

# Derivace a derivační strom, víceznačnost gramatiky

Thursday, May 30, 2013 8:40 AM

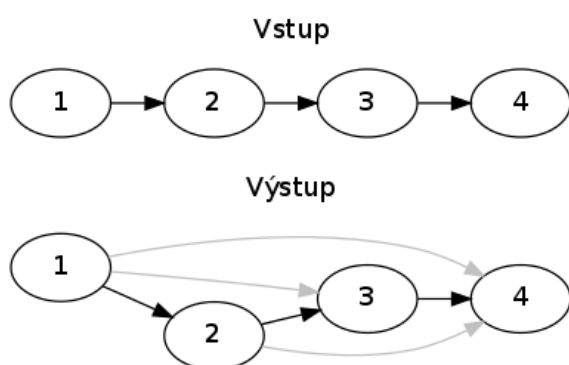
## Derivace

- Posloupnost kroků odvození terminálu pomocí přepisovacích pravidel gramatiky
- Derivační pohled odpovídá konstrukci parsovacího (syntaktického) stromu shora dolů (top-down).
- Parsování zdola nahoru (bottom-up) je spjato s pravými derivacemi.
- Podle toho, který neterminál nahradit v každém kroku derivace, se rozlišují **levá a pravá derivace**.

**DERIVACE** řetězce  $\alpha$  je posloupnost kroků odvození  $\alpha$  pomocí přepisovacích pravidel gramatiky

$$S = \alpha_1 \Rightarrow \alpha_2 \Rightarrow \dots \Rightarrow \alpha_n = \alpha$$

Dtto  $S \Rightarrow^* \alpha$  pozn.:  $\Rightarrow^*$  je uzávěr relace  $\Rightarrow$  (všechny přechody kam se dá transitivně dostat)



**PŘÍMÁ DERIVACE:**  $\alpha A \beta \Rightarrow \alpha \gamma \beta$ , kde  $A \rightarrow \gamma \in P$  (pozn.:  $P$  je množina pravidel, pomocí kterých lze odvodit jazyk.)

## Derivační strom

Derivační strom (parse tree) je orientovaný acyklický graf a je grafickou reprezentací, která říká, v jakém pořadí byla přepisovací pravidla uplatňována na neterminály, tedy jak vznikla věta jazyka.

- Kořen stromu je označen startovacím symbolem gramatiky
- Každý vnitřní uzel je ohodnocen neterminálními symboly.
- Listy jsou ohodnoceny terminálními symboly.
- Listy se čtou zleva doprava a dávají větu (generovanou gramatikou).
- Jestliže uzly  $n_1, n_2, \dots, n_k$  jsou bezprostřední následníci uzlu  $n$ , jsou ohodnoceny symboly  $A_1, A_2, \dots, A_k$  a uzel  $n$  je ohodnocen  $A$ , pak v množině pravidel gramatiky existuje pravidlo  $A \rightarrow A_1 A_2 \dots A_k$ .
- Není třeba značit orientaci hran.

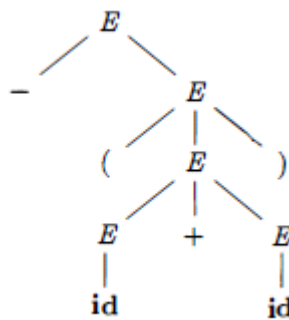
Jiná definice jazyka generovaného gramatikou je množina vět, které mohou být vytvořeny derivačním stromem. Proces hledání derivačního stromu pro danou větu (řetězec terminálů) se nazývá **parsování** tohoto řetězce

Derivační strom ignoruje variace v pořadí, v jakém jsou symboly přepisovány. Proto je mezi derivacemi a derivačními stromy vztah 1:N – např.:

$$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E + E) \Rightarrow -(\mathbf{id} + E) \Rightarrow -(\mathbf{id} + \mathbf{id})$$

$$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E + E) \Rightarrow -(E + \mathbf{id}) \Rightarrow -(\mathbf{id} + \mathbf{id})$$

Jsou různé derivace, jejich derivační strom ale vypadá stejně:



Pro získání jednoznačného derivačního stromu pro derivaci se proto užívá buď pravá a nebo levá derivace.

### Víceznačnost gramatik

Gramatika, která generuje větu, pro níž lze sestavit aspoň dva různé derivační stromy, je víceznačná. Jinak řečeno: je to taková gramatika, která produkuje více než jednu levou nebo víc než jednu pravou derivaci pro tutéž větu.

**Příklad:** Pro gramatiku

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

umožňuje vytvořit dvě levé derivace pro  $id + id * id$  (násobení není upřednostněno před sčítáním):

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

- **Nutnou podmínkou jednoznačnosti gramatiky je, aby pro žádný neterminální symbol neexistovalo jak pravidlo rekurzivní zprava, tak i pravidlo rekurzivní zleva**
- **Problém nejednoznačnosti bezkontextových jazyků je algoritmicky nerozhodnutelný.**

Je potřeba buďto vytvořit jednoznačné gramatiky pro kompilaci aplikací, nebo u nejednoznačných gramatik zavést dodatečná pravidla, která řeší případné nejednoznačnosti.

### Odstranění levé rekurze

- **Levorekurzivní gramatiku nelze použít k analýze shora dolů**

Odstranění pravidla rekurzivního zleva:

- Necht' je dána BKG  $G = (N, T, P, S)$ , ve které,
  - $A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \dots \mid A\alpha_m \mid \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$   
jsou všechna A pravidla v P a žádné z  $\beta$  nezačíná A.

- Pak  $G' = (N \cup \{A'\}, T, P', S)$ , kde  $P'$  obsahuje místo uvedených pravidel pravidla:

$$\circ A \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_n \mid \beta_1 A' \mid \beta_2 A' \mid \dots \mid \beta_n A'$$

$$A' \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_m \mid \alpha_1 A' \mid \alpha_2 A' \mid \dots \mid \alpha_m A'$$

From <<https://d.docs.live.net/e3534876709763a3/Dokumenty/ZCU/Statnice/Statnice.docx>>

