

Specyfikacja funkcjonalna – compressor

CEL PROJEKTU

Program ma na celu kompresję i dekompresję dowolnych plików przy użyciu algorytmu Huffmana. Compressor posiada możliwość regulacji stopnia kompresji – ośmio- , dwunasto- lub szesnastobitowa. Program może również szyfrować pliki.

SKŁADNIA

Aby skompresować plik:

compressor -x *input-file* -o *output-file* -compression-rate

Aby zdekompresować plik:

compressor -z *input-file* -o *output-file*

OPCJE

Konieczne:

-x *input-file*

Plik, który ma zostać skompresowany – musi istnieć.

-z *input-file*

Plik, który ma zostać zdekompresowany – musi istnieć.

-o *output-file*

Plik wyjściowy, będący wynikiem kompresji bądź dekompresji - nie musi istnieć.

Konieczne do kompresji:

-1

Ustaw ośmiobitowy stopień kompresji.

-2

Ustaw dwunastobitowy stopień kompresji.

-3

Ustaw szesnastobitowy stopień kompresji.

Opcjonalne:

-c password

Zaszyfruj dane wyjściowe hasłem „password”.

-v

Podaj dodatkowe informacje o procesach dekompresji oraz kompresji. Wyświetla wartość decymalną znaku oraz wygenerowany kod, odpowiadające temu znakowi.

-h

Wypisz pomoc, opisującą działanie programu.

TEORIA

Program implementuje algorytm Huffmana do kodowania znaków (ośmio-, dwunasto- lub szesnastobitowych). Kompresor znajduje dwa najrzadziej występujące węzły, a następnie łączy je do wspólnego „rodzica”. Jeśli w pliku, który chcemy skompresować znajduje się jeden znak, do słownika dodawany jest sztucznie „występujący raz” znak o wartości znaku „J”. Program jest również w stanie zaszyfrować plik podczas kompresji. W tym celu XORowane są wszystkie znaki hasła z każdym znakiem, wchodzącym w skład słownika oraz skompresowanej treści pliku. XORowanie wykorzystywane jest również przy obliczaniu sumy kontrolnej.

PLIK WEJŚCIOWY

Kompresja:

- dowolny niepusty plik o innej nazwie niż plik wejściowy.

Dekompresja:

- nieuszkodzony skompresowany plik.

PLIK WYJŚCIOWY

Kompresja:

- plik o zawartości:
 - 2 bity ‘1’,
 - 2 bity reprezentujące stopień kompresji:
 - ośmiobitowa – ‘01’,
 - dwunastobitowa – ‘10’,
 - szesnastobitowa – ‘11’.
 - 1 bit ‘1’,
 - 3 bity reprezentujące ile zer zostało dodanych na końcu pliku za pomocą zapisu binarnego,
 - 8 bitów reprezentujących sumę kontrolną,
 - 32 bity reprezentujące liczbę bitów słownika za pomocą zapisu binarnego,

- 32 bity reprezentujące liczbę unikalnych znaków w oryginalnym pliku za pomocą zapisu binarnego,
- słownik:
 - 8 / 12 / 16 bitów (zależnie od stopnia kompresji) reprezentujących wartość decymalną znaku za pomocą zapisu binarnego,
 - 8 bitów reprezentujących długość kodu znaku za pomocą zapisu binarnego,
 - x bitów (gdzie x to długość kodu znaku) reprezentujących kod znak.
- skompresowana treść pliku wejściowego,
- od 0 do 7 bitów uzupełniających skompresowaną treść pliku do wartości podzielnej przez 8.

Dekompresja:

- oryginalnie skompresowany plik.