

Character Sets

The tables that follow are taken from Reference 1. They list the characters as commonly used on modern computers and are intended as a reference for string drafting.

Reference

1. Griswold, Ralph E. and Madge T. Griswold, *The Icon Programming Language*, 3rd ed. Peer to Peer, 1997.

B

Characters

Characters serve two purposes: the representation of text using glyphs and control operations.

GLYPHS

The glyphs assigned to character codes associate meaning with the codes. Many sets of glyphs are used for purposes ranging from textual material to pictograms and printer's ornaments.

For textual material, on most computer platforms the underlying interpretation for letters, digits, and common punctuation marks is based on the 7-bit ASCII character set (American National Standards Institute, 1986) that assigns glyphs and other interpretations to the first 128 characters.

Various computer platforms extend ASCII in different ways, using different glyphs or associating them with different character codes.

It is now common to assign glyphs to the 128 remaining characters. This allows the use of characters from various languages, as well as various symbols. A collection of glyphs is called a font. Thousands of different fonts are available for various computer platforms.

One standard set of glyphs, which includes ASCII as a subset, is defined by ISO8859-1 (ISO, 1987) and is called Latin-1. This set is used on most UNIX workstations.

Another set of glyphs, called ECS (“extended character set”) (Microsoft, 1991) is used by MS-DOS in the absence of other fonts.

Finally, there is the EBCDIC character set (Ralston and Reilly, 1993) used on IBM mainframes. It assigns glyphs for letters, digits, and common punctuation marks to different character codes than ASCII does. Several different versions of EBCDIC are in use. The most commonly used one is shown in the table in this appendix.

In the table that follows, columns one through three show the decimal, octal, and hexadecimal values for codes. The Latin-1 encoding is shown in column four. Columns five and six show typical text fonts for the Macintosh and Microsoft Windows. The seventh column shows ECS, and the eighth EBCDIC. Finally, the last two columns show the Macintosh versions of symbols and printer ornaments (“dingbats”).

dec.	oct.	hex.	Latin-1	Macintosh	Windows	ECS	EBCDIC	symbols	dingbats
000	000	00							
001	001	01				☺			
002	002	02				●			
003	003	03				♥			
004	004	04				◆			
005	005	05				♣			
006	006	06				♠			
007	007	07				•			
008	010	08				■			
009	011	09				○			
010	012	0a				◼			
011	013	0b				♂			
012	014	0c				♀			
013	015	0d				♪			
014	016	0e				♫			
015	017	0f				✻			
016	020	10				▶			
017	021	11				◀			
018	022	12				↑			
019	023	13				!!			
020	024	14				¶			
021	025	15				§			
022	026	16				-			
023	027	17				↓			
024	030	18				↑			
025	031	19				↓			
026	032	1a				→			
027	033	1b				←			
028	034	1c				L			
029	035	1d				↔			
030	036	1e				▲			
031	037	1f				▼			

dec.	oct.	hex.	Latin-1	Macintosh	Windows	ECS	EBCDIC	symbols	dingbats
032	040	20	blank	blank	blank	blank		blank	blank
033	041	21	!	!	!	!		!	✂
034	042	22	"	"	"	"		∇	✂
035	043	23	#	#	#	#		#	✂
036	044	24	\$	\$	\$	\$		∃	✂
037	045	25	%	%	%	%		%	☐
038	046	26	&	&	&	&		&	⊙
039	047	27	'	'	'	'		₹	⊙
040	050	28	(((((✈
041	051	29)))))	✈
042	052	2a	*	*	*	*		*	☛
043	053	2b	+	+	+	+		+	☛
044	054	2c	,	,	,	,		,	☛
045	055	2d	-	-	-	-		-	☛
046	056	2e	☛
047	057	2f	/	/	/	/		/	☛
048	060	30	0	0	0	0		0	☛
049	061	31	1	1	1	1		1	☛
050	062	32	2	2	2	2		2	☛
051	063	33	3	3	3	3		3	☛
052	064	34	4	4	4	4		4	☛
053	065	35	5	5	5	5		5	☛
054	066	36	6	6	6	6		6	☛
055	067	37	7	7	7	7		7	☛
056	070	38	8	8	8	8		8	☛
057	071	39	9	9	9	9		9	☛
058	072	3a	:	:	:	:		:	☛
059	073	3b	;	;	;	;		;	☛
060	074	3c	<	<	<	<		<	☛
061	075	3d	=	=	=	=		=	☛
062	076	3e	>	>	>	>		>	☛
063	077	3f	?	?	?	?		?	☛
064	100	40	@	@	@	@	blank	≡	☛
065	101	41	A	A	A	A		A	☛
066	102	42	B	B	B	B		B	☛
067	103	43	C	C	C	C		X	☛
068	104	44	D	D	D	D		Δ	☛
069	105	45	E	E	E	E		E	☛
070	106	46	F	F	F	F		Φ	☛
071	107	47	G	G	G	G		Γ	☛
072	110	48	H	H	H	H		H	★
073	111	49	I	I	I	I		I	☆
074	112	4a	J	J	J	J	¢	∅	☛
075	113	4b	K	K	K	K	.	K	☆
076	114	4c	L	L	L	L	<	Λ	★
077	115	4d	M	M	M	M	(M	★
078	116	4e	N	N	N	N	+	N	★
079	117	4f	O	O	O	O		O	★

