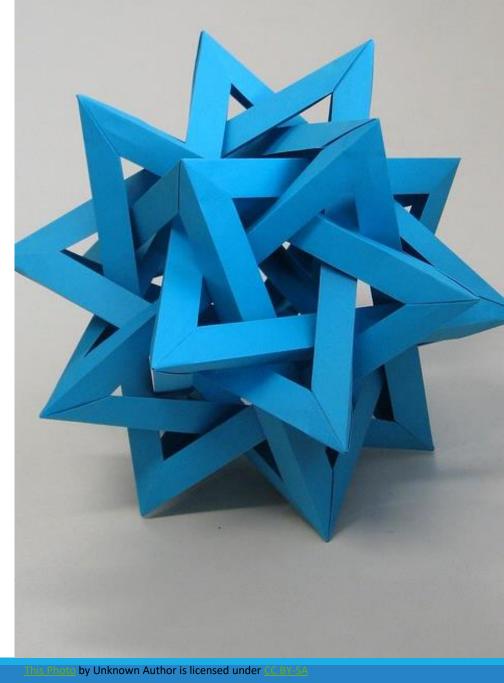


Unità T1: Rappresentazione dei dati



Politecnico di Torino, 2022/23 INFORMATICA / COMPUTER SCIENCES 1

Come contiamo?

- Il sistema di numerazione del mondo occidentale (sistema indo-arabo) è:
 - o decimale
 - posizionale

$$252 = 2 \times 100 + 5 \times 10 + 2 \times 1$$
$$= 2 \times 10^{2} + 5 \times 10^{1} + 2 \times 10^{0}$$

Sistemi di numerazione

- Non posizionali (additivi):
 - egiziano
 - o romano
 - o greco
- Posizionali:
 - babilonese (2 cifre, sessagesimale)
 - inuit, selti, maya (ventesimale)
 - indo-arabo (decimale)
- Ibridi:
 - cinese

Sistema di numerazione posizionale

- Occorre definire la base B da cui discendono varie caratteristiche:
 - o cifre = { 0, 1, 2, ..., B-1 }
 - o peso della cifra i-esima = Bi
 - o rappresentazione (numeri naturali) su N cifre
 - a_{N-1} a_{N-2} ... a₃ a₂ a₁ a₀

$$\mathbf{A} = \sum_{i=0}^{N-1} a_i \cdot B^i$$

Il sistema binario

- Base = 2
- Cifre = { 0, 1 }
- BIT = BInary DigiT

Esempio:

$$101_2 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$$

= 1 × 4 + 1 × 1
= 5₁₀

Binario e Decimale

ALCUNI NUMERI BINARI

0 0		1000 8	
1 1		10019	
10 2		101010	
11 3		101111	
100	4	110012	
101	5	110113	
110	6	111014	
111	7	111115	

ALCUNE POTENZE DI DUE

201	2 ⁹	512
212	2 ¹⁰	1024
224	2 ¹¹	2048
238	2 ¹²	4096
2416	2 ¹³	8192
2 ⁵ 32	2 ¹⁴	16384
2 ⁶ 64	2 ¹⁵	32768
2 ⁷ 128	2 ¹⁶	65536
2 ⁸ 256		

Conversione di numeri naturali da binario a decimale

Si applica direttamente la definizione effettuando la somma pesata delle cifre binarie:

$$1101_{2} = 1 \times 2^{3} + 1 \times 2^{2} + 0 \times 2^{1} + 1 \times 2^{0}$$

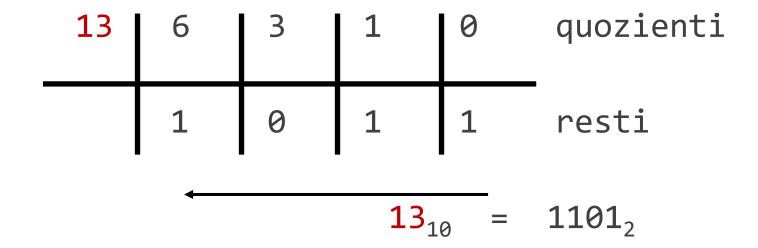
$$= 8 + 4 + 0 + 1$$

$$= 13_{10}$$

Conversione da sistema decimale a binario

Dall'interpretazione della codifica binaria

- Regola pratica:
 - Divisioni successive per due
 - Si prendono i resti in ordine inverso



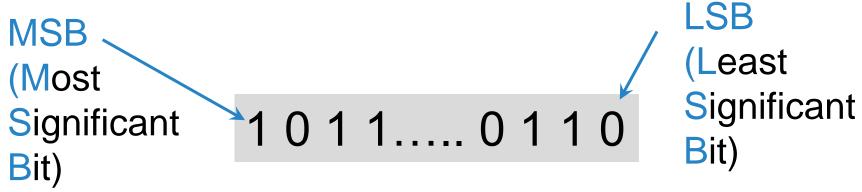
Limiti del sistema binario (rappresentazione naturale)

- Consideriamo numeri naturali in binario:
 - 1 bit ~ 2 numeri ~ { 0, 1 }₂ ~ [0 ... 1]₁₀
 - 2 bit ~ 4 numeri ~ { 00, 01, 10, 11}₂ ~ [0...3]₁₀
- Quindi in generale per numeri naturali a N bit:
 - o combinazioni distinte: 2^N
 - o intervallo di valori

```
0 \le x \le 2^{N} - 1 [ base 10 ] (000...0) \le x \le (111...1) [ base 2 ]
```

Terminologia

- Bit rappresenta una singola cifra
- Aggregazioni di bit rilevanti:
 - Byte = 8 bit
- Word = aggregazione di byte
 - 0 1,2,4,8
 - Utilizzate per le celle di memoria
- Dato un qualunque numero di bit



Limiti del sistema binario (rappresentazione naturale)

Bit	Simboli	Min10	Max10
4	16	0	15
8	256	0	255
16	65 536	0	65 535
32	4 294 967 296	0	4 294 967 295

