# Rappresentazioni di testo

Fondamenti di informatica

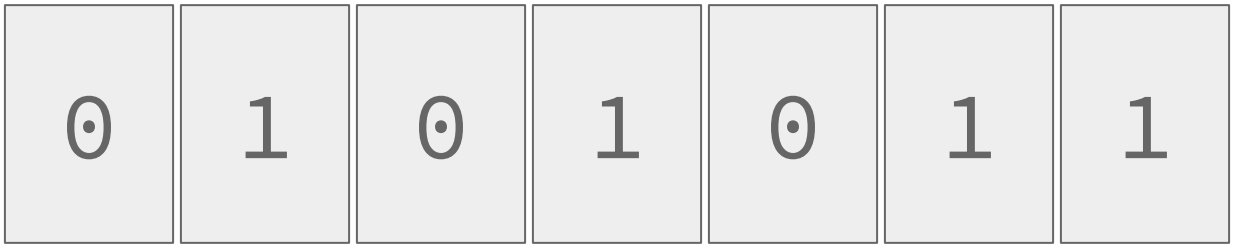
# Previously on "Fondamenti di informatica"

- Numeri in formato binario (e altre basi)
- Conversione di base
- Numeri razionali
- Numeri negativi

Ci sono solo 10 tipi di persone: quelle che conoscono la numerazione binaria e quelle che non la capiscono

