

Języki, automaty i obliczenia

Wykład 7: Języki bezkontekstowe

Sławomir Lasota

Uniwersytet Warszawski

13 kwietnia 2016

- 1 Gramatyki bezkontektowe
- 2 Języki bezkontekstowe a drzewa
- 3 Postaci normalne i problemy decyzyjne

$$\mathcal{G} = (A, N, S, \alpha)$$

- A to skończony zbiór symboli końcowych (*terminalnych*)
- N to skończony zbiór symboli niekończących (*nieterminalnych*), $A \cap N = \emptyset$
- $S \in N$ to symbol początkowy
- $\alpha \subseteq N \times (A \cup N)^*$ to skończony zbiór reguł przepisywania (*produkcji*)

Notacja

Reguły $(X, v) \in \alpha$ będziemy zapisywać $X \rightarrow_{\mathcal{G}} v$ albo $X \rightarrow v$.

Reguły rozszerzamy do relacji $\rightarrow_{\mathcal{G}} \subseteq (A \cup N)^* \times (A \cup N)^*$:

$$w \rightarrow_{\mathcal{G}} w' \iff \exists (X \rightarrow_{\mathcal{G}} v) \in \alpha, \exists u, t \in (A \cup N)^*. w = u X t, w' = u v t$$

i domyamy zwrotno-tranzytywnie: $w \twoheadrightarrow_{\mathcal{G}} w'$ wtw. gdy istnieje ciąg

$$w = v_0 \rightarrow_{\mathcal{G}} v_1 \rightarrow_{\mathcal{G}} \dots \rightarrow_{\mathcal{G}} v_n = w', \quad n \geq 0$$

zwany *wyprowadzeniem* (wywodem) w' z w (w gramatyce $\mathcal{G}$).

Język *bezkontekstowy* to język *generowany* przez gramatykę:

$$L(\mathcal{G}) = \{ w \in A^* : S \rightarrow_{\mathcal{G}}^* w \}$$

Przykład

Gramatyka $\mathcal{G} = (A, N, R, \alpha)$:

- $A = \{a, b, c\}$
- $N = \{P, R\}$
- $\alpha = \{ (R, RR), (R, P), (P, aPa), (P, bPb), (P, \varepsilon) \}$

$$R \rightarrow RR \mid P$$

$$P \rightarrow aPa \mid bPb \mid \varepsilon$$

Słowo $aaaaabbb \in L(\mathcal{G})$, ponieważ ma wyprowadzenie:

$$\begin{aligned} R &\rightarrow_{\mathcal{G}} RR \rightarrow_{\mathcal{G}} aPaR \rightarrow_{\mathcal{G}} aaPaR \rightarrow_{\mathcal{G}} aaaaR \rightarrow_{\mathcal{G}} \\ &aaaaP \rightarrow_{\mathcal{G}} aaaa bPb \rightarrow_{\mathcal{G}} aaaa bbb \end{aligned}$$

$$L(\mathcal{G}) = ?$$

Pytanie

Czy każde słowo ma co najwyżej jedno wyprowadzenie w gramatyce?

Nie, np.

$$S \longrightarrow S S \mid \varepsilon$$

Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka spełniająca ten warunek?

Przykłady języków bezkontekstowych

- palindromy nad $\{a, b\}$:

$$P \longrightarrow aPa \mid bPb \mid a \mid b \mid \varepsilon$$

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

$$X \longrightarrow aXb \mid \varepsilon$$

- $\{w \in \{a, b\}^* : \#_w(a) = \#_w(b)\}$

$$X \longrightarrow aXb \mid bXa \mid XX \mid \varepsilon$$

- wyrażenia arytmetyczne:

$$S \longrightarrow S + I \mid I$$

$$W \longrightarrow W \cdot W \mid W + W \mid (W) \mid 0 \mid 1$$

$$I \longrightarrow I \cdot C \mid C$$

$$C \longrightarrow (S) \mid 0 \mid 1$$

- poprawnie zbudowane wyrażenia nawiasowe:

$$W \longrightarrow WW \mid [W] \mid (W) \mid \langle W \rangle \mid \varepsilon$$

Fakt

Każdy język regularny jest bezkontekstowy.