dec.	oct.	hex.	Latin-1	Macintosh	Windows	ECS	EBCDIC	symbols	dingbats
080	120	50	P	P	P	P	&	Π	☆
081	121	51	Q	Q	Q	Q		Θ	*
082	122	52	R	R	R	R		Ρ	*
083	123	53	S	S	S	S		Σ	*
084	124	54	T	T	T	T		Τ	*
085	125	55	U	U	U	U		Υ	*
086	126	56	V	V	V	V		ς	*
087	127	57	W	W	W	W		Ω	*
088	130	58	X	X	X	X		Ξ	*
089	131	59	Y	Y	Y	Y		Ψ	*
090	132	5a	Z	Z	Z	Z	!	Ζ	*
091	133	5b	[[[[\$	[*
092	134	5c	\	\	\	\	*	∴	*
093	135	5d]]]])]]	*
094	136	5e	^	^	^	^	;	⊥	*
095	137	5f	`	`	`	`	┌	┌	☼
096	140	60	·	·	·	·	└	└	☼
097	141	61	a	a	a	a	/	α	☼
098	142	62	b	b	b	b		β	☼
099	143	63	c	c	c	c		χ	*
100	144	64	d	d	d	d		δ	*
101	145	65	e	e	e	e		ε	*
102	146	66	f	f	f	f		φ	☼
103	147	67	g	g	g	g		γ	*
104	150	68	h	h	h	h		η	*
105	151	69	i	i	i	i		ι	*
106	152	6a	j	j	j	j		φ	*
107	153	6b	k	k	k	k	,	κ	*
108	154	6c	l	l	l	l	%	λ	●
109	155	6d	m	m	m	m	└	μ	○
110	156	6e	n	n	n	n	>	ν	■
111	157	6f	o	o	o	o	?	ο	□
112	160	70	p	p	p	p		π	□
113	161	71	q	q	q	q		θ	□
114	162	72	r	r	r	r		ρ	□
115	163	73	s	s	s	s		σ	▲
116	164	74	t	t	t	t		τ	▼
117	165	75	u	u	u	u		υ	◆
118	166	76	v	v	v	v		ϖ	◇
119	167	77	w	w	w	w		ω	◐
120	170	78	x	x	x	x		ξ	
121	171	79	y	y	y	y		ψ	
122	172	7a	z	z	z	z	:	ζ	
123	173	7b	{	{	{	{	#	{	‘
124	174	7c					@		’
125	175	7d	}	}	}	}	,	}	“
126	176	7e	~	~	~	~	=	~	”
127	177	7f					"		