# ASCII

7 bit di codifica



# ASCII

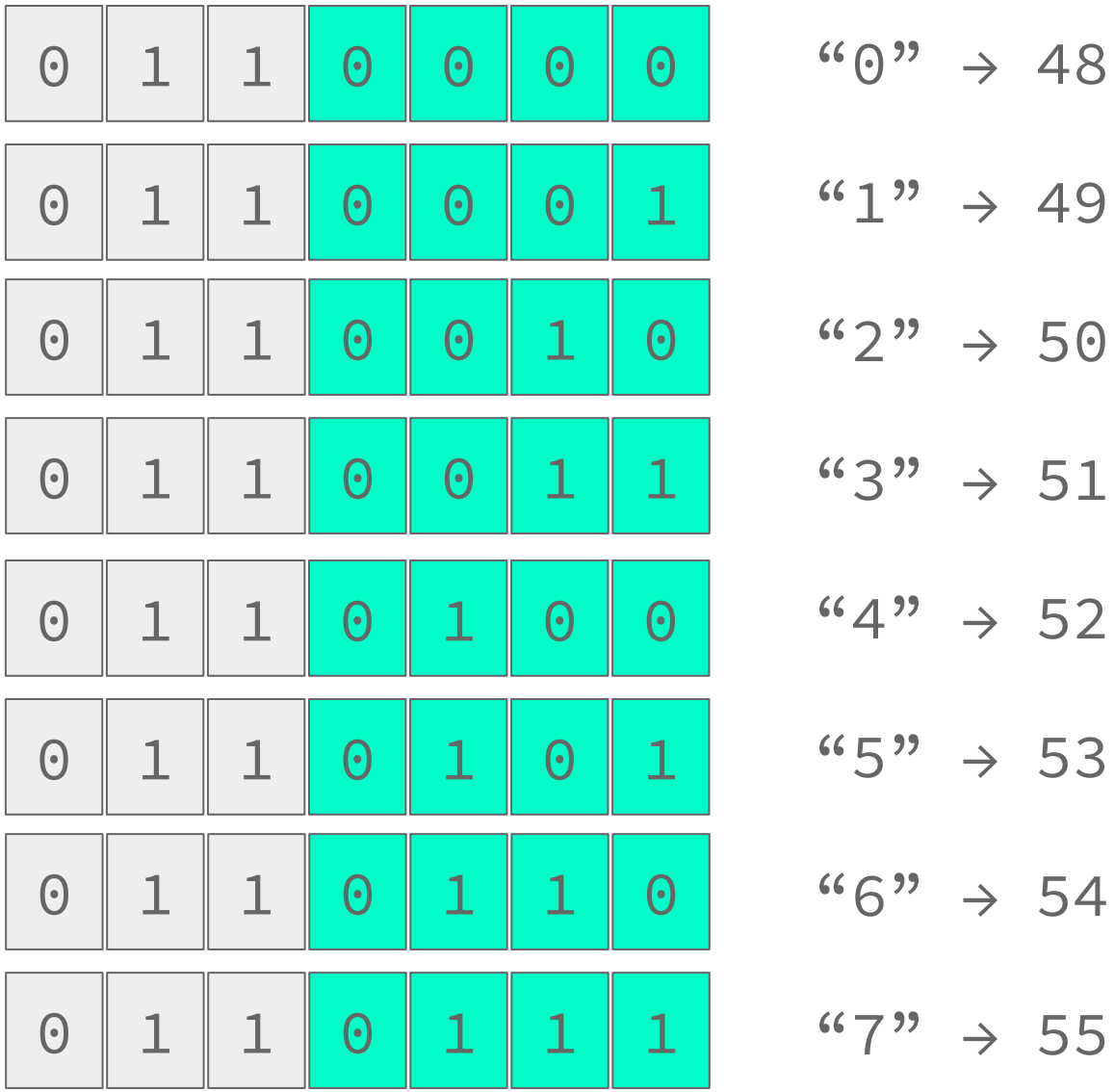
## ASCII Table

| Dec | Hex | Oct | Char | Dec | Hex | Oct | Char | Dec | Hex | Oct | Char | Dec | Hex | Oct | Char |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | | 32 | 20 | 40 | [space] | 64 | 40 | 100 | @ | 96 | 60 | 140 | ` |
| 1 | 1 | 1 | | 33 | 21 | 41 | ! | 65 | 41 | 101 | A | 97 | 61 | 141 | a |
| 2 | 2 | 2 | | 34 | 22 | 42 | " | 66 | 42 | 102 | B | 98 | 62 | 142 | b |
| 3 | 3 | 3 | | 35 | 23 | 43 | # | 67 | 43 | 103 | C | 99 | 63 | 143 | c |
| 4 | 4 | 4 | | 36 | 24 | 44 | $ | 68 | 44 | 104 | D | 100 | 64 | 144 | d |
| 5 | 5 | 5 | | 37 | 25 | 45 | % | 69 | 45 | 105 | E | 101 | 65 | 145 | e |
| 6 | 6 | 6 | | 38 | 26 | 46 | & | 70 | 46 | 106 | F | 102 | 66 | 146 | f |
| 7 | 7 | 7 | | 39 | 27 | 47 | ' | 71 | 47 | 107 | G | 103 | 67 | 147 | g |
| 8 | 8 | 10 | | 40 | 28 | 50 | ( | 72 | 48 | 110 | H | 104 | 68 | 150 | h |
| 9 | 9 | 11 | | 41 | 29 | 51 | ) | 73 | 49 | 111 | I | 105 | 69 | 151 | i |
| 10 | A | 12 | | 42 | 2A | 52 | * | 74 | 4A | 112 | J | 106 | 6A | 152 | j |
| 11 | B | 13 | | 43 | 2B | 53 | + | 75 | 4B | 113 | K | 107 | 6B | 153 | k |
| 12 | C | 14 | | 44 | 2C | 54 | , | 76 | 4C | 114 | L | 108 | 6C | 154 | l |
| 13 | D | 15 | | 45 | 2D | 55 | - | 77 | 4D | 115 | M | 109 | 6D | 155 | m |
| 14 | E | 16 | | 46 | 2E | 56 | . | 78 | 4E | 116 | N | 110 | 6E | 156 | n |
