எழுத்துக்களுக்கான குறி முறைகள் Text coding systems

கணினியில் பல்வேறு தரவுகளை உள்ளிடும் போது அத்தரவுகள் **0, 1** ஆகிய இலக்கங்களாலான பல்வேறு கோலங்களாக (**pattern**) மாற்றப்படும். இதற்கமைய நாம் தட்டச்சுச் செய்யும் இலக்கங்கள் (**numbers**) எழுத்துக்கள் (**alphabets**) விசேட குறியீடுகள˙ (**special characters-symbols**) போன்ற வடிவங்கள் மற்றும் ஒலி ஆகியன கணினியின் தேக்ககச் சாதனங்களில் களஞ்சியப்படுத்தி வைக்கப்பட இருமக் குறிமுறை பயன்படுத்தப்படும்.

A எனும் எழுத்தை உள்ளிடும்போது அது கணினியில் 1000001 என மாற்றிடு செய்யப்படும் அது **A** எனும் எழுத்துக்குரிய இருமக் குறிமுறையாகும். இதில் அடங்கியுள்ள பிற்றுகளின் எண்ணிக்கை **7** . இவ்வாறு ஒவ்வொரு) எழுத்துக்கள தரவுக்கெனவும் வெவ்வேறு பிற்றுக் கோலங்களினைக் (**pattern**) கொண்ட கோர்வைகள் பயன்படுத்தப்படும். ஒவ்வொரு குறிமுறைக்குமான பிற்றுகளின் எண்ணிக்கை மாறுபடும். இதற்கென வெவ்வேறு குறிமுறைகள் உள்ளன. அவை வருமாறு.

1. BCD (Binary Coded Decimal)
2. EBCDIC (Extended Binary Coded Decimal Interchange Code) 3. ASCII (American Standard Code for Information Interchange) 4. UNICODE

1. BCD - Binary Coded Decimal இரும குறிமுறைப் பதின்மம்

இக் குறிமுறை ஆரம்பகாலக் கணினிகளில் பயன்படுத்தப்பட்டது. இந்த குறிமுறையில் ஒரு இலக்கம் நான்கு பிற்றுகளால் (**bits**) வகைகுறிக்கப்படும். இது பதின்ம எண்களை (**decimal** வகைகுறிக்கப் பயன்படுத்தப்படும்.

இதன்மூலம் 16 குறியீடுகளை (2^4 = 16) மட்டுமே வகைகுறிக்க முடியும்

2. EBCDIC (Extended Binary Coded Decimal Interchange Code)

ASCII குறிமுறையைப் பயன்படுத்தி 128 வழி இயல்புகளை மட்டுமே எழுத முடியும். **EBCDIC** குறிமுறையைப் பயன்படுத்தி 256 வழிஇயல்புகளை எழுதலாம். இங்கு ஒரு குறி **8** பிற்றுக்களால் ஆக்கப்பட்ட இரும எண்களால் எழுதப்படும். ஆகவே, இந்த முறையைப் பயன்படுத்தி 256 வழி இயல்புகளை வகை குறிக்க முடியும். இந்த குறிமுறை **IBM main frame** கணினிகளில் பயன்படுத்தப்படும். இந்தக் குறிமுறையில் ஆங்கில **Capital** எழுத்துக்கள் **26** க்குமான ஒன்றுக்கொன்று வேறுபட்ட **EBCDIC** குறியீடுகளும் **Simple** எழுத்துக்களுக்கு ஒன்றுக்கொன்று

3. ASCII Code - American Standard Code for Information Interchange
தகவல் இடைப் பரிமாற்றத்துக்கான அமெரிக்க தரக் குறிமுறை

முதலில் இந்த குறிமுறையின்போது கணினியில் உள்ளிடப்படும் தபவுகள் கணினியில் **7** பிற்றுகள் கொண்ட இரும எண்களால் வகைகுறிக்கப்படும். இந்த குறிமுறையைப் பயன்படுத்தி **128** வழி இயல்புகளை வகை குறிக்கலாம். கணினித் தகவற˙ சாதனங்கள் போன்றவற்றில் எழுத்துக்களை (**text**) வகைகுறிக்க **ASCII** குறிமுறை அமெரிக்க தேசிய தர நிறுவனத்தினால் (**American National Standards Institute - ANSI**)) இனால் உருவாக்கப்பட்டு நியமப்படுத்தப்பட்டுள்ளது.

ஊதாரணம : **School** எனும் சொல்லை சாவிப்பலகையில் சாவிகளை அழுத்தி உள்ளிடும்போது கணினி அதனை வாங்கிக்கொள்ளும் விதத்தைக் காட்டுக.

1) முதலில் இந்த குறியீடுகளுக்குரிய பதின்ம எண்களை எழுதுக. **infotechtamil.com**
S - 83 c - 99 h - 104 o - 111 l - 108

2) இந்த ஒவ்வொரு பதின்ம எண்களுக்கும் உரிய இரும எண்களை எழுதுக. **S - 1010011 c - 1100011 h - 1101000 o - 1101111 l - 1101100**

3) குறிமுறையை முழுமையாக எழுதுக.

