SYSTEMS OF FORMAL MODELS AND THEIR APPLICATION

Martin Čermák
Master Degree Programme (2), FIT BUT
E-mail: xcerma16@stud.fit.vutbr.cz

Supervised by: Alexander Meduna
E-mail: meduna@fit.vutbr.cz

ABSTRACT

This paper introduces and discusses automata systems as a new way for formal languages processing. In the text we describe two formal models. The first one works in a sequentional mode. In the second one each component of automata system has its own input string. The computation step of each component is influenced each other by their states. The state of components of automata system can block or unblock some or all automata of the system.

1 ÚVOD

Překladače, resp. zpracování jazyků dnes patří k nejzákladnějším počítačovým aplikacím vůbec. Úloha spočívá v přijímání a zamítání vstupních řetězců tvořených např. sekvencí znaků. Množina všech přijatelných řetězců pak tvoří výsledný jazyk. Podle Chomského hierarchie jazyků rozeznáváme mezi čtyřmi základními typy jazyků, generovanými čtyřmi základními typy gramatik nebo přijímanými odpovídajícími typy automatů, viz [1] nebo [2].
V moderní teoretické informatice se snažíme zvýšit generativní sílu pomocí spolupráce více "slabších" gramatik, které se, ať už sekvenčně, či současně, podílí na generování řetězců. Pro podrobnější informace doporučuji [3, 4, 5].
Následující text bude pojednávat o problematice z jiného úhlu pohledu, přičemž bude zkoumat možnosti spolupráce automatů. Tu budeme označovat za práci tzv. "automatového systému", který je velice zajímavý nejen z pohledu teoretického, ale i praktického, kde využití může být nalezeno například při kombinované syntaktické analýze (viz níže), formálních verifikacích, vyhledávání vzorů v textu, či analýze biomedicínských dat.
Konvence použité v tomto textu jsou v souladu s [2].

2 AUTOMATOVÉ SYSTÉMY

Automatové systémy se musí skládat jednak ze samotných automatů, ale i z řízení, přičemž právě řízení je onen posilující faktor. Pro následující text uvažujme $I = \{1, \dots, n\}$ pro nějaké $n \geq 1$ a označme $M = (Q, \Sigma, \Gamma, \delta, F, s, z_0)$ zásobníkovým automatem, kde Q je neprázdná konečná množina stavů, Σ, Γ je vstupní, resp. zásobníková abeceda, $\delta : Q \times \Gamma \cup \{\varepsilon\} \times \Sigma \cup \{\varepsilon\} \to 2^{Q \times \Gamma^*}$ je funkcí přechodu a $F \subseteq Q$ je množinou koncových stavů. Precizní definici najdete v [2, 1].

