Gramatyki bezkontekstowe

Języki formalne i techniki translacji - Wykład 4

Maciek Gębala

29 października 2019

Notatki

Gramatyki bezkontekstowe

Definicja: Gramatyka bezkontekstowa G to czwórka $G = (N, T, P, S)$ gdzie

- N - skończony zbiór zmiennych (nieterminale),
- T - skończony zbiór symboli końcowych (terminale, alfabet),
- P - skończony zbiór produkcji postaci $A \to \alpha$, gdzie $A \in N$ i $\alpha \in (N \cup T)^*$,
- $S \in N$ - symbol początkowy.

Relacja wyprowadzenia $\underset{G}{\Rightarrow}$

Jeśli $A \to \beta$ jest produkcją w G i $\alpha, \gamma \in (N \cup T)^*$ to $\alpha A \gamma \underset{G}{\Rightarrow} \alpha\beta\gamma$ ($\alpha\beta\gamma$ jest bezpośrednio wyprowadzany z $\alpha A \gamma$ w gramatyce G).
Będziemy pisać tylko $\Rightarrow$ gdy gramatyka jest oczywista.
$\Rightarrow^*$ – zwrotne i przechodnie domknięcie $\Rightarrow$.

Notatki

Gramatyki bezkontekstowe

Definicja: Język generowany przez G to

$$L(G) = \{\ w\ :\ w \in T^* \ \wedge\ S \underset{G}{\Rightarrow}^* w\ \}.$$

Język L nazywamy bezkontekstowym jeśli jest identyczny z $L(G)$ dla pewnej gramatyki bezkontekstowej G.
G_1 i G_2 są równoważne jeżeli $L(G_1) = L(G_2)$.

Przykład

$$G = (\{S\}, \{0, 1\}, \{S \to \varepsilon, S \to 0, S \to 1, S \to 0S0, S \to 1S1\}, S)$$

lub inaczej (krócej) zapisując

$$S \to \varepsilon|0|1|0S0|1S1$$

Wyprowadzenie słowa 0110:

$$S \Rightarrow 0S0 \Rightarrow 01S10 \Rightarrow 0110$$

Notatki

Drzewo wyprowadzenia

Drzewo o następujących własnościach

1. każdy wierzchołek ma etykietę z $N \cup T \cup \{\varepsilon\}$,
2. korzeń ma etykietę S (symbol początkowy),
3. wierzchołki wewnętrzne mają etykiety z N,
4. jeżeli wierzchołek wewnętrzny ma etykietę A a jego synowie od lewej mają kolejno etykiety $X_1, \ldots, X_n$ to $A \to X_1 \ldots X_n$ jest produkcją w P.

Jeśli konkatenacja wszystkich liści czytanych od lewej do prawej daje α to drzewo nazywamy drzewem wyprowadzenia α.

Przykład drzewa wyprowadzenia (na tablicy)

Słowo $id + id * id$ dla gramatyki $E \to E + E|E * E|id$.

Notatki

Drzewo wyprowadzenia

Twierdzenie. Niech $G = (N, T, P, S)$ będzie gramatyką bezkontekstową. Wtedy $S \Rightarrow^* \alpha \iff$ istnieje drzewo wyprowadzenia α w gramatyce G.

Dowód

Indukcja względem ilości wierzchołków wewnętrznych w drzewie.

Definicja. Jeżeli w każdym kroku wyprowadzenia stosujemy produkcję do nieterminala leżącego najbardziej na lewo (prawo), to wyprowadzenie nazywamy lewostronnym (prawostronnym).
Jeżeli $w \in L(G)$ to w ma co najmniej jedno drzewo wyprowadzenia.
Każdemu drzewu wyprowadzenia odpowiada dokładnie jedno wyprowadzenie lewostronne (prawostronne).

Definicja. Jeśli w $L(G)$ istnieje słowo mające dwa różne drzewa wyprowadzenia to G nazywamy wieloznaczną.

Notatki

Usuwanie symboli bezużytecznych

Niech L – niepusty język bezkontekstowy. Wtedy L można wygenerować za pomocą gramatyki G o następujących własnościach

1. każdy symbol pojawia się w wyprowadzeniu jakiegoś słowa z L,
2. nie ma produkcji postaci $A \to B$ (produkcje jednostkowe), gdzie $A, B \in N$.

