

Kapitel 11

Zeichenverarbeitung

Zeichenverarbeitung

- Jedem Zeichen ist ein Zahlencode zugeordnet.
- Dadurch wird ermöglicht, zwischen verschiedenen Systemen Texte auszutauschen.
- Es werden Zahlen, Buchstaben, Steuerzeichen, Sonderzeichen, etc in eine Zahl codiert.

Zeichenverarbeitung

- Im Computer weit verbreitet ist der ASCII-Code.
(American Standard Code for Information Interchange – amerikanischer Standardcode für Informationsaustausch)
- ASCII-Code besteht aus 7-Bit
Jedes Zeichen wird durch einen 7-Bit Zahlencode repräsentiert.
- Dadurch können aber nicht alle verfügbaren Zeichen abgedeckt werden, weshalb man neue Code-Tabellen entwickelte.

Zeichenverarbeitung

- Die ASCII-Code-Tabelle umfasst das "einfache" Alphabet (!, ", #, \$, ..., 0, 1, 2, 3, ..., A, B, C, D, ..., a, b, c, d, ..., {, |, }) und Steuerungscode, die z.B. für die Druckersteuerung benötigt werden - insgesamt sind das 128 Zeichen.
- Nicht berücksichtigt sind darin Sonderzeichen - also deutsche Umlaute, das "ß" und andere nationale Sonderzeichen.
- Diese sind erst im erweiterten ASCII-Code enthalten. (128 – 255)

Zeichenverarbeitung

- Eine andere Zeichencode-Tabelle ist: ANSI-Code
- Der ANSI-Code ist ein 8-Bit Zeichensatz.
- In der Windows Konsole, die unsere C-Programme ausführt, gilt der ANSI-Code.
- Weiteres Beispiel für eine Zeichencode-Tabelle ist: Unicode
Durch Unicode (8-Bit, 16-Bit) wird momentan die umfassendste Code-Tabelle repräsentiert.

dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI
0	00	NUL	NUL	21	15	NAK	NAK	42	2A	*	*
1	01	SOH	SOH	22	16	SYN	SYN	43	2B	+	+
2	02	STX	STX	23	17	ETB	ETB	44	2C	,	,
3	03	ETX	ETX	24	18	CAN	CAN	45	2D	-	-
4	04	EOT	EOT	25	19	EM	EM	46	2E	.	.
5	05	ENQ	ENQ	26	1A	SUB	SUB	47	2F	/	/
6	06	ACK	ACK	27	1B	ESC	ESC	48	30	0	0
7	07	BEL	BEL	28	1C	FS	FS	49	31	1	1
8	08	BS	BS	29	1D	GS	GS	50	32	2	2
9	09	HT	HT	30	1E	RS	RS	51	33	3	3
10	0A	LF	LF	31	1F	US	US	52	34	4	4
11	0B	VT	VT	32	20	spc	spc	53	35	5	5
12	0C	FF	FF	33	21	!	!	54	36	6	6
13	0D	CR	CR	34	22	"	"	55	37	7	7
14	0E	SO	SO	35	23	#	#	56	38	8	8
15	0F	SI	SI	36	24	\$	\$	57	39	9	9
16	10	DLE	DLE	37	25	%	%	58	3A	:	:
17	11	DC1	DC1	38	26	&	&	59	3B	;	;
18	12	DC2	DC2	39	27	'	'	60	3C	<	<
19	13	DC3	DC3	40	28	((61	3D	=	=
20	14	DC4	DC4	41	29))	62	3E	>	>

dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI
63	3F	?	?	84	54	T	T	105	69	i	i
64	40	@	@	85	55	U	U	106	6A	j	j
65	41	A	A	86	56	V	V	107	6B	k	k
66	42	B	B	87	57	W	W	108	6C	l	l
67	43	C	C	88	58	X	X	109	6D	m	m
68	44	D	D	89	59	Y	Y	110	6E	n	n
69	45	E	E	90	5A	Z	Z	111	6F	o	o
70	46	F	F	91	5B	[[112	70	p	p
71	47	G	G	92	5C	\	\	113	71	q	q
72	48	H	H	93	5D]]	114	72	r	r
73	49	I	I	94	5E	^	^	115	73	s	s
74	4A	J	J	95	5F	¯	¯	116	74	t	t
75	4B	K	K	96	60	·	·	117	75	u	u
76	4C	L	L	97	61	a	a	118	76	v	v
77	4D	M	M	98	62	b	b	119	77	w	w
78	4E	N	N	99	63	c	c	120	78	x	x
79	4F	O	O	100	64	d	d	121	79	y	y
80	50	P	P	101	65	e	e	122	7A	z	z
81	51	Q	Q	102	66	f	f	123	7B	{	{
82	52	R	R	103	67	g	g	124	7C		
83	53	S	S	104	68	h	h	125	7D	}	}

dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI
126	7E	~	~	147	93	"	ô	168	A8	¨	¿
127	7F	DEL	DEL	148	94	"	ö	169	A9	©	®
128	80	Ã	P	149	95	•	ò	170	AA	a	¬
129	81	ü	ü	150	96	—	û	171	AB	«	½
130	82	,	é	151	97	—	ù	172	AC	¬	¼
131	83	f	â	152	98	~	ÿ	173	AD	-	i
132	84	"	ä	153	99	™	Ö	174	AE	®	«
133	85	...	à	154	9A	š	Ü	175	AF	—	»
134	86	†	å	155	9B	>	ø	176	B0	°	
135	87	‡	ç	156	9C	œ	£	177	B1	±	
136	88	^	ê	157	9D		Ø	178	B2	²	
137	89	‰	ë	158	9E	ž	×	179	B3	³	
138	8A	Š	è	159	9F	ÿ	f	180	B4	'	
139	8B	<	ï	160	A0		á	181	B5	μ	À
140	8C	Œ	î	161	A1	i	í	182	B6	¶	Â
141	8D		ì	162	A2	¢	ó	183	B7	·	Ã
142	8E	Ž	Ä	163	A3	£	ú	184	B8	¸	©
143	8F		Å	164	A4	¤	ñ	185	B9	¹	
144	90		É	165	A5	¥	Ñ	186	BA	º	
145	91	`	æ	166	A6		a	187	BB	»	+
146	92	'	Æ	167	A7	§	o	188	BC	¼	+

dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI	dez	hex	ASCII	ANSI
189	BD	½	¢	210	D2	Ò	Ê	231	E7	ç	þ
190	BE	¾	¥	211	D3	Ó	Ë	232	E8	è	þ
191	BF	¿	+	212	D4	Ô	È	233	E9	é	Ú
192	C0	À	+	213	D5	Õ	Ì	234	EA	ê	Û
193	C1	Á	-	214	D6	Ö	Í	235	EB	ë	Ü
194	C2	Â	-	215	D7	×	Î	236	EC	ì	Ý
195	C3	Ã	+	216	D8	Ø	Ï	237	ED	í	Ý
196	C4	Ä	-	217	D9	Ù	+	238	EE	î	—
197	C5	Å	+	218	DA	Ú	+	239	EF	ï	’
198	C6	Æ	ã	219	DB	Û	¡	240	F0	ð	-
199	C7	Ç	Ã	220	DC	Ü	—	241	F1	ñ	±
200	C8	È	+	221	DD	Ý	¡	242	F2	ò	=
201	C9	É	+	222	DE	Þ	—	243	F3	ó	¾
202	CA	Ê	-	223	DF	ß	—	244	F4	ô	¶
203	CB	Ë	-	224	E0	à	Ó	245	F5	õ	§
204	CC	Ì	¡	225	E1	á	ß	246	F6	ö	÷
205	CD	Í	-	226	E2	â	Ô	247	F7	÷	¸
206	CE	Î	+	227	E3	ã	Õ	248	F8	ø	¸
207	CF	Ï	¸	228	E4	ä	Ö	249	F9	ù	˙
208	D0	Ð	ð	229	E5	å	Õ	250	FA	ú	·
209	D1	Ñ	Ð	230	E6	æ	µ	251	FB	û	1

dez	hex	ASCII	ANSI
252	FC	ü	3
253	FD	ý	2
254	FE	þ	!
255	FF	ÿ	

Zeichen-Konstanten

- Zeichen-Konstanten werden in Hochkommata eingeschlossen.

Konstante	Bedeutung
'Z'	Zeichen Z
'\x31'	'1'
'\n'	Zeilenvorschub

```
cout << '\x43' << '\x48' << '\x41' << '\x52' << endl;
```

CHAR