

Learning Objectives

- Explain how computers encode characters using 7-bit ASCII
- Identify the difference between ASCII and Unicode



Character Set



ASCII and Unicode are the most widely used character sets.



The term Character set is used to describe the possible characters that can be represented in a computer system.



Representation of text

- When you press a key on the keyboard a stream of 1's and 0's is input to the processor
- The defined list of characters recognised by a computer's hardware and software is known as its **character set**.

ASCII uses 7 bits

- Used to represent lower-case and upper-case English characters and punctuation
- 65 to 90 upper-case
- 97 to 122 lower-case
- 48 to 57 0-9 numeric characters

However ASCII can not represent all languages



Computers represent characters in binary, these characters are represented by ASCII (American Standard Code for Information Interchange)

D	B	C
32	00100000	space
33	00100001	!
34	00100010	"
35	00100011	#
36	00100100	\$
37	00100101	%
38	00100110	&
39	00100111	'
40	00101000	(
41	00101001)
42	00101010	*
43	00101011	+
44	00101100	,
45	00101101	-
46	00101110	.
47	00101111	/
48	00110000	0
49	00110001	1
50	00110010	2
51	00110011	3
52	00110100	4
53	00110101	5
54	00110110	6
55	00110111	7
56	00111000	8

D	B	C
57	00111001	9
58	00111010	:
59	00111011	;
60	00111100	<
61	00111101	=
62	00111110	>
63	00111111	?
64	01000000	@
65	01000001	A
66	01000010	B
67	01000011	C
68	01000100	D
69	01000101	E
70	01000110	F
71	01000111	G
72	01001000	H
73	01001001	I
74	01001010	J
75	01001011	K
76	01001100	L
77	01001101	M
78	01001110	N
79	01001111	O
80	01010000	P
81	01010001	Q

D	B	C
82	01010010	R
83	01010011	S
84	01010100	T
85	01010101	U
86	01010110	V
87	01010111	W
88	01011000	X
89	01011001	Y
90	01011010	Z
91	01011011	[
92	01011100	\
93	01011101]
94	01011110	^
95	01011111	_
96	01100000	`
97	01100001	a
98	01100010	b
99	01100011	c
100	01100100	d
101	01100101	e
102	01100110	f
103	01100111	g
104	01101000	h
105	01101001	i
106	01101010	j

D	B	C
107	01101011	k
108	01101100	l
109	01101101	m
110	01101110	n
111	01101111	o
112	01110000	p
113	01110001	q
114	01110010	r
115	01110011	s
116	01110100	t
117	01110101	u
118	01110110	v
119	01110111	w
120	01111000	x
121	01111001	y
122	01111010	z
123	01111011	{
124	01111100	
125	01111101	}
126	01111110	~
127	01111111	DEL

KEY: D = denary
B = binary
C = character

Unicode

- Unicode was developed to account for every language in the world.
- It uses 2 bytes that give us 2^{16} possibilities (65,536).
- An example use of this would allow a user from any country to select their language when setting up an operating system.
- The Unicode character set would account for every language

