# Számok és karakterek ábrázolása

Wettl Ferenc

2006. szeptember 14.

Legfeljebb $n$-bites számokkal akarunk számolni.

$$x = \begin{cases} x & \text{ha } x \text{ nem negatív,} \\ 2^n - |x| & \text{ha } x \text{ negatív.} \end{cases}$$

A $2^n - |x|$ kiszámítása $n$-bites szavak közti bitműveletekkel: $|x|$ bitenkénti komplemense + 1, ugyanis
$2^n - |x| = (2^n - 1) - |x| + 1 = 11\ldots1_2 - |x| + 1$. Mivel
$|x| = 2^n - (2^n - |x|)$, ezért $x$ értékének meghatározása a bináris alakból ugyanígy történik, azaz ha az első bit egyes, $|x|$ értéke a bináris alak komplemense $+1$. A $-1$ alakja $11\ldots1_2$.

## Példa

legyen $n = 4$, $x = -5$: $-5 \rightarrow 16 - 5 \rightarrow 11 = 1011_2$
bitműveletekkel: $x = -5 \rightarrow |x| = 5 \rightarrow 0101_2 \rightarrow 1010_2 + 1_2 = 1011_2$
Visszaalakítás: $1011_2 \rightarrow 0100_2 + 1_2 = 0101_2 = 5$, tehát $x = -5$.

| | | | |
|---|---|---|---|
| 0 | 00 | | <control> |
| ... | | | |
| 31 | 1F | | <control> |
| 32 | 20 | | SPACE |
| 33 | 21 | ! | EXCLAMATION MARK |
| 34 | 22 | " | QUOTATION MARK |
| 35 | 23 | # | NUMBER SIGN |
| 36 | 24 | $ | DOLLAR SIGN |
| 37 | 25 | % | PERCENT SIGN |
| 38 | 26 | & | AMPERSAND |
| 39 | 27 | ' | APOSTROPHE |
| 40 | 28 | ( | LEFT PARENTHESIS |
| 41 | 29 | ) | RIGHT PARENTHESIS |
| 42 | 2A | * | ASTERISK |
| 43 | 2B | + | PLUS SIGN |
| 44 | 2C | , | COMMA |
| 45 | 2D | - | HYPHEN-MINUS |
| 46 | 2E | . | FULL STOP |
| 47 | 2F | / | SOLIDUS |
| 48 | 30 | 0 | DIGIT ZERO |
| ... | | | |
| 57 | 39 | 9 | DIGIT NINE |
| 58 | 3A | : | COLON |
| 59 | 3B | ; | SEMICOLON |
| 60 | 3C | < | LESS-THAN SIGN |
| 61 | 3D | = | EQUALS SIGN |
| 62 | 3E | > | GREATER-THAN SIGN |
| 63 | 3F | ? | QUESTION MARK |
| 64 | 40 | @ | COMMERCIAL AT |
| 65 | 41 | A | LATIN CAPITAL LETTER A |
| ... | | | |
| 90 | 5A | Z | LATIN CAPITAL LETTER Z |
| 91 | 5B | [ | LEFT SQUARE BRACKET |
| 92 | 5C | \ | REVERSE SOLIDUS |
| 93 | 5D | ] | RIGHT SQUARE BRACKET |
| 94 | 5E | ^ | CIRCUMFLEX ACCENT |
| 95 | 5F | _ | LOW LINE |
| 96 | 60 | ` | GRAVE ACCENT |
| 97 | 61 | a | LATIN SMALL LETTER A |
| ... | | | |
| 122 | 7A | z | LATIN SMALL LETTER Z |
| 123 | 7B | { | LEFT CURLY BRACKET |
| 124 | 7C | \| | VERTICAL LINE |
| 125 | 7D | } | RIGHT CURLY BRACKET |
| 126 | 7E | ~ | TILDE |
| 127 | 7F | | <control> |

1. ISO-8859-1 Latin1 (West European)
2. ISO-8859-2 Latin2 (East European)
3. ISO-8859-3 Latin3 (South European)
4. ISO-8859-4 Latin4 (North European)
5. ISO-8859-5 Cyrillic
6. ISO-8859-6 Arabic
7. ISO-8859-7 Greek
8. ISO-8859-8 Hebrew
9. ISO-8859-9 Latin5 (Turkish)
10. ISO-8859-10 Latin6 (Nordic)

