AUTOMATY A GRAMATIKY

10

Pokročilé uzávěrové vlastnosti
Rozhodovací problémy
Determinismus u ZA
Uzávěrové vlastnosti u DZA
Kontextové jazyky

Pavel Surynek

Univerzita Karlova v Praze

Matematicko-fyzikální fakulta
Katedra teoretické informatiky a matematické logiky

Pokročilé **uzávěrové vlastnosti** (1)

- bezkontextové jazyky jsou uzavřené na **inverzní homomorfismus**
  - homomorfismus je speciální případ bezkontextové (regulární substituce)
    - substituce $f: X \to 2^{\Sigma^*}$ je homomorfismus, jestliže $|f(x)|=1$ pro každé $x \in X$
  - $f: X \to 2^{\Sigma^*}$ homomorfismus, L, K bezkontextové jazyky nad X, resp. nad Σ
    - $f(L)=\{w_1w_2\dots w_n \mid (\exists x_1x_2\dots x_n \in L) w_1 \in f(x_1) \wedge w_2 \in f(x_2) \wedge \dots \wedge w_n \in f(x_n)\}$
      - víme, že $f(L)$ je bezkontextový
    - $f^{-1}(K)=\{x_1x_2\dots x_n \mid (\exists w_1w_2\dots w_n \in K) w_1 \in f(x_1) \wedge w_2 \in f(x_2) \wedge \dots \wedge w_n \in f(x_n)\}$
      - $f^{-1}(K)$ je bezkontextový
  - existuje ZA $Z = (Q, \Sigma, Y, \delta, q_0, z_0, F)$, že $L(Z) = K$
    - definujeme ZA $Z' = (Q', X, Y, \delta', [q_0, \lambda], z_0, F \times \{\lambda\})$, kde
      - $Q' = \{[q, u] \mid q \in Q \wedge u \in \Sigma^* \wedge (\exists x \in X, \exists v \in \Sigma^*)\ f(x) = vu\}$
        - použitých slov u,v je konečně mnoho
      - $\delta'([q,u], \lambda, y) = \{([p, u], w) \mid (p, w) \in \delta(q,\lambda,y)\}$
        $\cup \{([p, v], w) \mid (p, w) \in \delta(q,\sigma,y) \wedge u=\sigma v\}$
      - $\delta'([q, \lambda], x, y) = \{[q, f(x)], y\}$
        - po přečtení páskového symbolu x je do druhé komponenty stavu vloženo $f(x)$
        - nad druhou komponentou stavu proběhne simulace Z, přičemž pásku nečteme
    - platí, že $L(Z') = f^{-1}(L(Z)) = f^{-1}(K)$

Pokročilé **uzávěrové vlastnosti** (2)

- bezkontextové jazyky jsou uzavřené na **průnik s regulárním** jazykem
  - **paralelně** simulujeme zásobníkový a konečný automat
    - R regulární jazyk, že R = L(A) pro konečný automat $A = (Q_A, X, \delta_A, q_{A0}, F_A)$
    - L bezkontextový jazyk, že L = L(Z) pro zásobníkový automat $Z = (Q_Z, X, Y, \delta_Z, q_{Z0}, z_0, F_Z)$
  - definujeme zásobníkový automat $Z' = (Q_A \times Q_Z, X, Y, \delta, [q_{A0}, q_{Z0}], z_0, F_A \times F_Z)$, kde
    - $\delta([p,q], x, y) \ni ([r,s], u)$, jestliže
      - $p=r$ a $(s,u) \in \delta_Z(q,x,y)$ pro $x = \lambda$
        - Z nečte pásku, pracuje na zásobníku, A nepracuje
      - $r = \delta_A(p,x)$ a $(s,u) \in \delta_Z(q,x,y)$ pro $x \neq \lambda$
        - Z a A současně zpracují symbol x
    - $L(Z') = L(A) \cap L(Z) = R \cap L$

