Automaty a gramatiky (BI-AAG)

7. Zásobníkové automaty

Jan Holub

Katedra teoretické informatiky

Fakulta informačních technologií

ČVUT v Praze

Zásobníkový automat

Definice

Zásobníkový automat je sedmice $R = (Q, T, G, \delta, q_0, Z_0, F)$, kde:

- Q je konečná množina vnitřních stavů,
- T je konečná vstupní abeceda,
- G je konečná abeceda zásobníku,
- δ je zobrazení z $Q \times (T \cup \{\varepsilon\}) \times G^*$ do množiny konečných podmnožin $Q \times G^*$,
- $q_0 \in Q$ je počáteční stav,
- $Z_0 \in G$ je počáteční symbol v zásobníku,
- $F \subseteq Q$ je množina koncových stavů.

Zásobníkový automat

Konfigurace ZA R: $(q, w, \alpha) \in Q \times T^* \times G^*$, kde
q je okamžitý vnitřní stav,
w je dosud nezpracovaná část vstupního řetězce,
α je obsah zásobníku.

Počáteční konfigurace ZA R: $(q_0, w, Z_0), w \in T^*$

$\delta(q, a, \alpha) = \{(p_1, \gamma_1), (p_2, \gamma_2), ..., (p_m, \gamma_m)\}$: ZA ve stavu q přečte symbol a, přejde do stavu p_i $(i = 1, 2, ..., m)$ a řetězec α na vrcholu zásobníku nahradí řetězcem γ_i.

$\delta(q, \varepsilon, \alpha) = \{(p_1, \gamma_1), (p_2, \gamma_2), ..., (p_m, \gamma_m)\}$: přechod do nového stavu a změnu obsahu zásobníku bez čtení vstupního symbolu

Zásobníkový automat

Definice

Přechod zásobníkového automatu R je relace na množině konfigurací:
$(q, aw, \alpha\beta) \vdash (p, w, \gamma\beta)$, když $(p, \gamma) \in \delta(q, a, \alpha)$,
$a \in T \cup \{\varepsilon\}, \alpha, \beta, \gamma, \in G^*$.

$\vdash^k$: k-tá mocnina relace $\vdash$,
$\vdash^+$: tranzitivní uzávěr relace $\vdash$,
$\vdash^*$: tranzitivní a reflexivní uzávěr relace $\vdash$

Zásobníkový automat

Definice

Jazyk definovaný (přijímaný) ZA $R = (Q, T, G, \delta, q_0, Z_0, F)$:

1. *přechodem do koncového stavu*
   $L(R) = \{w : (q_0, w, Z_0) \vdash^* (q, \varepsilon, \gamma),\ \gamma \in G^*, q \in F, w \in T^*\}$,
2. s *prázdným zásobníkem*
   $L_\varepsilon(R) = \{w : (q_0, w, Z_0) \vdash^* (q, \varepsilon, \varepsilon),\ q \in Q, w \in T^*\}$.

Zásobníkový automat

Příklad

ZA, který přijímá jazyk $L(\text{ZA}) = \{ww^R : w \in \{a,b\}^*\}$:

$R = (\{q,p\}, \{a,b\}, \{a,b,S,Z\}, \delta, q, Z, \{p\})$, kde

$\delta(q,a,\varepsilon) = \{(q,a)\}$,

$\delta(q,b,\varepsilon) = \{(q,b)\}$,

$\delta(q,\varepsilon,\varepsilon) = \{(q,S)\}$,

$\delta(q,\varepsilon,aSa) = \{(q,S)\}$,

$\delta(q,\varepsilon,bSb) = \{(q,S)\}$,

$\delta(q,\varepsilon,SZ) = \{(p,\varepsilon)\}$.

