Języki, automaty i obliczenia

Wykład 5: Języki bezkontekstowe

Sławomir Lasota

Uniwersytet Warszawski

3 kwietnia 2024

Plan

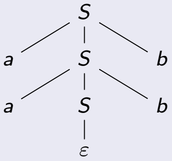
Gramatyki bezkontekstowe

Przykład

$L = \{ a^n b^n \ : \ n \in \mathbb{N} \}$

$S \longrightarrow a\,S\,b \mid \varepsilon$

$S \longrightarrow a\,S\,b \longrightarrow a\,a\,S\,b\,b \longrightarrow a\,a\,b\,b$

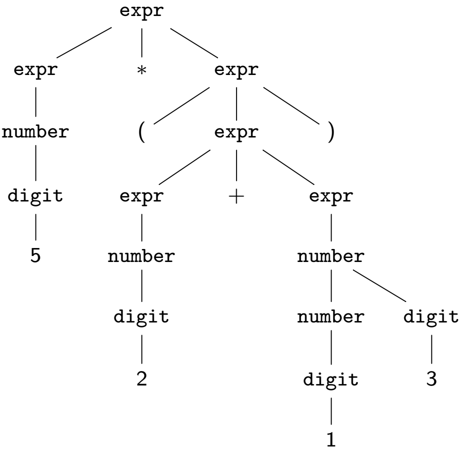


Analiza składniowa

$$\begin{aligned}
\texttt{expr} &\longrightarrow \texttt{expr} + \texttt{expr} \mid \texttt{expr} * \texttt{expr} \mid (\texttt{expr}) \mid \texttt{number} \\
\texttt{number} &\longrightarrow \texttt{digit} \mid \texttt{number}\ \texttt{digit} \\
\texttt{digit} &\longrightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{aligned}$$

Drzewo składniowe wyrażenia

$$5 * (2 + 13)$$



$$
\begin{aligned}
\texttt{expr} &\longrightarrow \texttt{expr} + \texttt{expr} \mid \texttt{expr} * \texttt{expr} \mid (\texttt{expr}) \mid \texttt{number} \\
\texttt{number} &\longrightarrow \texttt{digit} \mid \texttt{number digit} \\
\texttt{digit} &\longrightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
\end{aligned}
$$

- Analizator składniowy - gramatyka bezkontekstowa:

```
expr :
        expr  '+'  expr      { $$ := $1 + $2; }
      | expr  '*'  expr      { $$ := $1 * $2; }
      | '(' expr ')'         { $$ := $2; }
      | number               { $$ := $1; };
```

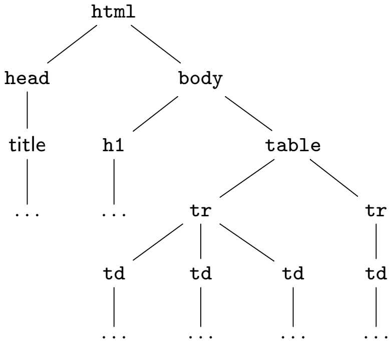
- Analizator leksykalny - wyrażenie regularne:

```
number :
        digit                { $$ := $1; }
      | number  digit        { $$ := 10 * $1 + $2; }
```

- Analizatory składniowe dla języka naturalnego

Drzewa

```
<html>
    <head>
        <title> ...</title>
    </head>
    <body>
        <h1> ...</h1>
        <table>
            <tr>
                <td> ...</td>
                <td> ...</td>
                <td> ...</td>
            </tr>
            <tr>
                <td> ...</td>
            </tr>
        </table>
    </body>
</html>
```



Plan

Gramatyki bezkontekstowe

$$\mathcal{G} \;=\; (A, N, S, \alpha)$$

- A to skończony zbiór symboli końcowych (*terminalnych*)
- N to skończony zbiór symboli niekońcowych (*nieterminalnych*), $A \cap N = \emptyset$
- $S \in N$ to symbol początkowy
- $\alpha \subseteq N \times (A \cup N)^*$ to skończony zbiór reguł przepisywania (*produkcji*)

Notacja

Reguły $(X, v) \in \alpha$ będziemy zapisywać $X \longrightarrow_{\mathcal{G}} v$ albo $X \longrightarrow v$.