Pokročilé **uzávěrové vlastnosti** (3)

- bezkontextové jazyky jsou uzavřené na **kvocienty s regulárním jazykem**
  - pro levý kvocient paralelně simulujeme konečný a zásobníkový automat, jakmile se KA dostane do přijímajícího stavu, ZA začne zpracovávat vstup
    - R regulární jazyk, že R = L(A) pro konečný automat $A = (Q_A, X, \delta_A, q_{A0}, F_A)$
    - L bezkontextový jazyk, že L = L(Z) pro zásobníkový automat $Z = (Q_Z, X, Y, \delta_Z, q_{Z0}, z_0, F_Z)$
  - připomenutí
    - $R\backslash L = \{ v \mid (\exists u \in R)\ uv \in L \}$
    - $L/R = \{ u \mid (\exists v \in R)\ uv \in L \}$
  - definujeme zásobníkový automat $Z' = (Q_A \times Q_Z \cup Q_Z, X, Y, \delta, [q_{A0}, q_{Z0}], z_0, F_Z)$, kde
    - $\delta([p,q], \lambda, y) = \{([r,s], u) \mid (\exists x \in X)(r = \delta_A(p,x) \wedge (s,u) \in \delta_Z(q,x,y))\} \cup \{([p,s],u) \mid (s,u) \in \delta_Z(q,\lambda,y) \} \cup \{(q,y) \mid p \in F_A\}$
    - $\delta(q, x, y) = \delta_Z(q,x,y)$ pro $x \in X \cup \{\lambda\}, q \in Q_Z$
  - $L(Z') = L(A)\backslash L(Z) = R\backslash L$

Pavel Surynek, 2015

Rozhodovací problémy

- dán bezkontextový jazyk L nad X a slovo $w \in X^*$, chceme **testovat**, zda **$w \in L$**
  - předpokládejme, že máme gramatiku $G = (V_N, V_T, S, P)$ v <u>Chomského normálním tvaru</u>, že $L(G) = L$ a $w = x_1 x_2 \ldots x_n$
- algoritmus **CYK**
  - najdeme $X_{i,j}$ pro i,j=1,2,...,n, kde
    - $X_{i,j} = \{ X \mid X \in V_N \wedge X \Rightarrow_G^* x_i x_{i+1} \ldots x_j \}$
    - když $S \in X_{1,n}$, máme $S \Rightarrow_G^* w$ a tedy $w \in L$
  - dynamické programování
    - $X_{i,i} = \{ X \mid X \in V_N \wedge X \Rightarrow_G x_i \}$
      - jeden odvozovací krok stačí díky Chomského tvaru
    - $X_{i,j} = \{ X \mid (\exists k \in \{i, i+1, \ldots j-1\})[(\exists Y \in X_{i,k}, \exists Z \in X_{k+1,j}) X \Rightarrow_G YZ] \}$
  - čas $O(|w|^3)$
- test, zda **L je nekonečný**
  - pomocí pumping lemmatu
    - existuje slovo $w \in L$, že $|w| \geq n$ a $|w| \leq 2n-1$
- test, zda **$L = \emptyset$**
  - regularizací

Př.: $G = (V_N, V_T, S, P)$, kde
$V_N = \{ S, A, B, C \}$
$V_T = \{ a, b, c \}$
$P = \{$ **$S \rightarrow AB \mid BC$**
$A \rightarrow BA \mid a$
$B \rightarrow CC \mid b$
$C \rightarrow AB \mid a$ $\}$
w = baaba

$X_{i,j}$	j				
i	B	S,A	∅	∅	S,A,C
	∅	A,C	B	B	S,A,C
	∅	∅	A,C	S,C	B
	∅	∅	∅	B	S,A
	∅	∅	∅	∅	A,C

Determinismus u ZA (1)

