6. přednáška LS 2006/2007

- Shrnutí předchozí přednášky.
- Témata dalších čtyř referátů (redukované gramatiky, Chomského normální forma, pumping lemmata, převod vícestavového ZA na 1-stavový).
- Zásobníkové automaty, ekvivalence s bezkontextovými gramatikami.
- Uzávěrové vlastnosti třídy bezkontextových jazyků (CFL)
- Deterministické zásobníkové automaty
- Další diskuse 2. zápočtové písemky

Shrnutí předchozí přednášky

- Probrali jsme pojem bezkontextová gramatika (BG). Víme, jak je definována, jak reprezentuje (generuje) jazyk, máme 'před očima' konkrétní příklady, např. z oblasti programovacích jazyků. Dobře rozumíme pojmu derivační strom (a odpovídající levé derivaci).
- Víme, že BG jsou prostředkem pro popis tzv. bezkontextových jazyků, což je širší třída než třída regulárních jazyků.
- Máme docela dobrou představu o tom, že gramatika má obvykle zachycovat nejen syntaxi jazyka (čili popisovat, která slova do jazyka patří), ale i sémantiku (např. vyhodnocování výrazů [interpret], překlad do jiného (programovacího) jazyka [kompilátor], ...).
- V souvislosti se sémantikou je nám jasná důležitost pojmu jednoznačných gramatik (a jazyků). Máme představu o tom, jak se některé jednoduché ('programátorské') nejednoznačné gramatiky převedou na jednoznačné.

Shrnutí předchozí přednášky - pokrač.

- Máme dobrou představu o tom, že implementace algoritmů, na nichž je založena syntaktická analýza (a vyhodnocení sémantiky), je obvykle založena na konstrukci (a vyhodnocení) derivačního stromu pro zadané vstupní slovo.
- Přitom není nutné celý derivační strom konstruovat, stačí vhodně realizovat např. levý průchod stromem; k tomu se pěkně hodí struktura, kterou nazýváme zásobníkový automat.

Redukované gramatiky (referát na cvičení)

Bezkontextová gramatika
$G = (\Pi, \Sigma, S, P)$ je *redukovaná* $\Leftrightarrow_{df}$ každý $X \in \Pi$ splňuje:

1. $\exists w \in \Sigma^* : X \Rightarrow^* w$,
2. $\exists \alpha, \beta \in (\Pi \cup \Sigma)^* : S \Rightarrow^* \alpha X \beta$.

Konstrukce $\mathcal{T} = \{ X \in \Pi \mid X$ splňuje 1. $\}$:
$\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3 \ldots$, kde
$\mathcal{T}_1 = \{ X \mid \exists w \in \Sigma^* : (X \to w) \in P \}$
$\mathcal{T}_{i+1} = \mathcal{T}_i \cup \{ X \mid \exists \alpha \in (\Sigma \cup \mathcal{T}_i)^* : (X \to \alpha) \in P \}$
$\mathcal{T}_n = \mathcal{T}_{n+1} \Longrightarrow \mathcal{T}_n = \mathcal{T} \quad (n \leq |\Pi|)$

Konstrukce $\mathcal{D} = \{ X \in \Pi \mid X$ splňuje 2. $\}$:
$\mathcal{D}_1 = \{ S \}$
$\mathcal{D}_{i+1} = \mathcal{D}_i \cup \{ X \mid \exists Y \in D_i, \alpha_1, \alpha_2 : \quad (Y \to \alpha_1 X \alpha_2) \in P \}$

Redukované gramatiky - pokr.

Gramatiky G_1, G_2 jsou *ekvivalentní* $\Leftrightarrow_{df}$ $L(G_1) = L(G_2)$.

Věta. Ke každé BG G, pro niž platí $L(G) \neq \emptyset$, lze sestrojit ekvivalentní redukovanou gramatiku.

Důkaz. Odstraníme všechna pravidla obsahující neterminály nesplňující podmínku 1. ($\exists w \in \Sigma^* : X \Rightarrow^* w$).
Pak odstraníme všechna pravidla obsahující neterminály nesplňující podmínku 2. ($\exists \alpha, \beta \in (\Pi \cup \Sigma)^* : S \Rightarrow^* \alpha X \beta$).
Proč je postup korektní ?

Věta. Existuje algoritmus, který pro zadanou BG G rozhodne, zda $L(G) = \emptyset$.

Poznámka. Ekvivalenci dvou BG nelze algoritmicky rozhodovat.
(Není obdoba konstrukce minim. KA.)

Příklad redukce gramatiky

Zredukujme gramatiku

$S \longrightarrow aSb \mid aAbb \mid \varepsilon$

$A \longrightarrow aAB \mid bB$

$B \longrightarrow aAb \mid BB$

$C \longrightarrow CC \mid cS$

Nevypouštějící bezkontextové gramatiky

Bezkontextová *gramatika* se nazývá
nevypouštějící $\Leftrightarrow_{df}$ neobsahuje pravidlo typu $X \longrightarrow \varepsilon$.

Věta. Ke každé BG G lze sestrojit nevypouštějící gramatiku G' tž.
$L(G') = L(G) - \{\varepsilon\}$.

