

1 Wzory

Informacja zdarzenia A :

$$I(A) = -\log_x P(A)$$

Entropia zdarzenia A :

$$H(X) = \sum_{i=1}^n P(A_i) I(A_i)$$

Średnia długość kodu C :

$$I(C) = \sum_{i=1}^n P(C_i) \cdot l_i$$

Nierówność Krafta (warunek konieczny jednoznacznej dekodowalności):

$$K(C) = \sum_{i=1}^n 2^{-l_i} \leq 1$$

2 Kod Huffmana

Znajdź dwa najrzadziej występujące elementy i połącz je w jeden element o prawdopodobieństwie $p_1 + p_2$. Rozróżnij je 0 lub 1. Powtórz ten krok na liście $n - 1$ długiej aż zostanie jeden element.

Jeśli nie znamy prawdopodobieństw, to możemy drzewo tworzyć dynamicznie, traktując ilość wystąpień jako wagę, które łączymy tworząc poddrzewa.

3 Kodowanie Eliasa

$$n = \lfloor \log_2(x) \rfloor + 1$$

3.1 γ

$$\gamma(x) = 0^{n-1}(x)_2$$

3.2 δ

$$\delta(x) = \gamma(n) + (x)_2$$

3.3 ω

Na koniec umieszczane jest 0, potem kodowana jest liczba $k = x$. Potem ten krok jest powtarzany dla $k = n - 1$ gdzie n to liczba bitów z poprzedniego kroku.

$$\omega(x) = \omega(n - 1) + (x)_2 + 0$$

4 Kodowanie Fibonacciego

$$f_0 = f_1 = 1$$

$$f_n = f_{n-1} + f_{n-2} : n \geq 2$$

$$x = \sum_{i=0} a_i \cdot f_i, a_i \in \{0, 1\}$$

5 Kodowanie arytmetyczne

- $[l, p) = [0, 1)$
- $d = p - l$
- $p = l + d \cdot F(j + 1)$
- $l = l + F(j) d$

6 Kodowanie słownikowe

6.1 LZ77

$$(o, l, k) = C_{i-o} \cdots C_{i-o+l} k$$

6.2 LZ78

1. Szukaj w słowniku najdłuższy prefiks aktualnego okna, jeśli nie znajdziesz to użyj ϵ .
2. Dodaj prefiks + znak do słownika.

$$(i, k) = s(i) + k$$

6.3 LZW

Podobne do LZ78, tylko że zaczynamy ze słownikiem.

$$(i) = s(i)$$

7 bzip2/BWT

Układamy tabelę z dwoma kolumnami. Pierwsza kolumna to słowo posortowane leksykograficznie. Druga kolumna to poprzedni znak. Na podstawie tej tabeli zapisujemy ostatnią kolumnę, i numer wiersza w którym w pierwszej kolumnie znajduje się początek słowa, a w drugiej kolumnie jego koniec.

e	h				
h	o	0	1	2	3
ll	e	e	h	l	l
lo	l	2	0	3	4
o	l				1

8 Move To Front

Jest to transformacja zmniejszająca entropię. Zaczynamy od tabeli liter ze słowa posortowanych alfabetycznie. Następnie dla każdej litery ze słowa kodujemy jej pozycję w tabeli, a następnie przesuwamy ją na początek tabeli. W ten sposób hello to 11203.