Co więcej, jeśli $\varepsilon \notin L$ to w G nie ma produkcji postaci $A \to \varepsilon$.

Symbol X jest użyteczny jeśli istnieje wyprowadzenie postaci $S \Rightarrow^* \alpha X \gamma \Rightarrow^* w$ $(w \in T^*)$, w p.p. X jest bezużyteczny.

Notatki

Usuwanie symboli bezużytecznych

Lemat. Dla dowolnej gramatyki bezkontekstowej $G = (N, T, P, S)$ z $L(G) \neq \emptyset$ można efektywnie znaleźć równoważną gramatykę bezkontekstową $G' = (N', T, P', S)$ t.że dla dowolnego $A \in N'$ istnieje $w \in T^*$ t.że $A \Rightarrow^* w$.

Dowód (szkic)

1. $N_s \leftarrow \emptyset$
   $N_n \leftarrow \{A : (A \to w) \in P \wedge w \in T^*\}$
2. `while` $N_s \neq N_n$ `do`
   $N_s \leftarrow N_n$
   $N_n \leftarrow N_s \cup \{A : (A \to \alpha) \in P \wedge \alpha \in (T \cup N_s)^*\}$
3. $N' \leftarrow N_n$
   $P' \leftarrow \{(A \to \alpha) \in P : A \in N_n \wedge \alpha \in (T \cup N_s)^*\}$

Notatki

Usuwanie symboli bezużytecznych

Lemat. Dla dowolnej gramatyki bezkontekstowej $G = (N, T, P, S)$ można efektywnie znaleźć równoważną gramatykę bezkontekstową $G' = (N', T', P', S)$ t.że dla każdego $X \in (N' \cup T')$ istnieją $\alpha, \beta \in (N' \cup T')^*$ t.że $S \Rightarrow^* \alpha X \beta$.

Dowód (szkic)

1. $N' \leftarrow \{S\}$
2. `while` można zmienić N' `do`
   Jeśli $A \in N'$ i $A \to \alpha_1 | \ldots | \alpha_n$ to dodaj wszystkie nieterminale z $\alpha_1, \ldots, \alpha_n$ do N' a terminale do T'.
3. Do P' przenieś tylko te produkcje z P które zawierają symbole z $N' \cup T' \cup \{\varepsilon\}$.

Twierdzenie. Każdy niepusty język bezkontekstowy jest generowany przez gramatykę bezkontekstową nie zawierającą symboli bezużytecznych.

Notatki

Usuwanie ε-produkcji

Twierdzenie. Jeżeli $L = L(G)$ dla gramatyki bezkontekstowej $G = (N, T, P, S)$ to dla $L \setminus \{\varepsilon\}$ istnieje gramatyka bezkontekstowa G' nie zawierająca ε-produkcji i symboli bezużytecznych.

Dowód

Symbole bezużyteczne usunęliśmy dzięki poprzedniemu twierdzeniu. Dla każdego nieterminala A sprawdzamy czy $A \Rightarrow^* \varepsilon$. Jeśli tak to każdą produkcję $B \to \alpha A \beta$ zastępujemy produkcjami $B \to \alpha A \beta$ i $B \to \alpha\beta$ (ale nie dołączamy $B \to \varepsilon$). Następnie usuwamy wszystkie ε-produkcje.

Notatki

Usuwanie produkcji jednostkowych

Twierdzenie. Każdy język bezkontekstowy nie zawierający ε jest definiowany za pomocą gramatyki nie zawierającej symboli bezużytecznych, ε-produkcji i produkcji jednostkowych.

Dowód

Jeżeli dla nieterminala A mamy $A \Rightarrow^* B$ i $A \neq B$ to dla każdej niejednostkowej produkcji $B \to \alpha$ dodajemy produkcję $A \to \alpha$. Następnie usuwamy produkcje jednostkowe.

Notatki

Postać normalna Chomsky'ego

Twierdzenie. Dowolny język bezkontekstowy nie zawierający ε jest generowany przez gramatykę której wszystkie produkcje są postaci $A \to BC$ lub $A \to a$, gdzie $A, B, C \in N$ i $a \in T$.