2.1 SEKVENČNÍ AUTOMATOVÝ SYSTÉM

Nechť $M_1,\ldots,M_n \in \mathcal{M}$, kde $\mathcal{M}$ je množina všech konečných, zásobníkových, nebo Turingových automatů a $\forall i \in I$, Ψ_i je přechodová funkce (ekvivalentního typu s δ_i dle definice konkrétního typu automatu) prováděná na pozadí. Pak definujeme ***n*-sekvenční automatový systém** (n-SAS) jako $\vartheta = ((M_1,\Psi_1),\ \ldots,\ (M_n,\Psi_n))$. **Konfiguraci** takového systému pak tvoří $(n+1)$-tice $\chi = (c_1,\ldots,\ c_n,\ \omega)_{|l}$, kde c_i pro $i \in I$ značí konfiguraci dílčí komponenty (automatu) až na vstupní řetězec, ω značí dosud nepřečtenou část vstupního řetězce a $l \in \{(i,j) : j \in (\mathbb{N}_0 \cup \{*\})$ určuje počet kroků provedených nad aktivní komponentou systému, kde $*$ zastupuje význam slova jakýkoliv a i určuje aktivní komponentu systému$\}$. Nechť $\chi = (c_1,\ldots,c_n,\omega)_{|l}$ a $\chi' = (c'_1,\ldots,c'_n,\omega')_{|l'}$ jsou dvě konfigurace sekvenčního automatového systému. Dále nechť $(c_i,\omega) \vdash_{M_i} (c'_i,\omega')[r_i]$ alespoň pro jedno $i \in I$, přičemž r_i označuje použité pravidlo z δ_i pro přechod mezi stavy. Pak můžeme pomocí komponenty M_i provést **přechod** n-SAS, zapsáno $\chi \vdash_{M_i} \chi'$, přičemž pro $j \neq i$, $c'_j = c_j$. Navíc pokud $r_i \notin \Psi_i$ a $l = (i,s)$, pak $l' = (i,s+1)$, pokud $r_i \in \Psi_i$ a $l = (i,s)$, pak $l' = l$, pokud $r_i \notin \Psi_i$ a $l = (i',s)$, kde $i' \in I \wedge i \neq i'$, pak $l' = (i,1)$. Pokud $l = (i',*)$, pak $l' = (i,*)$. Pro výpočet jsou definovány čtyři **základní módy** (D). První z nich předá řízení z aktivní komponenty systému v případě, kdy neexistuje možnost dalšího přechodu aktivní komponenty ($t \in D$). Ostatní tři módy uvažují konkrétní počty přechodů nad komponentami ($=k, \geq k, \leq k \in D$). Nechť $\vartheta = ((M_1,\Psi_1),\ldots,(M_n,\Psi_n))$ je n-SAS nad množinou $\mathcal{M}$ a nechť $f \in D$, kde $D = \{*,t\} \cup \{=k, \leq k, \geq k | k \in \mathbb{N}^+\}$. Pak **jazyk přijímaný *s*-tou komponentou *n*-SAS** definujeme jako $L(\vartheta)^f_s$=$\{\omega | (\chi,\omega) \vdash^f_{M_{i_1}} (\chi_1,\omega_1) \vdash^f_{M_{i_2}} \ldots \vdash^f_{M_{i_m}} (\chi_{i_m},\omega_m)$, $m \geq 1, 1 \leq i_j \leq n,\ 1 \leq j \leq m,\ \omega,\omega_1,\ldots,\omega_m \in \Sigma^*$, χ je počáteční konfigurace, χ_{i_m} konfigurace, $\omega_m = \varepsilon$ a M_s řetězec přijal$\}$ a **jazyk přijímaný celým *n*-SAS** jako $L(\vartheta)^f_s$=$\{\omega | (\chi,\omega) \vdash^f_{M_{i_1}} (\chi_1,\omega_1) \vdash^f_{M_{i_2}} \ldots \vdash^f_{M_{i_m}} (\chi_{i_m},\omega_m), m \geq 1, 1 \leq i_j \leq n, 1 \leq j \leq m, \omega,\omega_1,\ldots,\omega_m \in \Sigma^*$, χ je počáteční konfigurace, χ_{i_m} konfigurace, $\omega_m = \varepsilon$ a $\forall i \in I$, M_i řetězec přijal$\}$.

Popsaný model je vhodný zejména k syntaktické analýze, kombinující několik klasických metod. Každá z komponent systému může pracovat jinou metodou, nad vlastní částí vstupního řetězce, což vede ke zjednodušení popisu syntaktického analyzátoru a zefektivnění jeho výpočtu. Síla modelu se odvíjí od typu a počtu komponent, přičemž při zvolení zásobníkových automatů platí, že CFL$\subset$ n-SASL pro $n > 1$, kde CFL, n-SASL, je třída bezkontextových, resp. n-SAS jazyků.