Somma in binario

Regole base:

```
0 + 0 = 0

0 + 1 = 1

1 + 0 = 1

1 + 1 = 0 (carry = 1)
```

Si effettuano le somme parziali tra i bit dello stesso peso, propagando gli eventuali riporti:

Sottrazione in binario

Regole base:

```
0 - 0 = 0

0 - 1 = 1 (borrow = 1)

1 - 0 = 1

1 - 1 = 0
```

Sottrazione in binario

Si effettuano le differenze parziali tra i bit dello stesso peso, gestendo gli eventuali prestiti:

	1			
1	0	0	1	-
0	1	1	0	=
0	0	1	1	

Overflow

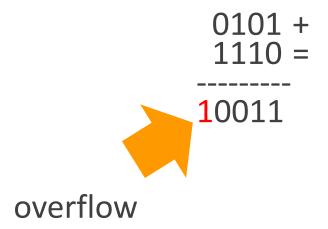
Si usa il termine overflow per indicare l'errore che si verifica in un sistema di calcolo automatico quando il risultato di un'operazione non è rappresentabile con la medesima codifica e numero di bit degli operandi.

- L'overflow è una condizione "dinamica"
 - Esiste solo come risultato di un'operazione

Politecnico di Torino, 2022/23 INFORMATICA / COMPUTER SCIENCES 15

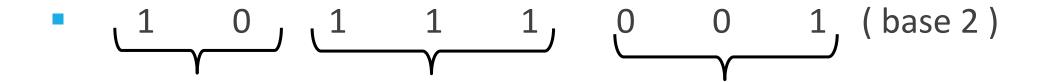
Overflow

- Nella somma in binario puro si ha overflow quando:
 - o si lavora con numero fisso di bit
 - si ha carry sul MSB
- Esempio: numeri da 4 bit codificati in binario puro



Il sistema ottale

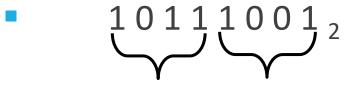
- base = 8 (talvolta indicata con Q per Octal)
 - o cifre = { 0, 1, 2, 3, 4, 5, 6, 7 }
 - o utile per scrivere in modo compatto i numeri binari (3:1)



2 7 1 (base 8)

Il sistema esadecimale

- base = 16 (talvolta indicata con H per Hexadecimal)
 - o cifre = { 0, 1, ..., 9, A, B, C, D, E, F }
 - o utile per scrivere in modo compatto i numeri binari (4:1)



B 9 16

Rappresentazione dei numeri relativi

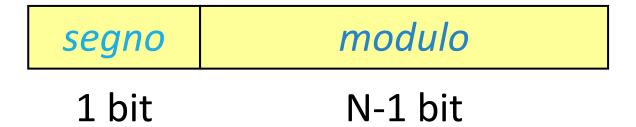
I numeri con segno

- Il segno dei numeri può essere solo di due tipi:
 - o positivo (+)
 - o negativo (−)
- È quindi facile rappresentarlo in binario ... ma non sempre la soluzione più semplice è quella migliore!
- Varie soluzioni, le più usate sono
 - Modulo e segno
 - Complemento a due

Codifica "modulo e segno"

- un bit per il segno (tipicamente il MSB):
 - \circ 0 = segno positivo (+)
 - \circ 1 = segno negativo ()

N-1 bit per il valore assoluto (anche detto modulo)



Modulo e segno: esempi

Usando una codifica su quattro bit:

Modulo e segno

- Svantaggi:
 - doppio zero (+ 0, 0)
 - o operazioni complesse
 - o es. somma A+B

$$A > 0$$
 $A < 0$
 $B > 0$ $A + B$ $B - |A|$
 $B < 0$ $A - |B|$ $-(|A| + |B|)$

Modulo e segno: limiti

In una rappresentazione M&S su N bit:

$$-(2^{N-1}-1) \le x \le +(2^{N-1}-1)$$

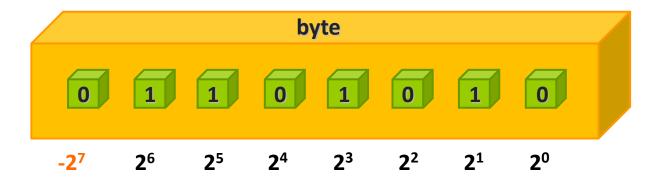
Esempi:

```
o 8 bit = [-127 ... +127]
```

○ 16 bit = [-32 767 ... +32 767]

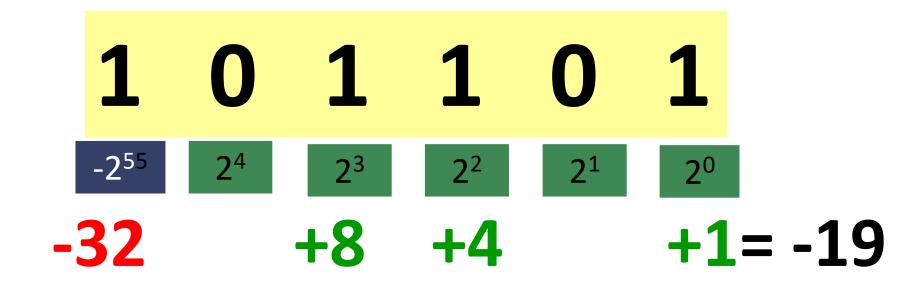
Codifica in complemento a due

- In questa codifica per un numero a N bit:
 - o il MSB ha peso negativo (pari a -2^{N-1})
 - o gli altri bit hanno peso positivo



Ne consegue che MSB indica sempre il segno:

Complemento a due (esempio)



Politecnico di Torino, 2022/23 INFORMATICA / COMPUTER SCIENCES 26

Da decimale a complemento a 2

- Per convertire un numero decimale in complemento a 2:
- Se positivo, si effettua la solita conversione
- Se negativo:
 - Si converte il modulo in binario
 - Si complementa ogni bit (0->1, 1->0)
 - Si somma 1 (sul corrispondente numero di bit)

Da decimale a complemento a 2

Esempio

```
• +15 su 5 bit in c.a.2 \Rightarrow +15 = 01111<sub>2</sub> \Rightarrow 01111
```

■ -12 su 5 bit in c.a.2 \Rightarrow +12 = 01100₂ complementiamo i bit sommiamo +1 (su bit) \Rightarrow 10011 -

10011 10011 + 00001 = 10100

Complemento a 2 e operazioni

- La rappresentazione in complemento a due è oggi la più diffusa perché semplifica la realizzazione dei circuiti per eseguire le operazioni aritmetiche
- Possono essere applicate le regole binarie a tutti i bit, segno compreso!