Dowód:

Niech $\mathcal{A} = (A, Q, \{q_0\}, F, \delta)$. Definiujemy gramatykę $\mathcal{G} = (A, N, S, \alpha)$:

- $N = Q$
- $S = q_0$
- $\alpha = \{(q, a, q') : (q, a, q') \in \delta\} \cup \{(q, \varepsilon) : q \in F\}$

$$w \in L(\mathcal{G}) \iff q_0 \twoheadrightarrow_{\mathcal{G}} w \iff \exists q \in F. \hat{\delta}(q_0, w, q) \iff w \in L(\mathcal{A})$$

Wyprowadzenie:

$$w \rightarrow_G w' \iff \exists (X \rightarrow_G v) \in \alpha, u, t \in (A \cup N)^*. w = uXt, v = uv t$$

$$w \twoheadrightarrow_G w' \iff \exists n \geq 0, v_0, \dots, v_n. w = v_0 \rightarrow_G v_1 \rightarrow_G \dots \rightarrow_G v_n = w'$$

Wyprowadzenie *lewostronne*:

$$w \xrightarrow{L}_G w' \iff \exists (X \rightarrow_G v) \in \alpha, u \in A^*, t \in (A \cup N)^*. w = uXt, v = uv t$$

$$w \xrightarrow{L}\twoheadrightarrow_G w' \iff \dots$$

Przykład

$$\begin{aligned} R &\xrightarrow{L}_G R R \xrightarrow{L}_G a P a R \xrightarrow{L}_G a a P a a R \xrightarrow{L}_G a a a a R \xrightarrow{L}_G \\ a a a a P &\xrightarrow{L}_G a a a a b P b \xrightarrow{L}_G a a a a b b \end{aligned}$$

Wyprowadzenie *prawostronne*?

Fakt

Jeśli $w \twoheadrightarrow_G w' \in A^*$ to $w \xrightarrow{L}\twoheadrightarrow_G w'$.

Przykład

Język palindromów nad $A = \{a, b\}$, generowany przez gramatykę

$$X \rightarrow_{\mathcal{G}} aXa \mid bXb \mid \varepsilon,$$

jest najmniejszym językiem $L \subseteq A^*$ spełniającym warunek:

$$L \supseteq \{a\}L\{a\} \cup \{b\}L\{b\} \cup \{\varepsilon\}.$$

Czyli L jest najmniejszym punktem stałym funkcji monotonicznej

$$L \mapsto \{a\}L\{a\} \cup \{b\}L\{b\} \cup \{\varepsilon\} : \mathcal{P}(A^*) \rightarrow \mathcal{P}(A^*).$$

Pytanie

Jak tę obserwację uogólnić do gramatyk o więcej niż jednym nieterminalu?

- 1 Gramatyki bezkontekstowe
- 2 Języki bezkontekstowe a drzewa
- 3 Postaci normalne i problemy decyzyjne

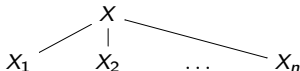
Drzewo to $T \subseteq \mathbb{N}^*$

- zamknięty na prefiksy: $\forall w, v \in \mathbb{N}^*. \quad wv \in T \implies w \in T$
- zamknięty „w lewo”: $\forall w \in \mathbb{N}^*, i, j \in \mathbb{N}. \quad wi \in T, j < i \implies wj \in T$

Drzewo nad X to para (T, l) , gdzie $l: T \rightarrow X$.

Niech $\mathcal{G} = (A, N, S, \alpha)$. **Drzewo wyprowadzenia** (wyvodu) w gramatyce $\mathcal{G}$ to drzewo nad $A \cup N \cup \{\varepsilon\}$ t.ż.

- S jest etykietą korzenia
- dla każdego wierzchołka drzewa o etykiecie X ,



albo



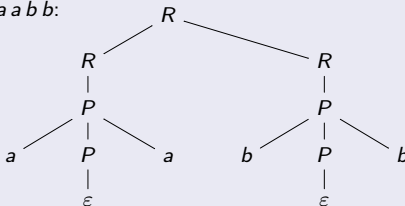
$$\begin{aligned} X &\longrightarrow_{\mathcal{G}} X_1 X_2 \dots X_n \\ X_1, \dots, X_n &\in A \cup N \end{aligned}$$

$$X \longrightarrow_{\mathcal{G}} \varepsilon$$

Plon drzewa wyprowadzenia to słowo $w \in (A \cup N)^*$ otrzymane z etykiet liści, w kolejności leksykograficznej (od lewej do prawej), pomijając ε .

