

# Kodowanie i kompresja danych 2020

## Laboratorium nr 3 i 4 (na ocenę)

### Zadanie na laboratorium

Napisz program kodujący i program dekodujący dany plik wejściowy do pliku wyjściowego. Program kodujący powinien dodatkowo na koniec zwrócić na ekran odpowiednią entropię kodowanych danych, średnią długość kodowania i stopień kompresji. Alfabetem wejściowym są 8-bitowe kody ASCII.

Należy wybrać jeden z wariantów:

**Ocena 3** Programy mają używać klasycznego kodowania Huffmana i w czasie kompresji dodawać odpowiednio zapamiętany słownik umożliwiający dekompresję.

**Ocena 4** Programy mają używać dynamicznych kodów Huffmana.

**Ocena 5** Programy mają używać adaptacyjnego kodowania arytmetycznego ze skalowaniem.

### Zadania przygotowawcze do kolokwium

#### Zadanie 1

Skonstruuj kod Huffmana dla następującego ciągu prawdopodobieństw występowania symboli:  $P(a) = 0.1$ ,  $P(b) = 0.1$ ,  $P(c) = 0.15$ ,  $P(d) = 0.25$  i  $P(e) = 0.4$ .

#### Zadanie 2

Prześledź jaką mają postać kody Huffmana dla:

1.  $k$  jednakowo prawdopodobnych symboli;
2.  $k$  symboli takich, że  $P(a_i) = \frac{1}{2^i}$ , dla  $i \in \{1, \dots, k-1\}$ , i  $P(a_k) = \frac{1}{2^{k-1}}$ .

#### Zadanie 3

W algorytmie Huffmana można w zależności od kolejności kodowania symboli z jednakowym prawdopodobieństwem, uzyskać różne (nieizomorficzne) drzewa kodów. Jaką strategię należałoby użyć, aby odchylenie standardowe długości kodów było jak najmniejsze? (Oczekiwana długość kodów będzie oczywiście jednakowa.)

#### Zadanie 4

Znajdź sytuacje, w których średnia długość kodu Huffmana jest

1. równa entropii;
2. przybliża się do entropii plus jeden (jest jak największa).

#### Zadanie 5

Prześledź działanie dynamicznego kodowania Huffmana. Wybierz w tym celu kilka losowych słów o długości 8, utworzonych z 4 literowego alfabetu  $\{a, b, c, d\}$ .

### Zadanie 6

Dla symboli z Zadania 1 skonstruuj 4-o i 5-o bitowe kody Tunstalla. Porównaj średnią ilość bitów na jeden symbol wejściowy tych kodów i kodu Huffmana.

### Zadanie 7

Dla kodu o następujących prawdopodobieństwach wystąpienia:  $P(a) = 0.2$ ,  $P(b) = 0.3$  i  $P(c) = 0.5$ , znajdź znacznik w kodowaniu arytmetycznym dla ciągu  $aabcba$ . Następnie odkoduj ciąg o długości 10 i znaczniku 0.63215699.

### Zadanie 8

Napisz pseudo-kod kodowania i dekodowania arytmetycznego ze skalowaniem. Prześledź pracę obu algorytmów dla danych z poprzedniego zadania.