- La somma e sottrazione si effettuano direttamente, senza badare ai segni degli operandi
- La sottrazione si può effettuare sommando al minuendo il CA2 del sottraendo

Somma in CA2 - esempio

00100110 + 11001011

verifica: 38 + (-53) = -15

Sottrazione in CA2 - esempio

00100110 - 11001011

00100110 -11001011 = 01011011

verifica: 38 - (-53) = 91

Overflow nella somma in CA2

Operandi con segno discorde: non si può mai verificare overflow.

 Operandi con segno concorde: c'è overflow quando il risultato ha segno discorde.

In ogni caso, si trascura sempre il carry sul MSB.

Complemento a 2: limiti

In una rappresentazione c.a 2 su N bit:

$$-(2^{N-1}) \le x \le +(2^{N-1}-1)$$

Esempi:

```
\circ 8 bit = [-128 ... +127]
```

○ 16 bit = [-32 768 ... +32 767]

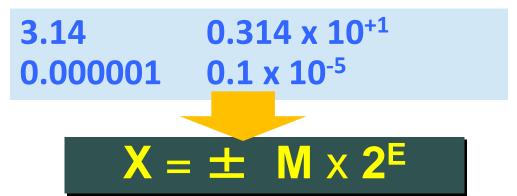
Rappresentazione di numeri reali

Rappresentazione di numeri reali

- Due opzioni:
 - 1. Dati N bit disponibili riservarne M per la parte frazionaria e N-M per la parte intera (VIRGOLA FISSA)



Implementare negli N bit la notazione esponenziale ("scientifica")
 (VIRGOLA MOBILE)



Perche' virgola mobile?

- Virgola fissa = si riserva un numero di posizioni (bit) predefinite alla parte intera ed alla parte frazionaria
 - Precisione fissa
- NOTA: I bit della parte frazionaria hanno peso 2⁻ⁱ

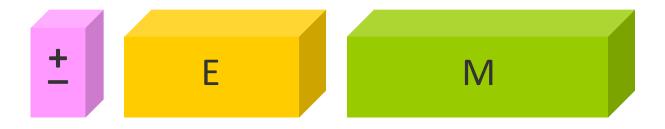
- Virgola mobile = precisione variabile
 - Nella stessa rappresentazione possiamo rappresentare sia numeri molto grandi (esponenti grandi) sia molto piccoli (esponenti piccoli)

Politecnico di Torino, 2022/23 INFORMATICA / COMPUTER SCIENCES 36

Rappresentazione in virgola mobile (Floating Point)

Nella memoria del calcolatore si memorizzano:

- Segno
- Esponente (con il suo segno)
- Mantissa



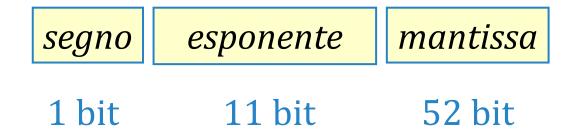
 $X = \pm M \times 2^{E}$

Formato IEEE-754

- Mantissa nella forma '1,...' (valore max < 2)</p>
- Base dell'esponente pari a 2
- IEEE 754 SP: (float)



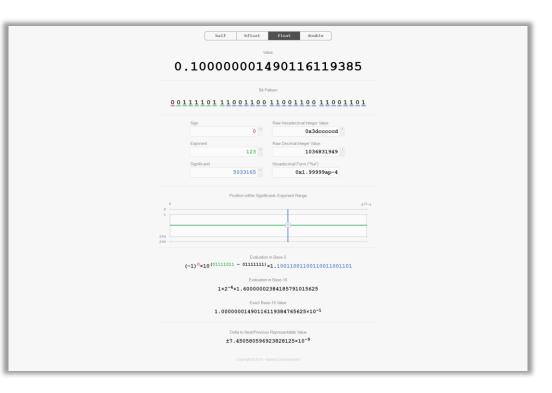
• IEEE 754 DP: (double)



Esempi

https://float.exposed/

	half bfloat float double
	Value
	10.0
	Bit Pattern
<u>0 1</u>	000001001000000000000000000000000000000
	Sign
	Exponent Raw Decimal Integer Value 1.30 1 1092616192 1
	Significand Hosadecirus Form ("Na") 0x1.4p+3
	Position within Significand-Exponent Range
0	273.4
254	
205	546000
	Emiliation in Base-2 (-1) 0×10 (10000010 - 01111111) ×1.010000000000000000000000000000000000
	Evaluation in Base-10 1x2 ³ x1.25
	1×2°×1.25 Exact Base-10 Value
	1.0×10 ¹
	Delta to NextPrevious Representable Value
	±9.5367431640625×10 ⁻⁷