Przykład

Gramatyka $\mathcal{G} = (A, N, S, \alpha)$:

- $A \;=\; \{a, b, c\}$
- $N \;=\; \{S, P\}$
- $\alpha \;=\; \{\, (S, S\,S),\; (S, P),\; (P, a\,P\,a),\; (P, b\,P\,b),\; (P, \varepsilon)\, \}$

$$S \longrightarrow S\,S \mid P$$
$$P \longrightarrow a\,P\,a \mid b\,P\,b \mid \varepsilon$$

Reguły rozszerzamy do relacji $\longrightarrow_{\mathcal{G}} \ \subseteq \ (A \cup N)^* \times (A \cup N)^*$:

$$w \longrightarrow_{\mathcal{G}} w' \iff \exists (X \longrightarrow_{\mathcal{G}} v) \in \alpha, \ \exists u, t \in (A \cup N)^*. \ \ w \ = \ u\,X\,t, \ \ w' \ = \ u\,v\,t$$

i domykamy zwrotno-tranzytywnie: $w \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w'$ wtw. gdy istnieje ciąg

$$w \ = \ v_0 \longrightarrow_{\mathcal{G}} v_1 \longrightarrow_{\mathcal{G}} \ldots \longrightarrow_{\mathcal{G}} v_n \ = \ w', \qquad n \geq 0$$

zwany *wyprowadzeniem* (wywodem) w' z w (w gramatyce $\mathcal{G}$).

Język *generowany* przez gramatykę: $L(\mathcal{G}) \ = \ \{\, w \in A^* \ : \ S \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w \,\}$.

Przykład

$$S \longrightarrow S\,S \mid P$$
$$P \longrightarrow a\,P\,a \mid b\,P\,b \mid \varepsilon$$

Słowo $a\,b\,b\,a\,b\,b \in L(\mathcal{G})$, ponieważ ma wyprowadzenie:

$$S \longrightarrow_{\mathcal{G}} S\,S \longrightarrow_{\mathcal{G}} P\,S \longrightarrow_{\mathcal{G}} a\,P\,a\,S \longrightarrow_{\mathcal{G}} a\ b\,P\,b\,a\,S \longrightarrow_{\mathcal{G}} a\ b\,b\,a\,S \longrightarrow_{\mathcal{G}}$$
$$a\,b\,b\,a\,P \longrightarrow_{\mathcal{G}} a\,b\,b\,a\,b\,P\,b \longrightarrow_{\mathcal{G}} a\,b\,b\,a\,b\,b$$

$L(\mathcal{G}) \ = \ ?$

Jednoznaczność

Pytanie

Czy każde słowo ma co najwyżej jedno wyprowadzenie w gramatyce?

Nie, np.

$$S \longrightarrow_{\mathcal{G}} S\,S \longrightarrow_{\mathcal{G}} P\,S \longrightarrow_{\mathcal{G}} a\,P\,a\,S \longrightarrow_{\mathcal{G}} a\;b\,P\,b\,a\,S \longrightarrow_{\mathcal{G}} a\;b\,b\,a\,S \longrightarrow_{\mathcal{G}}$$
$$a\,b\,b\,a\,P \longrightarrow_{\mathcal{G}} a\,b\,b\,a\,b\,P\,b \longrightarrow_{\mathcal{G}} a\,b\,b\,a\,b\,b$$

$$S \longrightarrow_{\mathcal{G}} S\,S \longrightarrow_{\mathcal{G}} S\,P \longrightarrow_{\mathcal{G}} S\,b\,P\,b \longrightarrow_{\mathcal{G}} S\,b\,b \longrightarrow_{\mathcal{G}}$$
$$P\,b\,b \longrightarrow_{\mathcal{G}} a\,P\,a\,b\,b \longrightarrow_{\mathcal{G}} a\;b\,P\,b\,a\,b\,b \longrightarrow_{\mathcal{G}} a\;b\,b\,a\,b\,b$$

Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka spełniająca ten warunek?

$$S \longrightarrow S\,S \mid P \qquad\qquad S \longrightarrow P\,S \mid P \mid \varepsilon$$
$$P \longrightarrow a\,P\,a \mid b\,P\,b \mid \varepsilon \qquad\qquad P \longrightarrow a\,P\,a \mid b\,P\,b$$

?

Przykłady języków bezkontekstowych

- $\{\, a^n\, b^n \;:\; n \in \mathbb{N} \,\}$

$$S \longrightarrow aSb \mid \varepsilon$$

- palindromy nad $\{a, b\}$:

$$P \longrightarrow aPa \mid bPb \mid a \mid b \mid \varepsilon$$

- $\{\, w \in \{a, b\}^* \;:\; \#_w(a) \;=\; \#_w(b) \,\}$

$$X \longrightarrow aXb \mid bXa \mid XX \mid \varepsilon$$

- wyrażenia arytmetyczne:

$$W \longrightarrow W \cdot W \mid W + W \mid (W) \mid 0 \mid 1$$

$$\begin{aligned} S &\longrightarrow S + I \mid I \\ I &\longrightarrow I \cdot C \mid C \\ C &\longrightarrow (S) \mid 0 \mid 1 \end{aligned}$$

- poprawnie zbudowane wyrażenia nawiasowe:

$$W \longrightarrow WW \mid [W] \mid (W) \mid \langle W \rangle \mid \varepsilon$$

Języki regularne są bezkontekstowe

Fakt

Każdy język regularny jest bezkontekstowy.