Důkaz.
Konstrukce $\mathcal{E} = \{\, X \in \Pi \mid X \Rightarrow^* \varepsilon \,\}$:
$\mathcal{E}_1 = \{\, X \mid (X \rightarrow \varepsilon) \in P \,\}$
$\mathcal{E}_{i+1} = \mathcal{E}_i \cup \{\, X \mid \exists \alpha \in \mathcal{E}_i^* : (X \rightarrow \alpha) \in P \,\}$
Pro každé pravidlo $(X \rightarrow \alpha) \in P$ zařadíme do P' všechna pravidla $X \rightarrow \beta$, kde $\beta \neq \varepsilon$ a β vznikne z α vypuštěním některých (třeba žádných) výskytů symbolů z $\mathcal{E}$.

Důsledek. Ke každé BG $G = (\Pi, \Sigma, S, P)$ existuje ekvivalentní BG $G_1 = (\Pi_1, \Sigma, S_1, P_1)$, kde ε může být pravou stranou pouze u pravidla $S_1 \rightarrow \varepsilon$; v takovém případě se pak S_1 nevyskytuje na pravé straně žádného z pravidel z P_1.

Příklad převodu na nevypouštějící gramatiku

K bezkontextové gramatice G dané uvedenými pravidly sestrojte nevypouštějící gramatiku G' takovou, že $L(G') = L(G) - \{\varepsilon\}$.

$$\begin{aligned}
S &\longrightarrow AB \mid \varepsilon \\
A &\longrightarrow aAAb \mid BS \mid CA \\
B &\longrightarrow BbA \mid CaC \mid \varepsilon \\
C &\longrightarrow aBB \mid bS
\end{aligned}$$

Chomského normální forma - referát na cvičení

BG je v *Chomského normální formě* (ChNF) $\Leftrightarrow_{df}$ každé pravidlo je tvaru $X \rightarrow YZ$ nebo $X \rightarrow a$.

Věta. Ke každé BG G lze sestrojit BG G' v CHNF tž. $L(G') = L(G) - \{\varepsilon\}$.

Důkaz.
- převod na nevypouštějící gramatiku
- odstranění pravidel typu $X \rightarrow Y$
- pro každé a přidáme A_a a pravidlo $A_a \rightarrow a$; na každé pravé straně delší než 1 nahradíme a neterminálem A_a
- každé pravidlo typu $X \rightarrow Y_1 Y_2 \dots Y_n$, kde $n \geq 3$, nahradíme $X \rightarrow Y_1 Z_1$, $Z_1 \rightarrow Y_2 Z_2$, ..., $Z_{n-3} \rightarrow Y_{n-2} Z_{n-2}$, $Z_{n-2} \rightarrow Y_{n-1} Y_n$
($Z_1, Z_2, \dots, Z_{n-2}$ nově přidané)

Odstranění pravidel typu $X \rightarrow Y$:
- Pro každý neterm. A zkonstruujeme

$$\mathcal{D}_A = \{\, B \mid A \Rightarrow^* B \,\}$$

- Pak pro každé $B \rightarrow \alpha$, kde $B \in \mathcal{D}_A$ a α není jeden neterminál, přidáme pravidlo $A \rightarrow \alpha$.
- Odstraníme všechna pravidla typu $X \rightarrow Y$.

Příklad převodu do Chomského normální formy

Následující gramatiku převeďte do Chomského normální formy:

$$
\begin{aligned}
&S \longrightarrow (E)\\
&E \longrightarrow F + F \mid F \times F\\
&F \longrightarrow a \mid S
\end{aligned}
$$

Pumping lemmata - referát na cvičení

Prostředky k důkazu neregularity či nebezkontextovosti daného jazyka.

U regulárních jazyků lze základní myšlenku neformálně vyjádřit takto:

> *V každém „dostatečně dlouhém“ slově regulárního jazyka L existuje „krátké“ neprázdné podslovo, zaručeně i „blízko začátku“, jehož vynecháním či „pumpováním“ dostáváme vždy slova jazyka L.*

(Např. $\{a^n b^n \mid n \geq 0\}$ není regulární.)

U bezkontextových jazyků lze základní myšlenku neformálně vyjádřit takto:

> *V každém „dostatečně dlouhém“ slově bezkontextového jazyka L existují někde „blízko sebe“ dvě „krátká“ (nepřekrývající se) slova, z nichž alespoň jedno je neprázdné a jejichž současným vynecháním či „pumpováním“ dostáváme zase slova jazyka L.*

(Např. $\{a^n b^n c^n \mid n \geq 0\}$ či $\{0^m 1^n 0^m 1^n \mid m, n \geq 0\}$ nejsou bezkontextové.)

Pumping lemma pro regulární jazyky

Věta. (Pumping lemma.) Nechť L je regulární jazyk. Pak nutně existuje $n \in \mathbb{N}$ tž. každé slovo $z \in L$, $|z| \geq n$, lze psát $z = uvw$, kde $|uv| \leq n$, $|v| \geq 1$ a pro vš. $i \geq 0$ je $uv^i w \in L$.

