

The older character sets of ASCII and extended ASCII use 7 and 8 bits. **Unicode** is another character set that maps binary combinations to characters.

The problem with ASCII is that it has a maximum of 128 characters that it can store. Japanese people need to know over 2000 characters and the Chinese alphabet contains around 50 000 characters. Therefore the character sets need to have more bits to store them. Unicode is an international system of storing these characters.

Unicode currently stores over 100 000 characters. There is a **16 bit (2 byte) version** of Unicode which contains 65536 (2^{16}) characters. A **32 bit (4 byte) version** can store over 4 billion (2^{32}) characters, far more than required for every language in the world.

Unicode uses the same codes as ASCII from 0~127.

The following table shows a sample of characters available in Unicode. The black lines in the table show sections of the character set which have been left out.

Binary	Hex	Decimal	Char	Binary	Hex	Decimal	Char
Basic Latin				0000 0000 0100 0100	0044	68	D
0000 0000 0010 0000	0020	32	SPACE	0000 0000 0100 0101	0045	69	E
0000 0000 0010 0001	0021	33	!	0000 0000 0100 0110	0046	70	F
0000 0000 0010 0010	0022	34	"	0000 0000 0100 0111	0047	71	G
Basic Latin				0000 0000 0100 1000	0048	72	H
0000 0000 0010 1100	002C	44	,	Basic Latin - Lowercase			
0000 0000 0010 1101	002D	45	-	0000 0000 0110 0001	0061	97	a
0000 0000 0010 1110	002E	46	.	0000 0000 0110 0010	0062	98	b
Basic Latin - Numeric				0000 0000 0110 0011	0063	99	c
0000 0000 0011 0000	0030	48	0	0000 0000 0110 0100	0064	100	d
0000 0000 0011 0001	0031	49	1	Greek - Lowercase			
0000 0000 0011 0010	0032	50	2	0000 0011 1011 0001	03B1	945	α
0000 0000 0011 0011	0033	51	3	0000 0011 1011 0010	03B2	946	β
0000 0000 0011 0100	0034	52	4	0000 0011 1011 0011	03B3	947	γ
0000 0000 0011 0101	0035	53	5	Japanese - Hiragana			
0000 0000 0011 0110	0036	54	6	0011 0000 0110 1001	3069	12393	ど
0000 0000 0011 0111	0037	55	7	0011 0000 0110 1010	306A	12394	な
0000 0000 0011 1000	0038	56	8	0011 0000 0110 1011	306B	12395	に
0000 0000 0011 1001	0039	57	9	Arabic			
Basic Latin - Uppercase				1111 1100 0010 1000	FC28	64552	ظ
0000 0000 0100 0001	0041	65	A	1111 1100 0010 1001	FC29	64553	ظ
0000 0000 0100 0010	0042	66	B	1111 1100 0010 1010	FC2A	64554	ظ
0000 0000 0100 0011	0043	67	C				

Question: What is the binary in Unicode that represents the letter D?

Answer: **0000 0000 0100 0100**

Question: What letter does the hexadecimal 3069 represent in Unicode?

Answer: ど

Q 47

Character Sets – Unicode - Questions

1. What is the benefit of using the Unicode character set over ASCII? Fill in **one** circle.

- It takes up less storage space It stores more characters
 There is no advantage It is a common character set

[1]

2. How many characters can be stored in 16 bit Unicode?

_____ characters

[1]

3. 32 bit Unicode can store how many characters? Fill in **one** circle.

- Approximately 3 billion 2^{24}
 1 677 7216 2^{32}

[1]

4. Convert the following characters to the decimal number in Unicode:

- a) E _____
 b) C _____
 c) 3 _____
 d) d _____

[4]

5. What characters are represented by the following binary in the Unicode character set?

- a) 0000 0000 0011 0100 _____
 b) 0000 0011 1011 0010 _____
 c) 0000 0000 0010 0010 _____

[3]

6. What is the binary used in Unicode to represent the following characters?

a) な _____

b) ع _____

c) γ _____

[3]

7. What is the hexadecimal used in Unicode to represent the following characters?

a) G _____

b) ظ _____

[2]

8. What is the hexadecimal used in Unicode to represent the following sequences of characters?

a) AH _____

b) 67 _____

c) cab _____

[3]

9. What is the binary code used to represent the following characters?

a) e _____

b) M _____

[2]

10

10

1. What is the benefit of using the Unicode character set over ASCII? Fill in **one** circle.

- It takes up less storage space It stores more characters
 There is no advantage It is a common character set

[1]

2. How many characters can be stored in 16 bit Unicode?

65 536 characters

[1]

3. 32 bit Unicode can store how many characters? Fill in **one** circle.

- Approximately 3 billion 2^{24}
 1 677 7216 2^{32}

[1]

4. Convert the following characters to the decimal number in Unicode:

- a) E 69
 b) C 67
 c) 3 51
 d) d 100

[4]

5. What characters are represented by the following binary in the Unicode character set?

- a) 0000 0000 0011 0100 4
 b) 0000 0011 1011 0010 β
 c) 0000 0000 0010 0010 "

[3]

10

6. What is the binary used in Unicode to represent the following characters?

a) 𐤃 0011 0000 0110 1010

b) ع 1111 1100 0010 1001

c) γ 0000 0011 1011 0011

[3]

7. What is the hexadecimal used in Unicode to represent the following characters?

a) G 0047

b) ظ FC28

[2]

8. What is the hexadecimal used in Unicode to represent the following sequences of characters?

a) AH 0041 0048

b) 67 0036 0037

c) cab 0063 0061 0062

[3]

9. What is the binary code used to represent the following characters?

a) e 0000 0000 0110 0101 (next in sequence after 'd')

b) M 0000 0000 0100 1101 (5 after 'H')

[2]

10