Dowód:

Niech $\mathcal{A} = (A, Q, \{q_0\}, F, \delta)$. Definiujemy gramatykę $\mathcal{G} = (A, N, S, \alpha)$:

- $N = Q$
- $S = q_0$
- $\alpha = \{\, q \longrightarrow a\, q' \;:\; (q, a, q') \in \delta \,\} \cup \{\, q \longrightarrow \varepsilon \;:\; q \in F \,\}$

$$w \in L(\mathcal{G}) \iff q_0 \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w \iff \exists q \in F.\ \widehat{\delta}(q_0, w, q) \iff w \in L(\mathcal{A})$$

Wyprowadzenia lewostronne

Wyprowadzenie:

$$w \longrightarrow_{\mathcal{G}} w' \iff \exists (X \longrightarrow_{\mathcal{G}} v) \in \alpha,\ u, t \in (A \cup N)^*.\ w = u\,X\,t,\quad w' = u\,v\,t$$

Wyprowadzenie *lewostronne*:

$$w \xrightarrow{L}_{\mathcal{G}} w' \iff \exists (X \longrightarrow_{\mathcal{G}} v) \in \alpha,\ u \in A^*,\ t \in (A \cup N)^*.\ w = u\,X\,t,\quad w' = u\,v\,t$$

Przykład

$$S \xrightarrow{L}_{\mathcal{G}} S\,S \xrightarrow{L}_{\mathcal{G}} P\,S \xrightarrow{L}_{\mathcal{G}} a\,P\,a\,S \xrightarrow{L}_{\mathcal{G}} a\,b\,P\,b\,a\,S \xrightarrow{L}_{\mathcal{G}} a\,b\,b\,a\,S \xrightarrow{L}_{\mathcal{G}}$$
$$a\,b\,b\,a\,P \xrightarrow{L}_{\mathcal{G}} a\,b\,b\,a\,b\,P\,b \xrightarrow{L}_{\mathcal{G}} a\,b\,b\,a\,b\,b$$

Wyprowadzenie *prawostronne*?

Fakt

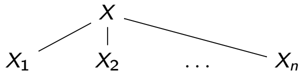
Jeśli $w \longrightarrow\!\!\!\!\!\rightarrow_{\mathcal{G}} w' \in A^*$ *to* $w \xrightarrow{L}\!\!\!\!\!\rightarrow_{\mathcal{G}} w'$.

Plan

Drzewa wyprowadzeń

Drzewo wyprowadzenia w gramatyce $\mathcal{G}$ to drzewo etykietowane $A \cup N \cup \{\varepsilon\}$ t.że

- S jest etykietą korzenia
- dla każdego wierzchołka drzewa o etykiecie X,



albo

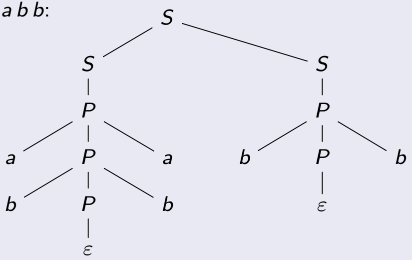
$$X \longrightarrow_{\mathcal{G}} X_1 X_2 \ldots X_n \qquad\qquad X \longrightarrow_{\mathcal{G}} \varepsilon$$

Przykład

Drzewo wyprowadzenia dla $a\,b\,b\,a\,b\,b$:

$S \longrightarrow S\,S \mid P$

$P \longrightarrow a\,P\,a \mid b\,P\,b \mid \varepsilon$



Wyprowadzenia a drzewa wyprowadzeń

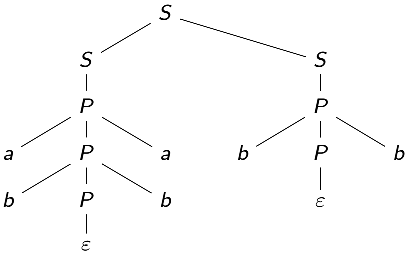
Plon drzewa wyprowadzenia to słowo $w \in (A \cup N)^*$ otrzymane z etykiet liści, w kolejności leksykograficznej (od lewej do prawej), pomijając ε.