$\big((\forall L$ tž. L je regulární $) \big)$

$(\exists n \in \mathbb{N})$
$(\forall z$ tž. $z \in L, |z| \geq n)$
$(\exists\ u, v, w$ tž. $z = uvw, |uv| \leq n, |v| \geq 1)$
$(\forall i \geq 0) \qquad uv^i w \in L$

'Zkratky':
$(\forall x$ tž. $A)\ B$ ..... $(\forall x : A \Rightarrow B)$
$(\exists x$ tž $A)\ B$ ..... $(\exists x : A \wedge B)$

Myšlenka důkazu: Konečný automat s n stavy během čtení (začátku) slova délky n projde nutně cyklem (který samozřejmě lze vypustit či opakovat).

Hra dvou hráčů pro důkaz neregularity zadaného L

Hráč A chce ukázat $(\exists n \in \mathbb{N})$ $(\forall z \text{ tž. } z \in L, |z| \geq n)$
$(\exists\ u, v, w \text{ tž. } z = uvw, |uv| \leq n, |v| \geq 1)$ $(\forall i \geq 0)$: $uv^iw{\in}L$

B chce ukázat negaci, tedy $(\forall n \in \mathbb{N})$ $(\exists z \text{ tž. } z \in L, |z| \geq n)$
$(\forall\ u, v, w \text{ tž. } z = uvw, |uv| \leq n, |v| \geq 1)$ $(\exists i \geq 0)$: $uv^iw \notin L$

1. A zvolí $n \in \mathbb{N}$
2. B zvolí slovo z tž. $z \in L$ a $|z| \geq n$ (nelze-li, A vyhrál)
3. A zvolí u, v, w tž. $z = uvw$, $|uv| \leq n$ a $|v| \geq 1$
4. B zvolí $i \geq 0$
5. *Výsledek:* je-li $uv^iw \in L$, vyhrál A, v případě $uv^iw \notin L$ vyhrál B.

Tvrzení. Má-li B vítěznou strategii, pak L není regulární.

Příklad důkazu neregularity

Pro $L_1 = \{a^j b^j \mid j \geq 0\}$ má B např. tuto vítěznou strategii :

1. A zvolí (libovolné) $n \in \mathbb{N}$
2. B zvolí $z = a^n b^n$
3. A zvolí libovolné u, v, w tž. $z = uvw$, $|uv| \leq n$ a $|v| \geq 1$
   (tedy $u = a^j$, $v = a^k$ pro nějaké j, k tž. $j + k \leq n$, $k \geq 1$)
4. B: zvolí $i = 0$ (lze kterékoli $i \neq 1$)
5. Jelikož $a^j a^{n-(j+k)} b^n \notin L$, B vyhrává.

Poznámka. Uvedená strategie funguje i pro jazyk
$L_2 = \{\, w \in \{a, b\}^* \mid |w|_a = |w|_b \,\}$.

(Ovšem už fakt, že $L_2 \cap [a^* b^*] = L_1$ spolu s tím, že L_1 je neregulární, implikuje, že L_2 je neregulární.)

Pumping lemma pro bezkontextové jazyky

Věta. ($uvwxy$-teorém)
Nechť L je bezkontextový jazyk. Pak existují přirozená čísla p, q tž. každé slovo $z \in L$ delší než p lze psát ve tvaru $z = uvwxy$, přičemž platí

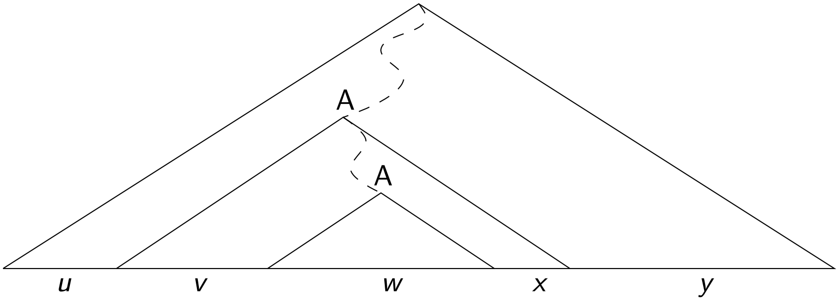
- $vx \neq \varepsilon$,
- $|vwx| \leq q$,
- pro vš. $i \geq 0$ je $uv^i wx^i y \in L$.

Lze uvažovat $n = max(p, q) + 1$:

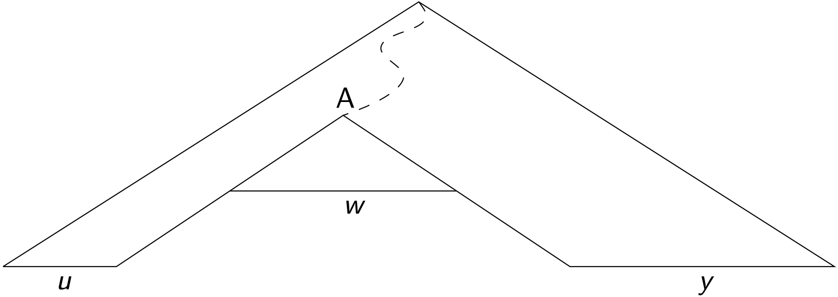
Věta. Nechť L je bezkontextový jazyk. Pak existuje přirozené číslo n tž. každé slovo $z \in L$, $|z| \geq n$, lze psát ve tvaru $z = uvwxy$, přičemž platí

- $vx \neq \varepsilon$,
- $|vwx| \leq n$,
- pro vš. $i \geq 0$ je $uv^i wx^i y \in L$.

