## TIS (PV030), písemná zkouška 27. 5. 2003

1. **4 b.** Definujte vyhledávací algoritmus Shift-Or.
2. **4 b.** Je možné, aby rychlost vyhledávání v lexikonu všech slov obsažených na WWW stránkách implementovaného na Pentiu 4 přesahovala 10000 nalezených slov za vteřinu? Navrhněte implementaci nebo vyvraťte. Svá tvrzení zdůvodněte.
3. **4 b.** Uveďte příklad gramatiky $G$ a lokálního číslování pravidel $G$ takového, že derivační kódování zprávy $X = fiataif$ je 0123 a do jazyka $L(G)$ patří také `fafiifaf`.
4. Mějme regulární výraz $R = (a + b)^* c(a + b)^*$.
   **5 b.** Sestrojte DKA pro vyhledávání $R$ přímým postupem (postupným derivováním) a nakreslete přechodový diagram tohoto automatu.
   **2 b.** Napište regulární výraz podobný $R$ délky 17 případně dokažte, že neexistuje.
5. **3 b.** Dejte příklad alespoň dvou dokumentů, dotazu a jejich vrstvených signatur, aby při vyhodnocení došlo k chybnému výběru dokumentu (false drop).

Dohromady **22 bodů**.

### Řešení

1. Viz slidy předmětu a skripta prof. Pokorného a kol., strana 31.
2. Je to možné: i když počet slov v lexikonu je několik desítek milionů, při implementaci pomocí datové struktury trie je počet instrukcí nutných k vyhledání slova úměrný *pouze* délce slova (ne počtu slov v lexikonu), tj. maximálně desítky instrukcí. Běžná Pentia přitom zvládají hodně přes $10^6$ instrukcí za sekundu.
3. G= ($\{S\}$, $\{f,i,a,t\}$, $\{ S \xrightarrow{0} fSf, S \xrightarrow{1} iSi, S \xrightarrow{2} aSa, S \xrightarrow{3} t, S \xrightarrow{4} \varepsilon\}$, $S$)
4. DKA má dva stavy, jeden, počateční odpovídá $R = \frac{dR}{a} = \frac{dR}{b}$ a druhý, koncový $\frac{dR}{c} = (a + b)^*$. Z prvního stavu se jde do druhého pod $c$, a v každém se zůstává pod $a$ i $b$.
   Regulární $R'$ podobný $R$ délky 17 je například $(R)$.
5. Dokumenty $D_1$ = slunce svítí, $D_2$ = venku. $f(\text{slunce}) = 100100$, $f(\text{svítí}) = 000011$, $f(\text{venku}) = 100010$. $f(D_1) = 100111$, $f(D_2) = 100010$. Dotaz $Q$ = venku. $f(Q) = f(\text{venku}) = 100010$, tedy chybný výběr $D_1$.