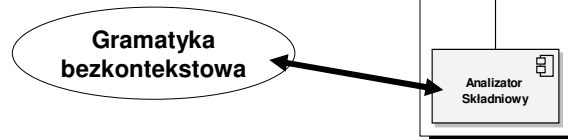


Analiza składniowa – struktura, definicje

- Gramatyki bezkontekstowe
- Przykłady wyprowadzeń
- Struktura analizatora składniowego
- Zbiory „lookahead”, „first” i „follow”



Języki formalne i kompilatory, © by Michał Śmiełek

Gramatyki bezkontekstowe

Gramatyka jest gramatyką bezkontekstową, jeśli wszystkie jej produkcje mają formę:

- $A \rightarrow \beta$, gdzie A jest symbolem nieterminalnym (należy do zbioru Γ , a β należy do zbioru $(\Sigma \cup \Gamma)^*$
- Zwróćmy uwagę, że to oznacza możliwość istnienia produkcji o postaci $A \rightarrow \lambda$, inaczej: produkcji pustej

Poznamy 2 rodzaje gramatyk bezkontekstowych - gramatyki LL i LR.

- Obydwa rodzaje gramatyk umożliwiają przetwarzanie zdania wejściowego od lewej do prawej strony (stąd pierwsze L).
- Gramatyki LL umożliwiają analizę składniową od góry w dół - zaczynamy od symbolu startowego i dochodzimy do zdania końcowego poprzez lewostronne (L) wyprowadzenie
- Gramatyki LR umożliwiają analizę składniową z dołu do góry - zaczynamy od zdania końcowego i dochodzimy do symbolu startowego poprzez prawostronne (R) wyprowadzenie

Wyprowadzenie dla gramatyki bezkontekstowej

Przykładowy zbiór produkcji dla gramatyki bezkontekstowej:

- $S \rightarrow \text{sta } B \text{ stp} \quad B \rightarrow D, B \quad B \rightarrow E ; B$
- $D \rightarrow \text{defv} : \text{id} \quad D \rightarrow \text{defc} : n$
- $E \rightarrow \text{add} \quad E \rightarrow \text{mul} \quad E \rightarrow \lambda$

S to symbol startowy, {S, B, D, E} to zbiór symboli nieterminalnych, {sta, stp, defv, defc, :, n, add, mul} to zbiór symboli terminalnych

Wyprowadzenie zdania polega na kolejnym stosowaniu produkcji. Jeżeli w produkcji napotykamy na symbol nieterminalny, rozwijamy go stosując odpowiednią produkcję (patrz wyprowadzenie podczas wykładu).

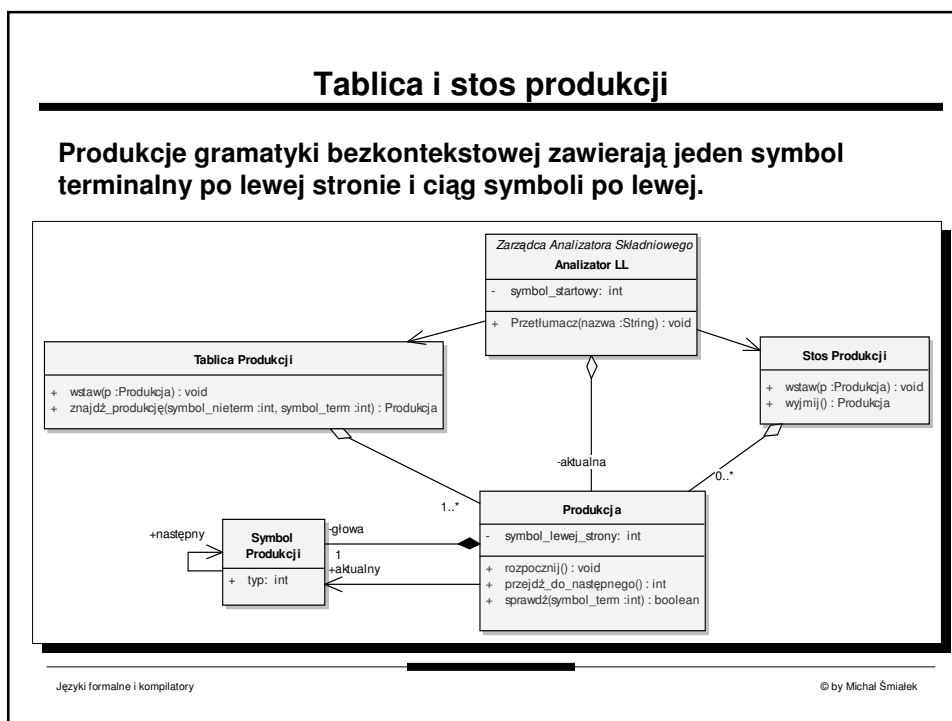
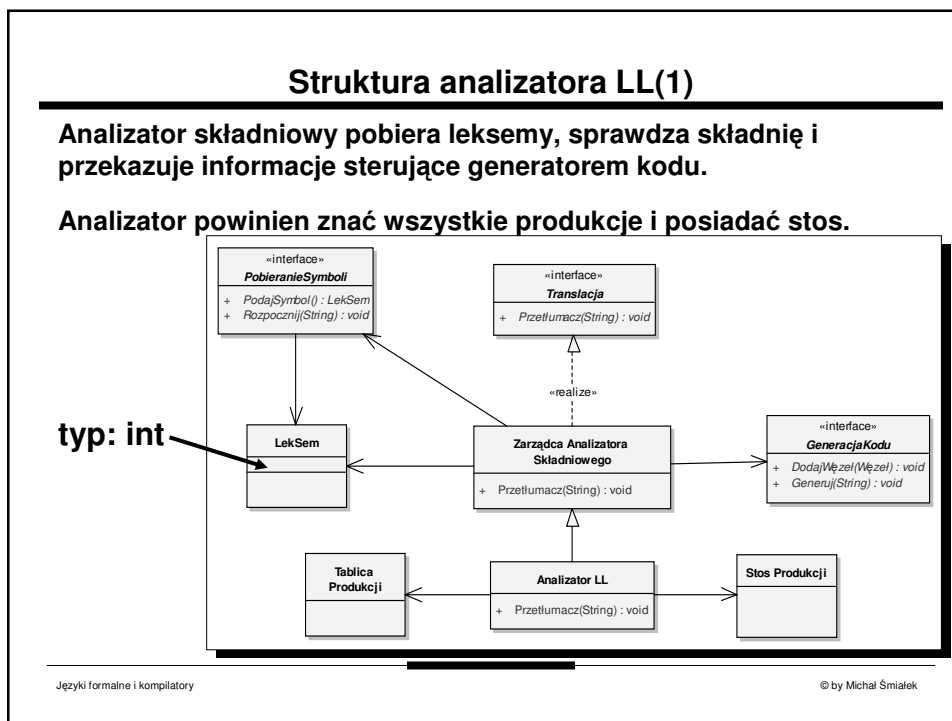
Sprawdzanie składni zdania