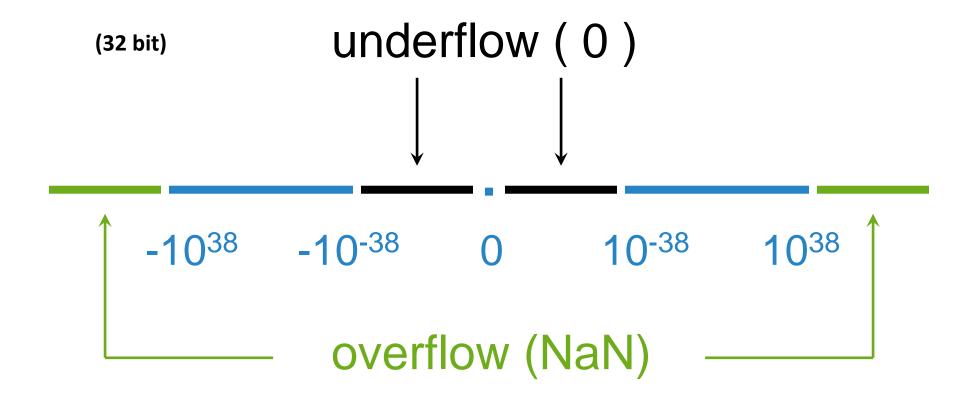


Floating point ed approssimazioni

 La limitatezza della precisione porta ad avere problemi con le operazioni aritmetiche

- Alcuni numeri NON sono rappresentabili in modo esatto
 - E non sono numeri 'strani'...
 - Valori quali 0.1, 0.6 sono approssimati

IEEE-754 SP: intervallo di valori

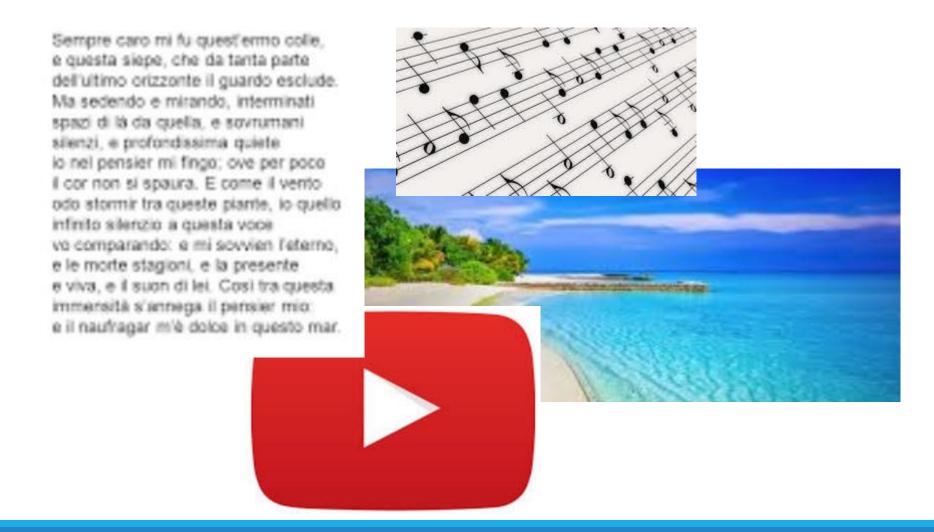


Floating point ed approssimazioni

- La limitatezza della precisione porta ad avere problemi con le operazioni aritmetiche
- Esempio: in FP, la somma NON e' associativa!!!
 - x+(y+z) puo' essere diverso da (x+y)+z!
- Esempio:
 - $x = -1.5_{10} * 10^{38}$
 - $y = +1.5_{10} * 10^{38}$
 - $z = 1.0_{10}$
 - Eseguendo su calcolatore
 - $x+(y+z) = -1.5_{10} * 10^{38} + (1.5_{10} * 10^{38} + 1) =$ = $-1.5_{10} * 10^{38} + 1.5_{10} * 10^{38} = \mathbf{0}$
 - $(x+y)+z = (-1.5_{10} * 10^{38} + 1.5_{10} * 10^{38}) + 1 = 1$

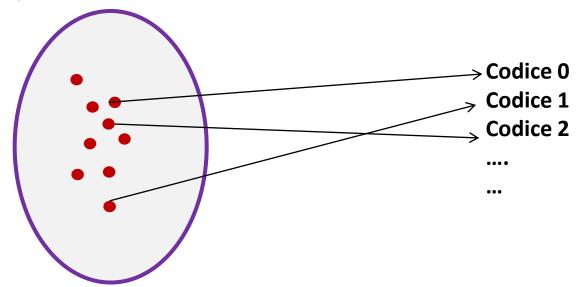
Rappresentazione di dati non numerici

Elaborazione dell'informazione non numerica



Informazione non numerica

- Il calcolatore è in grado di manipolare SOLO numeri!
- Per gestire dati non numerici l'unica possibilità è creare una corrispondenza tra oggetti e numeri
 - Ad ogni oggetto si assegna un codice univoco
 - Questo codice diventa la rappresentazione dell'oggetto
 - Nel calcolatore, il codice sara' binario...



Oggetti e numeri

- Assumendo di assegnare codici binari, dati N bit si possono codificare 2N «oggetti» distinti
- Esempio (3 bit):

Codici binari	000	001	010	011	100	101	110	111
oggetti	0	1	2	3	4	5	6	7

- Se viceversa ho M oggetti, per codificarli tutti dovrò usare un numero di bit N pari a N = log2 M
 - \circ In pratica, la prima potenza di 2 tale che $2^{N} > M$

Codifica dei caratteri: codice ASCII

- Occorre una codifica standard perché è il genere di informazione più scambiata:
 - codice ASCII (American Standard Code for Information Interchange)
- Usa 8 bit (originariamente 7 bit per US-ASCII) per rappresentare:
 - 52 caratteri alfabetici (a...z A...Z)
 - 10 cifre (0...9)
 - segni di interpunzione (,;!?...)
 - o caratteri di controllo