dec.	oct.	hex.	Latin-1	Macintosh	Windows	ECS	EBCDIC	symbols	dingbats
128	200	80		À		Ç			(
129	201	81		Á		ü	a)
130	202	82		Ç	,	é	b		(
131	203	83		È	f	â	c)
132	204	84		Ñ	"	ä	d		(
133	205	85		Ö	...	à	e)
134	206	86		Ü	†	á	f		<
135	207	87		á	‡	ç	g		>
136	210	88		à	^	ê	h		{
137	211	89		â	‰	ë	i		}
138	212	8a		ä	§	è			(
139	213	8b		ã	¢	ï)
140	214	8c		â	€	î			{
141	215	8d		ç		ï			}
142	216	8e		é		Ä			
143	217	8f		è		É			
144	220	90		ë		æ	j		
145	221	91		ë	,	Æ	k		
146	222	92		í	,	ø	l		
147	223	93		ì	"	ô	m		
148	224	94		î	"	ö	n		
149	225	95		ï	•	ò	o		
150	226	96		ñ	—	ù	p		
151	227	97		ó	—	û	q		
152	230	98		ò	™	ÿ			
153	231	99		ô	š	Ü			
154	232	9a		õ	›	ø			
155	233	9b		ú	œ	ƒ			
156	234	9c		ù		¥			
157	235	9d		ù		℞			
158	236	9e		ù	ÿ	ƒ			
159	237	9f		ü		á			
160	240	a0		†		°	~	Y	⌚
161	241	a1	¡	¢	¡	°	s	,	⌚
162	242	a2	£	£	£	°	t	≤	⌚
163	243	a3	¤	¤	¤	°	u	/	♥
164	244	a4	¥	•	¥	Ñ	v	∞	♥
165	245	a5	¦	¶	¦	°	w	∫	♥
166	246	a6	§	ß	§	°	x	♣	♥
167	247	a7	¨	®	¨	°	y	♦	♥
168	250	a8	©	©	©	°	z	♥	♥
169	251	a9	ª	™	ª	°		♠	♥
170	252	aa	«	™	«	°		↔	♠
171	253	ab	»	™	»	½		←	♠
172	254	ac	¼	™	¼	¾		↔	①
173	255	ad	½	≠	≠	≠	[↑	②
174	256	ae	¾	Æ	Æ	»		→	③
175	257	af	Ø	Ø	Ø	»		↓	④

dec.	oct.	hex.	Latin-1	Macintosh	Windows	ECS	EBCDIC	symbols	dingbats
176	260	b0	°	∞	°	∞		°	⑤
177	261	b1	±	±	±	±		±	⑥
178	262	b2	²	≤	²	≤		”	⑦
179	263	b3	³	≥	³	≥		≥	⑧
180	264	b4	´	¥	´	†		×	⑨
181	265	b5	μ	μ	μ	μ		∞	⑩
182	266	b6	¶	∂	¶	∂		∂	①
183	267	b7	·	∑	·	∑		•	②
184	270	b8		∏		∏		+	③
185	271	b9	˙	π	˙	π		≠	④
186	272	ba	◦	∫	◦	∫		≡	⑤
187	273	bb	»	ª	»]]		≈	⑥
188	274	bc	¼	◦	¼]]		∴	⑦
189	275	bd	½	Ω	½	Ω	}		⑧
190	276	be	¾	æ	¾	ƒ		—	⑨
191	277	bf	¿	ø	¿	⌋		⌋	⑩
192	300	c0	À	¿	À	⌋	{	⌋	①
193	301	c1	Á	¿	Á	⌋	A	⌋	②
194	302	c2	Â	¿	Â	⌋	B	⌋	③
195	303	c3	Ã	¿	Ã	⌋	C	⌋	④
196	304	c4	Ä	¿	Ä	⌋	D	⌋	⑤
197	305	c5	Å	≈	Å	⌋	E	⌋	⑥
198	306	c6	Æ	Δ	Æ	⌋	F	⌋	⑥
199	307	c7	Ç	«	Ç	⌋	G	⌋	⑧
200	310	c8	È	»	È	⌋	H	⌋	⑨
201	311	c9	É	...	É	⌋	I	⌋	⑩
202	312	ca	Ê		Ê	⌋		⌋	①
203	313	cb	Ë	À	Ë	⌋		⌋	②
204	314	cc	Ì	Ã	Ì	⌋		⌋	③
205	315	cd	Í	Ö	Í	⌋		⌋	④
206	316	ce	Î	œ	Î	⌋		⌋	⑤
207	317	cf	Ï	œ	Ï	⌋		⌋	⑥
208	320	d0	Ð	—	Ð	⌋	{	⌋	⑦
209	321	d1	Ñ	—	Ñ	⌋	J	⌋	⑧
210	322	d2	Ò	“	Ò	⌋	K	⌋	⑨
211	323	d3	Ó	”	Ó	⌋	L	⌋	⑩
212	324	d4	Ô	’	Ô	⌋	M	⌋	→
213	325	d5	Õ	’	Õ	⌋	N	⌋	→
214	326	d6	Ö	÷	Ö	⌋	O	⌋	↕
215	327	d7	×	ø	×	⌋	P	⌋	↕
216	330	d8	Ø	ÿ	Ø	⌋	Q	⌋	↗
217	331	d9	Ù	ÿ	Ù	⌋	R	⌋	↗
218	332	da	Ú	/	Ú	⌋		⌋	↘
219	333	db	Û	□	Û	■		⌋	↘
220	334	dc	Ü	◁	Ü	■		⌋	↘
221	335	dd	Ý	▷	Ý	■		⌋	↘
222	336	de	Þ	fi	Þ	■		⌋	↘
223	337	df	ß	fi	ß	■		⌋	↘

