

Programovací jazyk SVGER

October 30, 2016

Contents

1	Gramatiky	3
1.1	Identifikátory	3
1.2	Čísla	3
1.3	Řetězce	3
1.4	Závorky	3
1.5	Gramatika bílých znaků	4
1.6	Komentáře	4
2	Lekální analyzátor	4

1 Gramatiky

V této sekci se nachází popis jednotlivých gramatik, které jsou zpracovány lexikálním analyzátozem.

1.1 Identifikátory

Identifikátor začíná na kterýkoliv znak z množiny $z = \{a..z, A..Z, +, -, /, *, <\}$ a pokračuje v kterémkoliv ze znaků z množiny $m = z \cup \{0..9\}$

Z tohoto popisu nám vyplyne následující gramatika $G_{identifikatory}(\{S, A, B\}, \{z, m\}, P, S)$

$$\begin{aligned} S &\rightarrow zA|z \\ A &\rightarrow mA|m \end{aligned}$$

1.2 Číslo

Validním číslem je jakákoliv sekvence znaků, které se nachází v množině $d = 0..9$. Které jsou

Ná základě tohoto můžeme vytvořit následující gramatiku $G_{cisla}(\{S, B, C\}, \{d, .\}, P, S)$

$$\begin{aligned} S &\rightarrow dB|d \\ B &\rightarrow dB|d|d.C \\ C &\rightarrow dC|d \end{aligned}$$

1.3 Řetězce

Řetězce začínají a končí na znak ". V samotném řetězci se pak může nacházet jakýkoliv znak kterému předchází \ nebo cokoliv v množině ■ , která reprezentuje všechny tisknutelné znaky, které nejsou \ ". Toto lze reprezentovat následující gramatikou $G_{retiezce}(\{S, D, E\}, \{\backslash, ", \blacksquare\}, P, S)$

$$\begin{aligned} S &\rightarrow "D \\ D &\rightarrow "\backslash E| \blacksquare E \\ E &\rightarrow "D\backslash D| \blacksquare D \end{aligned}$$

1.4 Závorky

Závorky jsou důležitou součástí programovacího jazyků, které patří do rodiny lispu. Z důvodu přehlednosti jsem se rozhodl, že programovací jazyk bude podporovat nejen "kulaté "závorky tedy (a) ale i [] a {}. Popisuje je následující gramatika $G_{zavorky}(\{S\}, \{(\, , \, [\, , \, \{ \, , \, \} \}, P, S)$

$$S \rightarrow ([)][]\{\}$$

1.5 Gramatika bílých znaků

Je gramatika popisující všechny bílé znaky $G_{bileznaky}(\{S\}, \{\leftarrow | \mapsto | \sqcup\}, P, S)$

$$S \rightarrow \leftarrow | \mapsto | \sqcup$$

1.6 Komentáře

Jazyk má jednoduché komentáře začínající na ; a končící novým řádkem $G_{komentare}(\{S, F\}, \{\leftarrow, ;, \sqcup\}, P, S)$

$$S \rightarrow ; F$$

$$F \rightarrow \sqcup F | \leftarrow S$$

2 Lekální analyzátor

Nyní je třeba z předem definované gramatiky spojit a převést na automat. Pokud tento automat nebude deterministický je třeba jej determinizovat.

Samotný jazyk se skládá z jazyku oddělovačů (komentáře, bílé znaky) $L_{od} = G_{komentare} \cup G_{bileznaky}$ a jazyku významových tokenů

$$L_{vt} = G_{identifikatory} \cup G_{cisla} \cup G_{retezce} \cup G_{zavorky}$$

Celý jazyk pak lze zapsat tímto způsobem

$$L = (L_{od}^* \cdot L_{vt})^* \cdot L_{od}^*$$

Po spojení nám vzniká gramatika

$G_{vt}(\{S, S_{identifikatory}, S_{cisla}, S_{retezce}, S_{zavorky}, A, B, C, D, E, F, G\}, \{d, z, m, (,), \{, \}, [,], \backslash\}, P, S)$, která reprezentuje významové tokeny

$$S \rightarrow S_{identifikatory} | S_{cisla} | S_{retezce} | S_{zavorky} | S_{od}$$

kde S_{od} je definované v gramatice oddělovačů $G_{od}(\{S_{od}\}, \{; | \leftarrow | \mapsto | \sqcup | \sqcup | \leftarrow\}, P, S)$

$$S \rightarrow \leftarrow | \mapsto | \sqcup | ; A | \sqcup$$

$$A \rightarrow \sqcup A | ; A | \leftarrow$$

Po odstranění jednoduchých symbolů dostaneme následující gramatiku

$$S \rightarrow zA|z|dB|d|''D|;G|\leftarrow|\mapsto|\sqcup|(|)|[|]|{\{}|$$

$$A \rightarrow mA|m$$

$$B \rightarrow dB|d|d.C$$

$$C \rightarrow dC|d$$

$$D \rightarrow "|\backslash E| \blacksquare E$$

$$E \rightarrow "D|\backslash D| \blacksquare D$$

$$F \rightarrow \Box F| \leftarrow S$$

[illegible]