

CS 381 - A5

Martin Mueller

Due: April 3rd

1. Design a CFG G with terminal symbols $\Sigma = \{ (,), [,], \{, \} \}$ that generates exactly the set of strings w , where w satisfies the following conditions:

- w is a string over Σ^* where all/any (pairs of) parentheses, brackets and curly braces are balanced,
- all/any parentheses are contained within brackets (if present), or, in the absence of brackets, all/any parentheses are contained within curly braces (if present),
- and all/any brackets are contained within curly braces (if present).

Note that strings in $L(G)$ need not contain parentheses, brackets and braces. For example, a string in $L(G)$ could contain just braces and parentheses. We will assume that λ is in $L(G)$. Here is an example of string that **should be generated** by G :

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

Notice how, in the previous string, balanced parentheses are contained within balanced brackets, which are contained within balanced braces. Also notice that curly braces are not contained within brackets or parentheses. Likewise, brackets are not contained within parentheses.

Here are a couple of examples of strings that should **not** be generated by G :

$\{ \} ([])$

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

In the second string, there is a pair of parentheses that is not contained within brackets. Recall that brackets are not required to be present in every string generated by G , but if they are, then all parentheses must be contained within brackets. Obviously, any string in which the parentheses, brackets and/or braces are not balanced should not be generated by G .

I haven't given you a formally precise definition of $L(G)$ (and I don't plan to do so), but hopefully you get the idea of exactly what strings are in $L(G)$. Here are more strings that **should be generated** by G :

$[()] []$

$\{ [] \}$

$() (())$

$\{ () \}$

Here's what you need to do:

First, design such a CFG G , as described above.

Then, find an equivalent grammar G' that has the same language as G , but G' is in Chomsky Normal Form. Show how you find such a grammar G' .

Finally, implement and test a modified version of your grammar G' in the Automata-Simulator, where in your modified grammar, curly braces should be replaced with angle brackets $<$ and $>$. The reason for the modification is because the curly brace characters are reserved characters within the Automata-Simulator. Upload your grammar in the file **A5-1.cfg**.

The syntax for implementing a CFG in the Automata-Simulator is pretty straightforward. Here is an example

grammar:

$$S \rightarrow [S] \mid SS \mid ;$$

In your L^AT_EX write-up, give both G and G' . Your grammar G should be accompanied by a brief yet precise high-level explanation of how it works.

This grammar starts at S and has 3 options for its next choice. Whatever it selects, it cannot go "back up the tree" so to speak. A would start near the top of the tree, B would be further down, and so on. This is so that the precedence rules of the language can be obeyed. This grammar also allows for an arbitrarily deep nested and/or chained structure of symbols while still obeying the rules of the language. Finally, the grammar also accepts the empty string.

$$\begin{aligned} S &\rightarrow A \mid B \mid C \\ A &\rightarrow \{BA\}A \mid \{CA\}A \mid \lambda \\ B &\rightarrow [CB]B \mid \lambda \\ C &\rightarrow (C)C \mid \lambda \end{aligned}$$

Chomsky Normal Form Conversion:

New Start Symbol

$$\begin{aligned} S_0 &\rightarrow S \\ S &\rightarrow A \mid B \mid C \\ A &\rightarrow \{BA\}A \mid \{CA\}A \mid \lambda \\ B &\rightarrow [CB]B \mid \lambda \\ C &\rightarrow (C)C \mid \lambda \end{aligned}$$

Eliminate λ -rules

$$\begin{aligned} S_0 &\rightarrow S \mid \lambda \\ S &\rightarrow A \mid B \mid C \\ A &\rightarrow \{BA\}A \mid \{CA\}A \mid \{A\}A \mid \{B\} \mid \{BA\} \mid \{B\}A \mid \{\} \mid \{CA\} \mid \{C\}A \mid \{C\} \\ B &\rightarrow [CB]B \mid [B]B \mid [C]B \mid [C] \mid []B \mid [] \mid [B] \\ C &\rightarrow (C)C \mid (C) \mid ()C \mid () \end{aligned}$$

Eliminate unit rules

$$\begin{aligned} S_0 &\rightarrow \{BA\}A \mid \{CA\}A \mid \{A\}A \mid \{B\} \mid \{BA\} \mid \{B\}A \mid \{\} \mid \{CA\} \mid \{C\}A \mid \{C\} \\ &\mid [CB]B \mid [B]B \mid [C]B \mid [C] \mid []B \mid [] \mid [B] \\ &\mid (C)C \mid (C) \mid ()C \mid () \\ &\mid \lambda \\ S &\rightarrow \{BA\}A \mid \{CA\}A \mid \{A\}A \mid \{B\} \mid \{BA\} \mid \{B\}A \mid \{\} \mid \{CA\} \mid \{C\}A \mid \{C\} \\ &\mid [CB]B \mid [B]B \mid [C]B \mid [C] \mid []B \mid [] \mid [B] \\ &\mid (C)C \mid (C) \mid ()C \mid () \\ A &\rightarrow \{BA\}A \mid \{CA\}A \mid \{A\}A \mid \{B\} \mid \{BA\} \mid \{B\}A \mid \{\} \mid \{CA\} \mid \{C\}A \mid \{C\} \\ B &\rightarrow [CB]B \mid [B]B \mid [C]B \mid [C] \mid []B \mid [] \mid [B] \\ C &\rightarrow (C)C \mid (C) \mid ()C \mid () \end{aligned}$$