Przykład

Drzewo wyprowadzenia dla $a a b b$:



Fakt

Niech $w \in A^*$.

$w \in L(\mathcal{G})$ wtw. gdy w jest plonem drzewa wyprowadzenia w gramatyce $\mathcal{G}$.

Dowód:

$\Leftarrow$: indukcja ze względu na głębokość drzewa

$\Rightarrow$: indukcja ze względu na długość wyprowadzenia, korzystamy z obserwacji:

jeśli $w v \twoheadrightarrow_{\mathcal{G}} u$ to $u = u_1 u_2$ dla pewnych $u_1, u_2 \in (A \cup N)^*$ t.ż.

$$w \twoheadrightarrow_{\mathcal{G}} u_1 \quad \text{ i } \quad v \twoheadrightarrow_{\mathcal{G}} u_2.$$

Pytanie

Czy każde słowo ma co najwyżej jedno drzewo wyprowadzenia w gramatyce?

Nie, np.

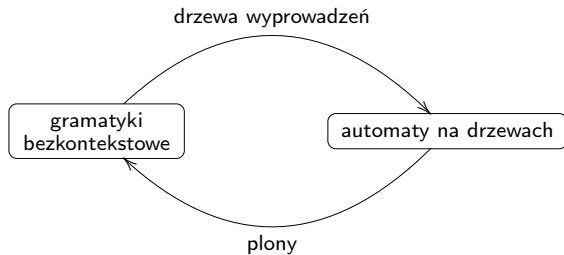
$$S \longrightarrow S S \mid \varepsilon$$

Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka *jednoznaczna*, tzn. taka, w której każde słowo ma co najwyżej jedno drzewo wyprowadzenia?

Nie, oto *język niejednoznaczny*:

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



Twierdzenie

Niech $\mathcal{A} = (A, Q, I, F, \delta)$ automat na drzewach. Język

$$\text{plony}(L(\mathcal{A})) = \{ \text{plon}(t) : t \in L(\mathcal{A}) \}$$

jest językiem bezkontekstowym.

Dowód:

Definiujemy gramatykę bezkontekstową $\mathcal{G} = (A, N, S, \alpha)$:

- $N = Q \times A \cup \{S\}$
- zbiór reguł α :

$$S \rightarrow_{\mathcal{G}} \varepsilon \iff I \cap F \neq \emptyset$$

$$S \rightarrow_{\mathcal{G}} (q, a) \iff q \in I$$

$$(q, a) \rightarrow_{\mathcal{G}} (q_1, a_1) \dots (q_n, a_n) \iff (q, a, q_1 \dots q_n) \in \delta \quad (\text{arność}(a) = n)$$

$$(q, a) \rightarrow_{\mathcal{G}} a \iff \exists w \in F^*. (q, a, w) \in \delta$$

$w \in L(\mathcal{G}) \iff$ istnieje drzewo wyprowadzenia słowa w w gramatyce $\mathcal{G} \iff$

$\exists t. \mathcal{A}$ ma bieg akceptujący nad $t \wedge \text{plon}(t) = w \iff w \in \text{plony}(L(\mathcal{A}))$

Twierdzenie

Niech $\mathcal{G} = (A, N, S, \alpha)$ gramatyka bezkontekstowa. Zbiór

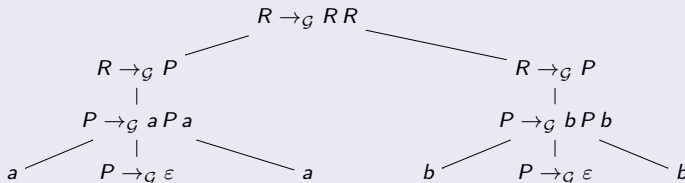
$$dw(\mathcal{G}) = \{\text{wszystkie drzewa wyprowadzeń słów } w \in A^* \text{ w gramatyce } \mathcal{G}\}$$

jest regularnym językiem drzew.

Konieczna poprawka!

W drzewie wyprowadzenia zastąp każdą etykietę $X \in N$ przez odpowiednią regułę.

