

Języki, automaty i obliczenia

Wykład 9: Własności języków bezkontekstowych

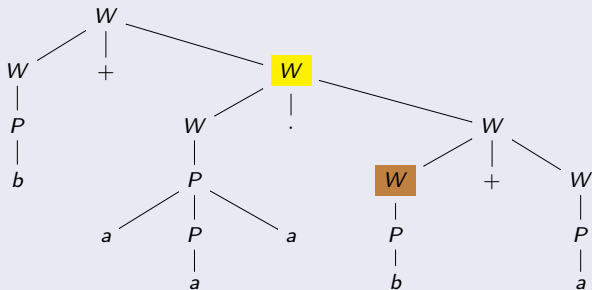
Sławomir Lasota

Uniwersytet Warszawski

22 kwietnia 2015

- 1 Pompowanie języków bezkontekstowych
- 2 Własności domknięcia
- 3 Obrazy przemienne języków bezkontekstowych są semiliniowe

Przykład


 $b + \mathbf{aaa} \cdot \mathbf{b} \mathbf{+ a} \varepsilon \in L(\mathcal{G})$
 $b + \mathbf{aaa} \cdot \mathbf{aaa} \cdot \mathbf{b} \mathbf{+ a} \mathbf{+ a} \varepsilon \in L(\mathcal{G})$
 $b + \mathbf{b} \varepsilon \in L(\mathcal{G})$

Lemat o pompowaniu

Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

$$x = w y u z v, \text{ gdzie } |y z| > 0 \text{ i } |y u z| \leq n$$

i dla każdego $m \in \mathbb{N}$, $w y^m u z^m v \in L$.

$$L \text{ bezkontekstowy} \implies \exists n \forall x \exists x = wyuzv \forall m. w y^m u z^m v \in L$$

Wniosek

L jest niebezkontekstowy, jeśli dla każdego n , istnieje słowo $x \in L$, $|x| \geq n$ takie, że dla każdego przedstawienia x jako

$$x = w y u z v, \text{ gdzie } |y z| > 0 \text{ i } |y u z| \leq n,$$

istnieje $m \in \mathbb{N}$ takie, że $w y^m u z^m v \notin L$.

$$L \text{ niebezkontekstowy} \iff \forall n \exists x \forall x = wyuzv \exists m. w y^m u z^m v \notin L$$

Lemat o pompowaniu

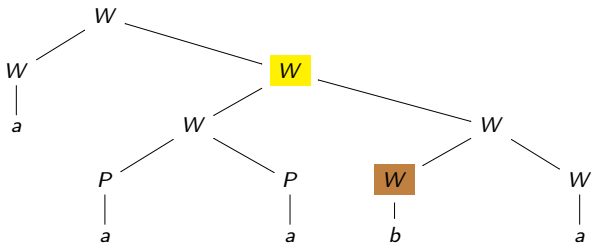
Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

$$x = w y u z v, \text{ gdzie } |y z| > 0 \text{ i } |y u z| \leq n$$

i dla każdego $m \in \mathbb{N}$, $w y^m u z^m v \in L$.

Dowód:

Niech $\mathcal{G} = (A, N, S, \alpha)$ gramatyka w postaci Chomskiego dla $L - \{\varepsilon\}$. Niech n takie, że każde drzewo wyprowadzenia dla słowa długości $\geq n$ ma ścieżkę dłuższą niż $|N|$.



$$L \text{ niebezkontekstowy} \iff \forall n \exists x \forall x=wyuzv \exists m. w y^m u z^m v \notin L$$

Przykład

Pokażemy, że język

$$L = \{w w : w \in \{a, b\}^*\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo

$$x_n = a^n b^n a^n b^n \in L.$$

Rozważmy dowolne słowa w, y, u, z, v takie, że

$$x_n = w y u z v, \text{ gdzie } |y z| > 0 \text{ i } |y u z| \leq n.$$

Dwa przypadki: albo obydwie litery a, b występują w którymś ze słów y, z , albo nie.

Przykład

Język

$$\{a^n b^n c^n : n \in \mathbb{N}\}$$

nie jest bezkontekstowy.

