Derivační stromy

V kapitole **2.1 Pojem gramatiky, odvození** jsme se mimojiné zabývali odvozením neboli derivací. Vysvětlili jsme si, že derivace je vlastně relace na množině $(N \cup \Sigma)^*$, která se řídí pravidly v dané gramatice. Značili jsme si ji dvojitou šipkou $\Rightarrow_G$, kde G je zadaná gramatika. Ukažme si to raději na jednoduchém příkladě:

Příklad 1.

Mějme gramatiku $G = (\{S, A, B\}, \{a, b, \varepsilon\}, P, S)$, kde množina $P = \{$

$$\begin{array}{lcl} S & \rightarrow & aABb \mid bbB \\ A & \rightarrow & aa \mid bBb \mid aAa \\ B & \rightarrow & b \mid bB \;\} \end{array}$$

Mějme například řetězec $aaAaBb$. Z toho řetězce můžeme odvodit v gramatice G řetězec $aabBbaBb$, protože v P je pravidlo $A \rightarrow bBb$.

Zdá se vám to nepřehledné? Zkusme si to v obou řetězcích raději naznačit, a to tak, že část, na kterou bude pravidlo použito, označíme hvězdičkou z obou stran:

$$aa{*}\mathbf{A}{*}aBb \Rightarrow_G aa{*}\mathbf{bBb}{*}aBb$$

Všimněte si, že neterminál A v původním řetězci jsme nahradili řetězcem bBb, protože v množině P existuje pravidlo $A \rightarrow bBb$. A to je přesně princip odvození neboli derivace.

Jak to ale vše souvisí s derivačními stromy? Určitě se takto musíte ptát. Odpovím opět otázkami. Proč jsme v řetězci $aaAaBb$ vybrali k odvození zrovna neterminál A? Proč jsme nejdříve nederivovali neterminál B? Má pořadí použitých pravidel nějaký vliv na výsledek – tj. slovo, které je odvozeno z gramatiky G? K odpovědi na tuto otázku nám pomůže znalost derivačních stromů.

Definice – derivační strom

Nechť $G = (N, \Sigma, P, S)$ je bezkontextová gramatika. Strom T nazveme derivačním stromem v G právě tehdy, když platí následující podmínky:

1. každý uzel má návěští (označení) z $(\Sigma \cup N)$,
2. kořen stromu T má návěští S,
3. vnitřní uzel má návěští z N,
4. list má návěští patřící do $N \cup \Sigma \cup \{\varepsilon\}$,
5. má-li vnitřní uzel A následníky $X_1, X_2, X_3, \ldots, X_n$ (v pořadí zleva doprava), pak $A \rightarrow X_1X_2X_3 \ldots X_n \in P$,
6. má-li uzel návěští ε, pak je listem a je jediný následník svého otce.

Pokud zřetězíme všechny listy po řadě zleva doprava, dostaneme slovo $w \in L(G)$, které je výsledkem derivačního stromu T.

Příklad 2.

