



Architectures des Ordinateurs

Chapitre 5

(Représentation des caractères)

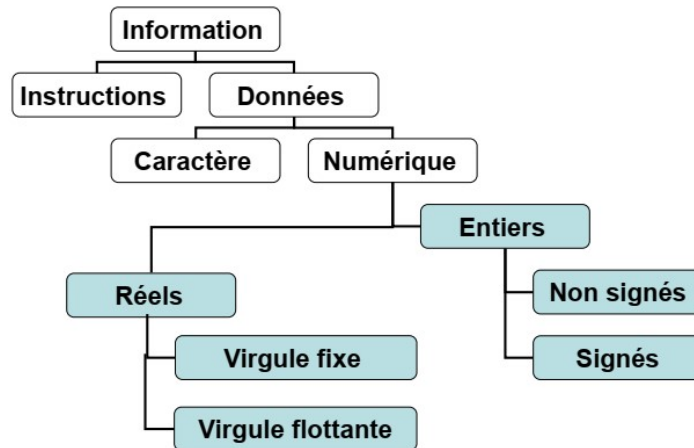
Mokrani Hocine
dr.mokrani@gmail.com

1

Représentation des caractères

2

Rappel des types de base



3

Caractères ASCII

▪ **L'ASCII** (American Standard Code for Information Interchange), créé en 1961, associe un nombre à chaque caractère.

▪ Les caractères sont codés sur 7 bits.

▪ Il y a 128 caractères différents (de 00 à 7F en hexadécimal).

▪ Contient les caractères nécessaires pour écrire en anglais.

▪ Les caractères de numéro 0 à 31 et le 127 ne sont pas affichables.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|-----|-----|----|---|---|---|---|-----|
| 0 | NUL | DLE | SP | 0 | @ | P | ` | p |
| 1 | SOH | DC1 | ! | 1 | A | Q | a | q |
| 2 | STX | DC2 | " | 2 | B | R | b | r |
| 3 | ETX | DC3 | # | 3 | C | S | c | s |
| 4 | EOT | DC4 | \$ | 4 | D | T | d | t |
| 5 | ENQ | NAK | % | 5 | E | U | e | u |
| 6 | ACK | SYN | & | 6 | F | V | f | v |
| 7 | BEL | ETB | ' | 7 | G | W | g | w |
| 8 | BS | CAN | (| 8 | H | X | h | x |
| 9 | HT | EM |) | 9 | I | Y | i | y |
| A | LF | SUB | * | : | J | Z | j | z |
| B | VT | ESC | + | ; | K | [| k | { |
| C | FF | FS | , | < | L | \ | l | |
| D | CR | GS | - | = | M |] | m | } |
| E | SO | RS | . | > | N | ^ | n | ~ |
| F | SI | US | / | ? | O | _ | o | DEL |

4

Limites de code ASCII

- Absence des caractères des langues étrangères à l'anglais.
- Les ordinateurs travaillant sur un multiple de huit bits (multiple d'un octet) depuis les années 1970. (Chaque caractère d'un texte en ASCII est souvent stocké dans un octet dont le 8^e bit est 0).
- Pourquoi l'utilisé alors?

=> Héritage des programmes qui fonctionne encore sur le code ASCII.

=> certains systèmes de messagerie électronique et de SMS fonctionnent avec des mots composés de seulement sept bits.

5

ASCII étendu

Il existe des extensions (sur 8 bits, de 80 à FF). Cependant, ces extensions ne sont pas standardisées.

| | | | | | | | | | | | | | | | |
|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|-----|---|
| 128 | Ç | 144 | É | 160 | á | 176 | ⌘ | 193 | ⌞ | 209 | ⌞ | 225 | ß | 241 | ± |
| 129 | à | 145 | æ | 161 | í | 177 | ⌘ | 194 | ⌞ | 210 | ⌞ | 226 | Γ | 242 | ≥ |
| 130 | é | 146 | Æ | 162 | ó | 178 | ⌘ | 195 | ⌞ | 211 | ⌞ | 227 | π | 243 | ≤ |
| 131 | â | 147 | ô | 163 | ú | 179 | ⌞ | 196 | — | 212 | ⌞ | 228 | Σ | 244 | ∫ |
| 132 | ä | 148 | ö | 164 | ñ | 180 | ⌞ | 197 | ⌞ | 213 | ⌞ | 229 | σ | 245 | ∫ |
| 133 | à | 149 | ò | 165 | Ñ | 181 | ⌞ | 198 | ⌞ | 214 | ⌞ | 230 | μ | 246 | ÷ |
| 134 | â | 150 | û | 166 | ª | 182 | ⌞ | 199 | ⌞ | 215 | ⌞ | 231 | τ | 247 | ≈ |
| 135 | ç | 151 | ù | 167 | º | 183 | ⌞ | 200 | ⌞ | 216 | ⌞ | 232 | Φ | 248 | ° |
| 136 | ê | 152 | — | 168 | ¿ | 184 | ⌞ | 201 | ⌞ | 217 | ⌞ | 233 | Ω | 249 | . |
| 137 | ë | 153 | Ö | 169 | — | 185 | ⌞ | 202 | ⌞ | 218 | ⌞ | 234 | Ω | 250 | . |
| 138 | è | 154 | Û | 170 | ¬ | 186 | ⌞ | 203 | ⌞ | 219 | ■ | 235 | δ | 251 | √ |
| 139 | ï | 156 | £ | 171 | ½ | 187 | ⌞ | 204 | ⌞ | 220 | ■ | 236 | ∞ | 252 | — |
| 140 | î | 157 | ¥ | 172 | ¾ | 188 | ⌞ | 205 | = | 221 | ■ | 237 | φ | 253 | ² |
| 141 | ì | 158 | — | 173 | ¡ | 189 | ⌞ | 206 | ⌞ | 222 | ■ | 238 | ε | 254 | ■ |
| 142 | Ä | 159 | ƒ | 174 | « | 190 | ⌞ | 207 | ⌞ | 223 | ■ | 239 | ∩ | 255 | |
| 143 | Å | 192 | Ł | 175 | » | 191 | ⌞ | 208 | ⌞ | 224 | α | 240 | ≡ | | |

