CS 4510 Automata and Complexity 2/20/2022

Worksheet 4: Context-Free Grammars

Worksheets are provided for your study purposes only. They are not graded. The answers to some of the questions will be uploaded to Canvas on Saturday, Feb 26th. For the rest of the questions, however, you are encouraged to discuss with others on Piazza and try to solve them collaboratively. You are also welcomed to discuss the problems during office hours. (Note - * denotes trickier problems)

- **Problem 1**

  Give a context-free grammar for the following languages where $\Sigma = \{a, b, c\}$ and i, j, k are non-negative integers:

  a. $L_1 = a^i b^j c^k$ where $i + j >= k$.

  b. $L_2 = a^i b^j c^k$ where $a^i b^j c^k$ is a palindrome.

- **Problem 2**

  Give a context-free grammar for the following languages where $\Sigma = \{0, 1\}$:

  a. $L_3 = \{w \mid w$ contains more 1s than 0s$\}$.

  b. $L_4 =$ All palindromes over Σ.

  c. $L_4 =$ All non-palindromes over Σ.

  d. $L_5 = \{x \in \Sigma \mid$ symbol at position i is same as symbol at position $i + 2$ and $|x| \geq 2\}$.

  e. $L_6 = 0^i 1^j$ where $i \leq j$ or $i > 2j$ (i, j are non-negative integers).

- **Problem 3**

  $L_7 = \{a^i b^j \mid 0 \leq j < i < 2j\}$ (i, j are non-negative integers)

  a. Write a context free grammar for this language L_7.

  b. Is your grammar ambiguous? Why or why not?

  c. Write your grammar in Chomsky Normal Form.

- **Problem 4**

  Give a context-free grammar for the following languages (i, j, k, m are non-negative integers):

a. $L_8 = a^i b^j c^k d^m$ such that $i = k$ or $j = m$.

*b. $L_9 = a^i b^j c^k d^m$ such that $i + k = j + m$.

- **Problem 5**

  Is the given grammar ambiguous? If yes, prove it and construct an unambiguous grammar for the same language.

  $$S \to aS \mid aSbS \mid c$$

- **Problem 6**

  Regular expressions form a context-free language!

  $L_{10} = \{r \mid r$ is a valid regular expression over alphabet $\{0, 1\}$ (all concatenations and spaces are omitted for simplicity) $\}$

  For example: "$(0 \cup 1)^*$" $\in L$, "$0 \cup \varepsilon$" $\in L$, while "(0" $\notin L$, "$*$" $\notin L$

  Create a CFG for L. To reduce confusion, please quote all terminals and use uppercase letters for non-terminals.

- ***Problem 7**

  Simplified HTML

  There is a famous meme for programmers that HTML cannot be parsed with regular expressions (https://stackoverflow.com/a/1732454/). The reason is that HTML is not a regular language. To decide whether HTML in its entirety is context-free is quite complex. However, if we simplify HTML drastically the problem becomes much simpler, In this question, you will show a simplified version of HTML is a context-free language by creating a CFG for it.

  Here is the specification for our language:

  The alphabet only contains 6 letters: "a", "b", "c", "<", ">", "/"

  A tag looks like this: <tag name>children</tag name>, where valid tag names are "aaa", "bbb", "ccc", "abc", and "cba". The opening tag name must match the closing tag name. The children can contain text mixed with tags. A tag can also be self-closing, like so: <tag name/>

  For example, the following are examples of our simplified HTML:

  - aabbcc
  - <aaa>bbabc</aaa>ccc
  - <abc>ccc<bbb></bbb></abc>
  - <aaa/><bbb/>
  - <cba>abc<ccc/>abc</cba>bbb

The following are NOT valid strings:

- <aaa
- aaa>
- aaa/
- <aaa> (Not closed)
- </aaa> (Not opened)
- <aaa></bbb> (Not matching)

Create a CFG for our simplified HTML.

- **Problem 8**

  Give a PDA for the following languages, where $\Sigma = \{0, 1\}$:

  a. $L_{11} = \{x\#y \mid x, y \in \Sigma$, Number of '100' substrings in x = Number of '001' substrings in y$\}$.

  b. $L_{12} = \{x\#y \mid x, y \in \Sigma$, Number of '111' substrings in x = Number of '101' substrings in y$\}$.