Przykład



Dodatkowo, dla uproszczenia pozbadźmy się liści ϵ .

Twierdzenie

Niech $\mathcal{G} = (A, N, S, \alpha)$ gramatyka bezkontekstowa. Zbiór

$$\text{dw}(\mathcal{G}) = \{\text{wszystkie drzewa wyprowadzeń słów } w \in A^* \text{ w gramatyce } \mathcal{G}\}$$

jest regularnym językiem drzew.

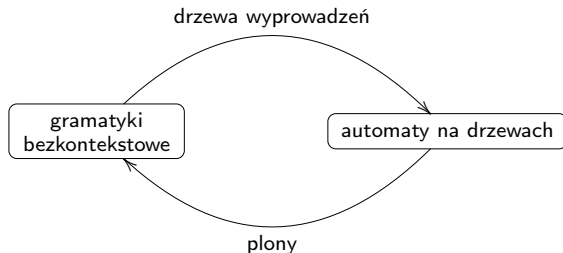
Dowód:

Definiujemy automat na drzewach $\mathcal{A} = (B, Q, I, F, \delta)$:

- $B = A \cup \alpha$ $\text{arność}(a) = 1, \text{ arność}((X, w)) = \begin{cases} |w|, & |w| > 0 \\ 1, & \text{w p.p.} \end{cases}$
- $Q = A \cup N \cup \{q_f\}$
- $I = \{S\}$
- $F = \{q_f\}$
- $\delta = \begin{aligned} &\{(a, a, q_f) : a \in A\} \cup \\ &\{(X, (X, w), w) : (X, w) \in \alpha, |w| > 0\} \cup \\ &\{(X, (X, \varepsilon), q_f) : (X, \varepsilon) \in \alpha\} \end{aligned}$

$$t \in L(\mathcal{A}) \iff t \in \text{dw}(\mathcal{G})$$

$\mathcal{A}$ można uczynić deterministycznym „z góry w dół”.



Czyli dla dowolnego regularnego języka drzew L , język

drzewa wyprowadzeń($plony(L)$)

jest deterministycznym „z góry w dół” regularnym językiem drzew. Sprzeczność?

- 1 Gramatyki bezkontektowe
- 2 Języki bezkontekstowe a drzewa
- 3 Postaci normalne i problemy decyzyjne

Założmy, że $\varepsilon \notin L(\mathcal{G})$.

Postać normalna Chomskiego

$$X \longrightarrow a \quad (a \in A)$$

$$X \longrightarrow YZ \quad (Y, Z \in N)$$

Dowód:

$$X \longrightarrow a \quad (a \in A)$$

$$X \longrightarrow w \quad (w \in N^*)$$

Postać normalna Greibach

$$X \longrightarrow a w \quad (a \in A, w \in N^*)$$

Pytanie

Czy można ograniczyć się do $|w| \leq 2$?

Pustość

Dane: język bezkontekstowy L

Wynik: czy $L = \emptyset$?

Uniwersalność

Dane: język bezkontekstowy L

Wynik: czy $L = A^*$?

Równość

Dane: języki bezkontekstowe L, M

Wynik: czy $L = M$?

Przynależność

Dane: Gramatyka bezkontekstowa $\mathcal{G} = (A, N, S, \alpha)$ w postaci Chomskiego
i słowo $a_1 \dots a_n$

Wynik: czy $a_1 \dots a_n \in L$?

$$V_{i,j} = \{X \in N : X \rightarrow_{\mathcal{G}}^* a_i \dots a_j\} \quad (i \leq j)$$

Algorytm

dla $i \in \{1 \dots n\}$, $V_{i,i} := \{X \in N : X \rightarrow_{\mathcal{G}} a_i\}$

dla $i, j \in \{1 \dots n\}$, $i < j$, $V_{i,j} := \emptyset$

powtarzaj dla $1 \leq i \leq j < k \leq n$

$$V_{i,k} := V_{i,k} \cup \{X \in N : \exists Y, Z \in N. (X, YZ) \in \alpha, Y \in V_{i,j}, Z \in V_{j+1,k}\}$$

aż do stabilizacji

wynik := $(S \in V_{1,n})$

W następnym odcinku:

automaty dla języków bezkontekstowych