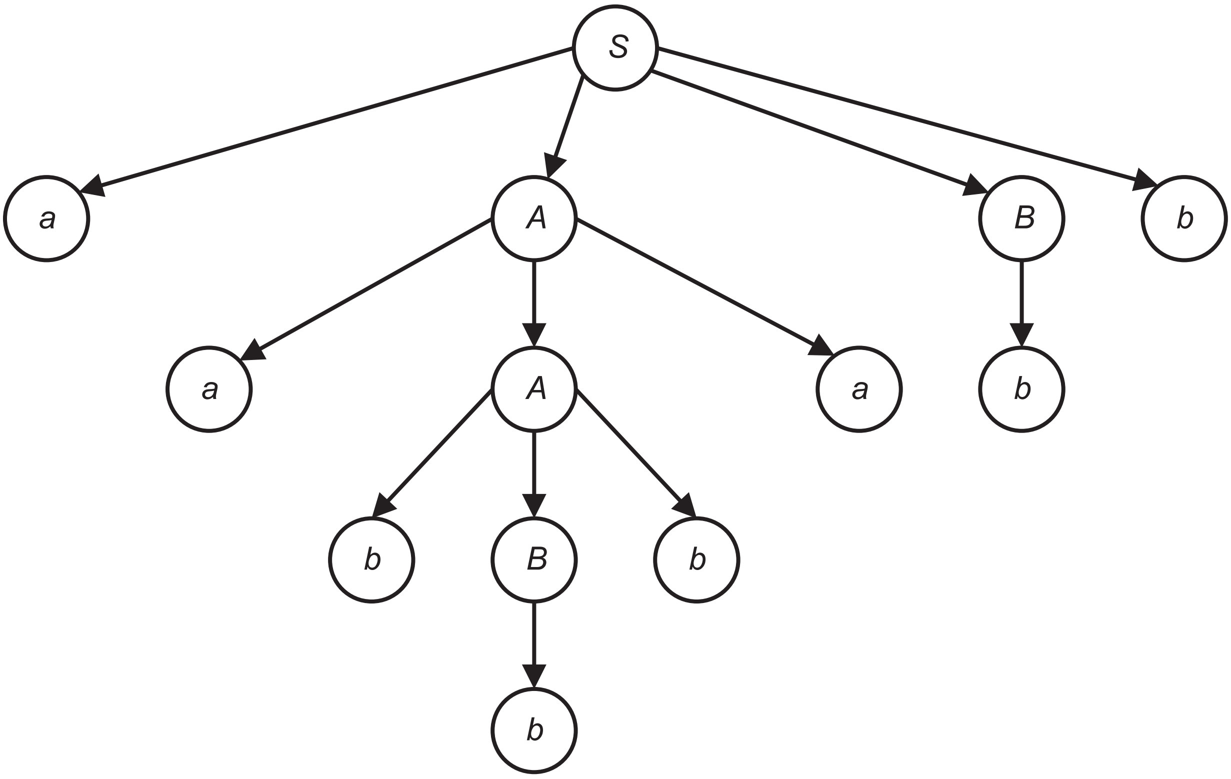
Mějme stejnou gramatiku $G = (\{S, A, B\}, \{a, b, \varepsilon\}, P, S)$ jako v příkladu 1. Množina pravidel $P = \{$

$$\begin{array}{rcl} S & \rightarrow & aABb \mid bbB \\ A & \rightarrow & aa \mid bBb \mid aAa \\ B & \rightarrow & b \mid bB \ \} \end{array}$$

Vezmeme např. slovo $aabbbabb$ jeho odvození v gramatice G:

$$S \Rightarrow aABb \Rightarrow aaAaBb \Rightarrow aabBbaBb \Rightarrow aabbbaBb \Rightarrow aabbbabb$$

Pokud bychom přesně sledovali derivace, které jsme provedli, získáme následující derivační strom:



Poznámka.

Všimněte si důležitého faktu, vedou-li šipky z nějakého vnitřního uzlu X do uzlů $X_1, X_2, ..., X_n$ (v pořadí zleva doprava), pak opravdu pravidlo $X \rightarrow X_1X_2 \ldots X_n$ patří do P. Např.

- ve 2. úrovni stromu jsou po řadě uzly a, A, B, b spojené s kořenem S a skutečně: $S \rightarrow aABb$.
- Nebo ve 3. úrovni jsou uzly a, A, a spojeny šipkami s vnitřním uzlem 2. úrovně A – pravidlo $A \rightarrow aAa \in P$.
- Ještě jeden příklad: ve 4. úrovni jsou uzly b, B, b spojeny šipkami s vnitřním uzlem A 3. úrovně. Je to proto, že jsme použili pravidlo $A \rightarrow bBb$.

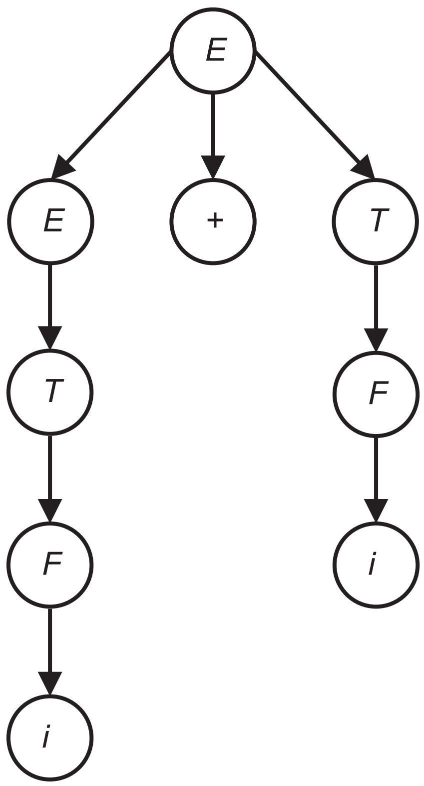
Navíc pro derivační strom opravdu platí, že vnitřní uzly mají vždy v návěští neterminál, kdežto listy naopak vždy terminál.

