



AKADEMIA GÓRNICZO-HUTNICZA  
IM. STANISŁAWA STASZICA W KRAKOWIE

## Algorytmy parsingu

### Teoria automatów i języków formalnych

Dr inż. Janusz Majewski  
Katedra Informatyki



## Metody rozbioru gramatycznego

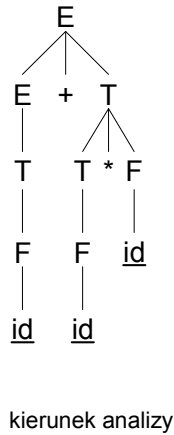
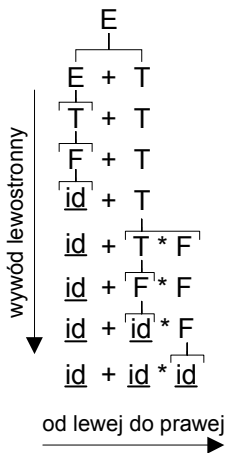
Problem rozbioru gramatycznego jest próbą odpowiedzi na pytanie: „czy dany ciąg symboli terminalnych jest słowem należącym do języka  $L(G)$  generowanego przez daną gramatykę?”. Istota postępowania jest następująca: jeśli dla danego słowa można zbudować drzewo rozbioru w danej gramatyce, to słowo to należy do języka generowanego przez gramatykę, jeśli zaś dla badanego słowa nie można zbudować drzewa rozbioru, wówczas słowo nie należy do języka. W zależności od strategii budowania drzewa rozbioru stosujemy jeden z dwóch algorytmów:

- algorytm wyprowadzający „top-down” (budujemy drzewo w kolejności: od korzenia ku liściom, odtwarzając wyprowadzenie lewostronne)
- algorytm redukujący „bottom-up” (budujemy drzewo w odwrotnej kolejności: od liści ku korzeniowi, odtwarzając wyprowadzenie prawostronne)

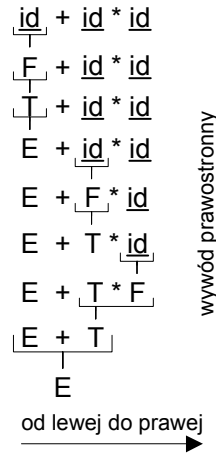
Rozważany ciąg symboli badany jest zawsze od lewej strony, w tym sensie oba algorytmy są lewostronne.

## Przykład

### Top-down



### Bottom-up



### Przykład:

Gramatyka:

$E \rightarrow E + T \mid T$

$T \rightarrow T * F \mid F$

$F \rightarrow (E) \mid id$

Analizowane słowo:

$id + id * id$

## Metoda „top-down”

Symbol początkowy gramatyki



Analizowane słowo

Rozpoczynamy od symbolu wyróżnionego gramatyki. W każdym kroku najbardziej lewy symbol nieterminalny zastępujemy prawą stroną jakiejś produkcji. Jeśli w wyniku kolejnego kroku otrzymamy symbol terminalny, to sprawdzamy, czy jest on identyczny z odpowiednim symbolem analizowanego słowa. Jeśli nie, to możemy „cofnąć się” i spróbować od nowa, korzystając w ponownym wyprowadzeniu z jakiejś innej produkcji (lub innych produkcji). Byłby to algorytm z powrotami (mało efektywny). Wynik: akceptacja słowa lub odrzucenie w wyniku przebadania wszystkich możliwości.

Problem: „jak uniknąć konieczności powrotów” może być pozytywnie i jednoznacznie rozstrzygnięty dla pewnej klasy gramatyk bezkontekstowych zwanych gramatykami LL(k). „Top-down” odtwarza wyprowadzenie lewostronne.



## Metoda „bottom-up”

Analizowane słowo



Symbol początkowy  
gramatyki

Analizując słowo od lewej strony bierzemy symbol bądź grupę symboli i zastępujemy ją lewą stroną jakiejś odpowiedniej produkcji. Postępujemy tak długo, aż dojdziemy do symbolu wyróżnionego gramatyki, co jest równoznaczne z akceptacją słowa. W każdym kroku powinna być zredukowana osnowa, czyli najbardziej na lewo położona fraza prosta. Także w tej metodzie możliwe jest postępowanie z powrotami.

Problem: „jak znaleźć osnowę nie mając jeszcze drzewa rozbioru syntaktycznego i czym ją zastąpić” może być pozytywnie i jednoznacznie rozstrzygnięty dla pewnych klas gramatyk bezkontekstowych: gramatyk precedencyjnych, czy tzw. gramatyk LR(k). „Bottom-up” odtwarza wprowadzenie prawostronne.