| 15 | F | 17 | | 47 | 2F | 57 | / | 79 | 4F | 117 | O | 111 | 6F | 157 | o |
| 16 | 10 | 20 | | 48 | 30 | 60 | 0 | 80 | 50 | 120 | P | 112 | 70 | 160 | p |
| 17 | 11 | 21 | | 49 | 31 | 61 | 1 | 81 | 51 | 121 | Q | 113 | 71 | 161 | q |
| 18 | 12 | 22 | | 50 | 32 | 62 | 2 | 82 | 52 | 122 | R | 114 | 72 | 162 | r |
| 19 | 13 | 23 | | 51 | 33 | 63 | 3 | 83 | 53 | 123 | S | 115 | 73 | 163 | s |
| 20 | 14 | 24 | | 52 | 34 | 64 | 4 | 84 | 54 | 124 | T | 116 | 74 | 164 | t |
| 21 | 15 | 25 | | 53 | 35 | 65 | 5 | 85 | 55 | 125 | U | 117 | 75 | 165 | u |
| 22 | 16 | 26 | | 54 | 36 | 66 | 6 | 86 | 56 | 126 | V | 118 | 76 | 166 | v |
| 23 | 17 | 27 | | 55 | 37 | 67 | 7 | 87 | 57 | 127 | W | 119 | 77 | 167 | w |
| 24 | 18 | 30 | | 56 | 38 | 70 | 8 | 88 | 58 | 130 | X | 120 | 78 | 170 | x |
| 25 | 19 | 31 | | 57 | 39 | 71 | 9 | 89 | 59 | 131 | Y | 121 | 79 | 171 | y |
| 26 | 1A | 32 | | 58 | 3A | 72 | : | 90 | 5A | 132 | Z | 122 | 7A | 172 | z |
| 27 | 1B | 33 | | 59 | 3B | 73 | ; | 91 | 5B | 133 | [ | 123 | 7B | 173 | { |
| 28 | 1C | 34 | | 60 | 3C | 74 | < | 92 | 5C | 134 | \ | 124 | 7C | 174 | \| |
| 29 | 1D | 35 | | 61 | 3D | 75 | = | 93 | 5D | 135 | ] | 125 | 7D | 175 | } |
| 30 | 1E | 36 | | 62 | 3E | 76 | > | 94 | 5E | 136 | ^ | 126 | 7E | 176 | ~ |
| 31 | 1F | 37 | | 63 | 3F | 77 | ? | 95 | 5F | 137 | _ | 127 | 7F | 177 | |