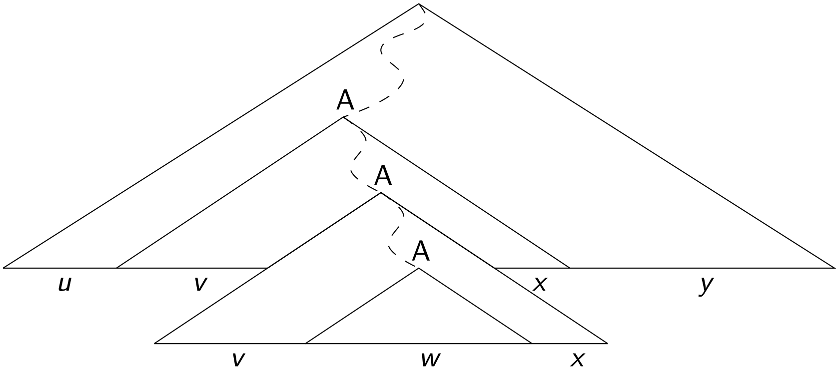
'Obrázkový' důkaz



Vypuštění



Napumpování



Precizace obrázkového důkazu (pro hlubší zájemce)

Důkaz. Nechť $L = L(G)$ pro BG
$G = (\Pi, \Sigma, S, P)$ v Chomského NF.
Všimněme si: když na větvi deriv. stromu pro $z \in L$ jsou dva výskyty téhož neterminálu, řekněme A, pak
$S \Rightarrow^* uAy \Rightarrow^* uvAxy \Rightarrow^* uvwxy = z$
pro nějaké $u, v, w, x, y \in \Sigma^*$, $vx \neq \varepsilon$.
Nechť $|\Pi| = k$. Všimněme si dále:
- na větvi délky alespoň $k+1$ jsou jistě aspoň dva výskyty téhož neterminálu;
- má-li derivační strom pro $z \in \Sigma^*$ všechny větve kratší než $k+1$, pak nutně $|z| \leq 2^{k-1}$ (počet listů binárního stromu hloubky $k-1$);
- v deriv. stromě pro $z \in L$, $|z| > 2^{k-1}$, určitě existují dva různé vrcholy v_1, v_2 na stejné větvi (v_1 blíž ke kořeni) označené stejným neterminálem, přičemž podstrom s kořenem v_1 má hloubku nejvýš $k+1$ – tedy má nejvýše 2^k listů.

Můžeme proto vzít: $p = 2^{k-1}$, $q = 2^k$.

Pumping lemmma - hra dvou hráčů

Když L bezkontextový, tak:
$(\exists n \in \mathbb{N})$
$(\forall z$ tž. $z \in L, |z| \geq n)$
$(\exists\ u, v, w, x, y$ tž. $\ z = uvwxy, |vwx| \leq n, vx \neq \varepsilon)$
$(\forall i \geq 0) : uv^i wx^i y \in L$

Tvrzení. Má-li B vítěznou strategii v následující hře, pak L není bezkontextový.

1. A zvolí $n \in \mathbb{N}$
2. B zvolí slovo z tž. $z \in L$ a $|z| \geq n$
3. A zvolí u, v, w, x, y tž.
   $z = uvwxy$, $|vwx| \leq n$ a $vx \neq \varepsilon$
4. B zvolí $i \geq 0$
5. *Výsledek:* je-li $uv^i wx^i y \in L$, vyhrál A, je-li $uv^i wx^i y \notin L$, vyhrál B.

Příklady důkazu nebezkontextovosti

Strategie B pro $L = \{ w \in \{a, b, c\}^* \mid |w|_a = |w|_b = |w|_c \}$:

1. A zvolí (libovolné) $n \in \mathbb{N}$
2. B zvolí $z = a^n b^n c^n$
3. A zvolí libovolné u, v, w, x, y
   tž. $z = uvwxy$, $|vwx| \leq n$ a $vx \neq \varepsilon$,
4. B zvolí $i = 0$ (lze kterékoli $i \neq 1$)
5. Jelikož $|vwx| \leq n$, slova v, x neobsahují aspoň jeden ze symbolů a, b, c. Proto uwy nepatří do L; vyhrává B.

Příklady důkazu nebezkontextovosti

Strategie B pro $L = \{ww \mid w \in \{0,1\}^*\}$

1. A zvolí (libovolné) $n \in \mathbb{N}$
2. B zvolí $z = 0^n 1^n 0^n 1^n$
3. A zvolí libovolné u, v, w, x, y
   tž. $z = uvwxy$, $|vwx| \leq n$ a $vx \neq \varepsilon$,
4. B zvolí $i = 0$ (lze kterékoli $i \neq 1$)
5. Jelikož $|vwx| \leq n$, slova v, x zasahují do nejvýš jednoho úseku nul a nejvýš jednoho úseku jedniček. Tedy $uwy = 0^{k_1} 1^{k_2} 0^{\ell_1} 1^{\ell_2}$,
   kde $k_1 \neq \ell_1$ nebo $k_2 \neq \ell_2$.
   Tedy $uwy \notin L$; vyhrává B.