Příklad 3.

Buď $G = (\{E, T, F\}, \{i, (,), *, +\}, P, E)$ gramatika s následujícími pravidly:

$$\begin{aligned} P = \{E &\rightarrow E + T \mid T \\ T &\rightarrow T * F \mid F \\ F &\rightarrow (E) \mid i\ \} \end{aligned}$$

Úkol: Vypište alespoň tři různé derivace $E \Rightarrow^* i + i$. K práci vám může pomoci derivační strom s výsledkem i+i:



Poté nalezněte odvození pro slova $i + i * i$ a $(i + i) * i$.

Řešení. Existuje 10 různých odvození slova $i + i$, všem přísluší derivační strom uvedený v zadání. Nabízím alespoň tři z nich.

1. $E \Rightarrow E + T \Rightarrow T + T \Rightarrow F + T \Rightarrow i + T \Rightarrow i + F \Rightarrow i + i$
2. $E \Rightarrow E + T \Rightarrow E + F \Rightarrow E + i \Rightarrow T + i \Rightarrow F + i \Rightarrow i + i$
3. $E \Rightarrow E + T \Rightarrow T + T \Rightarrow T + F \Rightarrow F + F \Rightarrow i + F \Rightarrow i + i$

2. úkolem bylo nalézt odvození slov $i + i * i$ a $(i + i) * i$ v gramatice G. Zde je:

1. $E \Rightarrow E + T \Rightarrow T + T \Rightarrow F + T \Rightarrow i + T \Rightarrow i + T * F \Rightarrow i + F * F \Rightarrow i + i * F \Rightarrow i + i * i$
2. $E \Rightarrow T \Rightarrow T * F \Rightarrow T * i \Rightarrow F * i \Rightarrow (E) * i \Rightarrow (E + T) * i \Rightarrow (E + F) * i \Rightarrow (E + i) * i \Rightarrow (T + i) * i \Rightarrow (F + i) * i \Rightarrow (F + i) * i \Rightarrow (i + i) * i$

Poznámka.

Předchozí příklad nám poslouží jako ukázka jednoho důležitého faktu. Ve chvíli, kdy jste měli nalézt alespoň 3 různé derivace slova $i + i$, tak jste svým odvozováním „sledovali“ derivační strom uvedený v zadání. Ten zůstává v našem příkladu pro libovolnou derivaci vždy stejný. Ne vždy tomu tak musí být – o tom si však budeme povídat až příště, můžu však prozradit, že se budeme bavit o jednoznačnosti nebo víceznačnosti bezkontextových gramatik.

Definice – kanonická derivace

Buď $G = (N, \Sigma, P, S)$ bezkontextová gramatika. Levou (respektive pravou) derivací řetězce $\alpha \in (N \cup \Sigma)^*$ na řetězec $\beta \in (N \cup \Sigma)^*$ rozumíme derivaci takovou, že je v každém odvození nahrazen vždy ten nejlevější (nejpravější) neterminál. Souhrně se takovým derivacím říká kanonické derivace.

Poznámka.

Příklad levé i pravé derivace jsme mohli vidět v příkladu 3. Například odvození

- $E \Rightarrow E + T \Rightarrow T + T \Rightarrow F + T \Rightarrow i + T \Rightarrow i + F \Rightarrow i + i$
- $E \Rightarrow E + T \Rightarrow T + T \Rightarrow F + T \Rightarrow i + T \Rightarrow i + T * F \Rightarrow i + F * F \Rightarrow i + i * F \Rightarrow i + i * i$

jsou levé derivace (přepisujeme vždy nejlevější neterminál). Naopak odvození

- $E \Rightarrow E + T \Rightarrow E + F \Rightarrow E + i \Rightarrow T + i \Rightarrow F + i \Rightarrow i + i$
- $E \Rightarrow T \Rightarrow T * F \Rightarrow T * i \Rightarrow F * i \Rightarrow (E) * i \Rightarrow (E + T) * i \Rightarrow (E + F) * i \Rightarrow (E + i) * i \Rightarrow (T + i) * i \Rightarrow (F + i) * i \Rightarrow (F + i) * i \Rightarrow (i + i) * i$

jsou pravé derivace (přepisujeme vždy nejpravější neterminál).