- zásobníkový automat Z = $(Q, X, Y, \delta, q_0, z_0, F)$ je **deterministický** (DZA), jestliže
  - $|\delta(q, x, y)| \leq 1$ pro všechna $q \in Q$, $x \in X \cup \{\lambda\}$, $y \in Y$
  - když $\delta(q,\lambda,y) \neq \emptyset$ pro $q \in Q$ a $y \in Y$, pak $\delta(q,x,y) = \emptyset$ pro všechna $x \in X$
- jazyk L je **deterministický bezkontextový**, jestliže existuje deterministický ZA Z, že $L = L(Z)$
  - $K = \{ww^R \mid w \in \{a,b\}^*\}$ je bezkontextový, ale nikoli deterministický
    - je třeba nedeterministicky <u>uhádnout střed</u> vstupního slova
- jazyk L se nazývá **bezprefixový**, jestliže pro každé $u,v \in L$ platí
  - když u je prefix v, pak $u = v$

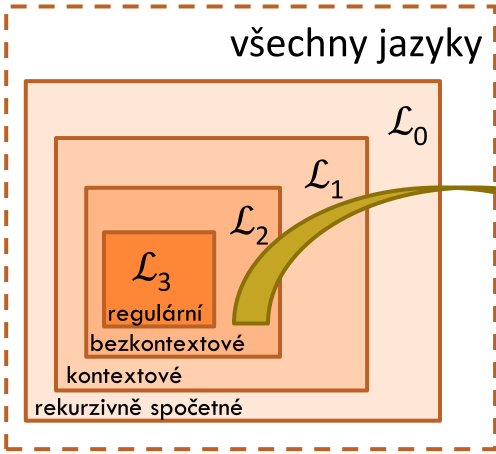
Pavel Surynek, 2015

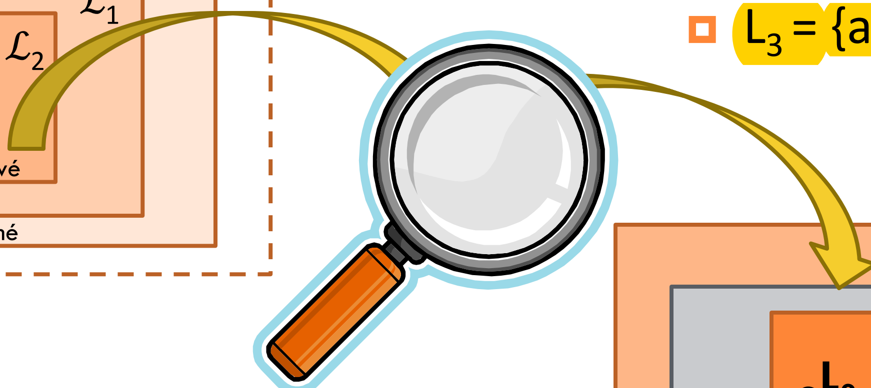
Determinismus u ZA (2)