Příklad z ukázkové 2. zápočtové písemky

3. Vyjmenujte, které z daných jazyků ($|w|$ označuje délku slova w, $|w|_a$ označuje počet výskytů symbolu a ve w; w^R značí zrcadl. obraz slova w)

$$L_1 = \{w \in \{a\}^* \mid |w| \textit{ je druhou mocninou celého čísla } \},$$
$$L_2 = \{0^m 1^n 0^m \mid n \leq m\},$$
$$L_3 = \{w \in \{a,b\}^* \mid |w|_a \textit{ je dělitelné pěti } \},$$
$$L_4 = \{w \in \{a,b\}^* \mid w \textit{ obsahuje podslovo abba nebo končí bbb}\}$$

jsou regulární:

jsou bezkontextové, ale ne regulární:

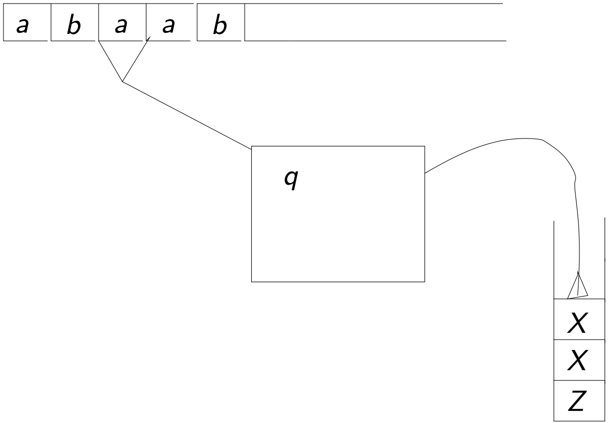
nejsou bezkontextové:

Zařazení konkrétních jazyků

Zjistěte, které z daných jazyků
jsou regulární:
jsou bezkontextové, ale ne regulární:
nejsou bezkontextové:

$L_1 = \{w \in \{a,b\}^* \mid |w|_a = |w|_b\}$
$L_2 = \{w \in \{a,b\}^* \mid |w|_a \text{ je sudé } \}$
$L_3 = \{w \in \{a,b\}^* \mid w \text{ obsahuje podslovo } abba\}$
$L_4 = \{w \in \{a,b,c\}^* \mid |w|_a = |w|_b = |w|_c\}$
$L_5 = \{w \in \{a,b\}^* \mid |w|_a \text{ je prvočíslo}\}$
$L_6 = \{\, 0^m 1^n \mid m \leq 2n \,\}$
$L_7 = \{\, 0^m 1^n 0^m \mid m = 2n \,\}$

Zásobníkové automaty



Zásobníkový automat - definice

Zásobníkový automat (ZA) (nedeterministický !)
$M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0)$:
Q ... konečná množina *stavů*,
Σ ... konečná *vstupní abeceda*,
Γ ... konečná *zásobníková abeceda*,
$q_0 \in Q$ ... *počáteční stav*,
$Z_0 \in \Gamma$ ... *poč. zásobníkový symbol*
$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \to \mathcal{P}_{fin}(Q \times \Gamma^*)$.

Konfigurace ZA M je trojice (q, w, α), kde $q \in Q$, $w \in \Sigma^*$, $\alpha \in \Gamma^*$.
Na množině konfigurací definujeme relaci $\vdash_M$:
$(q, aw, X\beta) \vdash_M (q', w, \alpha\beta) \quad \Leftrightarrow_{df} \quad \delta(q, a, X) \ni (q', \alpha)$
$(a \in (\Sigma \cup \{\varepsilon\})$, $w \in \Sigma^*$, $\beta \in \Gamma^*)$

Jazyky rozpoznávané zásobníkovými automaty

Relace $\vdash_M^*$ je reflexivním a tranzitivním uzávěrem relace $\vdash_M$.

Slovo $w \in \Sigma^*$ *je přijímáno* ZA M (prázdným zásobníkem) $\Leftrightarrow_{df}$
$(q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \varepsilon)$ pro něj. $q \in Q$.

Jazykem rozpoznávaným ZA M rozumíme jazyk
$L(M) = \{\, w \in \Sigma^* \mid w \text{ je přijímáno ZA } M \,\}$.

Poznámka. Jedná se o "přijímání prázdným zásobníkem".
Přidáme-li k parametrům ZA M množinu přijímajících (neboli 'koncových') stavů $F \subseteq Q$, lze definovat jazyk "přijímaný koncovým stavem"
$L_{KS}(M) = \{\, w \in \Sigma^* \mid (q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \alpha) \text{ pro něj. } q \in F \text{ a } \alpha \in \Gamma^* \,\}$
Lze převádět jeden typ přijímání na druhý.

