

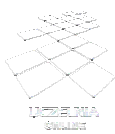
Wprowadzenie do analizy składniowej

Bartosz Bogacki

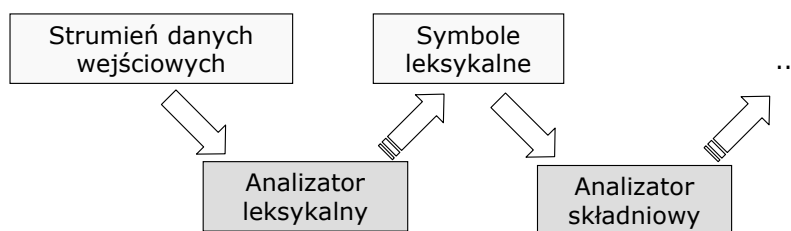
Bartosz.Bogacki@cs.put.poznan.pl



Witam Państwa. Wykład, który za chwilę Państwo wysłuchają dotyczy wprowadzenia do analizy składniowej. Zapraszam serdecznie do wysłuchania.

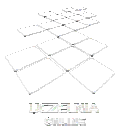


- Analiza składniowa stanowi kolejną po analizie leksykalnej fazę kompilacji
- Wejściem dla analizatora składniowego są symbole leksykalne a wyjściem drzewo rozbioru lub jego uproszczona wersja zwana abstrakcyjnym drzewem składniowym

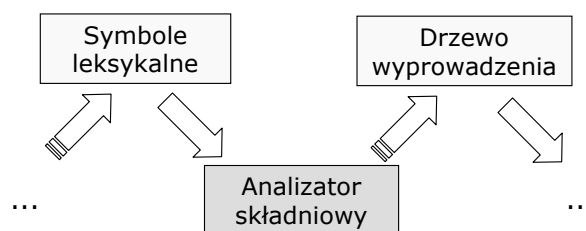


Analiza składniowa (2)

Analiza składniowa stanowi kolejną po analizie leksykalnej fazę kompilacji. Symbole leksykalne wytworzone przez analizator leksykalny trafiają na wejście analizatora składniowego. Tu sprawdzana jest poprawność kompilowanego kodu pod kątem składni.

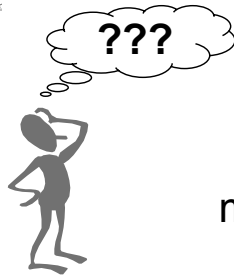
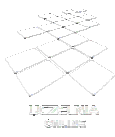


- Analiza składniowa stanowi kolejną po analizie leksykalnej fazę kompilacji
- Wejściem dla analizatora składniowego są symbole leksykalne a wyjściem drzewo rozbioru lub jego uproszczona wersja zwana abstrakcyjnym drzewem składniowym



Analiza składniowa (3)

W wyniku przeprowadzonej analizy składniowej powstaje drzewo wyprowadzenia obrazujące strukturę składniową analizowanego kodu. Przykłady drzew wyprowadzenia przedstawione zostaną w dalszej części wykładu.

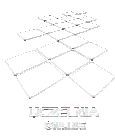


mleko małe dziecko wypilo.

Aby intuicyjnie zrozumieć istotę analizy składniowej przyjrzyjmy się zdaniu przedstawionemu na slajdzie.

mleko małe diecok wypilo.

Już na pierwszy rzut oka widać, że prezentowane zdanie zbudowane jest w sposób nieprawidłowy.



mleko małe dziecko wypilo.

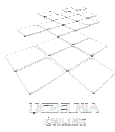


małe dziecko wypilo mleko.

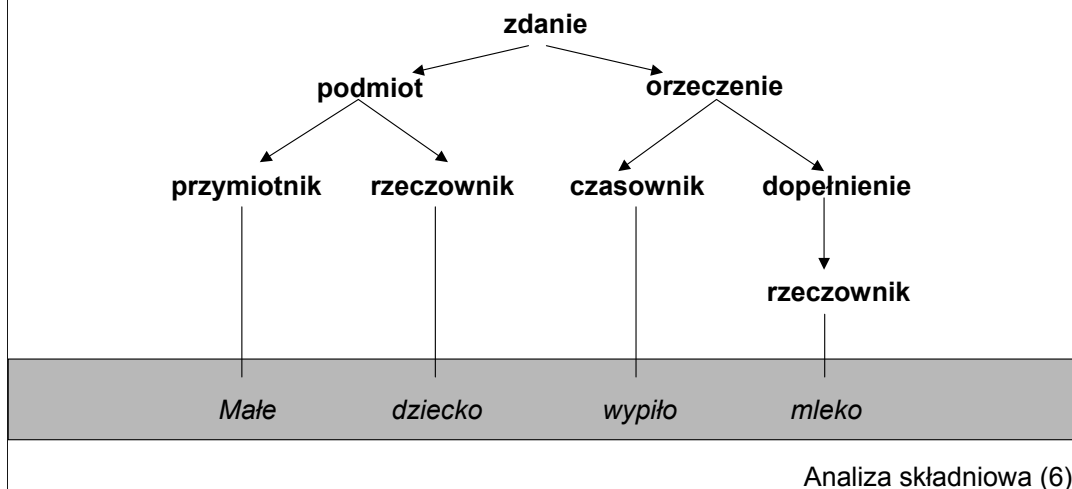
Analiza składniowa (5)

Zdanie przedstawione w dolnej części slajdu jest poprawne i zbudowane zgodnie z naszymi oczekiwaniami. Proszę zwrócić uwagę, iż oba zdania składają się dokładnie z tych samych wyrazów a każdy wyraz rozpatrywany z osobna jest poprawny. Nasza zdolność do oceny poprawności zdania jest intuicyjna i oparta na wykorzystaniu pewnych reguł, które wpojone zostały nam jeszcze w szkole podstawowej. Jeżeli jednak naszym celem jest automatyzacja procesu rozbioru gramatycznego zdań będziemy musieli podać ścisłe reguły rządzące tym procesem.

