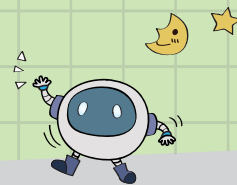


資料數位化



1 正整數數位化

在上一節中，我們學到 8 位元可以表示 0 ~ 255 共 256 個數字，若以 8 位元（1 個位元組）的容量來儲存正整數，每個正整數會固定占用 8 位元的容量。以十進位的數字「23」為例，轉換成二進位是「10111」，由於必須使用 8 個位元來儲存，因此沒有使用到的位元要補 0，記錄為「00010111」。

23 以二進位表示：10111

23 以 8 位元儲存：00010111

→ 共 8 個位元

若以 16 位元（2 個位元組）來儲存正整數，則共可表示 $2^{16} = 65536$ 個正整數，也就是 0 ~ 65535。以十進位的數字「147」為例，轉換成二進位是「10010011」，若以 16 位元表示，沒用到的位元一樣要補 0。

147 以二進位表示：10010011

147 以 8 位元儲存：10010011

147 以 16 位元儲存：0000000010010011

→ 共 16 個位元



手腦並用

想一想，十進位的正整數「1988」，若使用 16 位元儲存的結果為何？

2 文字數位化

我們用鍵盤在電腦中輸入的數字、符號、中文、英文等都屬於文字資料，將這些文字資料轉換成電腦可以理解的內容之過程，就稱為「編碼」。

1 ASCII 編碼系統

就像設定通關密語或暗號一樣，每個人都能制定自己的編碼方法，但若要让不同電腦內的資料可以交互使用，就必須制定相同的編碼規則，因此美國國家標準協會（ANSI）制定了「ASCII 編碼系統¹」。

ASCII 編碼系統最初是以 7 個位元來表示 1 個字元，共可表示 $2^7 = 128$ 個字元，包含英文大小寫、數字、標點符號等²（表 1-2-4）。但許多歐洲國家並非使用純英文的字元，於是後期擴充為 8 個位元，共可表示 256 個字元，適用於所有拉丁文字字母。



知識快遞

1. 美國資訊交換標準碼 (American Standard Code for Information Interchange, ASCII)。
2. ASCII 編碼系統中除了文字與符號，還包含一些控制字元，例如：BS 退格 (backspace)、CR 確認 (carriage return, 同 enter)、CAN 取消 (cancel)、DEL 刪除 (delete)。



手腦並用

參考下頁表 1-2-4，想一想，下列 ASCII 編碼的翻譯結果為何？

- (1) 01001000 = H
- (2) 01100101 = e
- (3) 01101100 = l
- (4) 01101100 = l
- (5) 01101111 = o

表 1-2-4 ASCII 編碼系統

十進位	二進位	字元	十進位	二進位	字元	十進位	二進位	字元	十進位	二進位	字元
000	00000000	NUL	032	00100000	(space)	064	01000000	@	096	01100000	`
001	00000001	SOH	033	00100001	!	065	01000001	A	097	01100001	a
002	00000010	STX	034	00100010	"	066	01000010	B	098	01100010	b
003	00000011	ETX	035	00100011	#	067	01000011	C	099	01100011	c
004	00000100	EOT	036	00100100	\$	068	01000100	D	100	01100100	d
005	00000101	ENQ	037	00100101	%	069	01000101	E	101	01100101	e
006	00000110	ACK	038	00100110	&	070	01000110	F	102	01100110	f
007	00000111	BEL	039	00100111	'	071	01000111	G	103	01100111	g
008	00001000	BS	040	00101000	(	072	01001000	H	104	01101000	h
009	00001001	HT	041	00101001	)	073	01001001	I	105	01101001	i
010	00001010	LF	042	00101010	*	074	01001010	J	106	01101010	j
011	00001011	VT	043	00101011	+	075	01001011	K	107	01101011	k
012	00001100	FF	044	00101100	,	076	01001100	L	108	01101100	l
013	00001101	CR	045	00101101	-	077	01001101	M	109	01101101	m
014	00001110	SO	046	00101110	.	078	01001110	N	110	01101110	n
015	00001111	SI	047	00101111	/	079	01001111	O	111	01101111	o
016	00010000	DLE	048	00110000	0	080	01010000	P	112	01110000	p
017	00010001	DC1	049	00110001	1	081	01010001	Q	113	01110001	q
018	00010010	DC2	050	00110010	2	082	01010010	R	114	01110010	r
019	00010011	DC3	051	00110011	3	083	01010011	S	115	01110011	s
020	00010100	DC4	052	00110100	4	084	01010100	T	116	01110100	t
021	00010101	NAK	053	00110101	5	085	01010101	U	117	01110101	u
022	00010110	SYN	054	00110110	6	086	01010110	V	118	01110110	v
023	00010111	ETB	055	00110111	7	087	01010111	W	119	01110111	w
024	00011000	CAN	056	00111000	8	088	01011000	X	120	01111000	x
025	00011001	EM	057	00111001	9	089	01011001	Y	121	01111001	y
026	00011010	SUB	058	00111010	:	090	01011010	Z	122	01111010	z
027	00011011	ESC	059	00111011	;	091	01011011	[	123	01111011	{
028	00011100	FS	060	00111100	<	092	01011100	\	124	01111100	
029	00011101	GS	061	00111101	=	093	01011101	]	125	01111101	}
030	00011110	RS	062	00111110	>	094	01011110	^	126	01111110	~
031	00011111	US	063	00111111	?	095	01011111	_	127	01111111	DEL

2 Big-5 code (大五碼)

「大五碼」是由臺灣財團法人資訊工業策進會（簡稱「資策會」）制定的編碼系統，採用 16 位元來編碼，以 2 個位元組表示 1 個中文字，共收錄了 13060 個中文字，是繁體中文最常用的中文編碼，普及於臺灣、香港與澳門。

3 Unicode (萬國碼)

因世界各國制定的文字編碼都不相同，為了讓電腦能同時處理多國語言混用的情況，「Unicode」將世界上大部分的語言和符號都統一整理成一套編碼系統。

Unicode 中的每一個字元都有獨一無二的「編碼」，常以 16 位元來表示 1 個字元，現今的作業系統都是採用 Unicode 編碼。



手腦並用

試試看，利用「全字庫中文標準交換碼」網站，查詢自己姓名的 Big-5 code、Unicode 並記錄下來：

（網址：<https://www.cns11643.gov.tw/>）

- 姓名為 王大明
- Big-5 code 為 A4FD A46A A9FA
- Unicode 為 738B 5927 660E