Codice ASCII

Dec Hx Oct Char Oec Hx Oct Hrm Chr Oec	Dec	Н	Oct	Cha	r I	Dec	H×	Oct	Html	Chr	Dec	H×	Oct	Html	Chr	Dec	H×	Oct	Html Ch	hr
1 1 001 SOH (start of heading) 2 2 002 STK (start of text) 3 3 21 041 !! 65 41 101 7 A 97 61 141 a a 34 22 042 "" 66 42 102 B B 98 62 142 b b 67 43 103 C C 99 63 143 c C 68 44 104 D D 99 63 143 c C 68 60 06 ACK (acknowledge) 88 60 10 BS (backspace) 99 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NF form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 1) 18 12 022 DC2 (device control 1) 18 12 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 23 17 027 FTB (end of trans. block) 23 17 027 FTB (end of medium) 25 30 EX (scape) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 29 1D 035 GS (group separator) 30 1E Ceroof separator) 30 1E Ceroof separator) 30 1E Ceroof separator) 30 1E O40 ACK (acknowledge) 31 A 22 042 " " 34 22 042 " " 34 22 042 " " 34 22 042 " " 34 22 042 " " 34 2 102 \$ B 34 30 55 ENG 66 44 104 D D 46 44 104 D D 46 45 105 E E 46 44 104 D D 46 45 105 E E 46 45 106 F F 46 45 105 E E 47 147 107 G G 48 110 H H 47 107 G G 48 110 H H 47 107 G G 48 110 H H 47 107 G G 48 110 H H 49 11 H G 48 110 H H 40 116 H L 40 116 I L 40 116 I L 41 101 K A 41 110 H B 40 111 H B 41 110 H B 41 110 H B 41 110 H B 42 110 H B 43 111 H B 44 110; B 45 2D 055 , A 47 17 4D 17 #71; G 48 110 H B 48 111 H B 49 11 H B 49 11 I L 40 115 J J 40 115 J J 41 11 H B 41 11 H B 42 11 H B 43 11 I L 44 11 04 #65; A 41 101 A																				
2 2 002 STX (start of text) 3 3 003 STX (end of text) 4 4 004 EOT (end of transmission) 5 5 005 ENQ (enquiry) 6 6 006 ACK (acknowledge) 7 7 007 BEL (bell) 8 8 010 BS (backspace) 9 9 011 TAB (horizontal tab) 10 A 012 LF (ML line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 23 17 027 ETB (end of trans. block) 25 18 033 ESC (escape) 26 10 036 RS (group separator) 27 18 031 EM (end of medium) 28 20 042 e#34; " 36 42 044 e#36; # 37 25 045 e#35; # 36 24 044 e#36; # 37 25 045 e#37; * 38 26 046 e#38; 6 39 44 104 e#68; D 100 64 144 e#100; d 66 44 104 e#68; D 100 65 145 e#100; d 69 45 105 e#99; E 101 65 145 e#101; e 70 46 106 e#70; F 102 66 146 e#102; f 71 47 107 e#71; G 103 67 147 e#103; g 71 47 107 e#71; G 103 67 147 e#103; g 72 48 110 e#72; H 104 68 150 e#104; h 73 49 111 e#73; I 74 4A 112 e#74; J 75 4B 113 e#73; H 106 6A 152 e#106; j 74 4A 112 e#74; J 75 4B 113 e#73; H 106 6A 152 e#106; j 77 4D 115 e#77; M 109 6D 155 e#109; m 14 5 D 055 e#44; ' 15 40 114 e#76; L 108 6C 154 e#101; n 109 6D 155 e#109; m 14 72 F 057 e#47; ' 17 4D 115 e#77; M 109 6D 155 e#109; m 14 75 057 e#47; ' 17 4D 115 e#77; M 109 6D 155 e#109; m 14 76 4C 114 e#76; L 108 6C 154 e#101; n 109 6D 155 e#101; n 119 11 021 DC1 (device control 1) 18 10 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 55 37 067 e#55; 7 38 57 127 e#79; W 119 77 167 e#112; V 22 16 036 RS (group separator) 25 18 033 ESC (escape) 26 18 036 RS (file separator) 26 18 036 RS (group separator) 37 05 EE 06 046 E#306; V 118 76 CF 176 e#112; V 129 1D 035 GS (group separator) 38 12 044 e#36; L 17 17 17 17 12	_																			a
3 3 003 ETX (end of text) 4 4 004 EOT (end of transmission) 5 5 005 ENQ (enquiry) 6 6 006 ACK (acknowledge) 7 7 007 BEL (bell) 8 8 010 BS (backspace) 9 9 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 S0 (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 10 02 DLE (data link escape) 18 10 12 DC1 (device control 1) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 25 19 031 EM (end of medium) 25 10 036 CR (group separator) 26 16 036 RS (group separator) 27 10 03 GS (group separator) 28 10 036 CR (cercord separator) 28 10 036 CR (group separator) 29 10 036 CR (group separator) 20 12 02 057 (group separator) 20 16 036 RS (group separator) 20 16 057 SR (group separator) 20 16 058 CR (group separator) 21 10 05 GS (group separator) 22 10 05 GS (group separator) 23 10 05 GS (group separator) 24 12 10 058 6496; C																				
4 4 004 EOT (end of transmission) 5 5 005 ENQ (enquiry) 37 25 045 6#37; 6 68 44 104 6#68; D 100 64 144 6#100; d 6 6 006 ACK (acknowledge) 8 8 010 BS (beckspace) 9 9 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 1 1 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SVN (synchronous idle) 23 17 027 ETB (end of trans. block) 25 19 031 EM (end of medium) 25 10 035 C8 (group separator) 26 16 036 RS (group separator) 27 16 036 RS (group separator) 28 10 036 RS (group separator) 29 10 036 RS (group separator) 20 12 6 036 RS (group separator) 20 12 6 036 RS (group separator) 26 16 036 ACK (acknowledge) 27 047 6#39; ' 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 46 106 6#70; F 102 66 146 6#102; f 70 47 11; G 103 67 147 6#103; G 70 48 110 6#72; H 104 6# 150 6#104; h 105 6#105; j 105 6#103; g 105 6#116; j 105 6#117; j 106 6A 152 6#106; j 106 6A 152 6#10																				c
5 5 005 ENQ (enquiry) 6 6 006 ACK (acknowledge) 7 7 007 BEL (bell) 8 8 010 B3 (backspace) 9 9 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 S0 (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 25 19 032 SUB (substitute) 26 1A 032 SUB (substitute) 27 10 60 60 RS (group separator) 28 1C 034 FS (file separator) 29 1D 035 GS (group separator) 20 1E (dato feed, new line) 37 25 045 c#33; 5 (483; 6) 38 26 046 c#338; 6 (70 46 60 6 470; F (102 66 146 c#102; F																				
38 26 046 6#38; 6																				