6

Caractères ISO-8859-1

- souvent appelée **Latin-1** ou **Europe occidentale**.
- reprend le codage des caractères imprimables d'US-ASCII.
- cette norme était utilisée par de nombreux système d'exploitation, dont Unix et Windows.

| ISO-8859-1 | | | | | | | | | | | | | | | | |
|------------|------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|
| | x0 | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | xA | xB | xC | xD | xE | xF |
| 0x | NUL | SOH | STX | ETX | EOT | ENO | ACK | BEL | BS | HT | LF | VT | FF | CR | SO | SI |
| 1x | DLE | DC1 | DC2 | DC3 | DC4 | NAK | SYN | ETB | CAN | EM | SUB | ESC | FS | GS | RS | US |
| 2x | SP | ! | " | # | \$ | % | & | ' | (|) | * | + | , | - | . | / |
| 3x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | : | ; | < | = | > | ? |
| 4x | @ | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
| 5x | P | Q | R | S | T | U | V | W | X | Y | Z | [| \ |] | ^ | _ |
| 6x | ` | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o |
| 7x | p | q | r | s | t | u | v | w | x | y | z | { | | } | ~ | DEL |
| 8x | PAD | HOP | BPH | NBH | IND | NEL | SSA | ESA | HTS | HTJ | VTS | PLD | PLU | RI | SS2 | SS3 |
| 9x | DCS | PU1 | PU2 | STS | CCH | MW | SPA | EPA | SOS | SGCI | SCI | CSI | ST | OSC | PM | APC |
| Ax | NBSP | ı | ç | £ | ¤ | ¥ | ¦ | § | ¨ | © | ª | « | ¬ | ® | ¯ | ° |
| Bx | ± | ² | ³ | ´ | µ | ¶ | · | ¸ | ¹ | º | » | ¼ | ½ | ¾ | ¿ | |
| Cx | À | Á | Â | Ã | Ä | Å | Æ | Ç | È | É | Ê | Ë | Ì | Í | Î | Ï |
| Dx | Ð | Ñ | Ò | Ó | Ô | Õ | Ö | × | Ø | Ù | Ú | Û | Ü | Ý | Þ | ß |
| Ex | à | á | â | ã | ä | å | æ | ç | è | é | ê | ë | ì | í | î | ï |
| Fx | ø | ñ | ò | ó | ô | õ | ö | ÷ | ø | ù | ú | û | ü | ý | þ | ÿ |

7

Quelques limites de ISO-8859-1

- Elle ignore quelques caractères européens:
 - Le symbole de l'euro € (qui n'existait pas encore lorsque ce jeu a été normalisé)
 - certaines des lettres normalement nécessaires à certaines langues théoriquement couvertes:
 - les lettres « œ », « Œ » et « Ÿ » en français.
 - les lettres « š » et « Š » en tchèque.
- La nécessité d'utiliser plusieurs jeux de caractères codés pour couvrir plusieurs alphabets. (utilisation des meta-données).

8

Caractère sur 16 bits (Unicode)

- Unicode propose de coder les caractères sur 16 bits (65536 valeurs possibles).
- Il permet de coder notamment les alphabets latins, cyrilliques, les caractères chinois, indiens et arabes.

| Char | Code | Char | Code | Char | Code | Char | Code | Char | Code | Char | Code | Char | Code |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| À | 100 | Đ | 110 | È | 118 | Ķ | 136 | Ń | 143 | Ó | d3 | Š | 15a |
| ā | 101 | đ | 111 | ē | 119 | ķ | 137 | ń | 144 | ó | f3 | š | 15b |
| Ā | 102 | Ď | 10e | Ě | 11a | Ĺ | 139 | Ň | 145 | Œ | 152 | Š | 160 |
| ǎ | 103 | ď | 10f | ě | 11b | ĺ | 13a | ň | 146 | œ | 153 | š | 161 |
| Ȧ | 104 | Ě | 112 | | | Ľ | 13b | Ñ | 147 | ƒ | 155 | Ț | 162 |
| ȧ | 105 | ē | 113 | Ĝ | 122 | ļ | 13c | ñ | 148 | ƣ | 156 | ț | 163 |
| Ĉ | 106 | ē | 115 | ĝ | 123 | Ļ | 13d | ō | 14c | ƣ | 157 | | z |
| ĉ | 107 | Ē | 116 | Ī | 12a | ļ | 13e | o | 14d | Ř | 158 | ť | 165 |
| Ċ | 10c | ē | 117 | ī | 12b | | | ō | 150 | ř | 159 | | z |
| č | 10d | | | ı | 12e | Ĺ | 141 | ô | 151 | Ş | 15e | | ž |
| | | | | ı | 12f | ı | 142 | | | ş | 15f | | ž |

- Limites: il existe différentes représentations binaires dont l'interprétation peut être très différente si la représentation effective est mal déterminée.

9

Conclusion

Nous savons maintenant,
Comment on peut représenter les éléments en binaire.

Ce qu'il reste à savoir?

Représentation des réels.

Représentation du son, vidéo ...

10