Pro vstupní řetězec $aabbaa$ provede automat R tuto posloupnost přechodů:

$$
\begin{array}{llll}
 & (q,aabbaa,Z) & \vdash (q,abbaa,aZ) & \vdash (q,bbaa,aaZ) \\
\vdash & (q,baa,baaZ) & \vdash (q,baa,SbaaZ) & \vdash (q,aa,bSbaaZ) \\
\vdash & (q,aa,SaaZ) & \vdash (q,a,aSaaZ) & \vdash (q,a,SaZ) \\
\vdash & (q,\varepsilon,aSaZ) & \vdash (q,\varepsilon,SZ) & \vdash (p,\varepsilon,\varepsilon)
\end{array}
$$

Základní vlastnosti ZA

Věta

Nechť $P = (Q, T, G, \delta, q_0, Z_0, F)$ je ZA. Jestliže $(q, w, A) \vdash_P^n (q', \varepsilon, \varepsilon)$, pak $(q, w, A\,\alpha) \vdash_P^n (q', \varepsilon, \alpha)$ pro všechna $A \in G$ a $\alpha \in G^*$.

Základní vlastnosti ZA

Věta

L je jazyk přijímaný ZA P_1 s prázdným zásobníkem právě tehdy, když L je jazyk přijímaný ZA P_2 přechodem do koncového stavu.

Důkaz: Nejdříve dokážeme, že z $L = L(P_2) \Rightarrow L = L_\varepsilon(P_1)$.
Nechť $P_2 = (Q, T, G, \delta, q_0, Z_0, F)$ je ZA takový, že $L = L(P_2)$.
Nechť $P_1 = (Q \cup \{q_\varepsilon, q_0'\}, T, G \cup \{X\}, \delta', q_0', X, \emptyset)$, kde δ' je definováno takto:

1. $\delta'(q_0', \varepsilon, X) = \{(q_0, Z_0X)\}$,
2. $\delta'(q, a, Z) = \delta(q, a, Z), \forall q, a, Z, q \in Q, a \in T \cup \{\varepsilon\}, Z \in G^*$,
3. $\delta'(q, \varepsilon, Z) = \{(q_\varepsilon, \varepsilon)\}, \forall q, Z, q \in F, Z \in G \cup \{X\}$,
4. $\delta'(q_\varepsilon, \varepsilon, Z) = \{(q_\varepsilon, \varepsilon)\}, \forall Z, Z \in G \cup \{X\}$.

Základní vlastnosti ZA

Věta
L je jazyk přijímaný ZA P_1 s prázdným zásobníkem právě tehdy, když L je jazyk přijímaný ZA P_2 přechodem do koncového stavu.

Důkaz (pokrač.): Nyní dokážeme, že z $L = L_\varepsilon(P_1) \Rightarrow L = L(P_2)$.
Nechť $P_1 = (Q, T, G, \delta, q_0, Z_0, \emptyset)$ je ZA takový, že $L = L_\varepsilon(P_1)$.
Nechť $P_2 = (Q \cup \{q_0', q_f\}, T, G \cup \{X\}, \delta, q_0', X, \{q_f\})$, kde δ' je definována takto:

1. $\delta'(q_0', \varepsilon, X) = \{(q_0, Z_0 X)\}$,
2. $\delta'(q, a, Z) = \delta(q, a, Z), \forall q, a, Z, q \in Q, a \in T \cup \{\varepsilon\}, Z \in G^*$,
3. $\delta'(q, \varepsilon, X) = \{(q_f, \varepsilon)\}, \forall q, q \in Q$.

Syntaktická analýza

Definice

Je dána $G = (N, T, P, S)$. Derivace $S \Rightarrow^* \alpha, \alpha \in (N \cup T)^*$ se nazývá *levá derivace* (*pravá derivace*), jestliže při každém kroku byl nahrazován nejlevější (nejpravější) neterminální symbol ve větné formě.

Derivačnímu stromu odpovídá jediná levá (pravá) derivace a naopak určité levé (pravé) derivaci odpovídá jediný derivační strom. Existuje proto lineární reprezentace derivačního stromu, nazývaná *rozklad*.