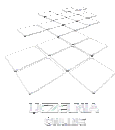
Powstaje więc pytanie - jak zweryfikować w sposób formalny czy zdanie jest poprawne?



Małe dziecko wypilo mleko.



Drzewo przedstawione na slajdzie nazywa się drzewem wyprowadzenia i opisuje składnię, czyli strukturę zdania. Ukazuje sposób w jaki to zdanie ulega rozbirowi na części składniowe. Widzimy, że zdanie składa się z dwóch części: podmiotu i orzeczenia. Podmiot również składa się z dwóch części oznaczonych symbolami przymiotnik oraz rzeczownik. Orzeczenie natomiast składa się z części opisanych symbolami: czasownik i dopełnienie. Aby opisać strukturę zdania wykorzystano specjalne symbole, takie jak: zdanie, podmiot, orzeczenie, dopełnienie, przymiotnik, rzeczownik oraz czasownik. Dla odróżnienia tych symboli od zwykłych słów oznaczone one zostały na slajdzie czcionką pogrubioną.



zdanie → **podmiot** **orzeczenie**
podmiot → **przymiotnik** **rzeczownik**
orzeczenie → **czasownik** **dopełnienie**
dopełnienie → **rzeczownik**
przymiotnik → *małe*
czasownik → *wypiło*
rzeczownik → *dziecko*
rzeczownik → *mleko*

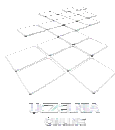
Analiza składniowa (7)

Oczywiście aby móc utworzyć drzewo opisujące strukturę zdania, musimy wprowadzić pewne reguły, które będziemy w tym celu konsekwentnie stosowali. Reguły te będą definiowały gramatykę w odniesieniu do której będziemy analizowali zdanie.

Na slajdzie przedstawiona została w formie reguł produkcji gramatyka wykorzystana do zbudowania drzewa prezentowanego na poprzednim slajdzie. Przyjrzyjmy się bliżej zastosowanej notacji.

Każdy wiersz to osobna produkcja. Jeśli w produkcji zastąpimy symbol strzałki sformułowaniem „może się składać z”, to możemy wykorzystać zaprezentowane reguły do konstrukcji drzewa składniowego.

Zauważmy, że gramatyka może posiadać więcej niż jedną regułę formowania symbolu znajdującego się po lewej stronie produkcji. Przykładowo symbol „rzeczownik” może zostać utworzony zarówno ze słowa „dziecko” jak i ze słowa „mleko”.



Wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

Bezpośrednio wyprowadzalny:

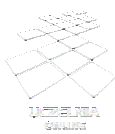
zdanie $\Rightarrow$ **podmiot orzeczenie**

Wyprowadzalny w co najmniej jednym kroku:

zdanie $\Rightarrow$ *małe dziecko wypilo mleko*

Analiza składniowa (8)

Drzewo wyprowadzenia może być traktowane jako graficzna forma wyprowadzenia. Wprowadzimy teraz dwa nowe pojęcia. Bezpośrednia wyprowadzalność oznacza przekształcenie ciągu znajdującego się z lewej strony symbolu $\Rightarrow$ polegające na zastąpieniu w nim jednego z symboli prawą stroną odpowiedniej reguły. W wyniku tego przekształcenia otrzymujemy ciąg znajdujący się po prawej stronie symbolu $\Rightarrow$. Sekwencję takich przekształceń pozwalającą udowodnić, że dane zdanie da się wyprowadzić z pewnego symbolu nazywamy „wyprowadzalnością w co najmniej jednym kroku” i oznaczamy symbolem $\Rightarrow$ z plusem. Slajd przedstawia bezpośrednią wyprowadzalność formy zdaniowej „podmiot orzeczenie” z symbolu „zdanie” oraz wyprowadzalność w co najmniej jednym kroku zdania „małe dziecko wypilo mleko”.



Bezpośrednia wyprowadzalność

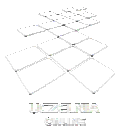
zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ podmiot orzeczenie

Analiza składniowa (9)

Przyjrzyjmy się teraz w jaki sposób można udowodnić wyprowadzalność w co najmniej jednym kroku zdania „małe dziecko wypilo mleko” z symbolu zdanie stosując do tego celu bezpośrednią wyprowadzalność.

W pierwszym kroku za pomocą pierwszej produkcji w gramatyce przedstawionej w górnej części slajdu dokonujemy bezpośredniego wyprowadzenia formy zdaniowej „podmiot orzeczenie”.



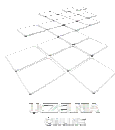
Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot** **orzeczenie**
 ⇒ **przymiotnik** **rzeczownik** **orzeczenie**

Analiza składniowa (10)

Teraz wykorzystując drugą produkcję rozbijamy podmiot na przymiotnik i rzeczownik. Uzyskujemy w ten sposób kolejne bezpośrednie wyprowadzenie będące formą zdaniową „przymiotnik rzeczownik orzeczenie”.



