Dwa twierdzenia pani Greibach dla jezykow bezkontekstowych

Gramatyka bezkontekstowa jest w postaci normalnej Greibach gdy każda produkcja jest jednej z dwóch postaci:

- $A \to a \cdot \alpha$;
- $A \to a$.

Na przykład następująca gramtyka dla języka D_1 (wyrażeń nawiasowych) jest w postaci normalnej Greibach.

$$S \to (\, S\,)\, S \mid (\,) \mid (\, S\,)$$

Twierdzenie (Greibach)

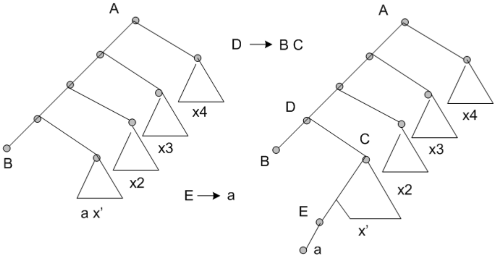
Każdą gramatyke bezkontekstową nie używającą słowa pustego można przekształcić do gramatyki w postaci normalnej Greibach. Rozmiar gramatyki rośnie wielomianowo.

Dowód 1-szego twierdzenia Greibach

Załóżmy, że gramatyka G jest w postaci norm. Chomsky'ego. Tworzymy nowe nieterminale (A, B) i nowe produkcje, tak by zachodziła własność że

$$((A, B) \to^* \alpha) \Leftrightarrow (A \to_G^* B \cdot \alpha)$$

dla każdego $\alpha \in T^*$.



Dekompozycja wyprowadzenia $A \to^* B\alpha$ na $E \to a$, $C \to^* E\, x_1'$ i $A \to^* D\, x_2 x_3 x_4$. Mamy cztery przypadki zależnie od tego, które ze słów x_1', $x_2 x_3 x_4$ są puste.

Początkowo w G' tworzymy produkcje *inicjalne*:
$S \to a \cdot (S, A)$, dla każdego $A \to a$.

Następnie tworzymy cztery typy produkcji :

1. $(A, B) \to a(C, E)(A, D)$ gdy $E \to A,\ D \to BC$
2. $(A, B) \to a \cdot (C, E)$ gdy $A \to BC$ i $E \to a$
3. $(A, B) \to a \cdot (A, D)$ gdy $D \to BC$ i $C \to a$
4. $(A, B) \to a$ gdy $A \to BC$ i $C \to a$.

Niech $\Sigma_0 = \{(,),[,]\}$ będzie alfabetem dwóch typów nawiasów .
Przypomnienie: D_2 jest językiem poprawnych nawiasów z Σ_0).
Rozszerzamy alfabet: $\Sigma = \Sigma_0 \cup \{\#, \$\}$.
Językiem Greibach nazywamy zbiór L_0 słów postaci:

$$x_1^1 \# x_2^1 \ldots \# x_{i_1}^1 \$ x_1^2 \# x_2^2 \ldots \# x_{i_2}^2 \$ x_1^3 \ldots \# x_{i_3}^3 \ldots \$ x_1^k \# x_3^k \ldots \# x_{i_k}^k$$

które spełniają:

$$x_j^i \in \Sigma_0^+ \;\; \& \;\; (\exists j_1, \ldots j_k) \, ''('' \cdot x_{j_1}^1 \cdot x_{j_2}^2 \cdot x_{j_3}^3 \ldots x_{j_k}^k \in D_2$$

Twierdzenie (Greibach)

L_0 *jest* **najtrudniejszym językiem bezkontekstowym**. *Jeśli* L_0 *można rozpoznać w czasie* $O(n^c)$, *dla pewnej stałej* $c \geq 1$, *to każdy język każdy język bezkontekstowy można rozpoznać w czasie* $O(n^c)$.

W dowodzie zakładamy, że zadana gramatyka bezkontekstowa jest w postaci normalnej Greibach.