Definice

Je dána $G = (N, T, P, S)$. Předpokládejme, že pravidla v P jsou očíslována $1, 2, \ldots, |P|$. *Rozkladem větné formy* α v G je posloupnost čísel pravidel použitých v derivaci $S \Rightarrow^* \alpha$.
Levým rozkladem větné formy α v G je pak posloupnost čísel pravidel použitých v levé derivaci $S \Rightarrow^* \alpha$.
Pravým rozkladem větné formy α v G je pak obrácená posloupnost čísel pravidel použitých v pravé derivaci $S \Rightarrow^* \alpha$.

Syntaktická analýza

Syntaktická analýza (rozklad) = konstrukce derivačního stromu

Metody syntaktické analýzy:

1. *shora dolů* (top down),
2. *zdola nahoru* (bottom up).

Definice

Syntaktická analýza metodou shora dolů je proces nalezení levého rozkladu dané věty v dané gramatice.

Definice

Syntaktická analýza metodou zdola nahoru je proces nalezení pravého rozkladu dané věty v dané gramatice.

Vztah BG a ZA

Věta
Je-li dána BG $G = (N, T, P, S)$, můžeme sestrojit ZA R takový, že $L(G) = L(R)$.

A. Konstrukce ZA (model syntaktické analýzy metodou shora dolů):
$R = (\{q\}, T, N \cup T, \delta, q, S, \emptyset)$, kde δ:

1. $\delta(q, \varepsilon, A) = \{(q, \alpha) : A \rightarrow \alpha \in P\}, \forall A, A \in N$ (**expanze**),
2. $\delta(q, a, a) = \{(q, \varepsilon)\}, \forall a, a \in T$ (**srovnání**).

Vrchol zásobníku u tohoto typu automatu bude vždy vlevo.

Vztah BG a ZA

Příklad

BG $G = (\{E, T, F\}, \{+, *, (,), a\}, P, E)$, kde P:

$(1)\ E \to E + T$ $(2)\ E \to T$ $(3)\ T \to T * F$

$(4)\ T \to F$ $(5)\ F \to (E)$ $(6)\ F \to a$.

ZA $R = (\{q\}, \{+, *, (,), a\}, \{+, *, (,), a, E, T, F\}, \delta, q, E, \emptyset)$, kde δ:

$$\begin{aligned}
\delta(q, \varepsilon, E) &= \{(q, E + T), (q, T)\} \\
\delta(q, \varepsilon, T) &= \{(q, T * F), (q, F)\} \\
\delta(q, \varepsilon, F) &= \{(q, (E)), (q, a)\} \\
\delta(q, b, b) &= \{(q, \varepsilon)\}, \forall b, b \in \{a, +, *, (,)\}.
\end{aligned}$$

Vztah BG a ZA

Příklad (pokrač.)

Věta $a + a * a$ má v gramatice G tuto levou derivaci:

$$\begin{aligned}
E &\Rightarrow E + T && (1)\\
&\Rightarrow T + T && (2)\\
&\Rightarrow F + T && (4)\\
&\Rightarrow a + T && (6)\\
&\Rightarrow a + T * F && (3)\\
&\Rightarrow a + F * F && (4)\\
&\Rightarrow a + a * F && (6)\\
&\Rightarrow a + a * a && (6)
\end{aligned}$$

Vztah BG a ZA

Příklad (pokrač.)

$$
\begin{array}{llrrrr}
(q, a+a*a, E) & \vdash & (q, & a+a*a, & E+T) & (1) \\
& \vdash & (q, & a+a*a, & T+T) & (2) \\
& \vdash & (q, & a+a*a, & F+T) & (4) \\
& \vdash & (q, & a+a*a, & a+T) & (6) \\
& \vdash & (q, & +a*a, & +T) & \\
& \vdash & (q, & a*a, & T) & \\
& \vdash & (q, & a*a, & T*F) & (3) \\
& \vdash & (q, & a*a, & F*F) & (4) \\
& \vdash & (q, & a*a, & a*F) & (6) \\
& \vdash & (q, & *a, & *F) & \\
& \vdash & (q, & a, & F) & \\
& \vdash & (q, & a, & a) & (6) \\
& \vdash & (q, & \varepsilon, & \varepsilon) &
\end{array}
$$

Vztah BG a ZA

Příklad (pokrač.)
Levý rozklad vstupní věty $a + a * a$: 1, 2, 4, 6, 3, 4, 6, 6.