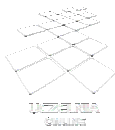
Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot** **orzeczenie**
 ⇒ **przymiotnik** **rzeczownik** **orzeczenie**
 ⇒ *małe* **rzeczownik** **orzeczenie**

Analiza składniowa (11)

Kolejnym krokiem jest zamiana symbolu „przymiotnik” na symbol „małe” zgodnie z produkcją piątą. Uzyskujemy formę zdaniową „małe rzeczownik orzeczenie”.



Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot** **orzeczenie**

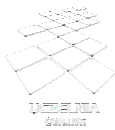
⇒ **przymiotnik** **rzeczownik** **orzeczenie**

⇒ *małe* **rzeczownik** **orzeczenie**

⇒ *małe dziecko* **orzeczenie**

Analiza składniowa (12)

Symbol „rzeczownik” zamieniamy na symbol „dziecko”. Forma zdaniowa uzyskana w bezpośrednim wyprowadzeniu to „małe dziecko orzeczenie”.



Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot** **orzeczenie**

⇒ **przymiotnik** **rzeczownik** **orzeczenie**

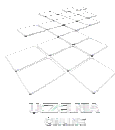
⇒ *małe* **rzeczownik** **orzeczenie**

⇒ *małe dziecko* **orzeczenie**

⇒ *małe dziecko* **czasownik** **dopełnienie**

Analiza składniowa (13)

Następnym krokiem jest zamiana symbolu **orzeczenie** na symbole „**czasownik** **dopełnienie**” zgodnie z produkcją trzecią. Uzyskana w ten sposób forma zdaniowa to: „*małe dziecko* **czasownik** **dopełnienie**”.



Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot orzeczenie**

⇒ **przymiotnik rzeczownik orzeczenie**

⇒ *małe* **rzeczownik orzeczenie**

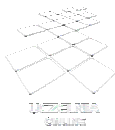
⇒ *małe dziecko* **orzeczenie**

⇒ *małe dziecko* **czasownik dopełnienie**

⇒ *małe dziecko wypilo* **dopełnienie**

Analiza składniowa (14)

Po zamianie symbolu czasownik na symbol „wypiło” uzyskujemy formę zdaniową: „małe dziecko wypilo dopełnienie”.



Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot** **orzeczenie**

⇒ **przymiotnik** **rzeczownik** **orzeczenie**

⇒ *małe* **rzeczownik** **orzeczenie**

⇒ *małe* *dziecko* **orzeczenie**

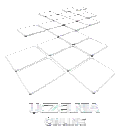
⇒ *małe* *dziecko* **czasownik** **dopełnienie**

⇒ *małe* *dziecko* *wypiło* **dopełnienie**

⇒ *małe* *dziecko* *wypiło* **rzeczownik**

Analiza składniowa (15)

Kolejnym krokiem jest zamiana symbolu „dopełnienie” na symbol „rzeczownik”.
Otrzymujemy: „małe dziecko wypilo rzeczownik”.



Bezpośrednia wyprowadzalność

zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie ⇒ **podmiot** **orzeczenie**

⇒ **przymiotnik** **rzeczownik** **orzeczenie**

⇒ *małe* **rzeczownik** **orzeczenie**

⇒ *małe* *dziecko* **orzeczenie**

⇒ *małe* *dziecko* **czasownik** **dopełnienie**

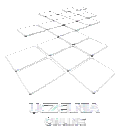
⇒ *małe* *dziecko* *wypiło* **dopełnienie**

⇒ *małe* *dziecko* *wypiło* **rzeczownik**

⇒ *małe* *dziecko* *wypiło* *mleko*

Analiza składniowa (16)

Ostatnim krokiem jest zamiana symbolu „rzeczownik” na symbol „mleko”. Uzyskaliśmy w ten sposób zdanie „małe dziecko wypilo mleko” i udowodniliśmy, że można je wyprowadzić w co najmniej jednym kroku z symbolu „zdanie”.



zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie $\Rightarrow$ *małe rzeczownik wypilo rzeczownik*

zdanie $\Rightarrow$ *małe dziecko wypilo mleko*

zdanie $\Rightarrow$ *małe mleko wypilo dziecko*

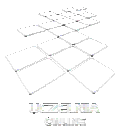
Analiza składniowa (17)

Proszę zwrócić uwagę, że dla rozpatrywanej gramatyki stosując bezpośrednio wyprowadzenie można dowieść poprawności składniowej większej liczby zdań.

Można udowodnić, że z symbolu „zdanie” istnieje wyprowadzenie w co najmniej jednym kroku formy zdaniowej „małe rzeczownik wypilo rzeczownik”.

Uwzględniając fakt, że symbol „rzeczownik” występuje w dwóch produkcjach poprawne w sensie składniowym są zdania:

- małe dziecko wypilo mleko
- małe mleko wypilo dziecko



zdanie → podmiot orzeczenie	przymiotnik → <i>małe</i>
podmiot → przymiotnik rzeczownik	czasownik → <i>wypiło</i>
orzeczenie → czasownik dopełnienie	rzeczownik → <i>dziecko</i>
dopełnienie → rzeczownik	rzeczownik → <i>mleko</i>

zdanie $\Rightarrow$ *małe* **rzeczownik** *wypiło* **rzeczownik**

zdanie $\Rightarrow$ *małe* *dziecko* *wypiło* *mleko*

zdanie $\Rightarrow$ *małe* *mleko* *wypiło* *dziecko*

zdanie $\Rightarrow$ *małe* *dziecko* *wypiło* *dziecko*

zdanie $\Rightarrow$ *małe* *mleko* *wypiło* *mleko*

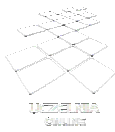
Analiza składniowa (18)

Poprawne składniowo są również następujące zdania, które z semantycznego punktu widzenia nie mają sensu:

- *małe* *dziecko* *wypiło* *dziecko*
- *małe* *mleko* *wypiło* *mleko*

Pamiętajmy jednak, że semantyka jest rozpatrywana osobno, a na tym etapie koncentrujemy się na składni.

