1.1 ASCII and Unicode

Bits: 0's and 1's

Computers are built from connected switches that, like light switches, are either on or off. On is represented as 1, and off is 0. A single 0 or 1 is called a ***bit***. 1011 is four bits. Eight bits, like 11000101, are called a ***byte***.

Humans represent information using characters and numbers like Z or 42. To present information that people can understand, computers need a way to represent characters and numbers using 0's and 1's.

PARTICIPATION ACTIVITY | 1.1.1: Bits. ✓

1) A 0 or 1 is called a _____ .

bit

Check Show answer

Correct

bit

The word "bit" is short for "binary digit." Binary means two values.

✓

2) Eight bits are called a _____ .

byte

Check Show answer

Correct

byte

The word refers to a chunk of something, like a "bite" of a sandwich, intentionally misspelled as "byte" to avoid confusion with "bit". In this case the chunk is eight bits.

✓

3) 101100 has _____ bits.

six

Check Show answer

Correct

6

Those six bits represent some kind of information to a computer.

✓

Feedback?

Characters as bits: ASCII

A ***character*** is a letter (a, b, ..., z, A, B, ..., Z), symbol (!, @, #, ...), or single-digit number (0, 1, ..., 9). Basically, each item on a computer keyboard is a character (though more characters exist). Each character can be given a unique bit code.

ASCII is a popular code for characters. ASCII stands for American Standard Code for Information Interchange, and was developed in 1963. ASCII uses 7 bits per code, and has codes for 128 characters. Ex: Using ASCII, the letter Z would be stored in a computer as 1011010. This material inserts a space for readability, as in: 101 1010. Each bit code is sometimes written as an equivalent decimal number (written as Dec below), discussed later.

Table 1.1.1: ASCII bit codes for common characters.

Bit code	Dec	Char	Bit code	Dec	Char	Bit code	Dec	Char
010 0000	32	space	100 0000	64	@	110 0000	96	`
010 0001	33	!	100 0001	65	A	110 0001	97	a
010 0010	34	"	100 0010	66	B	110 0010	98	b
010 0011	35	#	100 0011	67	C	110 0011	99	c
010 0100	36	$	100 0100	68	D	110 0100	100	d
010 0101	37	%	100 0101	69	E	110 0101	101	e
010 0110	38	&	100 0110	70	F	110 0110	102	f
010 0111	39	'	100 0111	71	G	110 0111	103	g
010	40	(	100	72	H	110	104	h

1000			1000			1000		
010 1001	41	)	100 1001	73	I	110 1001	105	i
010 1010	42	*	100 1010	74	J	110 1010	106	j
010 1011	43	+	100 1011	75	K	110 1011	107	k
010 1100	44	,	100 1100	76	L	110 1100	108	l
010 1101	45	-	100 1101	77	M	110 1101	109	m
010 1110	46	.	100 1110	78	N	110 1110	110	n
010 1111	47	/	100 1111	79	O	110 1111	111	o
011 0000	48	0	101 0000	80	P	111 0000	112	p
011 0001	49	1	101 0001	81	Q	111 0001	113	q
011 0010	50	2	101 0010	82	R	111 0010	114	r
011 0011	51	3	101 0011	83	S	111 0011	115	s
011 0100	52	4	101 0100	84	T	111 0100	116	t
011 0101	53	5	101 0101	85	U	111 0101	117	u
011 0110	54	6	101 0110	86	V	111 0110	118	v
011 0111	55	7	101 0111	87	W	111 0111	119	w

011 1000	56	8	101 1000	88	X	111 1000	120	x
011 1001	57	9	101 1001	89	Y	111 1001	121	y
011 1010	58	:	101 1010	90	Z	111 1010	122	z
011 1011	59	;	101 1011	91	[	111 1011	123	{
011 1100	60	<	101 1100	92	\	111 1100	124	|
011 1101	61	=	101 1101	93	]	111 1101	125	}
011 1110	62	>	101 1110	94	^	111 1110	126	~
011 1111	63	?	101 1111	95	_			

Feedback?

PARTICIPATION ACTIVITY 1.1.2: ASCII bit codes (and decimal number equivalents).

