Domácí úkol č.8 - Zdrojové kódy (komprese dat)

1. Rozhodněte, zda je daný kód jednoznačně dekódovatelný, či dokonce instantní. Případně najděte posloupnost znaků, kterou nelze jednoznačně dekódovat.

   a) Binární kód C_1

x	a	b	c	d
$C_1(x)$	0	01	011	111

   b) Ternární kód C_2

x	a	b	c	d	e	f	g
$C_2(x)$	0	1	20	21	220	221	222

   c) Binární kód C_3

x	a	b	c	d	e
$C_3(x)$	01	10	0100	001	110

2. Může existovat instantní binární kód pro vstupní abecedu $\{a, b, c, d, e\}$ s délkami slov $2, 2, 3, 3, 4$? Pokud ano, najděte nějaký takový kód. (Návod: Použijte Kraftovu nerovnost.)

   Bude tento kód optimálním kódem pro nějaké pravděpodobnostní rozdělení? Pokud ano, najděte nějaké takové pravděpodobnostní rozdělení na vstupní abecedě.

3. Je kód jednoznačně dekódovatelný? Pokud ano, nalezněte instantní kód se stejnými délkami slov. (Návod: Použijte McMillanovu větu.)

   a) Binární kód C

x	a	b	c	d	e	f
$C(x)$	00	10	001	101	011	111

   b) Ternární kód K

x	a	b	c	d	e	f
$K(x)$	0	02	1	12	20	21

4. Kolik znaků musí mít kódová abeceda,pokud chceme zakódovat 26 písmen anglické abecedy kódovými slovy délky nejvýše 2? A co kdyby navíc 6 samohlásek mělo kódová slova délky 1?

5. Pro informační zdroj X nalezněte binární i ternární Huffmanův kód. Spočtěte střední délky kódů a entropie zdroje.

x	a	b	c	d	e	f
$p(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

6. Pro stejný informační zdroj X jako v předchozím příkladu nalezněte nějaký Shannonův kód a nějaký (Shannon-Fano)-Eliasův kód. Porovnejte střední délky těchto kódů.

7. Pro informační zdroj X nalezněte dva binární Huffmanovy kódy tak, že jeden má všechna kódová slova délky nejvýše 4 a druhý má i kódová slova délky 5. Ověřte, že střední délka obou kódů je stejná.

x	a	b	c	d	e	f
$p(x)$	$0,04$	$0,4$	$0,06$	$0,3$	$0,1$	$0,1$

8. Pro stejný informační zdroj X jako v předchozím příkladu nalezněte nějaký Shannonův kód a nějaký (Shannon-Fano)-Eliasův kód. Porovnejte střední délky těchto kódů.

Výsledky

1. C_1 je jen jednoznačně dekódovatelný, C_2 je dokonce instantní, C_3 není jednoznačně dekódovatelný

2. Ano, neboť délky slov splňují Kraftovu nerovnost.

   Např. $C(a) = 00$, $C(b) = 01$, $C(c) = 100$, $C(d) = 101$, $C(e) = 1100$.

3. Kód C je jednoznačně dekódovatelný, neboť je suffixový. Obrátíme-li pořadí písmen, získáme instantní kód. Kód K není jednoznačně dekódovatelný, délky slov nesplňují Kraftovu nerovnost.

4. Pro kódová slova délky nejvýše dva potřebujeme 6 znaků, pokud navíc mají mít souhlásky kódová slova délky jedna, tak potřebujeme 8 znaků.

5. Binární Huffmanův kód má střední délku $L(C_H) = 2,5 = H_2(X)$.

x	a	b	c	d	e	f
$C_H(x)$	00	01	100	101	110	111

Ternární Huffmanův kód má střední délku $L(K) = 1,75 > H_3(X) \doteq 1,577$.

x	a	b	c	d	e	f
$K(x)$	1	2	01	02	001	002

6. Shannonův kód je tentokrát stejný jako binární Huffmanův kód, pravděpodobnosti jsou 2-diadické.

Eliasův kód má střední délku $L(C_E) = 3,5$.

x	a	b	c	d	e	f
$C_E(x)$	001	011	1001	1011	1100	1111

7. Binární Huffmanovy kódy mají střední délku $L(C_1) = L(C_2) = 2,2 > H_2(X)$.

x	a	b	c	d	e	f
$C_1(x)$	0011	1	0010	01	0001	0000
$C_2(x)$	00111	1	00110	01	0010	000

8. Binární Shannonův kód má střední délku $L(C_S) = 2,7 > H_2(X)$, Eliasův kód $L(C_E) = 3,7$.

x	a	b	c	d	e	f
$C_S(x)$	11110	00	11111	01	1000	1001
$C_E(x)$	000001	001	011101	101	11011	11110