Eliminate mixed rules

$$\begin{aligned} S_0 &\rightarrow U_0BAU_1A \mid U_0CAU_1A \mid U_0AU_1A \mid U_0BU_1 \mid U_0BAU_1 \\ &\mid U_0BU_1A \mid U_0U_1 \mid U_0CAU_1 \mid U_0CU_1A \mid U_0CU_1 \mid U_2CBU_3B \\ &\mid U_2BU_3B \mid U_2CU_3B \mid U_2CU_3 \mid U_2U_3B \mid U_2U_3 \mid U_2BU_3 \end{aligned}$$

$| U_4CU_5C \mid U_4CU_5 \mid U_4U_5C \mid U_4U_5$
 $| \lambda$
 $S \rightarrow U_0BAU_1A \mid U_0CAU_1A \mid U_0AU_1A \mid U_0BU_1 \mid U_0BAU_1$
 $| U_0BU_1A \mid U_0U_1 \mid U_0CAU_1 \mid U_0CU_1A \mid U_0CU_1$
 $| U_2CBU_3B \mid U_2BU_3B \mid U_2CU_3B \mid U_2CU_3 \mid U_2U_3B \mid U_2U_3$
 $| U_2BU_3 \mid U_4CU_5C \mid U_4CU_5 \mid U_4U_5C \mid U_4U_5$
 $A \rightarrow U_0BAU_1A \mid U_0CAU_1A \mid U_0AU_1A \mid U_0BU_1 \mid U_0BAU_1$
 $| U_0BU_1A \mid U_0U_1 \mid U_0CAU_1 \mid U_0CU_1A \mid U_0CU_1$
 $B \rightarrow U_2CBU_3B \mid U_2BU_3B \mid U_2CU_3B \mid U_2CU_3$
 $| U_2U_3B \mid U_2U_3 \mid U_2BU_3$
 $C \rightarrow U_4CU_5C \mid U_4CU_5 \mid U_4U_5C \mid U_4U_5$
 $U_0 \rightarrow \{$
 $U_1 \rightarrow \}$
 $U_2 \rightarrow [$
 $U_3 \rightarrow]$
 $U_4 \rightarrow ($
 $U_5 \rightarrow)$

Eliminate rules with more than 2 non-terminals

$S_0 \rightarrow S_1S_3 \mid S_4S_3 \mid U_0S_3 \mid S_1U_1 \mid S_1S_2$
 $| S_1S_5 \mid U_0U_1 \mid S_4S_2 \mid S_4S_5 \mid S_4U_1 \mid S_7S_8$
 $| U_2S_8 \mid S_7S_9 \mid S_7U_3 \mid U_2S_9 \mid U_2U_3 \mid U_2S_6$
 $| S_{11}S_{10} \mid S_{11}U_5 \mid U_4S_{10} \mid U_4U_5$
 $| \lambda$
 $S \rightarrow S_1S_3 \mid S_4S_3 \mid U_0S_3 \mid S_1U_1 \mid S_1S_2$
 $| S_1S_5 \mid U_0U_1 \mid S_4S_2 \mid S_4S_5 \mid S_4U_1$
 $| S_7S_8 \mid U_2S_8 \mid S_7S_9 \mid S_7U_3 \mid U_2S_9 \mid U_2U_3$
 $| U_2S_6 \mid S_{11}S_{10} \mid S_{11}U_5 \mid U_4S_{10} \mid U_4U_5$
 $A \rightarrow S_1S_3 \mid S_4S_3 \mid U_0S_3 \mid S_1U_1 \mid S_1S_2$
 $| S_1S_5 \mid U_0U_1 \mid S_4S_2 \mid S_4S_5 \mid S_4U_1$
 $B \rightarrow S_7S_8 \mid U_2S_8 \mid S_7S_9 \mid S_7U_3$
 $| U_2S_9 \mid U_2U_3 \mid U_2S_6$
 $C \rightarrow S_{11}S_{10} \mid S_{11}U_5 \mid U_4S_{10} \mid U_4U_5$
 $S_1 \rightarrow U_0B$
 $S_2 \rightarrow AU_1$
 $S_3 \rightarrow S_2A$
 $S_4 \rightarrow U_0C$
 $S_5 \rightarrow U_1A$
 $S_6 \rightarrow BU_3$
 $S_7 \rightarrow U_2C$
 $S_8 \rightarrow S_6B$
 $S_9 \rightarrow U_3B$
 $S_{10} \rightarrow U_5C$
 $S_{11} \rightarrow U_4C$
 $U_0 \rightarrow \{$
 $U_1 \rightarrow \}$
 $U_2 \rightarrow [$

$$\mathbb{U}_3 \rightarrow]$$

$$\mathbb{U}_4 \rightarrow ($$

$$\mathbb{U}_5 \rightarrow)$$