Każdemu symbolowi a_i przyporządkujemy pewien zbiór słów H_i.
Dla języka L' oznaczmy

$$Niedet(L') = \{(H_1, H_2, \ldots H_n) \ : H_i \subseteq \Sigma^*, \\ (\exists\ w_i \in H_i,\ 1 \leq i \leq n\ )\ w_1 w_2 \ldots w_n \in L\ \}$$

Niech $N = \{A_1, A_2, \ldots A_r\}$ będzie zbiorem nieterminali, definiujemy r typów nawiasów, gdzie i-ty nawias otwierający jest równy A_i, a zamykający jest równy $\bar{A}_i$. Oznaczmy przez D_r^N odpowiedający język r-nawiasowy.
Niech G będzie gramatyką bezk. w postaci normalnej Greibach.
Dla każdej produkcji $\pi = (A \rightarrow aA_1A_2 \ldots A_k)$ definujemy

$$akcja(\pi) = \bar{A}A_kA_{k-1}\ldots A_1.$$

$akcja(\pi)$ odpowiada jednemu krokowi niedet. automatu stosowego:

$$\bar{A}A_kA_{k-1}\ldots A_1 \quad \equiv \quad \text{pop(A)},\ push(A_k), push(A_{k-1})\ldots push(A_1).$$

W tym rozumowaniu wyobraźmy sobie, że wierzchołek stosu jest po prawej stronie stosu zapisanego poziomo jako s:łowo.

Dla symbolu $a_i \in \Sigma$ oznaczamy

$$H_i = \{\, akcja(\pi_1),\ akcja(\pi_2), \ldots akcja(\pi_i) \,\},$$

gdzie $\pi_1, \ldots \pi_i$ są wszystkimi produkcjami w których a występuje.

Niech S będzie symbolem startowym gramatyki i niech $w = a_1 a_2 \ldots a_n$, wtedy zachodzi:

$$w \in L(G) \Leftrightarrow (\{S\}, H_1, H_2, \ldots H_n) \in Niedet(D_r^N)$$

Powyższa równoważność znacza:

Fakt

$w \in L(G)$ wtw gdy istnieje poprawna historia automatu stosowego akceptującego w za pomocą pustego stosu, gdy początkowo na stosie jest początkowy symbol gramatyki.

Dla uproszczenia załóżmy, że alfabet składa się z liczb naturalnych. Rozważmy gramatykę

$$A \to aBA \mid a \quad B \to bBB \mid b$$

$$H_a = \{\bar{A}AB,\ \bar{A}\},\ H_b = \{\bar{B}BB, \bar{B}\}$$

Jeśli $w = abbba$ to $w \in L(G)$ ponieważ z ciągu zbiorów

$$\{A\}, H_a, H_b, H_b, H_b, H_a$$

można wybrać kolejno słowa

$$A \ \ \bar{A}AB \ \ \bar{B}BB \ \ \bar{B} \ \ \bar{B} \ \ \bar{A}$$

Konkatenacja tych słów daje poprawną historię działania automatu stosowego $A\bar{A}AB\bar{B}BB\bar{B}\bar{B}\bar{A}$. Jest to poprawne wyrażenie nawiasowe w $D_2^{\{A,B\}}$
(lepiej to widać gdy się zastąpi nawiasy A i B przez '[' i '(').

Zatem rozpoznawanie języka sprowadza się do problemu *membership* dla ciągów zbiorów $Niedet(D_r)$ dla r będących pewną stałą.

r typów nawiasów można zakodować dwoma typami nawiasów:

- i-ty nawias otwierający odpowiada słowu "$[\,(^{i-1}$".
- i-ty nawias zamykający odpowiada słowu "$)^{i-1}\,]$".

W ten sposób rozpoznawanie dowolnych języków bezkontekstowych sprowadza się do problemu *membership* dla $Niedet(D_2)$

Język L_0 koduje bezpośrednio problem $Niedet(D_2)$, zatem przynależność do dowolnego języka bezkontekstowego można zredukować do przynależność do L_0. Kończy to szkic dowodu twierdzenia Greibach o języku L_0.

Drugie twierdzenie Greibach można sformułować następująco (silniejsza forma):

Twierdzenie (Greibach)

Dla każdego języka bezkontekstowego L istnieje homomorfizm h taki że $L = h^{-1}(L'_0)$.

Podobnie dowodzi się następującego twierdzenia o reprezentacji.

Twierdzenie (Chomsky–Schützenberger)

Dla każdego języka bezkontekstowego L zachodzi $L = h(D_r \cap R)$, gdzie r zależy od L, h jest homomorfizmem, oraz R jest językiem regularnym.