16	10		48	30	0	80	50	P	112	76	p	144	96		176	B0	°	208	D0	Ð	240	F0	ð
17	11		49	31	1	81	51	Q	113	77	q	145	97		177	B1	±	209	D1	Ñ	241	F1	ñ
18	12		50	32	2	82	52	R	114	78	r	146	98		178	B2	²	210	D2	Ò	242	F2	ò
19	13		51	33	3	83	53	S	115	79	s	147	99		179	B3	³	211	D3	Ó	243	F3	ó
20	14		52	34	4	84	54	T	116	80	t	148	100		180	B4	¸	212	D4	Ô	244	F4	ô
21	15		53	35	5	85	55	U	117	81	u	149	101		181	B5	µ	213	D5	Õ	245	F5	õ
22	16		54	36	6	86	56	V	118	82	v	150	102		182	B6	¶	214	D6	Ö	246	F6	ö
23	17		55	37	7	87	57	W	119	83	w	151	103		183	B7	·	215	D7	×	247	F7	×
24	18		56	38	8	88	58	X	120	84	x	152	104		184	B8	¸	216	D8	Ø	248	F8	ø
25	19		57	39	9	89	59	Y	121	85	y	153	105		185	B9	¸	217	D9	Ù	249	F9	ù
26	1A		58	3A	:	90	5A	Z	122	86	z	154	106		186	BA	¸	218	DA	Ú	250	FA	ú
27	1B		59	3B	;	91	5B	[123	87	{	155	107		187	BB	»	219	DB	Û	251	FB	û
28	1C		60	3C	<	92	5C	\	124	88		156	108		188	BC	¼	220	DC	Ü	252	FC	ü
29	1D		61	3D	=	93	5D]	125	89	}	157	109		189	BD	½	221	DD	Ý	253	FD	ý
30	1E		62	3E	>	94	5E	^	126	90	~	158	110		190	BE	¾	222	DE	Þ	254	FE	þ
31	1F		63	3F	?	95	5F		127	91		159	111		191	BF	¸	223	DF	ß	255	FF	ÿ

Beispiele 1 + 2:

k → U+006B → 6B₁₆ → 0110 1011₂
 → die letzten 7Bit werden verwendet, um zu codieren, also ein „ASCII k“ in UTF-8
 UTF-8: **0110 1011**

ä → U+00E4 → E4₁₆ → **1110 0100**₂ → Eins am Start, also MultiByte Codierung.
 → 8 Bit Daten zu codieren, dafür braucht man 2 Byte.
 → Der UTF-8 Code beginnt also mit der Startsequenz **110**
 → Dann von hinten 6 Bit → **10 0100**, das Byte beginnt mit **10** (Regel)
 → Bleiben **11** der Daten → Auffüllen zu 6 Bit. **00 0011** + Startsequenz ergibt als
 erstes Byte: 1100 0011
 → UTF-8: **1100 0011 1010 0100**

UTF-8 Codierung mit bis zu 4 Bytes

Regeln:

- Ist das erste Bit eine 0, wird das Zeichen mit 1 Byte codiert
→ 7 Bit zur Codierung übrig, diese entspricht ASCII
- Sonst beginnt das erste Byte mit sovielen 1en, wie zur Codierung Bytes verwendet werden. Alle weiteren Bytes beginnen mit „10“ – **pro Byte** können dann also **6Bit** „Nutzdaten“ untergebracht werden.

	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f	
4E00	一	丁	丅	七	丄	冫	厂	万	丈	三	上	下	丌	不	与	丐	
4E10	丐	丑	刃	专	且	丕	世	卅	丘	丙	业	丛	东	丝	丞	丟	
4E20	北	兩	丟	卵	兩	严	並	喪	丨	凵	个	丫	冫	中	乳	丰	
4E30	丰	卵	串	弗	临	举	丶	丶	丸	丹	为	主	井	丽	举	丿	
4E40	丶	冫	乂	乃	メ	久	乚	毛	么	义	冫	之	乌	乍	乎	乏	
4E50	乐	禾	兵	兵	乔	厖	乖	乘	乘	乙	乚	一	乚	乚	九	乞	也
4E60	习	乡	屮	乱	乏	乏	书	𠂇	乱	乱	西	𠂇	𠂇	𠂇	𠂇	𠂇	

Beispiel 2:

- 乔 → U+4E54 → $4E54_{16} \rightarrow 0100\ 1110\ 0101\ 0010_2$
- 16 Bit Daten zu codieren, dafür braucht man 3 Byte ($3 \times 6 = 18$)
 - Der UTF-8 Code beginnt also mit der Startsequenz **1110**
 - Dann von hinten 6 Bit → **01 0010** , das Byte beginnt mit **10** (Regel) → **1001 0010**
 - Die nächsten 6 Bit → **1110 01** → **1011 1001**
 - Die fehlenden 4 Bit → **0100 mit Padding + Startsequenz** → **1110 0100**
 - UTF-8 Codiert: **1110 0100 1011 1001 1001 0010** (Nutzdaten grün)