



ASCII and Unicode



ASCII and Unicode

In the early days of computing, programmers would combine groups (sequences) of 0s and 1s to represent different things. For example, they might decide that 00000000 could be used to represent an A and 00000001 could be used to represent a B and so on. The problem was that different programmers used their own coding systems so the sequences meant different things to different people.



As a result of the confusion this caused, a standard was agreed upon for the representation of all the keyboard characters, including the numbers, and other commonly used functions.

This standard is called ASCII or the American Standard Code for Information Interchange. In fact, a 7-bit code was agreed upon as 7 bits gives 128 permutations, which is enough for the most commonly used characters.

More recently, extended ASCII was introduced which is an 8-bit code allowing for 256 characters.





Table 26.1 ASCII look-up table

Char	Decimal	Binary	Char	Decimal	Binary	Char	Decimal	Binary
!	33	00100001	0	48	00110000	F	70	01000110
"	34	00100010	1	49	00110001	G	71	01000111
#	35	00100011	2	50	00110010	H	72	01001000
\$	36	00100100	3	51	00110011	a	97	01100001
%	37	00100101	4	52	00110100	b	98	01100010
&	38	00100110	5	53	00110101	c	99	01100011
'	39	00100111	6	54	00110110	d	100	01100100
[40	00101000	7	55	00110111	e	101	01100101
]	41	00101001	8	56	00111000	f	102	01100110
*	42	00101010	9	57	00111001	g	103	01100111
+	43	00101011	A	65	01000001	h	104	01101000
,	44	00101100	B	66	01000010	i	105	01101001
-	45	00101101	C	67	01000011	j	106	01101010
.	46	00101110	D	68	01000100	k	107	01101011
/	47	00101111	E	69	01000101	l	108	01101100

ASCII does have certain limitations

ASCII was until recently the standard method of converting keyboard and other characters into binary codes. However, ASCII does have certain limitations:

- 256 characters are not sufficient to represent all of the possible characters, numbers and symbols.
- It was initially developed in English and therefore did not represent all of the other languages and scripts in the world.
- Widespread use of the web made it more important to have a universal international coding system.
- The range of platforms and programs has increased dramatically with more developers from around the world using a much wider range of characters.



As a result, a new standard called Unicode has emerged

ASCII codes have been subsumed within Unicode meaning that the ASCII code for a capital letter A is 65 and so is the Unicode code for the same character.

Unicode also includes international characters for over 20 countries and even includes conversions of classical and ancient characters.

To represent these extra characters it is obviously necessary to use more than 8 bits per character and there are two common encodings of Unicode in use today (UTF-8 and UTF-16).

As the name suggests the latter is a 16-bit code.