Z punktu widzenia analizy składniowej wszystkie cztery zdania są poprawne.



Gramatyka bezkontekstowa

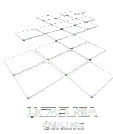
- Jeśli S_1 i S_2 są instrukcjami a E jest wyrażeniem, to:
"if E then S_1 else S_2 " jest instrukcją
- **instrukcja** $\rightarrow$ *if wyrażenie then instrukcja else instrukcja*

Analiza składniowa (19)

Konstrukcje w językach programowania często mają strukturę rekurencyjną. Ponieważ wyrażenia regularne, które poznaliśmy podczas omawiania analizy leksykalnej nie są wystarczające do opisu większości tych konstrukcji, potrzebny jest mocniejszy mechanizm.

Z pomocą pojawiają się tu gramatyki bezkontekstowe.

Przykładowo popularną w wielu językach instrukcję warunkową można zdefiniować za pomocą reguły: Jeśli S_1 i S_2 są instrukcjami a E jest wyrażeniem, to **"if E then S_1 else S_2 "** jest również instrukcją. Opisanie tej postaci instrukcji za pomocą wyrażenia regularnego nie jest możliwe. Można natomiast stworzyć odpowiednią gramatykę bezkontekstową.



- Produkcja

Produkcja (lub reguła zastępowania) to uporządkowana para (A, γ) zapisywana zwykle w postaci $A \rightarrow \gamma$, gdzie A jest symbolem a γ jest niepustym napisem.

Symbol A jest lewą stroną a γ prawą stroną produkcji.

Do tej pory mówiliśmy o gramatykach oraz wykorzystywanych mechanizmach w sposób dość nieformalny. Zajmiemy się więc teraz zdefiniowaniem potrzebnych pojęć. Zaczniemy od produkcji.

Produkcja (zwana również regułą zastępowania) to uporządkowana para (A, γ) zapisywana zwykle w postaci $A \rightarrow \gamma$, gdzie A jest symbolem a γ jest niepustym napisem. Symbol A jest lewą stroną produkcji, podczas gdy γ prawą stroną produkcji.



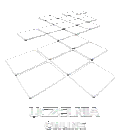
- Terminale i nieterminale

Jeżeli dana jest gramatyka G , to każdy symbol użyty po lewej stronie dowolnej produkcji nazywany jest *symbolem nieterminalnym (nieterminalem)*.

Pozostałe symbole są *symbolami terminalnymi (terminalami)*.

Jeżeli dana jest gramatyka G , to każdy symbol użyty po lewej stronie dowolnej produkcji nazywany jest symbolem nieterminalnym (nieterminalem).

Pozostałe symbole są symbolami terminalnymi (terminalami). Terminale są symbolami podstawowymi z których tworzone są napisy. Gdy mówi się o o gramatykach dla języków programowania - terminal jest synonimem symbolu leksykalnego.



- Symbol startowy

Wyróżniony jeden symbol nieterminalny w gramatyce.
Zbiór napisów definiowanych przez ten symbol jest
językiem definiowanym przez gramatykę.

Symbol startowy to wyróżniony, jeden symbol nieterminalny w gramatyce. Zbiór napisów definiowanych przez ten symbol (czyli wyprowadzalnych z tego symbolu) jest językiem definiowanym przez gramatykę.



Gramatyki (produkcyjne)

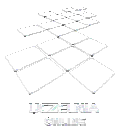
zdanie $\rightarrow$ podmiot orzeczenie
podmiot $\rightarrow$ przymiotnik rzeczownik
orzeczenie $\rightarrow$ czasownik dopełnienie
dopełnienie $\rightarrow$ rzeczownik
przymiotnik $\rightarrow$ małe
czasownik $\rightarrow$ wypić
rzeczownik $\rightarrow$ dziecko
rzeczownik $\rightarrow$ mleko

$E \rightarrow E + T$
 $E \rightarrow T$
 $T \rightarrow T * F$
 $T \rightarrow F$
 $F \rightarrow (E)$
 $F \rightarrow id$

Analiza składniowa (23)

Przyjrzyjmy się teraz przykładom. Na slajdzie znajdują się dwie gramatyki. Gramatyka po lewej stronie została zaprezentowana we wcześniejszej części wykładu. Po prawej stronie znajduje się gramatyka weryfikująca czy zdanie jest wyrażeniem arytmetycznym składającym się z pewnego identyfikatora, odpowiednio zastosowanych nawiasów oraz prawidłowo użytych operatorów mnożenia i dodawania.

Każdy wiersz gramatyki to oczywiście pojedyncza produkcja.



