

Závěrečný test do předmětu P030 Textové informační systémy

1. **7 b.** Necht' $N = 16$. Spočítejte Eliasovy kódy $\beta(N)$, $\gamma(N)$, $\omega(N)$.
2. **6 b.** Zakódujte binárně zprávu $F\mathbb{I}$ aritmetickým kódováním víte-li, že pravděpodobnost výskytu zdrojové jednotky \mathbb{I} je sedminásobná proti zdrojové jednotce F .
3. **3 b.** Najděte příklad vzorku délky alespoň dva a textu, na němž je KMP rychlejší než BM měřeno počtem porovnání znaků. Svě tvrzení zdůvodněte.
4. **3 b.** Najděte příklad vzorku délky nejvýše dva a textu délky n , při němž je složitost vyhledávání pomocí BM měřeno počtem porovnání znaků $n/2$. Svě tvrzení zdůvodněte.
5. **4 b.** Zakódujte zprávu $X = \text{má mě rád nemá mě rád má mě rád nemá mě rád}$ nad abecedou $A = \{\text{má, mě, rád, nemá}\}$ pomocí BSTW.
6. Mějme regulární výraz $R = (01)^*(00)^*1$ nad abecedou $A = \{0,1\}$.
 - 3 b.** Spočítejte $\frac{dR}{d001}$.
 - 6 b.** Sestrojte deterministický KA pro vyhledávání R a nakreslete přechodový diagram tohoto automatu.
7. **7 b.** Zakódujte zprávu $X = \text{abakadabra}$ nad abecedou $A = \{a,b,d,r,k\}$ pomocí Huffmanova kódování. Nakreslete kódový strom, který jste k tomu použili.
 - 4 b.** Existuje prefixový kód, který není Huffmanův? Pokud neexistuje, dokažte. Pokud existuje, dokažte příkladem prefixového kódu (kódového stromu) pro zprávu X , který není Huffmanův.
 - 4 b.** Jaká je entropie zdrojové zprávy $E(X)$, délka zakódované zprávy $l(X)$?
8. **3 b.** Vyjmenujte alespoň tři adaptivní slovníkové metody komprese a popište jejich hlavní myšlenku.

Dohromady **50 bodů**. Čas na vypracování 60 minut. Petr Sojka