# ASCII - LETTERE

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | “A” → 65 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | “a” → 97 |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | “B” → 66 |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | “b” → 98 |
| 1 | 0 | 0 | 0 | 0 | 1 | 1 | “C” → 67 |
| 1 | 1 | 0 | 0 | 0 | 1 | 1 | “c” → 99 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | “D” → 68 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | “d” → 100 |

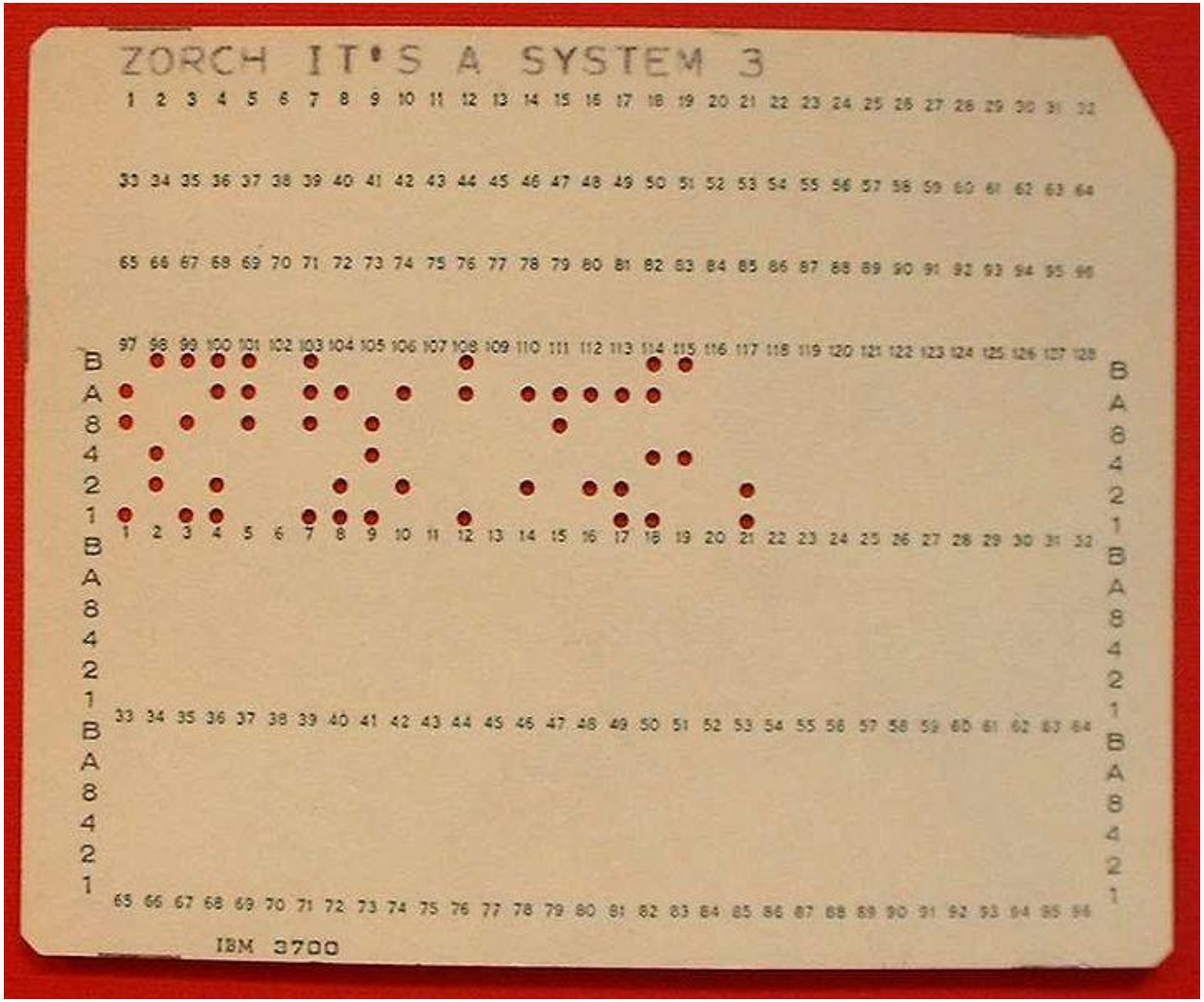
# ASCII - Numeri



# ASCII - Comandi

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | “ ” → $20_{16}$ |
| 0 | 0 | 0 | 1 | 0 | 1 | 1 | “\t” → $0B_{16}$ |
| 0 | 0 | 0 | 1 | 0 | 1 | 0 | “LF” → $0A_{16}$ |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | “CR” → $0D_{16}$ |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | “DEL”→ $7f_{16}$ |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | “NUL”→ 0 |

# ASCII



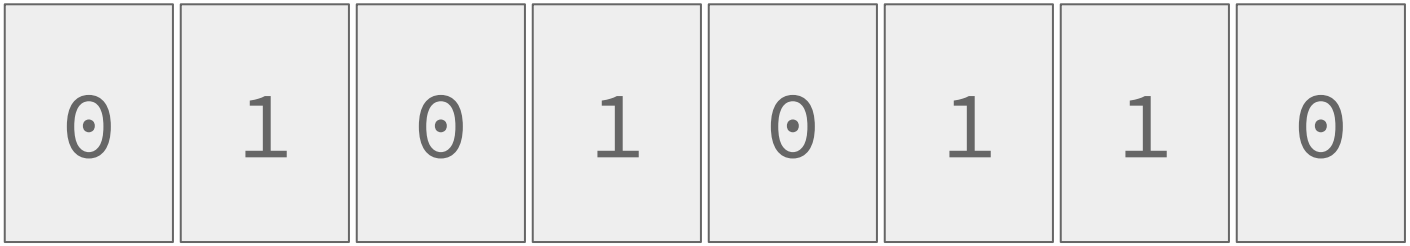
E LA LETTERA "È"?