Nieterminale

zdanie → podmiot orzeczenie
podmiot → przymiotnik rzeczownik
orzeczenie → czasownik dopełnienie
dopełnienie → rzeczownik
przymiotnik → małe
czasownik → wypilo
rzeczownik → dziecko
rzeczownik → mleko

E → **E** + **T**
E → **T**
T → **T** * **F**
T → **F**
F → (**E**)
F → id

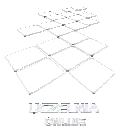
Analiza składniowa (24)

Na slajdzie pogrubioną czcionką oznaczone zostały nieterminale. W gramatyce po lewej stronie są to:

- zdanie, podmiot, orzeczenie, dopełnienie, przymiotnik, czasownik oraz rzeczownik.

W gramatyce przedstawionej po prawej stronie nieterminalami są:

- E, T oraz F.



Terminale

zdanie $\rightarrow$ podmiot orzeczenie
podmiot $\rightarrow$ przymiotnik rzeczownik
orzeczenie $\rightarrow$ czasownik dopełnienie
dopełnienie $\rightarrow$ rzeczownik
przymiotnik $\rightarrow$ **małe**
czasownik $\rightarrow$ **wypiło**
rzeczownik $\rightarrow$ **dziecko**
rzeczownik $\rightarrow$ **mleko**

$E \rightarrow E + T$
 $E \rightarrow T$
 $T \rightarrow T * F$
 $T \rightarrow F$
 $F \rightarrow (E)$
 $F \rightarrow \text{id}$

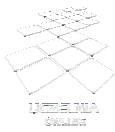
Analiza składniowa (25)

Tym razem na slajdzie pogrubioną czcionką oznaczone zostały terminale. W gramatyce po lewej stronie są to:

- **małe**, **wypiło**, **dziecko** oraz **mleko**.

W gramatyce przedstawionej po prawej stronie terminalami są:

- **+**, *****, **(**, **)** oraz **id**.



Symbol startowy

zdanie → podmiot orzeczenie
 podmiot → przymiotnik rzeczownik
 orzeczenie → czasownik dopełnienie
 dopełnienie → rzeczownik
 przymiotnik → małe
 czasownik → wypilo
 rzeczownik → dziecko
 rzeczownik → mleko

E → **E** + T
E → T
 T → T * F
 T → F
 F → (**E**)
 F → id

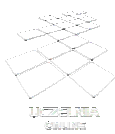
Analiza składniowa (26)

Na slajdzie pogrubioną czcionką oznaczony został dla każdej z gramatyk jej symbol startowy.

W gramatyce po lewej stronie jest to nieterminal „zdanie”.

W gramatyce po prawej stronie jest to nieterminal „E”.

Zwyczajowo symbolem startowym jest nieterminal znajdujący się po lewej stronie pierwszej produkcji w gramatyce.



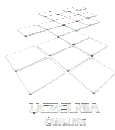
Gramatyka bezkontekstowa

- Gramatyka bezkontekstowa składa się z:
 - Terminali (np. *if*, *then*, *małe*, *mleko*, ...)
 - Nieterminali (np. **instrukcja**, **rzeczownik**, ...)
 - Symbolu startowego
 - Produkcji

Analiza składniowa (27)

Gramatyki, które będziemy rozważać to gramatyki bezkontekstowe. Gramatyka bezkontekstowa składa się z:

- Terminali
- Nieterminali
- Symbolu startowego
- Reguł produkcji



- Wyprowadzalność

$\alpha A \beta \Rightarrow \alpha \gamma \beta$, jeśli:

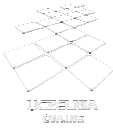
- $A \rightarrow \gamma$ jest produkcją
- α i β są dowolnymi ciągami symboli gramatyki.

Jeśli $\alpha_1 \Rightarrow \alpha_2 \Rightarrow \alpha_3 \Rightarrow \dots \Rightarrow \alpha_n$, to α_n **wyprowadza się** z α_1 , co zapisuje się $\alpha_1 \Rightarrow^* \alpha_n$

Przejdźmy teraz do definicji wyprowadzalności.

Alfa gamma beta jest bezpośrednio wyprowadzalne z alfa A beta jeśli istnieje produkcja dla A, która po prawej stronie zawiera symbol gamma, a alfa i beta są dowolnymi ciągami symboli gramatyki.

Jeśli istnieje sekwencja bezpośrednich wyprowadzeń alfa n z alfa jeden, to alfa n wyprowadza się z alfa 1. Sytuację taką oznaczamy zapisując symbol gwiazdki (*) nad symbolem wyprowadzalności ($\Rightarrow^*$).



Lewostronna wyprowadzalność

- Lewostronna wyprowadzalność

$\alpha A \beta \Rightarrow \alpha \gamma \beta$, jeśli:
Is

- $A \rightarrow \gamma$ jest produkcją
- α jest ciągiem zawierającym tylko symbole terminalne
- β jest dowolnym ciągiem symboli gramatyki.

Analiza składniowa (29)

W każdym kroku wyprowadzenia należy dokonać podwójnego wyboru. Po pierwsze, który nieterminal będzie zastępowany, a po drugie która produkcja zostanie użyta do zamiany nieterminala na odpowiadającą mu formę zdaniową.

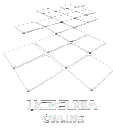
Jeśli w danym kroku do wymiany wybrany zostaje skrajnie lewy nieterminal, to takie wyprowadzenie nazywa się lewostronnym. Lewostronne wyprowadzenie oznaczamy poprzez umieszczenie liter Is pod symbolem wyprowadzalności ($\Rightarrow$).