L niebezkontekstowy $\iff \forall n \exists x \forall x=wyuzv \exists m. w y^m u z^m v \notin L$

Przykład

Spróbujmy pokazać, że język

$$L = \{a^i b^j c^k : i, j, k \text{ parami różne}\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo...

$$L \text{ niebezkontekstowy} \iff \forall n \exists x \forall x=wyuzv \exists m. w y^m u z^m v \notin L$$

Przykład

Spróbujmy pokazać, że język

$$L = \{a^i b^j c^k : i, j, k \text{ parami różne}\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo...

$$x_n = a^{f(n)} b^{g(n)} c^{h(n)} \in L.$$

L niebezkontekstowy $\iff \forall n \exists x \forall x=wyuzv \exists m. w y^m u z^m v \notin L$

Przykład

Spróbujmy pokazać, że język

$$L = \{a^i b^j c^k : i, j, k \text{ parami różne}\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo...

$$x_n = a^{f(n)} b^{g(n)} c^{h(n)} \in L.$$

B.u.o. $f(n) > g(n), h(n)$. Wtedy niech

$$w y u z \in a^+, \quad v \in a^* b^{g(n)} c^{h(n)}, \quad |yz| \neq f(n) - g(n), f(n) - h(n).$$

Dla każdego $m \in \mathbb{N}$, $w y^m u z^m v \in L$. Pompowanie nie działa...

L bezkontekstowy $\implies \exists n \forall x \exists x=wyuzv \forall m. w y^m u z^m v \in L$

Rozważmy słowa z *zaznaczonymi* pozycjami. Formalnie, rozszerzamy alfabet

$$A \cup \{\underline{a} : a \in A\}$$

Np. słowo $a\underline{a}a\underline{b}a\underline{a}a\underline{b}$ ma zaznaczone 3 pozycje.

Niech $|w|$ oznacza liczbę zaznaczonych pozycji w słowie w .

Lemat o pompowaniu (Ogden 1968)

Dla każdego języka bezkontekstowego L istnieje n takie, że każde słowo $x \in L$, $|x| \geq n$, można przedstawić jako

$$x = w y u z v, \text{ gdzie } |y z| > 0 \text{ i } |y u z| \leq n$$

i dla każdego $m \in \mathbb{N}$, $w y^m u z^m v \in L$.

$$L \text{ bezkontekstowy} \implies \exists n \forall x \exists x=wyuzv \forall m. w y^m u z^m v \in L$$

Dowód:

Tak samo.

$$L \text{ niebezkontekstowy} \iff \forall n \exists x \forall x=wyuzv \exists m. w y^m u z^m v \notin L$$

Przykład

Pokażemy, że język

$$L = \{a^i b^j c^k : i, j, k \text{ parami różne}\}$$

nie jest bezkontekstowy. Dla dowolnego $n \in \mathbb{N}$ rozważmy słowo

$$x_n = \underline{a}^n b^{n+n!} c^{n+2n!} \in L.$$

Rozważmy dowolne słowa w, y, u, z, v takie, że

$$x_n = w y u z v, \text{ gdzie } |y z| > 0 \text{ i } |y u z| \leq n.$$

Wtedy któreś ze słów y, z musi zawierać przynajmniej jedno a .

Trzy przypadki: albo $yz \in a^+$, albo $y \in a^+, z \in b^*$, albo $y \in a^+, z \in c^*$.

- 1 Pompowanie języków bezkontekstowych
- 2 Własności domknięcia**
- 3 Obrazy przemienne języków bezkontekstowych są semiliniowe

Fakt

Języki bezkontekstowe są zamknięte na sumę, konkatencję, iterację (gwiazdkę).

Dowód:

Gramatyki.

Fakt

Języki bezkontekstowe nie są zamknięte na przecięcie.

Dowód:

Język $\{a^n b^n c^m : n, m \in \mathbb{N}\} \cap \{a^n b^m c^m : n, m \in \mathbb{N}\}$ nie jest bezkontekstowy.

Wniosek

Języki bezkontekstowe nie są zamknięte na dopełnienie.