# ASCII ESTESO

Nell'ASCII mancano molti caratteri latini!

Soluzione: **ASCII esteso**

8 bit di codifica



# Ascii esteso

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | <NUL> | 32 | <SPC> | 64 | @ | 96 | ` | 128 | Ä | 160 | † | 192 | ¿ | 224 | ‡ |
| 1 | <SOH> | 33 | ! | 65 | A | 97 | a | 129 | Å | 161 | ° | 193 | ¡ | 225 | · |
| 2 | <STX> | 34 | " | 66 | B | 98 | b | 130 | Ç | 162 | ¢ | 194 | ¬ | 226 | ‚ |
| 3 | <ETX> | 35 | # | 67 | C | 99 | c | 131 | É | 163 | £ | 195 | √ | 227 | „ |
| 4 | <EOT> | 36 | $ | 68 | D | 100 | d | 132 | Ñ | 164 | § | 196 | ƒ | 228 | ‰ |
| 5 | <ENQ> | 37 | % | 69 | E | 101 | e | 133 | Ö | 165 | • | 197 | ≈ | 229 | Â |
| 6 | <ACK> | 38 | & | 70 | F | 102 | f | 134 | Ü | 166 | ¶ | 198 | Δ | 230 | Ê |
| 7 | <BEL> | 39 | ' | 71 | G | 103 | g | 135 | á | 167 | ß | 199 | « | 231 | Á |
| 8 | <BS> | 40 | ( | 72 | H | 104 | h | 136 | à | 168 | ® | 200 | » | 232 | Ë |
| 9 | <TAB> | 41 | ) | 73 | I | 105 | i | 137 | â | 169 | © | 201 | … | 233 | È |
| 10 | <LF> | 42 | * | 74 | J | 106 | j | 138 | ä | 170 | ™ | 202 | | 234 | Í |
| 11 | <VT> | 43 | + | 75 | K | 107 | k | 139 | ã | 171 | ´ | 203 | À | 235 | Î |
| 12 | <FF> | 44 | , | 76 | L | 108 | l | 140 | å | 172 | ¨ | 204 | Ã | 236 | Ï |
| 13 | <CR> | 45 | - | 77 | M | 109 | m | 141 | ç | 173 | ≠ | 205 | Õ | 237 | Ì |
| 14 | <SO> | 46 | . | 78 | N | 110 | n | 142 | é | 174 | Æ | 206 | Œ | 238 | Ó |
| 15 | <SI> | 47 | / | 79 | O | 111 | o | 143 | è | 175 | Ø | 207 | œ | 239 | Ô |
| 16 | <DLE> | 48 | 0 | 80 | P | 112 | p | 144 | ê | 176 | ∞ | 208 | – | 240 | |
| 17 | <DC1> | 49 | 1 | 81 | Q | 113 | q | 145 | ë | 177 | ± | 209 | — | 241 | Ò |
| 18 | <DC2> | 50 | 2 | 82 | R | 114 | r | 146 | í | 178 | ≤ | 210 | “ | 242 | Ú |
| 19 | <DC3> | 51 | 3 | 83 | S | 115 | s | 147 | ì | 179 | ≥ | 211 | ” | 243 | Û |
| 20 | <DC4> | 52 | 4 | 84 | T | 116 | t | 148 | î | 180 | ¥ | 212 | ‘ | 244 | Ù |
| 21 | <NAK> | 53 | 5 | 85 | U | 117 | u | 149 | ï | 181 | µ | 213 | ’ | 245 | ı |
| 22 | <SYN | 54 | 6 | 86 | V | 118 | v | 150 | ñ | 182 | ∂ | 214 | ÷ | 246 | ˆ |
| 23 | <ETB> | 55 | 7 | 87 | W | 119 | w | 151 | ó | 183 | ∑ | 215 | ◊ | 247 | ˜ |
| 24 | <CAN> | 56 | 8 | 88 | X | 120 | x | 152 | ò | 184 | ∏ | 216 | ÿ | 248 | ¯ |
| 25 | <EM> | 57 | 9 | 89 | Y | 121 | y | 153 | ô | 185 | π | 217 | Ÿ | 249 | ˘ |
| 26 | <SUB> | 58 | : | 90 | Z | 122 | z | 154 | ö | 186 | ∫ | 218 | / | 250 | ˙ |
| 27 | <ESC> | 59 | ; | 91 | [ | 123 | { | 155 | õ | 187 | ª | 219 | € | 251 | ° |
| 28 | <FS> | 60 | < | 92 | \ | 124 | \| | 156 | ú | 188 | º | 220 | ‹ | 252 | ¸ |
| 29 | <GS> | 61 | = | 93 | ] | 125 | } | 157 | ù | 189 | Ω | 221 | › | 253 | ˝ |
| 30 | <RS> | 62 | > | 94 | ^ | 126 | ~ | 158 | û | 190 | æ | 222 | fi | 254 | ˛ |
| 31 | <US> | 63 | ? | 95 | _ | 127 | <DEL> | 159 | ü | 191 | ø | 223 | fl | 255 | ˇ |

