

# **Teoria informacji i kodowania**

Sprawozdanie z ćwiczeń nr 1

Temat zajęć: „Kodowanie znaków (ASCII,  
UNICODE)”

Data ćwiczeń: 26.03.2024

Wykonawca: Kamil Borkowski 83374

Grupa: WCY22IY1S1

Prowadzący zajęcia: dr inż. Leszek Grad

### Treść zadania:

Zakoduj w formacie UTF-16BE oraz UTF-8 punkt kodowy, którego wartość dziesiętna wyznaczona jest jako reszta z dzielenia liczby MiesiącDzieńRok urodzenia przez liczbę wszystkich punktów kodowych Unicode.

Data urodzenia: 28/01/2003    ->    Liczba: 01282003

Punkty kodowe Unicode znajdują się w przedziale od  $0_{\text{hex}}$  do  $10FFFF_{\text{hex}}$ . W systemie dziesiętnym  $10FFFF_{\text{hex}}$  jest równe 1 114 111, więc razem z zerem wszystkich punktów kodowych jest 1 114 112.

Reszta dzielenia liczby MiesiącDzieńRok przez liczbę wszystkich punktów kodowych:

$$01282003 \bmod 1114112 = 167891$$

$$167891_{\text{dec}} = 28FD3_{\text{hex}}$$

Zatem punkt kodowy w Unicode ma wartość U+28FD3

$$167891_{\text{dec}} = 0010\ 1000\ 1111\ 1101\ 0011_{\text{bin}}$$

### UTF-8

Punkt kodowy U+28FD3 znajduje się w przedziale od U+10000 do U+10FFFF, więc w systemie UTF-8 będzie się składał z czterech bajtów, z czego pierwszy zaczyna się od 11110, a pozostałe bajty zaczynają się od 10

**11110000      10101000      10111111      10010011**

Pozostałe bity zostały uzupełnione bitami punktu kodowego od strony najmniej znaczącego bitu. Uzyskane bajty można zapisać w systemie UTF-8, po przekształceniu ich na system szesnastkowy.

11110000, 10101000, 10111111, 10010011 -> F0, A8, BF, 93

Zatem punkt kodowy U+28FD3 w UTF-8 wynosi: 0xF0 0xA8 0xBF 0x93

### UTF-16BE

Punkt kodowy U+28FD3 znajduje się w przedziale od U+10000 do U+10FFFF, więc w systemie UTF-16 będzie się składał z czterech bajtów, z czego pierwszy zaczyna się od 110110, a trzeci bajt zaczyna się od 110111, miejsca pozostałych bitów są uzupełniane różnicą punktu kodowego i  $0x10000_{\text{hex}}$ , która wynosi  $18FD3_{\text{hex}}$ .

$$18FD3_{\text{hex}} = 0001\ 1000\ 1111\ 1101\ 0011_{\text{bin}}$$

**11011000      01100011      11011111      11010011**

Uzyskane bajty można zapisać w systemie UTF-16, po przekształceniu ich na system szesnastkowy.

11011000      01100011      11011111      11010011 -> D8 63 DF D3

Każde słowo składa się z szesnastu bitów, czyli dwóch bajtów, w formie Big endian najbardziej znaczący bajt w słowie jest umieszczany na początku.

Zatem punkt kodowy U+28FD3 w UTF-16BE wynosi: D863 DFD3 FE FF