Příklad konstrukce zásobníkového automatu

Zkonstruujte zásobníkový automat rozpoznávající (prázdným zásobníkem) jazyk

$L = \{ww^R \mid w \in \{a, b\}^*\}$

Idea konstrukce zásobníkového automatu ke gramatice

ZA provádí (nedeterministicky) tzv. analýzu "shora dolů".
Např. ZA sestrojený ke gramatice

$$
\begin{array}{l}
1/\ A \longrightarrow A + B \\
2/\ A \longrightarrow B \\
3/\ B \longrightarrow B * C \\
4/\ B \longrightarrow C \\
5/\ C \longrightarrow (A) \\
6/\ C \longrightarrow a
\end{array}
$$

simuluje při úspěšném běhu na slově $a * (a + a)$ použití pravidel v pořadí 2,3,4,6,5,1,2,4,6,4,6 .
Analýza "zdola nahoru" ... pravá derivace pozpátku (6,4,6,4,2,6,4,1,5,3,2)

Poznámka. LL- či LR-analyzátory ... determ. verze v syntaktické analýze

Konstrukce zásobníkového automatu ke gramatice

Lemma. Ke každé BG G lze sestrojit ZA M (s 1 stavem) tž. $L(M) = L(G)$.

Důkaz. Pro BG $G = (\Pi, \Sigma, S, P)$ sestrojme
$M = (\{q_0\}, \Sigma, \Pi \cup \Sigma, \delta, q_0, S)$, kde

- pro $X \in \Pi$:
  $\delta(q_0, \varepsilon, X) = \{ (q_0, \alpha) \mid (X \to \alpha) \in P \}$,
- pro $a \in \Sigma$: $\delta(q_0, a, a) = \{ (q_0, \varepsilon) \}$
- (jinde δ přiřazuje $\emptyset$)

Indukcí se dá ukázat

$$S \Rightarrow_G^* u\alpha \iff (q_0, u, S) \vdash_M^* (q_0, \varepsilon, \alpha)$$

kde $u \in \Sigma^*$, $\alpha \in \{\varepsilon\} \cup \Pi(\Pi \cup \Sigma)^*$; $\Rightarrow_G^*$ zde označuje *levé* odvození

Důkaz indukcí (pro hlubší zájemce)

Probereme každou implikaci zvlášť a postupujeme indukcí podle délky odvození. (Relace $\Rightarrow_G^k$ je k-násobné složení relace $\Rightarrow_G$)

(A) Ukazujeme $S \Rightarrow_G^* u\alpha \implies (q_0, u, S) \vdash_M^* (q_0, \varepsilon, \alpha)$
neboli $S \Rightarrow_G^k u\alpha \implies (q_0, u, S) \vdash_M^* (q_0, \varepsilon, \alpha)$ pro $k = 0, 1, 2, \ldots$

1. Nechť $k = 0$. Pak $S \Rightarrow_G^k u\alpha$ znamená $u\alpha = S$; tedy $u = \varepsilon$ a $\alpha = S$. Pak $(q_0, u, S) \vdash_M^* (q_0, \varepsilon, \alpha)$ znamená $(q_0, \varepsilon, S) \vdash_M^* (q_0, \varepsilon, S)$, což očividně platí.

2. Nechť (A) platí pro $0, 1, \ldots, k$; dokažme pro $k+1$.

Nechť tedy $S \Rightarrow_G^{k+1} u\alpha$. Pak nutně $S \Rightarrow_G^k u'X\alpha' \Rightarrow_G^1 u'\gamma\alpha' = u\alpha$, pro nějaké pravidlo $X \longrightarrow \gamma$; nutně ex. $v \in \Sigma^*$ tak, že $u = u'v$, $\gamma\alpha' = v\alpha$.
Podle indukčního předpokladu $(q_0, u', S) \vdash_M^* (q_0, \varepsilon, X\alpha')$
a navíc máme $\delta(q_0, \varepsilon, X) \ni (q_0, \gamma)$; takže
$(q_0, u', S) \vdash_M^* (q_0, \varepsilon, \gamma\alpha') = (q_0, \varepsilon, v\alpha)$.
Z toho plyne $(q_0, u'v, S) \vdash_M^* (q_0, v, v\alpha)$.
Díky tomu, že $\delta(q_0, a, a) = \{ (q_0, \varepsilon) \}$ pro vš. $a \in \Sigma$, vyvodíme
$(q_0, u, S) = (q_0, u'v, S) \vdash_M^* (q_0, v, v\alpha) \vdash_M^* (q_0, \varepsilon, \alpha)$.

Důkaz indukcí - pokr.

(B) Ukazujeme $(q_0, u, S) \vdash_M^* (q_0, \varepsilon, \alpha) \Longrightarrow S \Rightarrow_G^* u\alpha$
neboli $(q_0, u, S) \vdash_M^k (q_0, \varepsilon, \alpha) \Longrightarrow S \Rightarrow_G^* u\alpha$ pro $k = 0, 1, 2, \ldots$
Tvrzení dokonce zesílíme a povolíme libovolné $\alpha \in (\Pi \cup \Sigma)^*$.

1. Nechť $k = 0$. Pak $(q_0, u, S) \vdash_M^k (q_0, \varepsilon, \alpha)$ znamená $u = \varepsilon$ a $\alpha = S$;
tedy zajisté $S \Rightarrow_G^* S = u\alpha$.