Alfa gamma beta jest bezpośrednio lewostronnie wyprowadzalne z alfa A beta jeśli:

-istnieje produkcja dla A, która po prawej stronie ma tylko symbol gamma

-Alfa jest ciągiem zawierającym tylko symbole terminalne

-Beta jest dowolnym ciągiem symboli gramatyki (czyli może zawierać zarówno terminale, jak i nieterminale)



Prawostronna wyprowadzalność

- Prawostronna wyprowadzalność

$\alpha A \beta \Rightarrow \alpha \gamma \beta$, jeśli:
ps

- $A \rightarrow \gamma$ jest produkcją
- α jest dowolnym ciągiem symboli gramatyki
- β jest ciągiem zawierającym tylko symbole terminalne

Analiza składniowa (30)

Jeśli do zastąpienia wybrany zostaje skrajnie prawy nieterminal, to wyprowadzenie nosi nazwę prawostronnego.

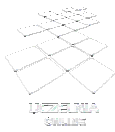
Prawostronne wyprowadzenie oznaczamy poprzez umieszczenie liter ps pod symbolem wyprowadzalności ($\Rightarrow$).

Alfa gamma beta jest bezpośrednio prawostronnie wyprowadzalne z alfa A beta jeśli:

-istnieje produkcja dla A, która po prawej stronie ma tylko symbol gamma

-Alfa jest dowolnym ciągiem symboli gramatyki (czyli może zawierać zarówno terminale, jak i nieterminale)

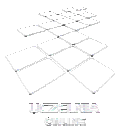
-Beta jest ciągiem zawierającym tylko symbole terminalne.



- Dla danej gramatyki G z symbolem startowym S :
 - $L(G)$ jest *językiem generowanym przez G*
- Ciąg terminali α należy do $L(G) \Leftrightarrow S \xRightarrow{*} \alpha$

Dla danej gramatyki G z pewnym wyróżnionym symbolem S (nazywanym symbolem startowym) relacja wyprowadzenia może zostać użyta do zdefiniowania języka $L(G)$, czyli języka generowanego przez gramatykę G w następujący sposób.

Ciąg terminali α należy do języka generowanego przez gramatykę G wtedy i tylko wtedy gdy jest wyprowadzalny z symbolu startowego S .



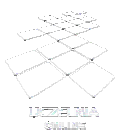
- Dla danej gramatyki G oraz $S \xRightarrow{*} \beta$:
 - Jeśli β zawiera tylko terminale, to β nazywamy *zdaniem*
 - Jeśli β zawiera terminale oraz nieterminale lub same nieterminale, to β nazywamy *formą zdaniową*
 - *Zdanie* jest *formą zdaniową* bez nieterminali

Niech dana będzie gramatyka G z symbolem startowym S , z którego wyprowadzalne jest β .

-Jeśli β zawiera tylko terminale, to β nazywamy zdaniem

-Jeśli β zawiera terminale oraz nieterminale lub same nieterminale, to β nazywamy formą zdaniową

Łatwo zauważyć, że zdanie jest formą zdaniową bez nieterminali.



Rekurencja w gramatykach

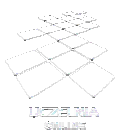
- Gramatyka rekursywna
 - $A \xRightarrow{*} \alpha A \beta \quad A \in N; \alpha, \beta \in (N \cup T)$
- Gramatyka lewostronnie rekursywna
 - $A \xRightarrow{*} A \alpha \quad A \in N; \alpha \in (N \cup T)$
- Gramatyka prawostronnie rekursywna
 - $A \xRightarrow{*} \alpha A \quad A \in N; \alpha \in (N \cup T)$

Analiza składniowa (33)

Gramatykę nazywamy rekurencyjną (lub rekursywną) jeśli w wyprowadzeniu dla danego symbolu nieterminalnego występuje ten sam symbol.

Jeśli symbol ten występuje na skrajnie lewej pozycji, to mamy do czynienia z lewostronną rekurencją.

Jeśli natomiast wspomniany symbol nieterminalny występuje na skrajnie prawej pozycji, to sytuację taką nazywamy prawostronną rekurencją.



Puste produkcje

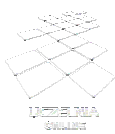
- Pusta produkcja (ε -produkcja)
 - $A \rightarrow \varepsilon \quad A \in N$
- Przykład:

$E \rightarrow (E)$
$E \rightarrow \varepsilon$

Analiza składniowa (34)

Produkcją pustą (epsilon produkcją) nazywamy produkcję, która po prawej stronie nie ma ani terminali ani nieterminali. Produkcja taka może być wykorzystana do utworzenia pewnego symbolu nieterminalnego w dowolnym momencie przetwarzania. Pozwala to na bardziej elastyczne stosowanie rekurencji.

Przykładem gramatyki wykorzystującej pustą produkcję jest przedstawiona na slajdzie gramatyka sprawdzająca czy występuje jednakowa liczba nawiasów otwierających co nawiasów zamykających.



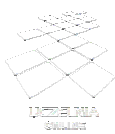
Drzewo wyprowadzenia

- Graficzną reprezentacją wyprowadzenia jest drzewo wyprowadzenia
 - Każdy węzeł wewnętrzny odpowiada nieterminalowi
 - Dzieci węzła są prawymi stronami produkcji
 - Liście w drzewie wyprowadzenia odpowiadają terminalom

Analiza składniowa (35)

Graficzną reprezentacją wyprowadzenia jest drzewo wyprowadzenia. W drzewie tym każdy węzeł wewnętrzny odpowiada nieterminalowi. Dzieci węzła są prawymi stronami produkcji, natomiast liście w drzewie wyprowadzenia odpowiadają terminalom.