Vztah BG a ZA

B. Konstrukce ZA (model syntaktické analýzy metodou zdola nahoru):
$R = (\{q, r\}, T, N \cup T \cup \{\#\}, \delta, q, \#, \{r\})$, kde δ:

1. $\delta(q, a, \varepsilon) = \{(q, a)\}$, $\forall a, a \in T$, (**přesun**),
2. $\delta(q, \varepsilon, \alpha) = \{(q, A) : A \to \alpha \in P\}$ (**redukce**),
3. $\delta(q, \varepsilon, \#S) = \{(r, \varepsilon)\}$ (**přijetí**).

Oproti definici zásobníkového automatu a jeho konfigurací bude u tohoto typu zásobníkového automatu **vrchol zásobníku vždy vpravo**.

Vztah BG a ZA

Příklad

Je dána BG $G = (\{E, T, F\}, \{+, *, (,), a\}, P, E)$, kde P:

$(1)\ E \rightarrow E + T$ $(2)\ E \rightarrow T$ $(3)\ T \rightarrow T * F$

$(4)\ T \rightarrow F$ $(5)\ F \rightarrow (E)$ $(6)\ F \rightarrow a$.

$R = (\{q, r\}, \{+, *, (,), a\}, \{E, T, F, +, *, (,), a, \#\}, \delta, q, \#, \{r\})$,

kde:

$\delta(q, b, \varepsilon) = \{(q, b)\}, \forall b, b \in \{a, +, *, (,)\}$

$\delta(q, \varepsilon, E + T) = \{(q, E)\}$

$\delta(q, \varepsilon, T) = \{(q, E)\}$

$\delta(q, \varepsilon, T * F) = \{(q, T)\}$

$\delta(q, \varepsilon, F) = \{(q, T)\}$

$\delta(q, \varepsilon, (E)) = \{(q, F)\}$

$\delta(q, \varepsilon, a) = \{(q, F)\}$

$\delta(q, \varepsilon, \#E) = \{(r, \varepsilon)\}$.

Vztah BG a ZA

Příklad (pokrač.)

Věta $a + a * a$ má v gramatice G tuto pravou derivaci:

$$
\begin{aligned}
E &\Rightarrow E + T && (1)\\
&\Rightarrow E + T * F && (3)\\
&\Rightarrow E + T * a && (6)\\
&\Rightarrow E + F * a && (4)\\
&\Rightarrow E + a * a && (6)\\
&\Rightarrow T + a * a && (2)\\
&\Rightarrow F + a * a && (4)\\
&\Rightarrow a + a * a && (6)
\end{aligned}
$$

Vztah BG a ZA

Příklad (pokrač.)

$$
\begin{array}{llrlrl}
(q, a+a*a, \#) & \vdash & (q, & +a*a, & \#a &) \\
 & \vdash & (q, & +a*a, & \#F &) \quad (6) \\
 & \vdash & (q, & +a*a, & \#T &) \quad (4) \\
 & \vdash & (q, & +a*a, & \#E &) \quad (2) \\
 & \vdash & (q, & a*a, & \#E+ &) \\
 & \vdash & (q, & *a, & \#E+a &) \\
 & \vdash & (q, & *a, & \#E+F &) \quad (6) \\
 & \vdash & (q, & *a, & \#E+T &) \quad (4) \\
 & \vdash & (q, & a, & \#E+T* &) \\
 & \vdash & (q, & \varepsilon, & \#E+T*a &) \\
 & \vdash & (q, & \varepsilon, & \#E+T*F &) \quad (6) \\
 & \vdash & (q, & \varepsilon, & \#E+T &) \quad (3) \\
 & \vdash & (q, & \varepsilon, & \#E &) \quad (1) \\
 & \vdash & (r, & \varepsilon, & \varepsilon &)
\end{array}
$$

