IB102 – úkol 5, příklad 1 – řešení **Odevzdání: 2. 11. 2015**

Vypracoval(a): UČO:
Skupina:

---

1. [2 body] Nechť L a R jsou regulární jazyky nad abecedou Σ a operace $inject(L, R)$ je definována následovně:

$$inject(L, R) = \{u \cdot v \cdot w \mid uw \in L, v \in R\}$$

Intuitivně je tedy $inject(L, R)$ jazyk obsahující všechna slova, která vzniknou vložením libovolného slova z jazyka R někam do libovolného slova z jazyka L. Například:

$$inject(\{a, aa\}, \{b, bb\}) = \{ba, ab, bba, abb, baa, aba, aab, bbaa, abba, aabb\}$$
$$inject(\{\varepsilon, a\}, \{a, ab\}) = \{a, ab, aa, aab, aba\}$$

Vaším úkolem je rozhodnout, zda je jazyk $inject(L, R)$ regulární, tedy že třída regulárních jazyků je uzavřená na operaci *inject*. Vaši odpověď dokažte, a to tak, že:

- Pokud rozhodnete, že není, najděte regulární jazyky L a R takové, že jazyk $inject(L, R)$ regulární není.
- Pokud rozhodnete, že je, dokažte tvrzení například s pomocí známých uzávěrových vlastností třídy regulárních jazyků prezentovaných na přednášce, nebo konstruktivně popsáním algoritmu na transformaci nějakého formalizmu pro popis regulárních jazyků.

Třída regulárních jazyků je uzavřená na operaci *inject*. To můžeme dokázat popisem algoritmu na konstrukci konečného automatu $\mathcal{A}$ rozpoznávajícího jazyk $inject(L, R)$.

Základní myšlenka konstrukce

Budeme konstruovat konečný automat pro výsledný jazyk. Vyjdeme z DFA pro jazyky L a R (ty existují, protože L, R jsou regulární). Kdykoli načteme nějaký prefix u slova $x = uw$ z jazyka L, budeme moci dále pokračovat slovem v z jazyka R a pak dokončit načítání příslušného sufixu w. Zároveň však musíme zajistit, že:

1. toto vložení půjde udělat právě jednou,
2. po vložení v bude možné pokračovat jen takovým sufixem w, že $uw \in L$.

To tedy znamená, že po dobu načítání v si musíme pamatovat, ve kterém stavu automatu pro L jsme začali v vkládat, a po skončení načítání v si musíme pamatovat, že již další slovo z R vkládat nelze.

IB102 – úkol 5, příklad 1 – řešení　　　Odevzdání: 2. 11. 2015

Vypracoval(a): 　　　　　　　　　　UČO:
Skupina:

Myšlenka konstrukce konečného automatu – podrobněji

Mějme konečný automat $\mathcal{A}_L$ pro jazyk L a konečný automat $\mathcal{A}_R$ pro jazyk R. Pro každý stav automatu $\mathcal{A}_L$ vytvoříme kopii $\mathcal{A}_R$. Dále si vytvoříme kopii automatu $\mathcal{A}_L$, kterou označíme $\mathcal{A}'_L$.

Nyní spojíme všechny uvedené automaty do jednoho automatu $\mathcal{A}$. Pro každý stav q automatu $\mathcal{A}_L$ opakujeme tento postup:

- z q vedeme přechod pod ε do vstupního stavu příslušné kopie $\mathcal{A}_R$,
- ze všech akceptujících stavů příslušné kopie $\mathcal{A}_R$ vedeme přechod pod ε do stavu q' automatu $\mathcal{A}'_L$, který odpovídá stavu q

Tímto jsme spojili všechny automaty a dále je potřeba zvolit počáteční stav a akceptující stavy $\mathcal{A}$. Jako počáteční stav $\mathcal{A}$ vybereme stav, který byl počátečním stavem automatu $\mathcal{A}_L$. Koncové stavy budou všechny koncové stavy automatu $\mathcal{A}'_L$.

Mohlo by se zdát, že automat $\mathcal{A}'_L$ není potřeba a z akceptujících stavů automatů pro jazyk R by se stačilo vracet do stavů automatu $\mathcal{A}_L$. Tím bychom však vytvořili jazyk, který dovoluje vložit více slov z jazyka R do slov z jazyka L, a to libovolně-krát.

Další chybnou úvahou je vytvoření pouze jednoho automatu pro jazyk R. To ale nestačí, protože se musíme vrátit právě do toho místa ve slově z L, kde bylo načítání přerušeno slovem z R.

Formální zápis konstrukce konečného automatu

DFA pro jazyk L je dán pěticí $\mathcal{A_L} = (Q_L, \Sigma, \delta_L, q_{L0}, F_L)$, DFA pro jazyk R pěticí $\mathcal{A_R} = (Q_R, \Sigma, \delta_R, q_{R0}, F_R)$. Pro jazyk *inject*$(L, R)$ vytvoříme NFA s ε-kroky $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$, kde

- $Q = (\{1, 2\} \times Q_L) \cup (Q_L \times Q_R)$
- $q_0 = (1, q_{L0})$
- $F = \{(2, q) \mid q \in F_L\}$
- $\delta((1, q), \varepsilon) = \{(q, q_{R0})\}$, kde $q \in Q_L$
- $\delta((p, q), \varepsilon) = \{(2, p)\}$, kde $p \in Q_L$ a $q \in F_R$
- $\delta((i, q), a) = \{(i, p) \mid p = \delta_L(q, a)\}$, kde $i \in \{1, 2\}$, $a \in \Sigma$ a $q \in Q_L$
- $\delta((p, q), a) = \{(p, r) \mid r = \delta_R(q, a)\}$, kde $a \in \Sigma$, $p \in Q_L$ a $q \in Q_R$

Poznámka: Stavy automatu $\mathcal{A}$ jsou označeny uspořádanými dvojicemi, kde první složka určuje, ve které části slova $u \cdot v \cdot w$ ($uw \in L, v \in R$) se nacházíme. Dvojice tvaru $(1, q)$ označují stavy pro část u, dvojice $(2, q)$ pro část w a dvojice (p, q), kde $p \in Q_L$, stavy pro část v.