Na przykład języki

$$\{w w : w \in \{a, b\}^*\} \quad \{a^i b^j c^k : i, j, k \text{ parami różne}\}$$

nie są bezkontekstowe, a ich dopełnienia są (dlaczego?).

Fakt (języki bezkontekstowe są zamknięte na podstawienia)

$L \subseteq A^*$ bezkontekstowy
 $h : A \rightarrow \mathcal{P}(B^*), \forall a \in A. h(a)$ bezkontekstowy $\implies \widehat{h}(L)$ bezkontekstowy.

W szczególności języki bezkontekstowe są zamknięte na obrazy homomorficzne:

$L \subseteq A^*$ bezkontekstowy
 $h : A \rightarrow B^*$ homomorfizm $\implies \vec{h}(L)$ bezkontekstowy.

Fakt (języki bezkontekstowe są zamknięte na przeciwobrazy homomorficzne)

$L \subseteq B^*$ bezkontekstowy
 $h : A \rightarrow B^*$ homomorfizm $\implies \vec{h}^{-1}(L)$ bezkontekstowy.

Twierdzenie

Deterministyczne języki bezkontekstowe są zamknięte na dopełnienie.

Na przykład dopełnienia języków

$$\{a^n b^n : n \in \mathbb{N}\} \quad \{w \$ w^R : w \in A^*\}$$

są deterministycznymi językami bezkontekstowymi, a język

$$L = \{a^i b^j c^k : i=j \vee j=k \vee i=k\}$$

nie (bo w p.p. język $(\{a, b, c\}^* - L) \cap a^* b^* c^*$ musiałby być bezkontekstowy).

Fakt

Deterministyczne języki bezkontekstowe nie są zamknięte na przecięcia.

Dowód:

Język $\{a^n b^n c^m : n, m \in \mathbb{N}\} \cap \{a^n b^m c^m : n, m \in \mathbb{N}\}$ nie jest bezkontekstowy.

Wniosek

Deterministyczne języki bezkontekstowe nie są zamknięte na sumy.

Zestawienie własności domknięcia

operacja	języki bezkontekstowe	deterministyczne języki bezkontekstowe
konkatenacja	✓	
iteracja (gwiazdka)	✓	
suma	✓	
przecięcie		
dopełnienie		✓
różnica		
podstawienia/obrazy	✓	
przeciwobrazy	✓	✓

Dopełnienie języków deterministycznych (idea dowodu)

Niech $\mathcal{A} = (A, Q, q_0, F, S, s_0, \delta)$ deterministyczny automat ze stosem.
Skonstruujemy automat deterministyczny \mathcal{A}'' t.ż. $L(\mathcal{A}'') = A^* - L(\mathcal{A})$.

Bez u.o. założmy, że symbol początkowy s_0 nie jest nigdy zdejmowany ze stosu.

Pytanie

Niech $\bar{\mathcal{A}}$ będzie identyczny z \mathcal{A} , ale stany akceptujące to $Q - F$. Czy $L(\bar{\mathcal{A}})$ to dopełnienie $L(\mathcal{A})$?

Uwaga! Nie możemy pozbyć się pustych przejść. Np. język

$$\{a^n b^m a^n : n, m \in \mathbb{N}\} \cup \{a^n b^m c b^m a^n : n, m \in \mathbb{N}\}$$

nie jest rozpoznawany przez deterministyczny automat ze stosem bez pustych przejść.

Konfigurację (q, w) nazywamy *niestabilną* jeśli $(q, w) \xrightarrow{\varepsilon}$, a w p.p. *stabilną*. Niech

$$L_s(\mathcal{A}) = \{w \in A^* : c_0 \xrightarrow{w} c \text{ dla jakiejś stabilnej konfiguracji akceptującej } c\}$$

Mówimy, że automat \mathcal{A} *nie pętli się*, jeśli $L_s(\mathcal{A}) \cup L_s(\bar{\mathcal{A}}) = A^*$.

