

Kodowanie i kompresja danych 2022

Laboratorium nr 5 i 6 (na ocenę)

Zadanie na laboratorium

Ocena 3 Napisz program który implementuje zmodyfikowany algorytm LZ77, który zamiast 3 liczb używa 2:

- $(0, kod_litera)$ - jeśli pierwszej litery w buforze kodowania nie ma w buforze słownika,
- (i, j) - dla najdłuższego możliwego prefiksu z bufora kodowania występującego w buforze słownika (przedłużonym o bufor kodowania) i to przesunięcie w słowniku a $j + 1$ ilość znaków do skopiowania (omijamy kod ostatniej litery z klasycznego LZ77).

W programie bufor słownika ma mieć 255 znaków a bufor kodowania 256 (stąd używane liczby są 8-bitowe).

Program ma mieć możliwość kodowania i dekodowania, dodatkowo podczas kodowania ma podawać długość kodowanego pliku, długość uzyskanego kodu, stopień kompresji, entropię kodowanego tekstu i entropię uzyskanego kodu.

Ocena 4 Napisz program kodujący i dekodujący dany plik za pomocą algorytmu LZW. Ciąg wartości indeksów słownika początkowo powinien być kodowany liczbami 9-bitowymi a następnie liczby powinny być automatycznie wydłużane o kolejne bity wraz ze wzrostem wielkości słownika.

Program dodatkowo podczas kodowania ma podawać długość kodowanego pliku, długość uzyskanego kodu, stopień kompresji, entropię kodowanego tekstu i entropię uzyskanego kodu.

Ocena 5 Napisz program kodujący i dekodujący dany plik za pomocą algorytmu LZW. Ciąg wartości indeksów słownika ma być zakodowany kodowaniem uniwersalnym. Alfabetem wejściowym są 8-bitowe kody. Domyślnie program powinien używać kodowania Eliasa ω oraz mieć możliwość opcjonalnego użycia pozostałych kodowań Eliasa oraz kodowania Fibonacciego.

Program dodatkowo podczas kodowania ma podawać długość kodowanego pliku, długość uzyskanego kodu, stopień kompresji, entropię kodowanego tekstu i entropię uzyskanego kodu.

Ocena programów może być przez prowadzącego laboratorium podwyższona o 0.5 za jakość wykonania.

Program może być napisany w dowolnym języku programowania, ale jego czas działania dla testów z pierwszej listy nie powinien przekraczać 3 minut. Programy powinny być wywoływane z linii poleceń z podanymi nazwami plików jako argumentami wywołania.

Zadania przygotowawcze do kolokwium

Zadanie 1

Użyj algorytmu LZ77 do zakodowania następującego ciągu:

barrrayar-bar-by-barrayar-bay

Założ, że wielkość okien wynosi odpowiednio 7 i 8. Jakie będą wyniki kompresji przy mniejszych wielkościach okien.

Zadanie 2

Wiedząc, że kodowanie poszczególnych liter jest następujące $C(a)=1$, $C(-)=2$, $C(r)=3$, $C(t)=4$, odkoduj podany trójką algorytmem LZ77:

$(0,0,3)$, $(0,0,1)$, $(0,0,4)$, $(2,8,2)$, $(3,1,2)$, $(0,0,3)$, $(6,4,4)$, $(9,5,4)$

Przyjmij, że wielkość obu okien wynosi 10.

Zadanie 3

Użyj algorytmu LZ78 do zakodowania następującego ciągu:

a-bar-array-by-barrayar-bay

Zadanie 4

Użyj algorytmu LZW do zakodowania następującego ciągu:

a-bar-array-by-barrayar-bay

Założ, że słownik początkowy wygląda następująco: 1 - a, 2 - b, 3 - r, 4 - y, 5 - -.

Zadanie 5

Dla słownika początkowego 1 - a, 2 - -, 3 - h, 4 - i, 5 - s, 6 - t, odkoduj algorytmem LZW następujący ciąg:

6,3,4,5,2,3,1,6,2,9,11,16,12,14,4,20,10,8,23,13

Sprawdź czy zakodowanie otrzymanego ciągu z tym samym słownikiem początkowym da ten sam kod.

Zadanie 6

Przedstaw następujące liczby: 157, 1604, 10345 w kodowaniach Eliasa γ , ω , δ oraz w kodowaniu Fibonacciego. Spróbuj określić od jakiej wartości jedno z tych kodowań jest najlepsze? Jak zachowuje się kodowanie Fibonacciego dla praktycznie stosowanych wartości?