7 7 007 BEL (bell) 8 8 010 BS (backspace) 9 9 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 39 27 047 6#39; ' 40 28 050 6#40; (72 48 110 6#72; H 104 68 150 6#105; h 41 29 051 6#41;) 41 29 051 6#41;) 42 2A 052 6#42; * 44 2C 054 6#42; * 45 2B 053 6#43; + 46 2E 056 6#46; . 47 4A 112 6#74; J 106 6A 152 6#106; j 48 2B 053 6#43; + 49 113 6#75; K 107 6B 153 6#107; k 40 114 6#76; L 106 6A 152 6#106; j 41 29 051 6#42; † 42 C 054 6#42; * 44 2C 054 6#44; , 45 2B 055 6#45; - 46 2E 056 6#46; . 47 4A 112 6#74; J 106 6A 152 6#105; j 48 113 6#75; K 107 6B 153 6#107; k 49 2B 055 6#45; - 40 2B 055 6#45; - 41 20 054 6#44; 1 45 2B 055 6#45; - 46 2E 056 6#46; . 47 4A 112 6#74; J 106 6A 152 6#106; j 48 10 6#74; J 106 6A 152 6#106; j 49 110 6B 153 6#107; k 40 114 6#76; L 106 6A 152 6#106; j 40 2B 055 6#45; - 47 4D 115 6#77; M 109 6D 155 6#109; m 46 2E 056 6#46; . 47 4B 113 6#75; M 109 6D 155 6#109; m 47 4B 112 6#74; J 106 6A 152 6#106; j 47 4B 112 6#74; J 106 6A 152 6#106; j 48 2B 053 6#43; - 48 2B 053 6#43; - 49 113 6#75; M 109 6D 155 6#105; j 49 31 061 6#49; l 115 6#75; M 110 6E 156 6#110; n 47 2F 057 6#47; J 77 4D 115 6#75; M 110 6E 156 6#110; n 47 2F 057 6#47; J 78 110 6 6B 153 6#109; n 48 2B 053 6#46; . 49 2B 053 6#46; . 40 2B 053 6#46; . 41 29 051 6#42; T 74 4A 112 6#74; J 106 6A 152 6#106; j 49 2B 112 6#78; N 110 6E 156 6#109; n 46 2E 056 6#46; . 47 44 Al 12 6#74; J 106 6A 152 6#106; j 49 2B 112 6#107; M 110 6E 156 6#108; l 40 2B 053 6#46; . 47 4A 112 6#74; J 106 6A 152 6#109; l 40 2B 053 6#48; l 40 2B 053 6#46; l 40 2B 053 6#46; l 40 2B 053 6#46; l 40 2B 053						38														
8 8 010 BS (backspace) 9 9 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 40 28 050 6#40; (72 48 110 6#72; H 105 69 151 6#105; 1 73 49 111 6#73; I 105 69 151 6#105; 1 74 44 112 6#74; J 106 6A 152 6#104; h 42 2A 052 6#43; + 75 4B 113 6#73; K 107 6B 153 6#107; k 42 2A 052 6#44; , 76 4C 114 6#76; L 108 6C 154 6#108; l 45 2D 055 6#45; - 78 4E 116 6#78; N 109 6D 155 6#108; l 109 6D 155											71	47	107	G	G					
9 9 011 TAB (horizontal tab) 10 A 012 LF (NL line feed, new line) 11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 3) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 10 036 RS (group separator) 27 18 033 ESC (escape) 28 10 036 RS (group separator) 29 10 036 RS (group separator) 30 1E 036 RS (group separator) 41 29 051 6#41;) 42 2A 052 6#42; * 43 2B 053 6#43; + 44 2C 054 6#43; + 45 2D 055 6#45; - 46 2C 056 6#46; . 47 4D 115 6#77; M 4D 115 6#77;	8					40	28	050	a#40;	(72	48	110	6#72;	H	104	68	150	a#104;	h
11 B 013 VT (vertical tab) 12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 25 19 032 SUB (substitute) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 31 22 C 014 FF (NP form feed, new page) 44 2C 054 c#44; 75 4B 113 c#75; K 107 6B 153 c#107; k 108 6C 154 c#108; l 109 6D 155 c#109; m 109 6D 15 c#109; m 109 6D		9	011	TAB		41	29	051	a#41;	j 🐘	73	49	111	a#73;	I	105	69	151	6#105;	i
12 C 014 FF (NP form feed, new page) 13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 23 17 027 ETB (end of medium) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 38 050 CAP (escape) 48 2C 054 6#45; - 46 2E 056 6#46; . 47 2F 057 6#47; / 48 30 060 6#48; 0 48 30 060 6#48; 0 48 30 060 6#48; 0 49 31 061 6#49; 1 49 31 061 6#49; 1 49 31 061 6#49; 1 40 30 060 6#48; 0 49 31 061 6#49; 1 40 31 061 6#49; 1 41 06 155 6#110; n 47 2F 057 6#47; / 48 116 6#78; N 49 110 6E 156 6#110; n 49 11 06 E 156 6#110; n 49 11 16 F 157 4#111; 0 40 11 6F 157 4#111; 0 40 11 6F 157 4#111; 0 41 11 06 E 156 6#111; 0 41 11 06 E 156 6#110; n 41 11 06 E 156 6#111; 0 41 11 06 E 156 6#11; 0 41 11 06 E 156 6#11; 0	10	A	012	LF	(NL line feed, new line)	42	2A	052	6#42;	*	74	4A	112	6#74;	J	106	6A	152	a#106;	j
13 D 015 CR (carriage return) 14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 23 17 027 ETB (end of trans. block) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 25 19 032 SUB (substitute) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 45 2D 055 6#45; - 46 2E 056 6#46; . 47 2F 057 6#47; / 48 2D 055 6#46; . 48 446; . 47 2F 057 6#47; / 48 2D 055 6#46; . 48 42 116 6#78; N 110 6E 156 6#110; n 10 111 6F 157 6#111; o 111 6F 157 6#112; o 111 6F 157 6#	11	В	013	VT	(vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
14 E 016 SO (shift out) 15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 25 18 033 ESC (escape) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 46 2E 056 . . 47 2F 057 / / 48 30 060 0 0 48 30 060 0 0 48 30 060 0 0 48 30 060 0 0 48 30 060 0 0 49 31 061 1 1 48 30 060 0 0 49 31 061 1 1 48 30 060 0 0 49 31 061 1 1 48 30 060 0 0 49 31 061 1 1 48 30 060 0 0 49 31 061 1 1 48 30 060 0 0 49 31 061 1 1 48 30 060 0 0 48 30 120 P P 112 70 160 p P 113 71 161 p P 113 71 161 p P 114 72 162 r r 12 14 72 162 r r 13 3 063 3 3 83 53 123 S S 115 73 163 s t 116 74 164 p P 117 75 165 u u 118 76 166 v v 119 77 167 w w 119	12	С	014	FF	(NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L					
15 F 017 SI (shift in) 16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 47 2F 057 6#47; / 48 30 060 6#48; 0 48 30 060 6#51; 2 48 25 2 122 6#82; R 114 72 162 6#114; L 48 30 060 6#51; 3 48 54 124 6#84; T 48 30 060 6#52; 4 48 454 124 6#84; T 48 54 124 6#84	13	D	015	CR	(carriage return)	45	2D	055	6#45;	-	77	4D	115	6#77;	М	109	6D	155	6#109;	\mathbf{m}
16 10 020 DLE (data link escape) 17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 25 19 031 EM (end of medium) 25 19 032 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 10 046 CAS (record separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 48 30 060 c#48; 0 49 31 061 c#49; 1 50 32 062 c#50; 2 50 34 064 c#52; 4 50 44 54 124 c#84; T 51 16 74 164 c#116; t 51 37 163 c#11; u 52 57 167 c#116; t 52 34 064 c#52; 4 52 34 064 c#52	14	E	016	S0	(shift out)						78	4E	116	6#78;	N	110	6E	156	6#110;	\mathbf{n}