# ISO-8859-2, Microsoft CP1250 (Windows Latin2), CP852 (DOSLatin2)

| | | | | |
|---|---|---|---|---|
| ISO-8859-1 | C1 | Á | U+00C1 | LATIN CAPITAL LETTER A WITH ACUTE |
| ISO-8859-1 | E1 | á | U+00E1 | LATIN SMALL LETTER A WITH ACUTE |
| ISO-8859-1 | D5 | Õ | U+00D5 | LATIN CAPITAL LETTER O WITH TILDE |
| ISO-8859-1 | DB | Û | U+00DB | LATIN CAPITAL LETTER U WITH CIRCUMFLEX |
| ISO-8859-1 | F5 | õ | U+00F5 | LATIN SMALL LETTER O WITH TILDE |
| ISO-8859-1 | FB | û | U+00FB | LATIN SMALL LETTER U WITH CIRCUMFLEX |
| ISO-8859-2 | D5 | Ő | U+0150 | LATIN CAPITAL LETTER O WITH DOUBLE ACUTE |
| ISO-8859-2 | DB | Ű | U+0170 | LATIN CAPITAL LETTER U WITH DOUBLE ACUTE |
| ISO-8859-2 | F5 | ő | U+0151 | LATIN SMALL LETTER O WITH DOUBLE ACUTE |
| ISO-8859-2 | FB | ű | U+0171 | LATIN SMALL LETTER U WITH DOUBLE ACUTE |
| CP1250 | 82 | ‚ | U+201A | SINGLE LOW-9 QUOTATION MARK |
| CP1250 | 84 | „ | U+201E | DOUBLE LOW-9 QUOTATION MARK |
| CP1250 | 85 | … | U+2026 | HORIZONTAL ELLIPSIS |
| CP1250 | 91 | ‘ | U+2018 | LEFT SINGLE QUOTATION MARK |
| CP1250 | 92 | ’ | U+2019 | RIGHT SINGLE QUOTATION MARK |
| CP1250 | 93 | “ | U+201C | LEFT DOUBLE QUOTATION MARK |
| CP1250 | 94 | ” | U+201D | RIGHT DOUBLE QUOTATION MARK |
| CP1250 | 96 | – | U+2013 | EN DASH |
| CP1250 | 97 | — | U+2014 | EM DASH |

- U+0000 - U+007F ASCII
- U+0080 - U+00FF Latin-1
- U+0100 - U+017F Latin Extended-A
- U+0180 - U+024F Latin Extended-B
- U+1E00 - U+1EFF Latin Extended

- UTF-8 minden karakter kódja 8, 16 vagy 32-bites.
- UTF-16 minden karakter kódja 16 vagy 32-bites.
- UTF-32 minden karakter 32-bites.

| Unicode | | UTF-8 | a karakter hivatalos neve |
|---|---|---|---|
| U+0020 | | 20 | SPACE |
| U+0030 | 0 | 30 | DIGIT ZERO |
| U+0040 | @ | 40 | COMMERCIAL AT |
| U+0041 | A | 41 | LATIN CAPITAL LETTER A |
| U+0061 | a | 61 | LATIN SMALL LETTER A |
| U+00C1 | Á | c3 81 | LATIN CAPITAL LETTER A WITH ACUTE |
| U+00C9 | É | c3 89 | LATIN CAPITAL LETTER E WITH ACUTE |
| U+00CD | Í | c3 8d | LATIN CAPITAL LETTER I WITH ACUTE |
| U+00D3 | Ó | c3 93 | LATIN CAPITAL LETTER O WITH ACUTE |
| U+00D6 | Ö | c3 96 | LATIN CAPITAL LETTER O WITH DIAERESIS |
| U+00DA | Ú | c3 9a | LATIN CAPITAL LETTER U WITH ACUTE |
| U+00DC | Ü | c3 9c | LATIN CAPITAL LETTER U WITH DIAERESIS |
| U+00E1 | á | c3 a1 | LATIN SMALL LETTER A WITH ACUTE |
| U+00E9 | é | c3 a9 | LATIN SMALL LETTER E WITH ACUTE |
| U+00ED | í | c3 ad | LATIN SMALL LETTER I WITH ACUTE |
| U+00F3 | ó | c3 b3 | LATIN SMALL LETTER O WITH ACUTE |
| U+00F6 | ö | c3 b6 | LATIN SMALL LETTER O WITH DIAERESIS |
| U+00FA | ú | c3 ba | LATIN SMALL LETTER U WITH ACUTE |
| U+00FC | ü | c3 bc | LATIN SMALL LETTER U WITH DIAERESIS |
| U+0150 | Ő | c5 90 | LATIN CAPITAL LETTER O WITH DOUBLE ACUTE |
| U+0151 | ő | c5 91 | LATIN SMALL LETTER O WITH DOUBLE ACUTE |
| U+0170 | Ű | c5 b0 | LATIN CAPITAL LETTER U WITH DOUBLE ACUTE |
| U+0171 | ű | c5 b1 | LATIN SMALL LETTER U WITH DOUBLE ACUTE |

| Kódtartomány (darab) | bináris alak | UTF-8 |
|---|---|---|
| 000000-00007F (128) | 0zzzzzzz | 0zzzzzzz |
| 000080-0007FF (1920) | 00000yyy yyzzzzzz | 110yyyyy 10zzzzzz |
| 000800-00FFFF (63488) | xxxxyyyy yyzzzzzz | 1110xxxx 10yyyyyy 10zzzzzz |
| 010000-10FFFF (1048576) | 000wwwxx xxxxyyyy yyzzzzzz | 11110www 10xxxxxx 10yyyyyy 10zzzzzz |

Á 00C1→1100 0001→00011 000001→11000011 10000001→C3 81

Õ 00D5→1101 0101→00011 010101→11000011 10010101→C3 95

Ő 0150→0001 0101 0000→00101 010000→11000101 10010000→C5 90