

Unicode

Unicode คือ รหัสคอมพิวเตอร์ที่ใช้แทนตัวอักษร ทั้งพยัญชนะ, สระ, วรรณยุกต์, ตัวเลข หรือสัญลักษณ์ต่างๆ ได้มากกว่า ASCII ซึ่งเป็นรหัสคอมพิวเตอร์แบบเก่าที่เก็บได้สูงสุดเพียง 256 อักษร โดยปัจจุบัน Unicode สามารถเก็บได้ 1 ล้านอักษร จาก 24 ภาษาทั่วโลก ซึ่งแต่ละภาษา จะมีการกำหนด Unicode แบบตายตัว ไม่มีการซ้ำกัน

ตารางรหัส Unicode ภาษาอังกฤษ และภาษาไทย

Codepage 874 - Thai

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-		0001	0002	0003	0004	0005	0006	0007	0008	0009	000A	000B	000C	000D	000E	000F
1-	0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F
2-	0020	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3-	0030	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>
4-	0040	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N
5-	0050	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^
6-	0060	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n
7-	0070	p	q	r	s	t	u	v	w	x	y	z	{		}	~
8-																
9-																
A-	0E48	ก	ข	ช	ค	ฅ	ฉ	ง	จ	ฉ	ซ	ซ	ฌ	ญ	ณ	ณ
B-	0E10	ฐ	ฑ	ฒ	ณ	ด	ต	ถ	ท	ธ	น	บ	ป	ผ	ฝ	พ
C-	0E20	ภ	ม	ย	ร	ฤ	ล	ภา	ว	ศ	ษ	ส	ห	ฬ	อ	ฮ
D-	0E30	ะ	ั	า	ำ	ิ	ี	ึ	ุ	ู	เ	อ	๑	๒	๓	๔
E-	0E40	๕	๖	๗	๘	๙	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙
F-	0E50	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙	๐	๑	๒	๓	๔

ตารางเลขฐานสิบ, เลขฐานสอง, เลขฐานแปด และเลขฐานสิบหก

Decimal (base 10)	Binary (base 2)	Octal (base 8)	Hexadecimal (base 16)
00	0000	00	0
01	0001	01	1
02	0010	02	2
03	0011	03	3
04	0100	04	4
05	0101	05	5
06	0110	06	6
07	0111	07	7
08	1000	10	8
09	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

ตัวอย่างการแปลงเลขฐานสิบ ไปเป็นเลขฐานสิบหก

$$3616 = 3616 / 16 \text{ ได้ } 226 \text{ เศษ } 0$$

$$= 226 / 16 \text{ ได้ } 14 \text{ เศษ } 2$$

$$= 14 / 16 \text{ ได้ } 0 \text{ เศษ } 14 \text{ (ผลลัพธ์เป็น 0 ให้หยุดคำนวณทันที)}$$

$$= 0E20 \text{ (หรือ E20)} \Rightarrow \text{'ภ'}$$

ตัวอย่างการแปลงเลขฐานสิบหก ไปเป็นเลขฐานสิบ

$$0E0D \text{ (หรือ E0D)} = (0 \times 16^3) + (14 \times 16^2) + (0 \times 16^1) + (13 \times 16^0)$$

$$= 0 + 3584 + 0 + 13$$

$$= 3597 \Rightarrow \text{'ญ'}$$

การแปลงอักขระ เป็นรหัส Unicode ด้วยภาษาไพธอน

นำอักขระมาเข้าฟังก์ชัน `ord()` เพื่อแปลงเป็นเลขฐานสิบ แล้วใช้ฟังก์ชัน `hex()` แปลงเป็นเลขฐานสิบหก

```
# Convert character to unicode (Eng)
eng_A = ord('A')
hex_code = hex(eng_A)
print(hex_code) # '0x41' หรือ '0x041' หรือ '0041'

# Convert character to unicode (Thai)
thai_9 = ord('๙')
hex_code = hex(thai_9)
print(hex_code) # '0xe59' หรือ '0e59'
```

การแปลงรหัส Unicode เป็นอักขระ ด้วยภาษาไพธอน

นำรหัสเลขฐานสิบหกมาเข้าฟังก์ชัน `int()` เพื่อแปลงชนิดข้อมูลเป็น `integer` แล้วใช้ฟังก์ชัน `chr()` แปลงเป็นอักขระ

```
# Convert unicode to character (Eng)
hex_char = '0x61'
dec_code = int(hex_char, 16)
char = chr(dec_code)
print(char) # 'a'

# Convert unicode to character (Thai)
hex_char = '0xe17'
dec_code = int(hex_char, 16)
char = chr(dec_code)
print(char) # 'ท'
```