17 11 021 DC1 (device control 1) 18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 49 31 061 6#49; 1 50 32 062 6#50; 2 50 32 063 6#51; 3 50 35 065 6#53; 5 50 55 125 6#85; U 50 74 064 6#52; 4 50 74 064 6#60; 4 50 74 064 6#52; 4 50 74 064 6#60; 4 50 74 064 6#52; 4 50 74 064 6#60; 4 50 74 064 6#52; 4 50 74 064 6#60; 4 50 74	15	F	017	SI	(shift in)						79	4F	117	O	0	111	6F	157	6#111;	0
18 12 022 DC2 (device control 2) 19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 30 2 062 2 2 50 32 062 2 2 51 33 063 3 3 52 34 064 4 4 53 35 065 5 5 54 36 066 6 6 55 37 067 7 7 56 38 070 7 7 57 39 071 9 9 58 3A 072 : : 59 3B 073 ; ; 60 3C 074 < < 70 92 5C 134 \ \ 124 7C 174 \ 125 7D 175 } \ 30 126 076 > > 71 16 74 164 t t	16	10	020	DLE	(data link escape)	48					80					112	70	160	p	p
19 13 023 DC3 (device control 3) 20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 31 023 DC3 (device control 3) 51 33 063 3 3 52 34 064 4 4 64 54 124 T T 116 74 164 t t 65 35 065 5 5 65 55 125 U U 117 75 165 u u 66 56 126 V V 118 76 166 v V 68 58 126 V V 119 77 167 w W 68 58 130 X X 120 78 170 x X 69 59 131 Y Y 121 79 171 y Y 69 58 132 Z Z 122 7A 172 z Z 60 3C 074 < < 60 3C 074 <	17	11	021	DC1													-			
20 14 024 DC4 (device control 4) 21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 32 34 064 c#52; 4 34 064 c#52; 4 35 065 c#53; 5 37 067 c#53; 5 38 070 c#54; 6 38 070 c#55; 7 39 071 c#57; 9 30 071 c#57; 9 30 071 c#57; 9 30 072 c#58; 1 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 31 064 c#52; 4 34 064 c#52; 4 34 064 c#52; 4 34 064 c#52; 4 35 124 c#84; T 116 74 164 c#116; t 117 75 165 c#117; u 118 76 166 c#118; v 119 77 167 c#129; u 119 77 167 c#119; u 119 77 167 c#129; u 119 77 167 c#119; u 119 77 167 c#19; u 119																				
21 15 025 NAK (negative acknowledge) 22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 29 1D 035 GS (group separator) 30 1E 036 RS (record separator) 31 35 065 4#53; 5 35 065 4#53; 5 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#54; 6 36 066 4#55; 7 37 067 4#55; 7 38 070 4#56; 8 38 070 4#56; 8 38 070 4#57; 9 39 071 4#57; 9 30																				
22 16 026 SYN (synchronous idle) 23 17 027 ETB (end of trans. block) 24 18 030 CAN (cancel) 25 19 031 EM (end of medium) 26 1A 032 SUB (substitute) 27 1B 033 ESC (escape) 28 1C 034 FS (file separator) 30 1E 036 RS (record separator) 30 1E 036 RS (record separator) 34 36 066 4#54; 6 54 36 066 4#54; 6 55 37 067 4#55; 7 87 57 127 4#87; W 119 77 167 4#119; W 120 78 170 4#120; X 121 79 171 4#121; Y 122 7A 172 4#122; Z 123 7B 173 4#123; { 124 7C 174 4#124; 125 7D 175 4#125; } 126 7E 176 4#126; ~																				
23 17 027 ETB (end of trans. block) 55 37 067 7 7 87 57 127 W W 119 77 167 w W 24 18 030 CAN (cancel) 56 38 070 8 8 8 58 130 X X 120 78 170 x X 25 19 031 EM (end of medium) 57 39 071 9 9 89 59 131 Y Y 121 79 171 y Y 26 1A 032 SUB (substitute) 58 3A 072 : : 90 5A 132 Z Z 122 7A 172 z Z 27 1B 033 ESC (escape) 59 3B 073 ; ; 91 5B 133 [[123 7B 173 { { 28 1C 034 FS (file separator) 60 3C 074 < < 92 5C 134 \ \ 124 7C 174 93 5D 135]] 125 7D 175 } } 30 1E 036 RS (record separator) 62 3E 076 > > 94 5E 136 ^ ^ 126 7E 176 ~ ~																				
24 18 030 CAN (cancel) 56 38 070 8 8 8 58 130 X X 120 78 170 x X 25 19 031 EM (end of medium) 57 39 071 9 9 89 59 131 Y Y 121 79 171 y Y 26 1A 032 SUB (substitute) 58 3A 072 : 90 5A 132 Z Z 122 7A 172 z Z 27 1B 033 ESC (escape) 59 3B 073 ; 91 5B 133 [[123 7B 173 { { 60 3C 074 < < 92 5C 134 \ \ 124 7C 174 93 5D 135]] 125 7D 175 } } 30 1E 036 RS (record separator) 62 3E 076 > > 94 5E 136 ^ ^ 126 7E 176 ~ ~																				
25 19 031 EM (end of medium) 57 39 071 4#57; 9 89 59 131 4#89; Y 121 79 171 4#121; Y 26 1A 032 SUB (substitute) 58 3A 072 4#58; 90 5A 132 4#90; Z 122 7A 172 4#122; Z 18 033 ESC (escape) 59 3B 073 4#59; 91 5B 133 4#91; [123 7B 173 4#123; { 60 3C 074 4#60; < 92 5C 134 4#92; \ 124 7C 174 4#124; 93 1D 035 GS (group separator) 61 3D 075 4#61; 93 5D 135 4#93;] 125 7D 175 4#125; } 30 1E 036 RS (record separator) 62 3E 076 4#62; > 94 5E 136 4#94; 126 7E 176 4#126; ~																				
26 1A 032 SUB (substitute) 58 3A 072 c#58;: 90 5A 132 c#90; Z 122 7A 172 c#122; Z 27 1B 033 ESC (escape) 59 3B 073 c#59; 91 5B 133 c#91; [123 7B 173 c#123; { 28 1C 034 FS (file separator) 60 3C 074 c#60; < 92 5C 134 c#92; \ 124 7C 174 c#124; 125 7D 175 c#125; \ 126 7E 176 c#126; ~																				
27 1B 033 ESC (escape) 59 3B 073 ; 91 5B 133 [[123 7B 173 { { 28 1C 034 FS (file separator) 60 3C 074 < < 92 5C 134 \ \ 29 1D 035 GS (group separator) 61 3D 075 = = 93 5D 135]] 125 7D 175 } } 30 1E 036 RS (record separator) 62 3E 076 > > 94 5E 136 ^ ^ 126 7E 176 ~ ~																				
28 1C 034 FS (file separator) 60 3C 074 6#60; < 92 5C 134 6#92; \ 124 7C 174 6#124; 29 1D 035 GS (group separator) 61 3D 075 6#61; = 93 5D 135 6#93;] 125 7D 175 6#125; } 30 1E 036 RS (record separator) 62 3E 076 6#62; > 94 5E 136 6#94; ^ 126 7E 176 6#126; ~																				
29 1D 035 GS (group separator) 61 3D 075 = = 93 5D 135]] 125 7D 175 } } 30 1E 036 RS (record separator) 62 3E 076 > > 94 5E 136 ^ ^ 126 7E 176 ~ ~										•										
30 1E 036 RS (record separator) 62 3E 076 > > 94 5E 136 ^ ^ 126 7E 176 ~ ~																				
															_					
31 1F 037 US (unit separator) 63 3F 077 ? ? 95 5F 137 _ _ 127 7F 177 DEL																				
	31	1F	037	US	(unit separator)	63	3F	077	6#63;	?	95	5F	137	_	_	127	7F	177		DEL