# Ascii esteso

- Tante varianti diverse, una per area geografica
- “CP-1252” è il codice usato in europa
- Molte varianti per soddisfare i segni diacritici
- **Problema**: il simbolo dell’euro

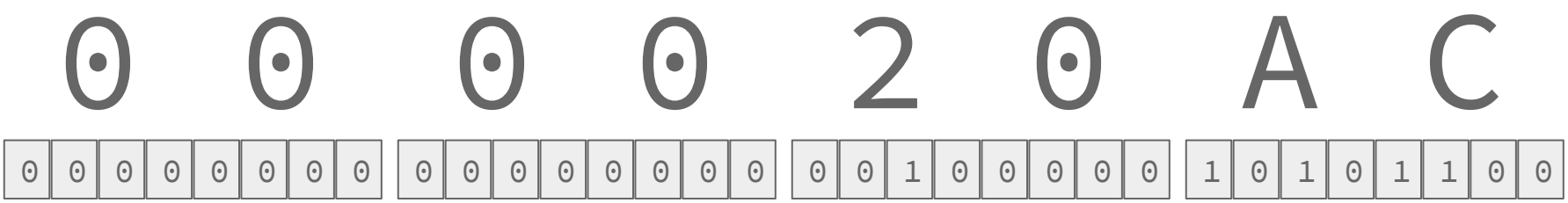
# Unicode


1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 130A 130B 130C 130D 130E 130F 1310 1311 1312 1313 1314
0
1
2
3
4
5
6
7
8

# UNICODE

Standard che associa ad ogni carattere di ogni lingua nota un numero

€ = U+20AC



# UTF-8

Unicode occupa occupa molto spazio per scrivere banali lettere

**UTF-8** è una rappresentazione più compatta:

- I caratteri più usati si scrivono con meno byte
- I caratteri meno usati si scrivono con più byte