Dowód (konstrukcja)

Niech G będzie gramatyką bez symboli bezużytecznych, ε-produkcji i produkcji jednostkowych. Wtedy jeśli prawa strona produkcji jest długości 1 to jest postaci $A \to a$. Dla pozostałych produkcji wykonujemy następujące operacje:

1. Jeśli po prawej stronie występuje terminal a to dodajemy do N nowy nieterminal C_a a do produkcji $C_a \to a$ i zastępujemy a przez C_a.
2. Teraz jeśli prawa strona produkcji jest dłuższa niż 1 to zawiera tylko nieterminale. Jeśli jest postaci $A \to B_1 \dots B_n$ dla $n > 2$, to tworzymy nowe nieterminale $D_1, \dots, D_{n-2}$ i zastępujemy tą produkcję przez $A \to B_1 D_1, D_1 \to B_2 D_2, \dots, D_{n-3} \to B_{n-2} D_{n-2}, D_{n-2} \to B_{n-1} B_n$.

Notatki

Przykład

$$\begin{aligned} S &\to bA|aB \\ A &\to bAA|aS|a \\ B &\to aBB|bS|b \end{aligned}$$

Przykład na tablicy.

Postać normalna Chomsky'ego

$$\begin{aligned} S &\to C_bA|C_aB \\ A &\to C_bD_A|C_aS|a \\ B &\to C_aD_B|C_bS|b \\ C_a &\to a \\ C_b &\to b \\ D_A &\to AA \\ D_B &\to BB \end{aligned}$$

Notatki

Postać normalna Greibach

Produkcje są postaci $A \to a\alpha$, gdzie $A \in N$, $a \in T$ i $\alpha \in N^*$.

Określmy jako A-produkcje wszystkie produkcje z nieterminalem A po lewej stronie.

Lemat 1. Niech $G = (N, T, P, S)$ będzie gramatyką bezkontekstową. Niech $A \to \alpha_1 B \alpha_2$ będzie produkcją w P i niech $B \to \beta_1 | \ldots | \beta_r$ będzie zbiorem wszystkich B-produkcji. Niech $G' = (N, T, P', S)$ będzie gramatyką otrzymaną z G przez usunięcie produkcji $A \to \alpha_1 B \alpha_2$ i dodanie produkcji $A \to \alpha_1 \beta_1 \alpha_2 | \ldots | \alpha_1 \beta_r \alpha_2$. Wówczas $L(G) = L(G')$.

Dowód

Dowód po strukturze drzewa wyprowadzenia.

Notatki

Postać normalna Greibach

Lemat 2. Niech $G = (N, T, P, S)$ będzie gramatyką bezkontekstową. Niech $A \to A\alpha_1 | \ldots | A\alpha_r$ będzie zbiorem tych A-produkcji których prawe strony zaczynają się od A. Niech $A \to \beta_1 | \ldots | \beta_s$ będzie zbiorem pozostałych A-produkcji. Niech $G' = (N \cup \{B\}, T, P', S)$ będzie gramatyką utworzoną poprzez dodanie nowego nieterminala B i zastąpienie wszystkich A-produkcji przez następujące produkcje

$$A \to \beta_i | \beta_i B, \qquad 1 \leqslant i \leqslant s$$

$$B \to \alpha_j | \alpha_j B, \qquad 1 \leqslant j \leqslant r$$

Wtedy $L(G) = L(G')$.

Notatki

Postać normalna Greibach

Dowód

W wyprowadzeniu lewostronnym ciąg produkcji postaci $A \to A\alpha_i$ musi kiedyś skończyć się produkcją $A \to \beta_j$

$$A \Rightarrow A\alpha_{i_1} \Rightarrow A\alpha_{i_2}\alpha_{i_1} \Rightarrow \ldots \Rightarrow A\alpha_{i_n} \ldots \alpha_{i_1} \Rightarrow \beta_j \alpha_{i_n} \ldots \alpha_{i_1}$$

Można to zastąpić przez

$$A \Rightarrow \beta_j B \Rightarrow \beta_j \alpha_{i_n} B \Rightarrow \ldots \Rightarrow \beta_j \alpha_{i_n} \ldots \alpha_{i_2} B \Rightarrow \beta_j \alpha_{i_n} \ldots \alpha_{i_1}$$

Ponieważ transformacja ta jest obustronna to $L(G) = L(G')$

Notatki

Postać normalna Greibach

Twierdzenie. Każdy język bezkontekstowy L nie zawierający ε jest generowany przez pewną gramatykę w której każda produkcja jest postaci $A \to a\alpha$, gdzie $a \in T$, $A \in N$ i $\alpha \in N^*$.