Caratteri di controllo

```
CR (13) Carriage Return
LF, NL (10) New Line, Line Feed
FF, NP (12) New Page, Form Feed
   HT (9) Horizontal Tab
   VT (11) Vertical Tab
   NUL (0) Null
   BEL (7) Bell
   EOT (4) End-Of-Transmission
```

UNICODE e UTF-8

- Unicode utilizza 21 bit per carattere ed esprime tutti i caratteri di tutte le lingue del mondo (più di un milione), oltre agli emoji 🥳.
- È il codice usato per rappresentare i caratteri in Python
- UTF-8 è la codifica di Unicode su file più usata:
 - 1 byte per caratteri US-ASCII (MSB=0)
 - 2 byte per caratteri Latini con simboli diacritici, Greco, Cirillico, Armeno, Ebraico, Arabo, Siriano e Maldiviano
 - 3 byte per altre lingue di uso comune
 - 4 byte per caratteri rarissimi



https://home.unicode.org/
https://unicode-table.com/it/

Codifiche o formati di testo/stampa

- Non confondere il formato di un file word, con il codice ASCII!!
- Un testo può essere memorizzato in due formati
 - Formattato: sono memorizzate sequenze di byte che definiscono l'aspetto del testo (e.g., font, spaziatura)
 - Non formattato: sono memorizzati unicamente i caratteri che compongono il testo

Codifiche audio, video, ...

- Molto più articolate, ma basate sul solito principio di associazione oggetti <-> codici
 - Per es: I colori sono codificati su 8 bit per canale (R,G,B), quindi fino a 256 sfumature di colore per canale

Oggetto di corsi più avanzati...