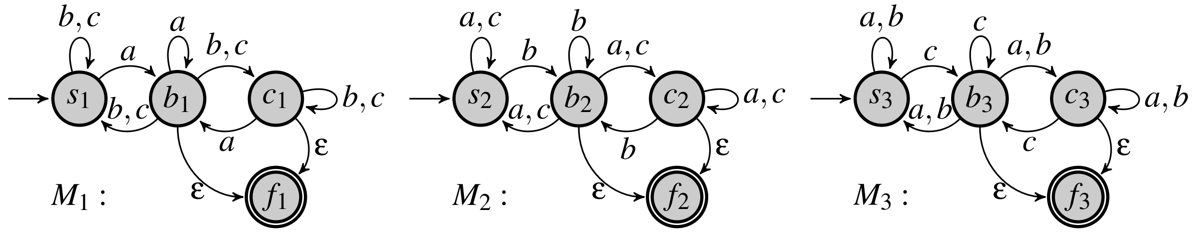
2.2 MULTIPŘIJÍMAJÍCÍ AUTOMATOVÝ SYSTÉM

Nechť $\forall i \in I$, M_i je zásobníkový automat. Pak definujeme ***n*-přijímající, stavem řízený, automatový systém** (n-MAS) jako $(n+2)$-tici $\vartheta = (M_1,\ldots,M_n,\Psi,S)$, kde Ψ je konečná množina stavových pravidel tvaru $(q_1,\ldots,q_n) \to (d_1,\ldots,\ d_n)$, $\forall i \in I$: $q_i \in Q_i$, $d_i \in \{e,d\}$, d, resp. e značí neaktivní (disable), resp. aktivní (enable) komponentu automatového systému a S je n-tice $(d^0_1,\ldots,d^0_n)$ a značí počáteční aktivitu komponent n-MAS. **Multikonfigurace** je definována jako n-tice $\chi = (x^{d_1}_1,\ldots,x^{d_n}_n)$, kde $\forall i \in I$: $x^{d_i}_i = (q_i,z_i,\omega_i)^{d_i} \in (Q_i \times \Gamma^*_i \times \Sigma^*)^{\{d,e\}}$, kde index d z x^d_i, e z x^e_i, říká, že jde o konfiguraci neaktivní, resp. aktivní komponenty n-MAS a $\omega_i \in \Sigma^*$ značí nezpracovanou část vstupního řetězce automatu M_i. Mějme n-MAS multikonfigurace $\chi = ((q_1,\gamma_1 z_1,a_1\omega_1)^{d_1},\ldots,(q_n,\gamma_n z_n,a_n\omega_n)^{d_n})$ a $\chi' = ((q'_1,z'_1,\omega'_1)^{d'_1},\ \ldots,(q'_n,z'_n,\omega'_n)^{d'_n})$, přičemž $\forall i \in I$: $q_i,q'_i \in Q_i$; $\gamma_i \in \Gamma_i \cup \{\varepsilon\}$; $z_i,z'_i,x_i \in \Gamma^*_i$; $d_i,d'_i \in \{e,d\}$, $\omega_i \in \Sigma^*$, $a_i \in \Sigma \cup \{\varepsilon\}$ a $d_i = e \Rightarrow \exists(q'_i,x_i) \in \delta_i(q_i,\gamma_i,a_i)$. Pak můžeme provést **přechod** z multikonfigurace χ do χ', píšeme $\chi \vdash \chi'$, kde $(q'_1,\ldots,q'_n) \to (d'_1,\ldots,d'_n)$, přičemž $\forall j \in I$ taková, že $d_j = d$, platí, že $q'_j = q_j$, $\omega'_j = a_j\omega_j$ a $z'_j = \gamma_j z_j$, jinak $q'_j \in Q_j$, $z'_j = x_j z_j$, kde $(q'_j,x_j) \in \delta_j(q_j,\gamma_j,a_j)$ a