S C h o o l

1010011 11000 1101000 1101111 1101111 1101100

Decimal	Binary	Value	Description	Decimal	Binary	Value	Decimal	Binary	Value
0	0	NUL	(Null character)	43	101011	+	86	1010110	V
1	1	SOH	(Start of Header)	44	101100	,	87	1011111	W
2	10	STX	(Start of Text)	45	101101	-	88	1011000	X
3	11	ETX	(End of Text)	46	101110	.	89	1011001	Y
4	100	EOT	(End of Text)	47	101111	/	90	1011010	Z
5	101	ENQ	(Enquiry)	48	110000	0	91	1011011	[
6	110	ACK	(Acknowledgment)	49	110001	1	92	1011100	\
7	111	BEL	(Bell)	50	110010	2	93	1011101	]
8	1000	BS	(Backspace)	51	110011	3	94	1011110	^
9	1001	HT	(Horizontal Tab	52	110100	4	95	1011111	_
10	1010	LF	(Line Feed)	53	110101	5	96	1100000	`
11	1011	VT	(Vertical Tab)	54	110110	6	97	1100001	a
12	1100	FF	(Form Feed)	55	110111	7	98	1100010	b
13	1101	CR	(Carriage Return)	56	111000	8	99	1100011	c
14	1110	SO	(Serial in)(Shift Out)	57	111001	9	100	1100100	d
15	1111	SI	(Serial Out)(Shift Out)	58	111010	:	101	1100101	e
16	10000	DLE	(Data Link Escape)	59	111011	;	102	1100110	f
17	10001	DC1	(XON) (device Control -1)	60	111100	<	103	1100111	g
18	10010	DC2	(Device Control-2)	61	111101	=	104	1101000	h
19	10011	DC3	(XOFF)(Device Control-3)	62	111110	>	105	1101001	i
20	10100	DC4	(Device Control-4)	63	111111	?	106	1101010	j
21	10101	NAK	(Negative Acknowledgement)	64	1000000	@	107	1101011	k
22	10110	SYN	(Synchronous Idle)	65	1000001	A	108	1101100	l
23	10111	ETB	(End of Trans.Block)	66	1000010	B	109	1101101	m
24	11000	CAN	(Cancel)	67	1000011	C	110	1101110	n
25	11001	EM	(End of Medium)	68	1000100	D	111	1101111	o
26	11010	SUB	(Substitute)	69	1000101	E	112	1110000	p
27	11011	ESC	(Escape)	70	1000110	F	113	1110001	q
28	11100	FS	(File Separator)	71	1000111	G	114	1110010	r
29	11101	GS	(Group Separator)	72	1001000	H	115	1110011	s
30	11110	RS	(Request to Send)	73	1001001	I	116	1110100	t
31	11111	US	(Unit Separator)	74	1001010	J	117	1110101	u
32	100000	SP	(Space)	75	1001011	K	118	1110110	v
33	100001	!		76	1001100	L	119	1110111	w
34	100010			77	1001101	M	120	1111000	x
35	100011	#		78	1001110	N	121	1111001	y
36	100100	$		79	1001111	O	122	1111010	z
37	100101	%		80	1010000	P	123	1111011	{
38	100110	&		81	1010001	Q	124	1111100	|
39	100111	'		82	1010010	R	125	1111101	}
40	101000	(		83	1010011	S	126	1111110	~
41	101001	)		84	1010100	T	127	1111111	DEL
42	101010	*		85	1010101	U			

Table 2.2 The ASCII table (7 BIT)

ஒற்றைக் குறிமுறை / ஒருங்குறிமுறை (UNICODE)

infotechtamil.com

கணினிகளில் தரவுகளை எடுத்துக்காட்ட, **ASCII** குறிமுறை மூலமாக **128** வழி இயல்புகளையும் **EBCDIC** குறிமுறையைப் பயன்படுத்தி **256** வழி இயல்புகளையும் பெற்றுக்கொள்ள முடியும். எனினும், சிங்களம், ஜப்பான், சீன, தமிழ் மொழிகளில் உள்ள எழுத்துக்கள் 256 வழி இயல்புகளிலும் அதிகமாகையால் இந்த குறிமுறைகளைப் பயன்படுத்த இயலாது. **16 Bits** பிற்றுகள் கொண்ட ஒன்றுக்கொன்று வேறுபட்ட குறிகள் **65,536** இனை(2^{16} = 65,536) தயார்படுத்தப்பட்ட குறிமுறையே ஒற்றைக் குறிமுறை எனப்படும்.

,இதன் மூலம் சிங்களம், தமிழ் உட்பட உலகின் அனைத்து மொழிகளிலுள்ள எழுதுக்களை வகைகுறிக்கவும் விசேட குறியீடுகள், உருக்கள் ஆகியவற்றை வகைகுறிக்கவும் முடிகிறது.