

# Kompresja danych 2020

Lista zadań nr 2

1. (1) Uzasadnij, że kody stałe przyporządkowane elementom alfabetu w dynamicznym kodowaniu Huffmana są prefiksowe i optymalne dla jednostajnego rozkładu prawdopodobieństw (czyli  $p_i = 1/n$  dla  $i = 1, \dots, n$ ).
2. (1) Porównaj długość zakodowanej postaci ciągu  $a^n b^{2n} c^{3n} def$  przy zastosowaniu kodów Huffmana i dynamicznych kodów Huffmana. Przyjmij, że alfabet jest równy  $\{a, b, c, d, e, f\}$  oraz, że w (statycznym) kodowaniu Huffmana korzystamy z prawdopodobieństw uzyskanych w oparciu o częstości wystąpień elementów w całym kodowanym ciągu.
3. (3) Zapoznaj się z kanonicznymi kodami Huffmana. Przedstaw tworzenie kodu, kodowanie i dekodowanie. Omów ich zalety w porównaniu ze „standardowym” kodowaniem Huffmana.  
Uwaga: p. notatka z pracy magisterskiej M.Cylkowskiej lub E.S.Schwartz, B.Kallick, Generating a canonical prefix encoding, Commun. ACM, 7(3)166-196, 1964.
4. (2) Pokaż, że kod Shannona jest kodem prefiksowym, ale nie jest optymalnym kodem prefiksowym.
5. (1) Porównaj wartość entropii i średniej długości kodu Shannona-Fano w przypadku, gdy prawdopodobieństwo wystąpienia każdego symbolu jest odwrotnością naturalnej potęgi liczby 2.
6. (1) Pokaż, że kod Shannona-Fano nie jest optymalnym kodem prefiksowym.
7. (1) Pokaż, że jeśli  $d_1, \dots, d_n$  oznaczają długości słów (rozszerzonych) kodów Huffmana, to

$$\sum_{i=1}^n d_i \leq (n^2 + n - 2)/2.$$

8. (3\*) Sprawdź czy zakodowana postać danych utworzona przy użyciu dynamicznego kodu Huffmana daje zawsze większy lub równy rozmiar danych niż statyczny kod Huffmana (przy założeniu, że prawdopodobieństwa dla statycznych kodów Huffmana wyznaczamy na podstawie częstości wystąpień symboli w całym tekście).

*Przypomnienie:* punkty za to zadanie nie są wliczane do sumy punktów do zdobycia na ćwiczeniach.

## Zadania dodatkowe (nie deklarujemy ich na ćwiczenia):

1. (0) Omów zmiany w drzewie kodu i proces kodowania tekstu **axyxyx** w dynamicznym kodowaniu Huffmana przy założeniu, że alfabet składa się z 26 małych liter alfabetu angielskiego.  
Przedstaw działanie dekodowania dla uzyskanej wcześniej zakodowanej postaci **axyxyx**.
2. (1) Pokaż sposób implementacji dynamicznych kodów Huffmana, który pozwala wykonywać modyfikację drzewa kodu w czasie proporcjonalnym do długości ścieżki na której zwiększamy wagi.

*Uwaga:* wystarczy uszczegółwić algorytm z wykładu.

Tomasz Jurdziński