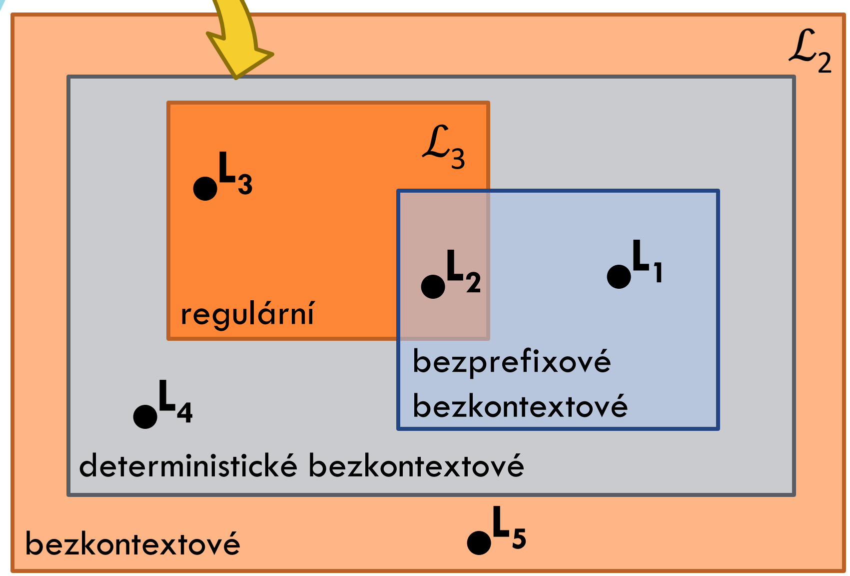
- nechť Z je deterministický zásobníkový automat, pak říkáme, že N(Z) je **bezkontextový bezprefixový jazyk**
  - pro $w \in X^*$ $w \in N(Z) \Rightarrow (\forall v \in X^*)(v \neq \lambda \Rightarrow wv \notin N(Z))$
    - po přečtení w je zásobník vyprázdněn a není možné pokračovat
  - jelikož konstrukce Z', že L(Z')=N(Z) zachovává determinismus, je N(Z) rovněž deterministický bezkontextový jazyk
- nikoli každý deterministický bezkontextový jazyk je bezprefixový
  - uvažme regulární L = {a, aa}
- když L nad abecedou X je deterministický bezkontextový jazyk, pak L.\$ pro $\$ \notin X$ je bezkontextový bezprefixový

Zjemnění Chomského hierarchie

- svědci
  - $L_1 = \{a^i b^i \mid i=1,2,\dots\}$
  - $L_2 = \{aba\}$
  - $L_3 = \{a, aa\}$







- další svědci
  - $L_4 = \{a^i b^j \mid i,j=1,2,\dots \wedge i \leq j\}$
  - $L_5 = \{ww^R \mid w \in \{a,b\}^*\}$

Uzávěrové vlastnosti pro **DZA** (1)

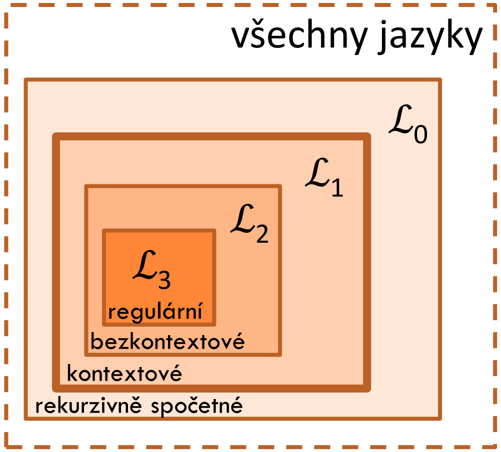
- deterministické bezkontextové jazyky jsou uzavřené na **doplněk**
  - připomeňme: bezkontextové na doplněk uzavřené nejsou
- důkaz je technicky náročný, uvedeme jen několik myšlenek
  - případy <u>nepřijetí stavem</u> u deterministického ZA
    - (i) přečte vstupní slovo a skončí v nepřijímajícím stavu
    - (ii) vyprázdní zásobník
    - (iii) dostane se do konfigurace, která nemá pokračování
    - (iv) zacyklí se v krocích na zásobníku
      - opakují se konfigurace
    - (v) dojde k nekontrolovatelnému růstu zásobníku
  - případy (i), (ii), (iii) lze poznat snadno, případy (iv) a (v) obtížněji

Pavel Surynek, 2015

Uzávěrové vlastnosti pro **DZA** (2)