Weźmy zdanie: sta defv : id , ; mul ; stp

Jak sprawdzić, czy jest to poprawne zdanie w języku zdefiniowanym gramatyką z poprzedniego slajdu?

Należy zacząć od symbolu startowego, stosując odpowiednią produkcję. Przetwarzając produkcję porównujemy symbole terminalne w produkcji z napotykanymi w zdaniu kolejnymi symbolami. Jeżeli w produkcji napotykamy na symbol nieterminalny – stosujemy odpowiednią produkcję. Jeżeli istnieje kilka produkcji, których lewą stronę stanowi dany symbol nieterminalny, dokonujemy wyboru na podstawie pierwszego znaku produkcji.

Pytanie: czy dla wszystkich gramatyk bezkontekstowych można zastosować powyższą procedurę?



Zbiór „first”

Zbiór „first” zawiera te symbole terminalne, które mogą się znaleźć na początku produkcji.

Definicja (rekurencyjna):

- $\text{first}(a\alpha) = \{a\}$
- $\text{first}(A) = \{a: \exists \alpha (A \rightarrow \alpha \in P; a \in \text{first}(\alpha))\}$
- gdzie: $a \in \Sigma, A \in \Gamma, \alpha \in (\Sigma \cup \Gamma)^*, P$ jest zbiorem produkcji

Zwróćmy uwagę, że:

- $\text{first}(\lambda) = \{\lambda\}$
- $\text{first}(A\alpha) = \text{first}(A)$, gdy $\lambda \notin \text{first}(A)$
 $= (\text{first}(A) - \{\lambda\}) \cup \text{first}(\alpha)$, gdy $\lambda \in \text{first}(A)$

Zbiór „follow”

Zbiór „follow” zawiera te symbole terminalne, które mogą się znaleźć po symbolu nieterminalnym we wszystkich produkcjach.

Definicja:

- $\text{follow}(A) = \{a: \exists X, \alpha, \beta (X \rightarrow \alpha A \beta \in P \text{ i } a \in \text{first}(\beta))\}$,
gdy $\lambda \notin \text{first}(\beta)$
- $\text{follow}(A) = \{a: \exists X, \alpha, \beta (X \rightarrow \alpha A \beta \in P \text{ i } a \in ((\text{first}(\beta) - \{\lambda\}) \cup \text{follow}(X)))\}$,
gdy $\lambda \in \text{first}(\beta)$
- $\text{follow}(S) = (\text{jak wyżej}) \cup \{\lambda\}$
- gdzie: $a \in \Sigma, A, X \in \Gamma, \alpha, \beta \in (\Sigma \cup \Gamma)^*, P$ jest zbiorem produkcji

Zbiór „lookahead”, gramatyka LL(1)

Zbiór „lookahead” dotyczy konkretnej produkcji. Określa on, pod wpływem jakich symboli terminalnych należy zastosować tę produkcję.

Definicja:

- $\text{lookahead}(A \rightarrow \alpha) = \text{first}(\alpha)$, gdy $\lambda \notin \text{first}(A)$
 $= (\text{first}(\alpha) - \{\lambda\}) \cup \text{follow}(A)$, gdy $\lambda \in \text{first}(A)$

Gramatyka jest LL(1), gdy zbiory „lookahead” dla produkcji z tą samą lewą stroną – są rozłączne.

Sprawdzanie gramatyki LL

Aby sprawdzić, czy gramatyka jest LL(1) należy porównać zbiory lookahead dla tych produkcji, które posiadają identyczne lewe strony.

Dla podanej na początku gramatyki należy sprawdzić produkcje:

- $B \rightarrow D, B \quad B \rightarrow E; B$
- $D \rightarrow \text{defv} : \text{id} \quad D \rightarrow \text{defc} : \text{n}$
- $E \rightarrow \text{add} \quad E \rightarrow \text{mul} \quad E \rightarrow \lambda$

Jakie są zbiory lookahead dla tych produkcji. Czy zbiory lookahead są wzajemnie rozłączne?