Zaletą takiej reprezentacji jest jej czytelność. Wadą natomiast jest strata informacji o tym w jakiej kolejności dokonane zostały wyprowadzenia.



Drzewo wyprowadzenia

Gramatyka

```

E → E + E
E → E * E
E → ( E )
E → id

```

Ciąg wejściowy

id * (id + id)

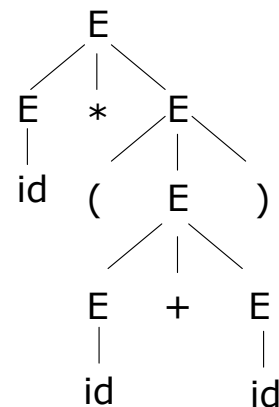
Wyprowadzenie

```

E ⇒ E * E
  ⇒ id * E
  ⇒ id * ( E )
  ⇒ id * ( E + E )
  ⇒ id * ( id + id )

```

Drzewo wyprowadzenia

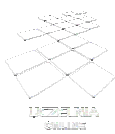


Analiza składniowa (36)

Przyjrzyjmy się elementom, które do tej pory poznaliśmy. W lewej części slajdu znajduje się gramatyka. Pozwala ona określić czy ciąg wejściowy jest poprawnym wyrażeniem arytmetycznym zawierającym poprawnie zastosowane nawiasy okrągłe oraz operatory dodawania i mnożenia. Przypomnijmy, że uzyskana odpowiedź brzmi – tak, jeśli z symbolu startowego gramatyki możemy dokonać wyprowadzenia ciągu wejściowego.

W środkowej części slajdu znajduje się wyprowadzenie. Wybierając odpowiednie produkcje dokonujemy bezpośredniego wyprowadzenia rozpoczynając od symbolu startowego.

W prawej części slajdu znajduje się drzewo wyprowadzenia dla rozważanego przykładu. Korzeniem drzewa jest oczywiście symbol startowy. Liśćmi są terminale, które występowały w ciągu wejściowym. Proszę zwrócić uwagę, że przechodząc przez liście drzewa od strony lewej do prawej otrzymamy zdanie będące ciągiem wejściowym.



Niejednoznaczność

Gramatyka

```

E → E + E
E → E * E
E → ( E )
E → id

```

Ciąg wejściowy

```
id + id * id
```

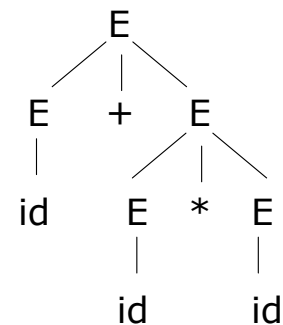
Wyprowadzenie

```

E ⇒ E + E
  ⇒ id + E
  ⇒ id + E * E
  ⇒ id + id * E
  ⇒ id + id * id

```

Drzewo wyprowadzenia



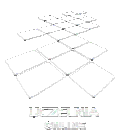
Analiza składniowa (37)

Rozpatrzmy kolejny przykład.

Po lewej stronie slajdu znajduje się poznana wcześniej gramatyka. Postaramy się przygotować wyprowadzenie oraz drzewo wyprowadzenia dla ciągu wejściowego: `id + id * id`.

Środkowa część slajdu przedstawia wyprowadzenie dla rozważanego ciągu wejściowego. Rozpoczyna się ono od zamiany nieterminala `E` na symbole znajdujące się po prawej stronie pierwszej produkcji, czyli `E + E`. Następnie dokonujemy zamiany pierwszego nieterminala `E` na `id` zgodnie z ostatnią produkcją występującą w gramatyce. Kontynuując proces zamieniania (czyli stosując bezpośrednie wyprowadzenie) uzyskujemy zdanie będące ciągiem danych wejściowych.

Po prawej stronie slajdu znajduje się drzewo wyprowadzenia dla rozpatrywanego ciągu wejściowego.



Niejednoznaczność

Gramatyka

$E \rightarrow E + E$
$E \rightarrow E * E$
$E \rightarrow (E)$
$E \rightarrow id$

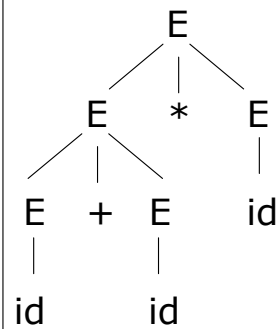
Ciąg wejściowy

id + id * id

Wyprowadzenie

$$\begin{aligned}
 E &\Rightarrow E * E \\
 &\Rightarrow E + E * E \\
 &\Rightarrow id + E * E \\
 &\Rightarrow id + id * E \\
 &\Rightarrow id + id * id
 \end{aligned}$$

Drzewo wyprowadzenia



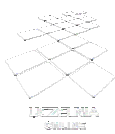
Analiza składniowa (38)

Dla ciągu wejściowego id + id * id można jednak dokonać wyprowadzenia również w inny sposób.

Środkowa część slajdu przedstawia alternatywne wyprowadzenie. Proszę zwrócić uwagę, że tym razem w pierwszym kroku wykorzystujemy drugą produkcję gramatyki. Dokonujemy zamiany E na E * E. Następnie pierwsze E zamieniamy na E + E.