Idea dowodu:

$$\mathcal{A} \longmapsto \text{niepętłący się } \mathcal{A}' \longmapsto \mathcal{A}'' \quad (L(\mathcal{A}'') = L_s(\mathcal{A}''))$$

$$\mathcal{A} \longmapsto \text{niepętłący się } \mathcal{A}' \longmapsto \mathcal{A}''$$

Mówimy, że para $(q, s) \in Q \times S$ *pętli się*, jeśli automat \mathcal{A} ma nieskończony bieg:

$$(q, s) \xrightarrow{\varepsilon} (p_1, w_1) \xrightarrow{\varepsilon} (p_2, w_2) \xrightarrow{\varepsilon} \dots$$

Stany $\{q, p_1, p_2, \dots\}$ nazwijmy osiągalnymi z (q, s) .

Automat \mathcal{A}' otrzymujemy przez eliminację pętających się par (q, s) :

$$\delta'(q, s, \varepsilon) = \begin{cases} (q_{\text{tak}}, s) & \text{jeśli jakiś stan akceptujący jest osiągalny z } (q, s) \\ (q_{\text{nie}}, s) & \text{w p.p.} \end{cases}$$

$L(\mathcal{A}') = L(\mathcal{A})$. Czy \mathcal{A}' nie pętli się? Gdyby się pętlił, to

$$(q_0, s_0) \xrightarrow{\nu} (p_0, w_0) \xrightarrow{\varepsilon} (p_1, w_1) \xrightarrow{\varepsilon} (p_2, w_2) \xrightarrow{\varepsilon} \dots$$

a wtedy para (p_i, w_i) , gdzie w_i jest najkrótsze spośród wszystkich w_j , pętliłaby się.

$$A \mapsto \text{niepełący się } A' \mapsto A''$$

- $Q'' = Q' \times \{t, n, ?\}$
- $q_0'' = (q_0', ?)$ (bez u.o. zakładamy, że $q_0 \notin F$)
- $F'' = \{(q, n) : q \in Q'\}$
- $S'' = S'$
- $s_0'' = s_0'$
- jeśli $\delta'(q, s, \varepsilon) = (p, w)$, to

- $\delta''((q, t), s, \varepsilon) = ((p, t), w)$

- $\delta''((q, ?), s, \varepsilon) = \begin{cases} ((p, ?), w) & \text{o ile } p \notin F' \\ ((p, t), w) & \text{o ile } p \in F' \end{cases}$

jeśli $\delta'(q, s, a) = (p, w)$, dla $a \in A$, to

- $\delta''((q, ?), s, \varepsilon) = ((q, n), w)$

- $\delta''((q, t), s, a) = \begin{cases} ((p, ?), w) & \text{o ile } p \notin F' \\ ((p, t), w) & \text{o ile } p \in F' \end{cases}$

- $\delta''((q, n), s, a) = \begin{cases} ((p, ?), w) & \text{o ile } p \notin F' \\ ((p, t), w) & \text{o ile } p \in F' \end{cases}$

Fakt

Jeśli L jest (deterministycznym) językiem bezkontekstowym a R jest regularny, to

$$L \cap R$$

$$L \cup R$$

$$L^{-1}R \quad \text{jest (deterministycznym) językiem bezkontekstowym.}$$

$$R^{-1}L$$

$$L \otimes R$$

Dowód:

Produkt automatu ze stosem i automatu skończonego.

- 1 Pompowanie języków bezkontekstowych
- 2 Własności domknięcia
- 3 Obrazy przemienne języków bezkontekstowych są semiliniowe

Dla $P \subseteq \mathbb{N}^k$, zdefiniujmy

$$P^\oplus = \{p_1 + p_2 + \dots + p_m : (p_1, p_2, \dots, p_m) \in P^*\}$$

Zbiór *liniowy* $X \subseteq \mathbb{N}^k$ to dowolny zbiór postaci

$$X = b + P^\oplus = \{b + p_1 + \dots + p_m : (p_1, \dots, p_m) \in P^*\}, \quad \text{dla } b \in \mathbb{N}^k, P \subseteq_{\text{fin}} \mathbb{N}^k.$$