- deterministické bezkontextové jazyky jsou uzavřené na **inverzní homomorfismus** a **průnik s regulárním jazykem**
  - předchozí konstrukce zachovají determinismus
- deterministické bezkontextové jazyky <u>nejsou</u> uzavřené na **průnik, sjednocení a homomorfismus**
  - důkaz pro sjednocení
    - $L = \{a^i b^j c^k \mid i,j,k = 0,1,2,\ldots \wedge (i \neq j \vee j \neq k \vee i \neq k)\}$ je bezkontextový (sjednocení tří deterministických)
      - nechť L je navíc deterministický, pak -L rovněž
      - $K = \{a^i b^j c^k \mid i,j,k = 0,1,2,\ldots\}$ je regulární a $(-L) \cap K = \{a^i b^i c^i \mid i = 0,1,2, \ldots\}$, který není bezkontextový
  - důkaz pro homomorfismus
    - $L = \{a^i b^i c^j \mid i,j = 0,1,2,\ldots\}$
    - $K = \{a^i b^j c^j \mid i,j = 0,1,2,\ldots\}$
      - K a L jsou deterministické bezkontextové
    - $0L \cup 1K$ je deterministický bezkontextový, $1L \cup 1K$ nikoli
      - nedeterminismus $1L \cup 1K$ rozborem případů
      - $f(0) = 1$, identita jinak, pak $f(0L \cup 1K) = 1L \cup 1K$

Kontextové jazyky (1)



- kontextové jazyky
  - generované **kontextovými** gramatikami (context sensitive) [typu 1]
    - pravidla tvaru $\alpha X\beta \rightarrow \alpha w\beta$, kde $\alpha, \beta \in (V_N \cup V_T)^*$, $X \in V_N$ a $w \in (V_N \cup V_T)^*$ a $w \neq \lambda$
    - nebo $S \rightarrow \lambda$, pokud S není na pravé straně žádného jiného pravidla
  - generované **nezkracujícími** gramatikami
    - pravidla tvaru $\alpha \rightarrow \beta$, kde $\alpha, \beta \in (V_N \cup V_T)^*$ a $|\alpha| \leq |\beta|$
    - nebo $S \rightarrow \lambda$, pokud S není na pravé straně žádného jiného pravidla
      - kontextová pravidla jsou nezkracující, naopak nikoli
- pro každou nezkracující gramatiku $G = (V_N, V_T, S, P)$ existuje kontextová gramatika $G' = (V_N', V_T, S, P')$, že $L(G) = L(G')$
  - ukazuje, že definice jsou **ekvivalentní**

Př.: $G = (V_N, V_T, S, P)$, kde
$V_N = \{ S, A, B, C \}$
$V_T = \{ a, b, c \}$
$P = \{$
$S \rightarrow aSBC \mid abC$
$CB \rightarrow BC$
$bB \rightarrow bb$
$bC \rightarrow bc$
$cC \rightarrow cc$
$\}$

G je nezkracující,
ale nikoli kontextová

$L(G) = \{a^i b^i c^i \mid i = 1,2,\dots \}$

Kontextové jazyky (2)

- $G = (V_N, V_T, S, P)$ nezkracující gramatika
  - nejprve **separace**
    - výsledkem je separovaná gramatika $G'' = (V_N'', V_T, S, P'')$
      - pro každý terminál $x \in V_T$ zavedeme nový neterminál X_x
      - v pravidlech nahradíme každý terminál x za X_x
      - zavedeme pravidla $X_x \rightarrow x$
    - $L(G'') = L(G)$
  - **náhrada nezkracujícího** pravidla $\alpha \rightarrow \beta$ sérií kontextových
    - nechť $\alpha \rightarrow \beta = A_1A_2 \ldots A_m \rightarrow B_1B_2 \ldots B_n$, kde $m \leq n$
      - $\mathbf{A_1}A_2 \ldots A_m \rightarrow \mathbf{C_1}A_2 \ldots A_m$
      - $C_1\mathbf{A_2} \ldots A_m \rightarrow C_1\mathbf{C_2} \ldots A_m$ …
      - $C_1 \ldots C_{m-1}\mathbf{A_m} \rightarrow C_1C_2 \ldots \mathbf{C_m}$
      - $\mathbf{C_1}C_2 \ldots C_m \rightarrow \mathbf{B_1}C_2 \ldots C_m$
      - $B_1\mathbf{C_2}C_2 \ldots C_m \rightarrow B_1\mathbf{B_2}C_3 \ldots C_m$ …
      - $B_1B_2 \ldots B_{m-1}\mathbf{C_m} \rightarrow B_1B_2 \ldots B_{m-1}\mathbf{B_mB_{m+1} \ldots B_n}$
    - $C_1, C_2, \ldots, C_m$ jsou pro každé nahrazované pravidlo <u>nové</u>