Dowód

Niech $G = (N, T, P, S)$ będzie gramatyką w postaci normalnej Chomsky'ego. Załóżmy, że $N = \{A_1, \ldots, A_n\}$.
Modyfikujemy produkcje tak aby jeśli produkcja jest postaci $A_i \to A_j\alpha$ to $i < j$.

`for` $k \to 1$ `to` n `do`
  `for` $j \to 1$ `to` $k - 1$ `do`

1. Za każdą produkcję postaci $A_k \to A_j\alpha$ wstaw produkcje $A_k \to \beta\alpha$ dla wszystkich produkcji $A_j \to \beta$ (Lemat 1).
2. Dla produkcji postaci $A_k \to A_k\alpha$ wykonaj Lemat 2 używając nowy nieterminal B_k.

Notatki

Postać normalna Greibach

Dowód cd.

Po wykonaniu tego algorytmu mamy gramatykę równoważną o produkcjach w postaci:

1. $A_i \to A_j\gamma$, gdzie zawsze $i < j$,
2. $A_i \to a\gamma$, gdzie $a \in T$,
3. $B_i \to \gamma$, gdzie $\gamma \in (N \cup \{B_1, \ldots, B_{i-1}\})^*$.

Zauważmy, że A_n-produkcje muszą zaczynać się terminalem. Teraz rozważmy A_{n-1}-produkcje: ich lewe strony muszą zaczynać się terminalem lub nieterminalem A_n więc możemy je z Lematu 1 zastąpić prawymi stronami A_n-produkcji (wszystkie zaczynają się terminalem). I tak do A_1.

Łatwo zauważyć że B-produkcje nigdy nie zaczynają się nieterminalem B więc też z Lematu 1 możemy je zastąpić prawymi stronami A-produkcji.

Notatki

Przykład

$$\begin{aligned} A_1 &\to A_2A_3 \\ A_2 &\to A_3A_1|b \\ A_3 &\to A_1A_2|a \end{aligned}$$

Nie pasuje $A_3 \to A_1A_2$ więc z Lematu 1 dostajemy $A_3 \to A_2A_3A_2$.

Dalej nie pasuje więc ponownie z Lematu 1 otrzymujemy $A_3 \to A_3A_1A_3A_2|bA_3A_2$.

Teraz mamy $A_3 \to A_3A_1A_3A_2|bA_3A_2|a$, korzystamy z Lematu 2 i otrzymujemy $A_3 \to a|aB_3|bA_3A_2|bA_3A_2B_3$ i $B_3 \to A_1A_3A_2|A_1A_3A_2B_3$.

Notatki

Przykład

Teraz odpowiednio podstawiając zgodnie z Lematem 1 otrzymujemy

$$\begin{aligned} A_3 &\to a|aB_3|bA_3A_2|bA_3A_2B_3 \\ A_2 &\to aA_1|aB_3A_1|bA_3A_2A_1|bA_3A_2B_3A_1|b \\ A_1 &\to aA_1A_3|aB_3A_1A_3|bA_3A_2A_1A_3|bA_3A_2B_3A_1A_3|bA_3 \\ B_3 &\to aA_1A_3A_3A_2|aB_3A_1A_3A_3A_2|bA_3A_2A_1A_3A_3A_2| \\ &\quad bA_3A_2B_3A_1A_3A_3A_2|bA_3A_3A_2|aA_1A_3A_3A_2B_3| \\ &\quad aB_3A_1A_3A_3A_2B_3|bA_3A_2A_1A_3A_3A_2B_3| \\ &\quad bA_3A_2B_3A_1A_3A_3A_2B_3|bA_3A_3A_2B_3 \end{aligned}$$

Notatki

Notatki