Fakt

Niech $w \in A^*$.
$S \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} w \iff w$ *jest plonem drzewa wyprowadzenia w gramatyce* $\mathcal{G}$.

$$S \longrightarrow_{\mathcal{G}} S\,S \longrightarrow_{\mathcal{G}} S\,P \longrightarrow_{\mathcal{G}} S\,b\,P\,b \longrightarrow_{\mathcal{G}} S\,b\,b \longrightarrow_{\mathcal{G}}$$
$$P\,b\,b \longrightarrow_{\mathcal{G}} a\,P\,a\,b\,b \longrightarrow_{\mathcal{G}} a\,b\,P\,b\,a\,b\,b \longrightarrow_{\mathcal{G}} a\,b\,b\,a\,b\,b$$

rysunek



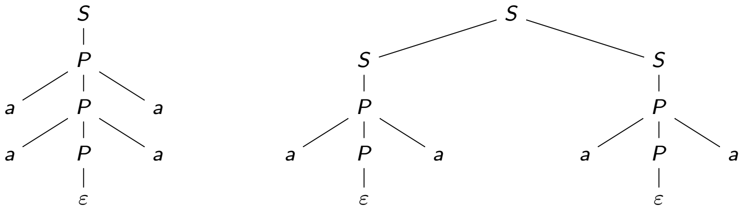
$$S \longrightarrow S\,S \mid P$$
$$P \longrightarrow a\,P\,a \mid b\,P\,b \mid \varepsilon$$

Jednoznaczność

Pytanie

Czy każde słowo ma co najwyżej jedno drzewo wyprowadzenia w gramatyce?

Nie, np.



Pytanie

Czy dla każdej gramatyki istnieje równoważna gramatyka *jednoznaczna*, tzn. taka, w której każde słowo ma co najwyżej jedno drzewo wyprowadzenia?

Nie, oto *język niejednoznaczny*:

$$L = \{ a^n b^n c^m d^m \ : \ n, m \in \mathbb{N} \} \cup \{ a^n b^m c^m d^n \ : \ n, m \in \mathbb{N} \}$$

Język bezkontekstowy jako najmniejszy punkt stały

Przykład

Język palindromów parzystych nad $A = \{a, b\}$, generowany przez gramatykę

$$P \longrightarrow_{\mathcal{G}} a\,P\,a \mid b\,P\,b \mid \varepsilon,$$

jest najmniejszym językiem $L \subseteq A^*$ spełniającym warunek:

$$L \;\supseteq\; \{a\}L\{a\} \;\cup\; \{b\}L\{b\} \;\cup\; \{\varepsilon\}.$$

Czyli L jest najmniejszym punktem stałym funkcji monotonicznej

$$L \;\mapsto\; \{a\}L\{a\} \;\cup\; \{b\}L\{b\} \;\cup\; \{\varepsilon\} \;:\; \mathcal{P}(A^*) \to \mathcal{P}(A^*).$$

Pytanie

Jak tę obserwację uogólnić do gramatyk o więcej niż jednym nieterminalu?

Plan

Postaci normalne

Załóżmy, że $\varepsilon \notin L(\mathcal{G})$.

Postać normalna Chomsky'ego

$$X \longrightarrow a \quad (a \in A) \qquad\qquad X \longrightarrow YZ \quad (Y, Z \in N)$$

Dowód:

$$X \longrightarrow a \quad (a \in A) \qquad\qquad X \longrightarrow w \quad (w \in N^*)$$

Postać normalna Greibach

$$X \longrightarrow a\, w \quad (a \in A,\ \ w \in N^*)$$

Problemy decyzyjne

Pustość

Dane: gramatyka bezkontekstowa $\mathcal{G}$ (np. w postaci Chomsky'ego)

Wynik: czy $L(\mathcal{G}) = \emptyset$?

Uniwersalność

Dane: gramatyka bezkontekstowa $\mathcal{G}$

Wynik: czy $L(\mathcal{G}) = A^*$?

Równość

Dane: gramatyki bezkontekstowe $\mathcal{G}$, $\mathcal{G}'$

Wynik: czy $L(\mathcal{G}) = L(\mathcal{G}')$?

Problem analizy składniowej (przynależności do języka)

Analiza składniowa

Dane: gramatyka bezkontekstowa $\mathcal{G} = (A, N, S, \alpha)$ i słowo $a_1 \dots a_n$

Wynik: czy $a_1 \dots a_n \in L(\mathcal{G})$?

$$V_{i,j} = \{ X \in N : X \longrightarrow\!\!\!\!\rightarrow_{\mathcal{G}} a_i \dots a_j \} \qquad (i \leq j)$$

rysunek

Algorytm

`dla` $i \in \{1 \dots n\}$, $V_{i,i} := \{ X \in N : X \longrightarrow_{\mathcal{G}} a_i \}$

`dla` $i, j \in \{1 \dots n\}$, $i < j$, $V_{i,j} := \emptyset$

`powtarzaj dla` $1 \leq i \leq j < k \leq n$

$V_{i,k} := V_{i,k} \cup \{ X \in N : \exists X \longrightarrow YZ \in \alpha .\ Y \in V_{i,j},\ Z \in V_{j+1,k} \}$

`aż do stabilizacji`

`wynik` $:= (S \in V_{1,n})$

W następnym odcinku:

automaty dla języków bezkontekstowych