2. Nechť (B) platí pro $0, 1, \ldots, k$; dokažme pro $k+1$.

Nechť tedy $(q_0, u, S) \vdash_M^{k+1} (q_0, \varepsilon, \alpha)$.
Pak nutně $(q_0, u, S) \vdash_M^k (q_0, u', \alpha') \vdash_M^1 (q_0, \varepsilon, \alpha)$.
a/ Když poslední krok využívá $\delta(q_0, a, a) \ni (q_0, \varepsilon)$, pak $u' = a$, $u = va$,
$\alpha' = a\alpha$, a vyvodíme $(q_0, v, S) \vdash_M^k (q_0, \varepsilon, a\alpha)$.
Podle indukčního předpokladu $S \Rightarrow_G^* va\alpha = u\alpha$.
b/ Když poslední krok využívá $\delta(q_0, \varepsilon, X) \ni (q_0, \gamma)$ (což odpovídá pravidlu
$X \longrightarrow \gamma$), pak $u' = \varepsilon$, $\alpha' = X\beta$ a $\gamma\beta = \alpha$.
Podle indukčního předpokladu $S \Rightarrow_G^* uX\beta$; použitím $X \longrightarrow \gamma$ vyvodíme
$S \Rightarrow_G^* u\gamma\beta = u\alpha$.

Konstrukce gramatiky k zásobníkovému automatu

Lemma. Ke každému ZA M s jedním stavem lze sestrojit bezkontextovou gramatiku G tž. $L(G) = L(M)$.

Důkaz. Mějme
$M = (\{q_0\}, \Sigma, \Gamma, \delta, q_0, Z_0)$, kde $\Sigma \cap \Gamma = \emptyset$.

Sestrojme $G = (\Gamma, \Sigma, Z_0, P)$ tak, že
$(A \to a\alpha) \in P \iff \delta(q_0, a, A) \ni (q_0, \alpha)$ $(a \in (\Sigma \cup \{\varepsilon\})$.

Indukcí se dá ukázat
$Z_0 \Rightarrow^*_G u\alpha \iff (q_0, u, Z_0) \vdash^*_M (q_0, \varepsilon, \alpha)$
kde $u \in \Sigma^*$, $\alpha \in \Gamma^*$; $\Rightarrow^*_G$ zde označuje *levé* odvození
Speciálně tedy platí
$Z_0 \Rightarrow^*_G u \iff (q_0, u, Z_0) \vdash^*_M (q_0, \varepsilon, \varepsilon)$ (neboli $L(G) = L(M)$)

Převod vícestavového ZA na 1-stavový - referát na cvičení

Lemma. Ke každému ZA M lze sestrojit ZA M' s 1 stavem tž. $L(M) = L(M')$.

Důkaz.
Pro $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0)$ konstruujeme $M' = (\{s\}, \Sigma, \Gamma', \delta', s, R)$,
kde $\Gamma' = (Q \times \Gamma \times Q) \cup \{R\}$,
$\delta'(s, \varepsilon, R) = \{\, (s, \langle q_0, Z_0, q\rangle) \mid q \in Q \,\}$
a δ' je dále dodefinována tak, že platí
$\forall w : (s, w, \langle p, X, q\rangle) \vdash^*_{M'} (s, \varepsilon, \varepsilon) \Longleftrightarrow (p, w, X) \vdash^*_M (q, \varepsilon, \varepsilon)$
Je tedy $\forall w : (s, w, R) \vdash^*_{M'} (s, \varepsilon, \varepsilon) \Longleftrightarrow (q_0, w, Z_0) \vdash^*_M (q, \varepsilon, \varepsilon)$ pro nějj.$q \in Q$
neboli $L(M') = L(M)$.

δ' je dodefinována takto:
- pro $(q', \varepsilon) \in \delta(q, a, X)$,
kde $a \in (\Sigma \cup \{\varepsilon\})$, zařadíme do
$\delta'(s, a, \langle q, X, q' \rangle)$ prvek (s, ε),
- pro $(q', A_1 A_2 \ldots A_n) \in \delta(q, a, X)$
$(n \geq 1)$ zařadíme do $\delta'(s, a, \langle q, X, \overline{q} \rangle)$ prvek
$(s, \langle q', A_1, q_1 \rangle \langle q_1, A_2, q_2 \rangle \ldots \langle q_{n-1}, A_n, \overline{q} \rangle)$ pro každé
$\overline{q}, q_1, q_2, \ldots, q_{n-1} \in Q$.

Příklad.

$M = (\{p, q, r\}, \{a, b\}, \{A, B\}, \delta, p, A)$, kde
$\delta(p, a, A) = \{(q, AA), (p, B)\}$,
$\delta(q, b, A) = \{(q, AA)\}$,
$\delta(p, \varepsilon, B) = \{(q, A)\}$, $\delta(q, \varepsilon, A) = \{(r, \varepsilon)\}$, $\delta(r, a, A) = \{(r, A)\}$,
$\delta(r, b, A) = \{(r, \varepsilon)\}$
(jinde je hodnota δ rovna $\emptyset$)

Příklad převodu vícestavového ZA na jednostavový

K zásobníkovému automatu M se vstupní abecedou $\{a, b\}$, zásobníkovou abecedou $\{A, B\}$, počátečním zásobníkovým symbolem A, množinou stavů $\{p, q, r\}$, počátečním stavem p a přechodovou funkcí δ definovanou následovně

$\delta(p, a, A) = \{(q, AA), (p, B)\}$,
$\delta(q, b, A) = \{(q, AA)\}$,
$\delta(p, \varepsilon, B) = \{(q, A)\}$,
$\delta(q, \varepsilon, A) = \{(r, \varepsilon)\}$,
$\delta(r, a, A) = \{(r, A)\}$,
$\delta(r, b, A) = \{(r, \varepsilon)\}$

(pro ostatní prvky def. oboru je funkční hodnota rovna $\emptyset$) sestrojte *jednostavový* ZA rozpoznávající jazyk $L(M)$

Poté sestrojte gramatiku generující tento jazyk. Použijte přitom konstrukce obsažené ve výše uvedených důkazech.