Type a character: S ASCII bit code: **1010011** ASCII number: **83**

Feedback?

PARTICIPATION ACTIVITY 1.1.3: ASCII.

1) What is the 7-bit code for a lower-case 'a'?

1100001

Check Show answer

✓ **Correct**

110 0001

Note: 'a' is different from 'A', which has bit code 1000001.

2) What is the 7-bit code for a blank space?

0100000

Check Show answer

✓ **Correct**

010 0000

A space is indeed a character, so requires a bit code.

3) What two-letter word does this sequence of bits represents in ASCII? Pay attention to upper/lower case. Use the above ASCII table.
1001000 1101001

Hi

Check Show answer

✓ **Correct**

Hi

1001000 is 'H'. 1101001 is 'i'.

4) Suppose an email message has 500 characters. How many bits would a computer use to store that email, using ASCII code having 7 bits per character?

3500
`bits`

Check Show answer

✓ **Correct**

3500

500 chars × (7 bits/char) = 3500.

Feedback?

CHALLENGE ACTIVITY | 1.1.1: ASCII code. ✓

547404.4091098.qx3zqy7

✓ 1 ✓ 2 ✓ 3 ✓ 4

Jump to level 1

Convert the ASCII code to a character.
Ex: 1100001 is a

0101001:)

1 2 3 **4**

Check Next

Done. Click any level to practice more. Completion is preserved.

✓ Expected:)
An ASCII table can be used to look up the character.

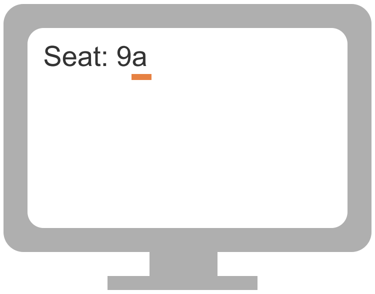
Feedback?

Text is a sequence of character codes

Computers commonly deal with text, consisting of a sequence of characters. The computer stores each character's ASCII code in successive locations in the computer's memory. Each location has at least enough bits (often more) to store an ASCII code.

PARTICIPATION ACTIVITY 1.1.4: Characters are encoded and stored in the computer's memory. ✓

■ 1 2 3 4 ◀ ☑ 2x speed



Memory

101 0011	S
110 0101	e
110 0001	a
111 0100	t
011 1010	:
010 0000	(space)
011 1001	9
110 0001	a

Here, 9 is just another character, and has code 011 1001.
Character a appears a second time, again with code 110 0001.

Captions ^

1. Each character has a unique bit code in ASCII. Uppercase S is 101 0011.

2. Text is stored as a sequence of bit codes in the computer's memory. S is 101 0011, e is 110 0101, a is 110 0001, and t is 111 0100.
3. Symbols and spaces are also characters, and stored in the memory as bit codes. A colon (:) is 011 1010. A space is 010 0000.
4. Here, 9 is just another character, and has code 011 1001. Character a appears a second time, again with code 110 0001.

Feedback?

PARTICIPATION ACTIVITY | 1.1.5: Characters in a computer's memory. ✓

Refer to the above animation.

1) How many characters are in the text displayed on the screen on the left side of the animation? ✓

8

Check **Show answer**

Correct

8

Here is each character and the count: S(1), e(2), a(3), t(4), :(5), space(6), 9(7), and a(8).

2) How many memory locations are used to store that text? ✓

8

Check **Show answer**

Correct

8

In the animation, each character is stored in a unique memory location. Thus, the 8 characters require 8 memory locations.

3) The first letter in the text is S. What bits are stored in the first memory location? ✓

1010011

Check **Show answer**

Correct

101 0011

101 0011 is the ASCII code for an uppercase S.

4) What is the minimum number of bits that each memory location must be able to store in the example above?

7

Check Show answer

✓

Correct

7

Each ASCII code for a character uses 7 0's and 1's. Because each memory location stores a character, each location must have at least 7 bits.

5) How many total bits are needed to store the text shown?

56

Check Show answer

✓

Correct

56

The text has 8 characters, and each character requires 7 bits, yielding 8 × 7 = 56 bits total.

Feedback?