Przykład

Dla $b = (1, 2)$, $P = \{(2, 0), (0, 3)\}$,

$$b + P^\oplus = \{(n, m) \in \mathbb{N}^2 : n \equiv 1 \pmod{2}, m \equiv 2 \pmod{3}\}.$$

Zbiór *semiliniowy* $X \subseteq \mathbb{N}^k$ to dowolna skończona suma zbiorów liniowych:

$$b_1 + P_1^\oplus \cup b_2 + P_2^\oplus \cup \dots \cup b_n + P_n^\oplus$$

Pytanie

Niech $B, P \subseteq_{\text{fin}} \mathbb{N}^k$. Czy $B + P^\oplus$ jest zbiorem (semi)liniowym?

Dla $k = 1$, zbiory semiliniowe to zbiory prawie okresowe.

- Dla $w = a_1 \dots a_n \in A^*$, zdefiniujmy funkcję (multizbiór nad A)

$$\mathcal{P}(w) : A \rightarrow \mathbb{N} \quad \mathcal{P}(w)(a) = |\{i \in \{1 \dots n\} : a_i = a\}|.$$

Obraz przemienny języka L nad A to:

$$\mathcal{P}(L) = \{\mathcal{P}(w) : w \in L\}.$$

- Dla ustalonego liniowego porządku na alfabecie $A = \{b_1, \dots, b_k\}$,

$$b_1 < b_2 < \dots < b_k,$$

obraz przemienny języka L można utożsamiać z podzbiorem \mathbb{N}^k :

$$\mathcal{P}(w)(i) = \mathcal{P}(w)(b_i).$$

Przykład

Niech $A = \{a, b\}$, $a < b$.

$$\mathcal{P}(\{a^n b^n a : n \in \mathbb{N}\}) = \mathcal{P}((ab)^* a) = \{(n+1, n) : n \in \mathbb{N}\}$$

$$\mathcal{P}(\text{palindromy}) = \{(n, m) : \text{przynajmniej jedno spośród } n, m \text{ parzyste}\}$$

Twierdzenie (Parikh 1961)

Obrazy przemienny języka bezkontekstowego jest semiliniowy.

Wniosek

Dla każdego języka bezkontekstowego L istnieje język regularny R t.żé

$$\mathcal{P}(L) = \mathcal{P}(R).$$

Dowód:

$$b_1 + P_1^\oplus \cup \dots \cup b_n + P_n^\oplus$$

$$v + \{v_1, \dots, v_m\}^\oplus \subseteq \mathbb{N}^k \quad \mapsto \quad w + (w_1 + \dots + w_m)^* \subseteq A^*$$

Wniosek

Dla alfabetów jednoliterowych, języki bezkontekstowe to języki regularne.

Niech $L = L(\mathcal{G})$, gdzie $\mathcal{G} = (A, N, S, \alpha)$. Pokażemy, że $\mathcal{P}(L)$ jest semiliniowy.

- Dla drzewa wyprowadzenia t , oznaczmy przez $N(t)$ zbiór nieterminali pojawiających się w t .

Dla $M \subseteq N$, niech $L_M \subseteq L$ zawiera słowa posiadające drzewo wyprowadzenia t takie, że $N(t) = M$. Ponieważ

$$L = \bigcup_{M \subseteq N} L_M,$$

wystarczy pokazać, że $\mathcal{P}(L_M)$ jest semiliniowy, dla dowolnego $M \subseteq N$.

- Drzewo wyprowadzenia nazwijmy *płatkiem*, jeśli na żadnej ścieżce żaden nieterminal nie pojawia się więcej niż $|N|$ razy.

$$B_M = \{\mathcal{P}(w) : w \text{ ma płatkie drzewo wyprowadzenia } t \text{ takie, że } N(t) = M\}$$

$$P_M = \{\mathcal{P}(wv) : wXv \text{ ma płatkie drzewo wyprowadzenia o korzeniu } X \text{ takie, że } N(t) \subseteq M, \text{ dla pewnego } X \in N\}$$

Pokażemy, że $\mathcal{P}(L_M) = B_M + P_M^\oplus$.

rysunek

maszyny Turinga