

Algorytmy Tekstowe

Laboratorium 3 - Statyczne i dynamiczne kodowanie Huffmana

14.04.2021

Magdalena Pastuła

Zadanie 1.

Opracowany format pliku:

Na początku pliku znajdują się dwie liczby 64 bitowe: pierwsza określa długość zakodowanego tekstu, druga długość zapisanego sposobu kodowania. Obydwie długości są bitach. Suma tych dwóch liczb jest równa długości pozostałej części pliku.

Następnie znajduje się część z zapisanym kodowaniem, którą można podzielić na sekcje dla każdego znaku o następującej konstrukcji:

- długość zapisu znaku w bitach - długość stała 5 bitów. Wynika to z założenia kodowania tekstu w plikach w UTF-8, w którym znak nie ma stałej długości kodu
- kod znaku w UTF-8 o długości zapisanej na poprzednich 5 bitach
- długość wagi znaku - długość stała 5 bitów
- waga znaku w tekście, to znaczy liczba wystąpień, na jej podstawie jest budowane drzewo

Na koniec pliku znajduje się zakodowany tekst.

Zadanie 3.

Podane w tabelach rozmiary są przybliżone. Zmierzone współczynniki kompresji dla odpowiednich rozmiarów plików:

Źródło pliku	Rozmiar	Kodowanie statyczne
Guttenberg	1 kB	17.81 %
Guttenberg	10 kB	41.30 %
Guttenberg	100 kB	45.35 %
Guttenberg	1 MB	45.76 %
Github	1 kB	9.20 %
Github	10 kB	36.13 %
Github	100 kB	36.47 %
Github	1 MB	36.88 %
Losowe znaki	1 kB	-8.20 %
Losowe znaki	10 kB	14.44 %
Losowe znaki	100 kB	16.84 %

Źródło pliku	Rozmiar	Kodowanie statyczne
Losowe znaki	1 MB	17.07 %

Ta sama tabela z porównaniem dla rozmiaru:

Źródło pliku	Rozmiar	Kodowanie statyczne
Guttenberg	1 kB	17.81 %
Github	1 kB	9.20 %
Losowe znaki	1 kB	-8.20 %
Guttenberg	10 kB	41.30 %
Github	10 kB	36.12 %
Losowe znaki	10 kB	14.44 %
Guttenberg	100 kB	45.35 %
Github	100 kB	36.47 %
Losowe znaki	100 kB	16.84 %
Guttenberg	1 MB	45.76 %
Github	1 MB	36.88 %
Losowe znaki	1 MB	17.07 %

Zadanie 4.

Czasy kompresji:

Źródło pliku	Rozmiar	Kodowanie statyczne	Dekodowanie statyczne
Guttenberg	1 kB	28.09 ms	4.69 ms
Guttenberg	10 kB	386.58 ms	18.61 ms
Guttenberg	100 kB	4 554.28 ms	180.69 ms
Guttenberg	1 MB	58.69 s	1 890.36 ms
Github	1 kB	20.87 ms	3.01 ms
Github	10 kB	525.54 ms	20.90 ms
Github	100 kB	5 533.13 ms	202.53 ms
Github	1 MB	61.86 s	2 180.09 ms
Losowe znaki	1 kB	61.87 ms	6.60 ms
Losowe znaki	10 kB	646.24 ms	41.27 ms
Losowe znaki	100 kB	7 057.09 ms	290.66 ms

Źródło pliku	Rozmiar	Kodowanie statyczne	Dekodowanie statyczne
Losowe znaki	1 MB	63.24 s	2 611.10 ms

Czasy kompresji posortowane względem rozmiaru pliku:

Źródło pliku	Rozmiar	Kodowanie statyczne	Dekodowanie statyczne
Guttenberg	1 kB	28.09 ms	4.69 ms
Github	1 kB	20.87 ms	3.01 ms
Losowe znaki	1 kB	61.87 ms	6.60 ms
Guttenberg	10 kB	386.58 ms	18.61 ms
Github	10 kB	525.54 ms	20.90 ms
Losowe znaki	10 kB	646.24 ms	41.27 ms
Guttenberg	100 kB	4 554.28 ms	180.69 ms
Github	100 kB	5 533.13 ms	202.53 ms
Losowe znaki	100 kB	7 057.09 ms	290.66 ms
Guttenberg	1 MB	58.69 s	1 890.36 ms
Github	1 MB	61.86 s	2 180.09 ms
Losowe znaki	1 MB	63.24 s	2 611.10 ms