Encoding more characters: Unicode

Unicode is another character encoding standard, published in 1991, whose codes can have more bits than ASCII and thus can represent over 100,000 items, such as symbols and non-English characters. Characters in Unicode are represented as a number, or ***code point***. Ex: In Unicode, the letter "H" is represented as U+0048. U+ means the character is encoded in Unicode, and 0048 is the corresponding code point. The code point is written in hexadecimal, which is discussed elsewhere.

Characters can range from U+0000 to U+10FFFF. The table below provides a very small subset of encodings.

Table 1.1.2: Unicode code points for control characters and basic Latin.

Code point	Char	Code point	Char	Code point	Char	Code point	Char	Code point	Char	Co pc
0020	space	0030	0	0040	@	0050	P	0060	`	00
0021	!	0031	1	0041	A	0051	Q	0061	a	00

0022	"	0032	2	0042	B	0052	R	0062	b	0C
0023	#	0033	3	0043	C	0053	S	0063	c	0C
0024	$	0034	4	0044	D	0054	T	0064	d	0C
0025	%	0035	5	0045	E	0055	U	0065	e	0C
0026	&	0036	6	0046	F	0056	V	0066	f	0C
0027	'	0037	7	0047	G	0057	W	0067	g	0C
0028	(	0038	8	0048	H	0058	X	0068	h	0C
0029	)	0039	9	0049	I	0059	Y	0069	i	0C
002A	*	003A	:	004A	J	005A	Z	006A	j	00
002B	+	003B	;	004B	K	005B	[	006B	k	00
002C	,	003C	<	004C	L	005C	\	006C	l	00
002D	-	003D	=	004D	M	005D	]	006D	m	00
002E	.	003E	>	004E	N	005E	^	006E	n	0C
002F	/	003F	?	004F	O	005F	_	006F	o	

Feedback?

UTF-8, UTF-16, and UTF-32 are encoding standards that indicate how the Unicode is stored. In the UTF-8 standard, characters are stored using variable widths and range from one to four bytes. Whereas in the UTF-32 encoding, all characters are stored as a single 32-bit value. An application that converts the encoding to the final characters viewed by the end user must know which standard is utilized. Emails, web pages, and other digital media frequently contain additional information, or metadata, to indicate how characters are stored. Ex: A webpage may contain the tag <meta charset='utf-8'> to indicate that the UTF-8 Unicode standard is used to encode text.

PARTICIPATION ACTIVITY | 1.1.6: Unicode. ✓

1) An uppercase letter P is represented as _____ in

Correct ✓

Unicode.

● U+0050

○ U+0070

Uppercase and lowercase letters are represented by different encodings. U+0050 corresponds to uppercase P, while U+0070 corresponds to lowercase p.

2) U+0020 represents _____ .

○ an uppercase A

● a space

✓

Correct

Unicode includes encodings for numbers, letters, symbols, and control characters.

3) Which text is represented by the following unicode? U+0061 U+0020 U+0062 U+0020 U+0063

● a b c

○ ABC

✓

Correct

U+0061, U+0062, and U+0063 correspond to lowercase letters a, b, and c, respectively. Each letter is separated U+0020, which corresponds to a space.

4) In Unicode, each character is stored as a 16-bit value.

○ True

● False

✓

Correct

UTF-8, UTF-16, and UTF-32 indicate how each code point is stored. A code point can be represented as a sequence of one to four 8-bit bytes, one or two 16-bit code units, or a single 32-bit code unit.

Feedback?

CHALLENGE ACTIVITY | 1.1.2: Unicode code point. ✓

547404.4091098.qx3zqy7

Jump to level 1

✓ 1

✓ 2

✓ 3

✓ 4

Convert the hexadecimal Unicode to a character.
Ex: U+0061 is a

U+002E: .

1	2	3	**4**

Check Next

Done. Click any level to practice more. Completion is preserved.

✓ Expected: .
A Unicode table can be used to look up the code.

Feedback?

This section provides a simple introduction to Unicode. We encourage the interested reader to Unicode Consortium for additional information on advanced topics and features.

Exploring further:

- Wikipedia: ASCII
- http://www.asciitable.com/
- Unicode 9.0 Character Code Charts

How was this section? 👍 👎 Provide section feedback