Př.: $AB \rightarrow BA$

zavedeme neterminály C_1, C_2

náhrada za **$AB \rightarrow BA$**

$AB \rightarrow C_1B$

$C_1B \rightarrow C_1C_2$

$C_1C_2 \rightarrow BC_2$

$BC_2 \rightarrow BA$

Kontextové **uzávěrové vlastnosti** (1)

- kontextové jazyky jsou uzavřené na konečná **sjednocení**
  - kontextové gramatiky $G_1=(V_N^1, V_T, S_1, P_1)$ a $G_2=(V_N^2, V_T, S_2, P_2)$, kde $V_N^1 \cap V_N^2 = \emptyset$
  - položme $G = (V_N^1 \cup V_N^2 \cup \{S'\}, V_T, S', P1 \cup P2 \cup \{S' \rightarrow S_1 | S_2 \})$
    - S' je nový neterminál
      - ošetřit přepis původních počátečních symbolů na λ
  - platí $L(G) = L(G_1) \cup L(G_2)$
- kontextové jazyky jsou uzavřené na **konkatenaci**
  - položme $G = (V_N^1 \cup V_N^2 \cup \{S'\}, V_T, S', P_1 \cup P_2 \cup \{S' \rightarrow S_1.S_2 \})$
    - S' je nový neterminál
      - ošetřit přepis původních počátečních symbolů na λ
  - platí $L(G) = L(G_1).L(G_2)$
- kontextové jazyky jsou uzavřené na **iteraci**
  - položme $G = (V_N^1 \cup \{S', S''\}, V_T, S', P_1 \cup \{S' \rightarrow \lambda \mid S'', \; S'' \rightarrow S''S_1 \mid S_1 \})$
    - S', S'' jsou nové neterminály
      - je třeba dávat pozor na výskyt počátečního neterminálu vpravo (u bezkontextových netřeba)
  - platí $L(G) = L(G_1)^*$
- kontextové jazyky jsou uzavřené na **zrcadlový obraz**
  - položme $G = (V_N^1, V_T, S_1, \{v^R \rightarrow w^R \mid v \rightarrow w \in P_1\})$
  - platí $L(G) = L(G_1)^R$

Kontextové **uzávěrové vlastnosti** (2)

- kontextové jazyky jsou uzavřené na **kontextovou substituci**
  - substituce f: $X \to 2^{Y^*}$ je kontextová substituce, jestliže $f(x)$ je kontextový jazyk pro každé $x \in X$
    - máme $G_x = (V_N^x, Y, S_x, P_x)$, že $f(x) = L(G_x)$ pro každé $x \in X$
      - V_N^x jsou po dvou disjunktní
  - mějme $G = (V_N, X, S, P)$, zajímá nás kontextovost $f(L(G))$
  - položíme $G' = (V_N \cup \bigcup_{x \in X} V_N^x \cup X, Y, S, P \cup \bigcup_{x \in X} P^x \cup \{x \to S_x \mid x \in X\})$
    - případně ošetříme přepis neterminálů jiných než S na λ
    - platí $L(G') = f(L(G))$
- kontextové jazyky jsou **uzavřené na doplňky** a **konečné průniky**
  - velmi hluboký výsledek (řešilo se asi 20 let)
    - věta Immerman–Szelepcsényi