# UTF-8

Se il primo bit è **0**, allora uso un solo byte

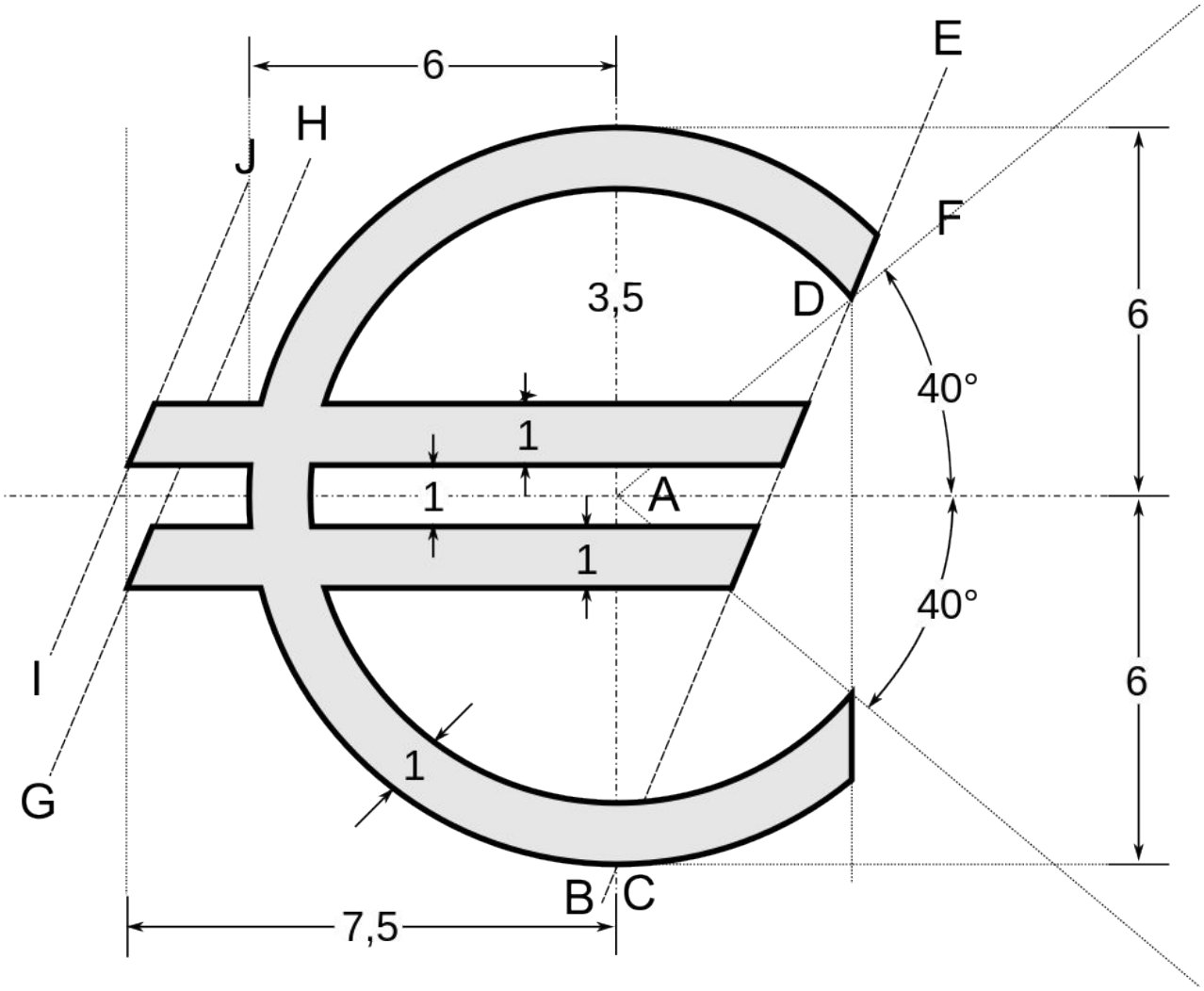
0XXXXXXX (7 bit utili)

Se il primo bit è **1**, i bit successivi mi dicono quanti ulteriori byte vanno letti

110XXXXX 10XXXXXX (11 bit utili)
1110XXXX 10XXXXXX 10XXXXXX (16 bit utili)
11110XXX 10XXXXXX 10XXXXXX 10XXXXXX (21 bit)

COMPATIBILE CON ASCII

# Font ttf



# "GRAZIE"

serif

sans serif

# MONOSPACED


78.76
+ 69.31
**148.07**