

TIS (P030), písemná zkouška 29. 5. 2002

1. Mějme regulární výraz $R = (b + c)^*(a + b)$.
 - 1 b. Spočtěte $\frac{dR}{ba}$ pozičním vektorem.
 - 4 b. Sestrojte DKA pro vyhledávání R přímým postupem (postupným derivováním) a nakreslete přechodový diagram tohoto automatu.
2. Mějme vyhledávací problém $SFOR_kCO, V = p_1 \dots p_m$.
 - (a) 2 b. Kolik stavů má vyhledávací automat?
 - (b) 3 b. Necht' $k = 2, V = P030, T = \text{test_P030_dam_za_30b}$. Kolik řetězců a jakých je nalezeno?
3. 5 b. Zakódujte binárně zprávu P030 aritmetickým kódováním. Pravděpodobnosti výskytů zdrojových jednotek se spočtou ze zprávy.
4. 4 b. Najděte příklad alespoň dvou vzorků a textu, na němž je algoritmus CW desetkrát rychlejší než naivní algoritmus, měřeno počtem porovnání. Pokud takový příklad neexistuje, dokažte. Svá tvrzení zdůvodněte.
5. 3 b. Dekódujte zprávu zakódovanou Fibonacciho kódem řádu 2:
10101011100011
6. 4 b. Uveďte příklad prefixového optimálního kódu, který není Huffmanův.
7. Zdůvodněte či vyvrat'te protipříklady.
 - (a) 2 b. Platí u kosinové míry měření podobnosti dokumentů ve vektorovém modelu implikace, že z nulové vzdálenosti plyne, že dokumenty mají stejné počty výskytů sledovaných termů?
 - (b) 2 b. Je každá průměrná délka kódového slova Huffmanova kodování zprávy X menší než průměrná entropie $AE(X)$?
8. Necht' $N = 32$. Spočtěte Eliasovy kódy
 - (a) 1 b. $\beta'(N)$,
 - (b) 2 b. $\gamma(N)$,
9. 3 b. Dejte příklad alespoň dvou dokumentů, dotazu a jejich vrstvených signatur, aby při vyhodnocení došlo k chybnému výběru dokumentu (false drop).
10. 2 b. Definujte koeficienty přesnosti a úplnosti.
11. 2 b. Spočtěte $\text{DIRT}(\text{FIdivadlo}, \text{FIfestival})$.

Dohromady 40 bodů.

Každý list s řešením podepište.