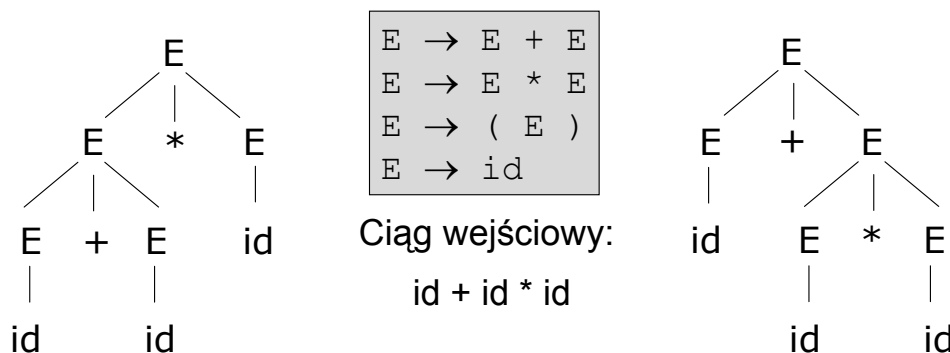
Po prawej stronie slajdu znajduje się alternatywne drzewo wyprowadzenia dla rozpatrywanego ciągu wejściowego.

Jak łatwo dostrzec drzewo wyprowadzenia różni się od drzewa prezentowanego na poprzednim slajdzie.



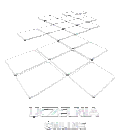
Niejednoznaczność

Gramatyka, w której zdanie ma więcej niż jedno drzewo wyprowadzenia nazywa się *gramatyką niejednoznaczną*.



Analiza składniowa (39)

Gramatyka, w której zdanie ma więcej niż jedno drzewo wyprowadzenia nazywa się *gramatyką niejednoznaczną*. Na slajdzie pokazano rozpatrywaną gramatykę, która jest przykładem gramatyki niejednoznacznej oraz dwa różne drzewa wyprowadzenia, które uzyskano dla tego samego ciągu wejściowego: id + id * id.



Równoważność gramatyk

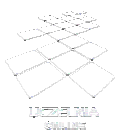
Dwie gramatyki *generujące* ten sam język nazywamy *gramatykami równoważnymi*.

```
E → E + E
E → E * E
E → ( E )
E → id
```

```
E → E + T
E → T
T → T * F
T → F
F → ( E )
F → id
```

Analiza składniowa (40)

Dwie gramatyki generujące ten sam język nazywamy gramatykami równoważnymi. Proszę zwrócić uwagę, że dla gramatyki niejednoznacznej znajdującej się po lewej stronie można stworzyć gramatykę równoważną, która nie jest niejednoznaczna. Przykład takiej gramatyki przedstawiony jest po prawej stronie slajdu.



Równoważność gramatyk

Gramatyka

$E \rightarrow E + T$
$E \rightarrow T$
$T \rightarrow T * F$
$T \rightarrow F$
$F \rightarrow (E)$
$F \rightarrow id$

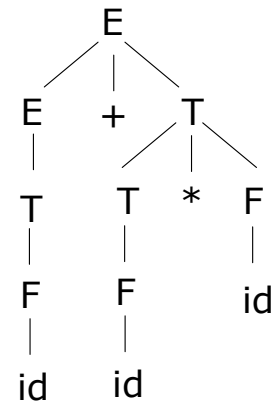
Ciąg wejściowy

id + id * id

Wyprowadzenie

$$\begin{aligned}
 E &\Rightarrow E + T \\
 &\Rightarrow T + T \\
 &\Rightarrow F + T \\
 &\Rightarrow id + T \\
 &\Rightarrow id + T * F \\
 &\Rightarrow id + F * F \\
 &\Rightarrow id + id * F \\
 &\Rightarrow id + id * id
 \end{aligned}$$

Drzewo wyprowadzenia



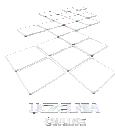
Analiza składniowa (41)

Przyjrzyjmy się w takim razie jak będzie wyglądało wyprowadzenie dla rozpatrywanego ciągu wejściowego (czyli id + id * id) jeśli wykorzystamy gramatykę, która nie jest niejednoznaczna. Gramatyka ta przedstawiona jest po lewej stronie slajdu.

W środkowej części slajdu dostrzec można wyprowadzenie dla zadanego ciągu wejściowego.

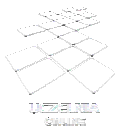
W prawej części slajdu znajduje się drzewo wyprowadzenia.

Łatwo sprawdzić, że korzystając z tej gramatyki dla każdego poprawnie zbudowanego ciągu wejściowego uzyskamy jedno drzewo wyprowadzenia.



- Gramatyka bezkontekstowa
- Produkcje
- Terminale i nieterminale
- Wyprowadzalność
- Drzewo wyprowadzenia
- Niejednoznaczność
- Równoważność gramatyk

Podsumowując, w wykładzie wprowadzono pojęcie gramatyki bezkontekstowej, produkcji oraz terminali i nieterminali. Omówiono kwestie wyprowadzalności, bezpośredniej wyprowadzalności oraz wyprowadzalności co najmniej w jednym kroku. Zademostrowano drzewa wyprowadzenia. Omówiono niejednoznaczność oraz równoważność gramatyk.



- Aho A. V., Sethi R., Ullman J. D., *Kompilatory Reguły, metody i narzędzia*, WNT 2002
- Tremblay J. P., Sorenson P. G., *The Theory and Practice of Compiler Writing*, McGraw-Hill, 1985
- Wilhelm R., Maurer D., *Compiler Design*, Addison-Wesley 1995

Więcej informacji na temat omówionych w wykładzie zagadnień można znaleźć w literaturze przedstawionej na slajdzie.