A-Level
Computer Science

Character sets

Lesson Objectives

Students will learn about:

- Different characters set, such as ASCII, extended ASCII, and Unicode.
- Applications of each of these character sets.

1.

Content

What is a character?

- A character or symbol that is present on the keyboard.
- It has a specific character code that consists of numbers.

What happens when a character is typed?

- A code is generated for each character or symbol while typing in a keyboard.
- This code is then converted to its character or symbol for displaying and printing purposes.

Character set

- A complete set of all the characters is called a character set.
- Different languages are represented using different character sets.
- These character sets are unique to meet global standards.

ASCII

- The ASCII (American Standard Code for Information Interchange) character set is a 7-bit set of codes that can represent 128 different characters.
- This consists of upper-case letters, lower-case letters, digits, punctuation marks, special characters, and control characters.
- ASCII code is used for English only.

ASCII

Category	Number of characters
Upper-case and lower-case letters	52 characters
Numbers (0-9)	10 characters
Punctuation, space, and other symbols	33 characters
Non-printable control codes	32 characters

Some ASCII codes

Character	Denary value	Binary value	Hex
A	065	01000001	41
DEL	127	11111111	7F
*	042	00101010	2A
4	052	00110100	34

Complete list of ASCII codes

Hex	Char	Hex	Char	Hex	Char	Hex	Char	Hex	Char	Hex	Char
20	<space>	31	1	42	B	53	S	64	d	75	u
21	!	32	2	43	C	54	T	65	e	76	v
22	“	33	3	44	D	55	U	66	f	77	w
23	#	34	4	45	E	56	V	67	g	78	x
24	$	35	5	46	F	57	W	68	h	79	y
25	%	36	6	47	G	58	X	69	i	7A	z
26	&	37	7	48	H	59	Y	6A	j	7B	{
27	‘	38	8	49	I	5A	Z	6B	k	7C	|
28	(	39	9	4A	J	5B	[	6C	l	7D	}
29	)	3A	:	4B	K	5C	\	6D	m	7E	~
2A	*	3B	;	4C	L	5D	]	6E	n	7F	<delete>
2B	+	3C	<	4D	M	5E	^	6F	o		
2C	,	3D	=	4E	N	5F	_	70	p		
2D	-	3E	>	4F	O	60	`	71	q		
2E	.	3F	?	50	P	61	a	72	r		
2F	/	40	@	51	Q	62	b	73	s		
30	0	41	A	52	R	63	c	74	t		

Extended ASCII

- Extended ASCII code consists of an 8-bit character set, and hence 256 different characters can be encoded.
- Characters used in European languages can also be represented in this coding.

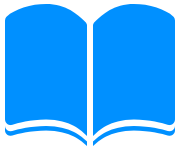
Unicode

- Unicode is the industrial standard for encoding characters in most of the world’s writing systems.
- Initially, this was a 16-bit system that permitted over 65 000 characters.
- The number of bits has now been extended up to 32, permitting coding of several billions of characters.

Unicode

- This system uses 8 to 32 bits per character.
- Because of a higher number of bits per character in Unicode, the files occupy more memory space.
- Facebook and Google also use the Unicode system as users communicate in different languages.
- The ASCII codes for the characters and symbols remained unchanged in Unicode.

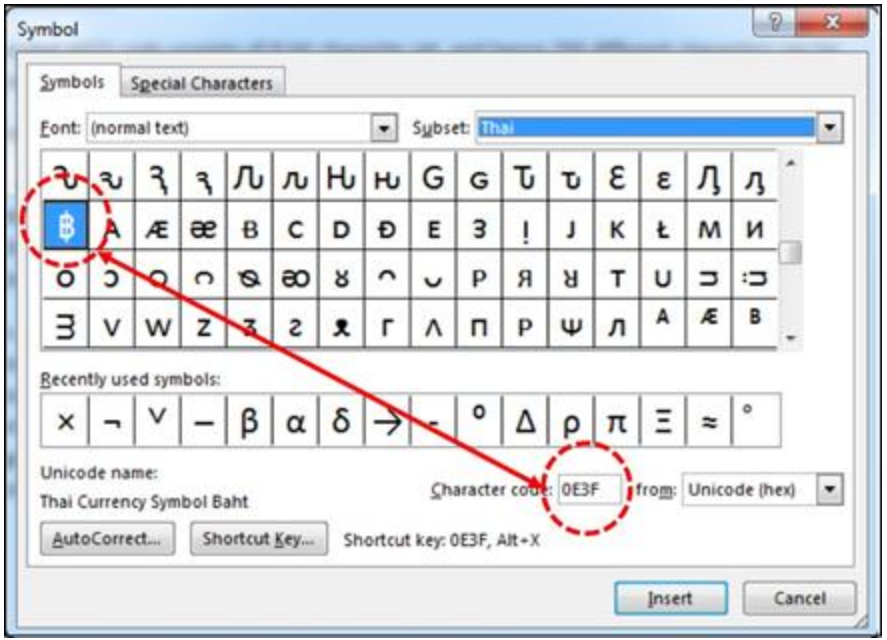
Unicode



- The codes for characters from other languages were added to the list.
- Unicode allocates character codes for languages all over the world.
- Several code pages are used to represent Unicode.

Unicode

- Microsoft word provides an option for users to select letters from other languages such as Thai, Greek, and Latin.
- A user can also type a specific character in a document.
- For example: to enter the character “฿”, its Unicode (OE3F) is typed, and then, ALT+X keys are typed.



Let’s review some concepts

Character

A character or symbol that is present on the keyboard has a specific character code that consists of numbers.

Character set

A complete set of all the characters is called a character set.

ASCII

The ASCII (American Standard Code for Information Interchange) character set is a 7-bit set of codes that can represent 128 different characters.

Extended ASCII

Extended ASCII code consists of an 8-bit character set, and hence 256 different characters can be encoded.

Unicode

Unicode is the industrial standard for encoding characters in most of the world’s writing system. This system uses 8 to 32 bits per character.

2.

Activity

Activity-1

Duration: 10 minutes

1. Hexadecimal values are used to represent web addresses or URLs (Uniform Resource Locator). The ASCII codes are used to represent the web address.

   For example, www.google.com becomes: (using the ASCII codes)

w	w	w	.	g	o	o	g	l	e	.	c	o	m
%77	%77	%77	%2E	%67	%6F	%6F	%67	%6C	%65	%2E	%63	%6F	%6D

   Similarly, use the ASCII code table given in this article to find out the hexadecimal representation of ASCII codes for the URL:

   www.facebook.com

Activity-1

Duration: 10 minutes

2. What number of bits are used to represent ASCII, extended ASCII, and Unicode character sets? Complete the table below.

	ASCII	Extended ASCII	Unicode
Number of bits			

3.

End of topic questions

End of topic questions

1. What is the different character sets available?
2. What are the advantages of extended ASCII character set over the ASCII character set?
3. Why is Unicode adopted as the international standard for character coding?
4. A sorting algorithm sorts the words: “Right, left, Zebra, apple” using the hexadecimal numerical value of the ASCII character set. In what order are these words sorted?
5. How are the ASCII character codes adapted to the Unicode character set?