Vztah BG a ZA

Příklad (pokrač.)
Pravý rozklad vstupní věty $a + a * a$: 6, 4, 2, 6, 4, 6, 3, 1.

Deterministický ZA

Definice

Zásobníkový automat $R = (Q, T, G, \delta, q_0, Z_0, F)$ je *deterministický*, jestliže platí:

1. $|\delta(q, a, \gamma)| \leq 1, \forall q, a, \gamma, q \in Q, a \in (T \cup \{\varepsilon\}), \gamma \in G^*$.
2. Je-li $\delta(q, a, \alpha) \neq \emptyset$, $\delta(q, a, \beta) \neq \emptyset$ a $\alpha \neq \beta$, pak α není příponou β a β příponou α (tzn. $\gamma\alpha \neq \beta, \alpha \neq \gamma\beta$).
3. Je-li $\delta(q, a, \alpha) \neq \emptyset$, $\delta(q, \varepsilon, \beta) \neq \emptyset$, pak α není příponou β a β příponou α (tzn. $\gamma\alpha \neq \beta, \alpha \neq \gamma\beta$).

Deterministický ZA

Konstrukce det. ZA metodou zhora dolů (**A**) pro BG v normálním tvaru podle Greibachové:

(normální tvar podle Greibachové: všechna pravidla ve tvaru $A \rightarrow a\alpha$, kde $a \in T, \alpha \in N^*$)

$R = (\{q\}, T, N, \delta, q, S, \emptyset)$, kde
$\delta(q, a, A) = \{(q, \alpha) : A \rightarrow a\alpha \in P\}$, $\forall A, A \in N$.

Deterministický ZA

Konstrukce det. ZA metodou zdola nahoru (**B**):

Příklad

BG $G = (\{S, A, B\}, \{a, b, c, d\}, P, S)$, kde P:

$S \to Aa \qquad A \to Bb \mid c \qquad B \to d$

$R = (\{q, r\}, \{a, b, c, d\}, \{S, A, B, a, b, c, d, \#\}, \delta, q, \#, \{r\})$, kde δ:

$$
\begin{array}{ll}
(1) & \delta(q, a, \varepsilon) = (q, a) \\
 & \delta(q, b, \varepsilon) = (q, b) \\
 & \delta(q, c, \varepsilon) = (q, c) \\
 & \delta(q, d, \varepsilon) = (q, d) \\
(2) & \delta(q, \varepsilon, Aa) = (q, S) \\
 & \delta(q, \varepsilon, Bb) = (q, A) \\
 & \delta(q, \varepsilon, c) = (q, A) \\
 & \delta(q, \varepsilon, d) = (q, B) \\
(3) & \delta(q, \varepsilon, \#S) = (r, \varepsilon)
\end{array}
$$

Vztah BG a ZA

Příklad (pokrač.)

ZA je nedeterministický vlivem přesunů podle (1). Tyto přesuny je možné provádět v závislosti na obsahu zásobníku:

$(1)'$ $\delta(q, a, A) = \{(q, Aa)\}$ – symbol a se vyskytuje ve větné formě jen za symbolem A,

$\delta(q, b, B) = \{(q, Bb)\}$ – symbol b se ve větné formě může vyskytnout jen za symbolem B,

$\delta(q, c, \#) = \{(q, \#c)\}$,

$\delta(q, d, \#) = \{(q, \#d)\}$ – symboly c, d se mohou vyskytnout pouze na začátku větné formy.