dec.	oct.	hex.	Latin-1	Macintosh	Windows	ECS	EBCDIC	symbols	dingbats
224	340	e0	à	‡	à	α	\	◊	↔
225	341	e1	á	·	á	β		<	↔
226	342	e2	â	,	â	Γ	S	®	↘
227	343	e3	ã	„	ã	π	T	©	↘
228	344	e4	ä	%	ä	Σ	U	™	↘
229	345	e5	å	Â	å	σ	V	Σ	↔
230	346	e6	æ	Ê	æ	μ	W	{	↔
231	347	e7	ç	Á	ç	τ	X	}	↔
232	350	e8	è	Ë	è	Φ	Y	}	↔
233	351	e9	é	È	é	θ	Z	}	↔
234	352	ea	ê	Í	ê	Ω			↔
235	353	eb	ë	Î	ë	δ			↔
236	354	ec	ì	Ï	ì	∞			↔
237	355	ed	í	Ì	í	∅			↔
238	356	ee	î	Ó	î	€			↔
239	357	ef	ï	Ô	ï	∩			↔
240	360	f0	ð	🍏	ð	≡	0		↔
241	361	f1	ñ	Ò	ñ	±	1	}	↔
242	362	f2	ò	Ú	ò	≥	2	}	↔
243	363	f3	ó	Û	ó	≤	3	}	↔
244	364	f4	ô	Ü	ô	∫	4		↔
245	365	f5	õ	ı	õ)	5		↔
246	366	f6	ö	ˆ	ö	÷	6	}	↔
247	367	f7	÷	˜	÷	≈	7	}	↔
248	370	f8	ø	˘	ø	°	8	}	↔
249	371	f9	ù	˙	ù	•	9		↔
250	372	fa	ú	˚	ú	·			↔
251	373	fb	û	°	û	√			↔
252	374	fc	ü	˘	ü	n			↔
253	375	fd	ý	˚	ý	²			↔
254	376	fe	þ	˘	þ	■			↔
255	377	ff	ÿ	˘	ÿ				↔

ASCII CONTROL CHARACTERS

The first 32 characters in ASCII are called control characters and are entered by depressing the control key while typing another character. These characters have associated names and functions, such as backspacing and tabbing. These are shown in the following table.

dec.	oct.	hex.	escape seq.	coding	function
000	000	00		control-@	null
001	001	01		control-a	
002	002	02		control-b	
003	003	03		control-c	
004	004	04		control-d	
005	005	05		control-e	
006	006	06		control-f	
007	007	07		control-g	bell
008	010	08	\b	control-h	backspace
009	011	09	\t	control-i	tab
010	012	0a	\n	control-j	linefeed
011	013	0b	\v	control-k	vertical tab
012	014	0c	\f	control-l	formfeed
013	015	0d	\r	control-m	return
014	016	0e		control-n	
015	017	0f		control-o	
016	020	10		control-p	
017	021	11		control-q	
018	022	12		control-r	
019	023	13		control-s	
020	024	14		control-t	
021	025	15		control-u	
022	026	16		control-v	
023	027	17		control-w	
024	030	18		control-x	
025	031	19		control-y	
026	032	1a		control-z	
027	033	1b	\e	control-[escape
028	034	1c		control-\	
029	035	1d		control-]	
030	036	1e		control-^	
031	037	1f		control-_	