Uzávěrové vlastnosti třídy bezkontextových jazyků (CFL)

Věta. CFL je uzavřena vůči sjednocení, zřetězení, iteraci, zrcadlovému obrazu, substituci (tedy i homomorfismu).

Důkaz. K BG $G_1 = (\Pi_1, \Sigma, S_1, P_1)$,
$G_2 = (\Pi_2, \Sigma, S_2, P_2)$ lze zkonstruovat
$G = (\Pi, \Sigma, S, P)$ tž.$L(G) = L(G_1) \cup L(G_2)$ takto (předp., že $\Pi_1 \cap \Pi_2 = \emptyset$):
$\Pi = \Pi_1 \cup \Pi_2 \cup \{S\}$, kde $S \notin \Pi_1 \cup \Pi_2$,
$P = P_1 \cup P_2 \cup \{ S \longrightarrow S_1, S \longrightarrow S_2 \}$.
Přímočará je i konstrukce BG pro
$L(G_1) \cdot L(G_2)$, $L(G_1)^*$, $L(G_1)^R$.
Podobně k BG $G = (\Pi, \Sigma, S, P)$ a gramatikám G_a (pro vš. $a \in \Sigma$) lze snadno zkonstruovat gramatiku, která generuje jazyk vzniklý z $L(G)$ substitucí $L(G_a)$ za každé a.

Další uzávěrové vlastnosti CFL

Věta. CFL je uzavřena vůči průniku s regulárním jazykem, i vůči kvocientu podle regulárního jazyka.
(Tj. pro každý bezkontextový L a regulární R, jsou $L \cap R$, $R \backslash L$, L/R bezkontextové.)

Věta. CFL není uzavřena vůči průniku a doplňku.

Důkaz. $L_1 = \{a^i b^j c^k \mid i = j\}$,
$L_2 = \{a^i b^j c^k \mid j = k\}$ jsou bezkontextové.
Přitom $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ bezkontextový není.
Z de Morganových pravidel ($L_1 \cap L_2 = \overline{(\overline{L_1} \cup \overline{L_2})}$) plyne, že kdyby byla CFL uzavřena vůči doplňku, tak by byla uzavřena i vůči průniku (což není).

Deterministické zásobníkové automaty

Deterministický zásobníkový automat (DZA) je
ZA $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$, pro nějž platí:

1. $\delta(q, a, X)$ je vždy nejvýše jednoprvková množina (pro $a \in \Sigma \cup \{\varepsilon\}$)
2. je-li $\delta(q, \varepsilon, X) \neq \emptyset$, pak $\delta(q, a, X) = \emptyset$ pro vš. $a \in \Sigma$.

Jazyk L je *deterministický bezkontextový jazyk*, jestliže $L = L_{KS}(M)$ pro nějaký DZA M.
$(w \in L_{KS}(M) \quad \Leftrightarrow_{df} \quad (q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \alpha)$ pro $q \in F)$

Poznámka. Využitím "bottom-symbolu" lze snadno ukázat, že k DZA M lze zkonstruovat DZA M' tž. $L_{PZ}(M) = L_{KS}(M')$.
K DZA M lze také sestrojit DZA M' tž. $L_{KS}(M) \cdot \{\$\} = L_{PZ}(M')$, kde $\$$ je nově přidaný znak.

Uzávěrové vlastnosti DCFL

Věta. Třída DCFL je uzavřena vůči doplňku. Na druhé straně není uzavřena vůči průniku ani vůči sjednocení.

Poznámky.
Uzavřenost vůči doplňku ... nestačí prohození přijímajících a nepřijímajících stavů (ale dá se "dotáhnout")
Neuzavřenost vůči průniku:
$L_1 = \{a^i b^j c^k \mid i = j\}$,
$L_2 = \{a^i b^j c^k \mid j = k\}$
jsou *deterministické* bezkontextové.
Přitom $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ není (ani) bezkontextový.
Neuzavřenost vůči sjednocení ... de Morganova pravidla.

DCFL je vlastní podtřídou CFL

Příklad. Bezkontextový jazyk
$L = \{a^i b^j c^k \mid (i \neq j) \vee (j \neq k)\}$
není v DCFL:
jinak by i $\overline{L}$ byl v DCFL; pak by ovšem
$\overline{L} \cap [a^* b^* c^*] = \{\, a^n b^n c^n \mid n \geq 0\}$
byl bezkontextový (což není).

Příklady dalších nedeterministických bezkontextových jazyků:

$\{\, ww^R \mid w \in \{a, b\}^* \,\}$,

$\{\, a^i b^j c^k \mid (i = j) \vee (j = k) \,\}$

Poznámka. Existuje algoritmus, který rozhoduje ekvivalenci dvou zadaných deterministických zásobníkových automatů.
(Byl to otevřený problém od cca 1965; v r. 1997 vyřešil G. Sénizergues).