$\omega'_j = \omega_j$. Nechť $\vartheta = (M_1, \dots, M_n, \Psi, S)$, kde $M_1, \dots, M_n$ jsou zásobníkové automaty přijímající v koncovém stavu (pro zásobníkové automaty přijímající s vyprázdněním zásobníku, nebo v koncovém stavu s vyprázdněním zásobníku, je definice analogická), je n-MAS a nechť $\chi_0 = ((q_1, z_1, \omega_1)^{d_1}, \dots, (q_n, z_n, \omega_n)^{d_n})$ je jeho počáteční a $\chi_f = ((q'_1, z'_1, \varepsilon)^{d'_1}, \dots, (q'_n, z'_n, \varepsilon)^{d'_n})$ koncová n-MAS multikonfigurace. Pak definujeme n-MAS **jazyk první komponenty**, $L_{first}(\vartheta)$, jako $L_{first}(\vartheta) = \{(\omega_1, \dots, \omega_n) | \chi_0 \vdash^* \chi_f; q'_1 \in F_1\}$, **jazyk sjednocení**, $L_\cup(\vartheta)$, jako $L_\cup(\vartheta) = \{(\omega_1, \dots, \omega_n) | \chi_0 \vdash^* \chi_f; q'_j \in F_j$ alespoň pro jedno $j \in I$ a **jazyk průniku**, $L_\cap(\vartheta)$, jako $L_\cap(\vartheta) = \{(\omega_1, \dots, \omega_n) | \chi_0 \vdash^* \chi_f; q'_j \in F_j$ pro všechna $j \in I$.
Definice n-MAS nad konečnými automaty, resp. Turingovými stroji jsou analogické předchozím definicím.
Tento formální model má poměrně velký potenciál. n-MAS složený z konečných automatů, muže být vhodný, mimo jiné, např. pro alternativní popis C/E Petriho sítí, čímž může přispět do oblasti verifikací. Takto navržený systém je schopen zpracovávat celou třídu regulárních jazyků a jistou podtřídu bezkontextových a kontextových jazyků. Naopak n-MAS zásobníkových automatů dosahuje síly Turingova stroje a uplatnění by jsme hledali zejména v matematické lingvistice, vyhledávání vzorů, či statické analýze.

2.2.1 PŘÍKLAD

Mějme 3-MAS konečných automatů $\vartheta = (M_1, M_2, M_3, \Psi, (e, e, e))$:



$P = \{(s_1, s_2, s_3) \to (e, e, e)$, $(b_1, s_2, s_3) \to (d, e, e)$, $(s_1, b_2, s_3) \to (e, d, e)$, $(s_1, s_2, b_3) \to (e, e, d)$, $(b_1, b_2, s_3) \to (d, d, e)$, $(b_1, s_2, b_3) \to (d, e, d)$, $(s_1, b_2, b_3) \to (e, d, d)$, $(b_1, b_2, b_3) \to (e, e, e)$, $(f_1, c_2, c_3) \to (d, e, e)$, $(f_1, f_2, c_3) \to (d, d, e)$, $(c_1, f_2, c_3) \to (e, d, e)$, $(c_1, f_2, f_3) \to (e, d, d)$, $(f_1, c_2, f_3) \to (d, e, d)\}$, $\Psi = P \cup \{(q_1, q_2, q_3) \to (d, d, d) : q_1 \in Q_1, q_2 \in Q_2, q_3 \in Q_3$, kde $\forall (a_1, a_2, a_3) \in \{e, d\} \times \{e, d\} \times \{e, d\}$, $(q_1, q_2, q_3) \to (a_1, a_2, a_3) \notin P\}$. Je snadné ukázat, že jazyk průniku, $L_\cap(\vartheta)$, je pak definován jako $L_\cap(\vartheta) = \{(\omega_1, \omega_2, \omega_2) : |\omega_1|_a = |\omega_2|_b = |\omega_3|_c\}$.

REFERENCE

[1] Meduna, A.: Automata and Languages: Theory and Applications, London, Springer, 2000.

[2] Češka, M.: Teoretická informatika 1. Učební texty, Brno, CZ, 2002.

[3] Dassow, J., Paun, C., Rozenberg, G.: Grammar Systems, Handbook of Formal Languages, Rozenberg, G. and Salomaa, A. (eds.), Volume 2. 1997.

[4] Lukáš, R.: Multigenerativní gramatické systémy. Dizertační práce, Brno, CZ, 2006.

[5] Meduna, A., Lukas, R.: Multigenerative Grammar Systems, Schedae Informaticae, Volume 15, str. 175-188, 2006.