CSCI 3130: Formal Languages and Automata Theory — Homework 3

The Chinese University of Hong Kong, Fall 2015 — due 11:59pm Thursday October 22

Collaborating on homework is encouraged, but you must write your own solutions in your own words and list your collaborators. Copying someone else's solution will be considered plagiarism and may result in failing the whole course.

Please answer clearly and concisely. Explain your answers. Unexplained answers will get lower scores or even no credits.

(1) (12 points) For each of these languages, give both a context-free grammar *and* a pushdown automaton. Briefly explain how your CFG and PDA work; answers without sufficient explanation will get no points. Your CFG should be relatively simple and contain at most 6 variables.

(a) $L_1 = \{x\#y \mid y^R \text{ is a substring of } x, \text{ and } x \in \{\mathtt{a},\mathtt{b}\}^*\}$, $\Sigma = \{\mathtt{a},\mathtt{b},\#\}$
Recall that w^R is w written backwards, and u is a substring of v if $v = sut$ for some strings s, t

(b) $L_2 = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k \mid i = j \text{ or } j = k\}$, $\Sigma = \{\mathtt{a},\mathtt{b},\mathtt{c}\}$

(c) $L_3 = \{\mathtt{0}^i\mathtt{1}^j\mathtt{2}^{i+j} \mid i \geqslant 0, j \geqslant 0\}$

(d) $L_4 = \{w \in \{\mathtt{a},\mathtt{b}\}^* \mid w \text{ contains more } \mathtt{a}\text{'s than } \mathtt{b}\text{'s}\}$

(2) (8 points) Consider the following context-free grammar G that describes strings consisting of well-nested parentheses and square brackets:

$$S \to SS \mid \mathtt{(}S\mathtt{)} \mid \mathtt{[}S\mathtt{]} \mid \varepsilon$$

The alphabet of G consists of `(`, `)`, `[`, `]`.

(a) Convert G to Chomsky Normal Form.

(b) Apply the Cocke–Younger–Kasami algorithm to obtain parse trees for the following strings: `()[]()` and `[()]`. Show the table of variables that generate every substring. Also draw the parse trees you get.

(c) Give a CFG G' that describes the same language as G but is not ambiguous. (Your G' needs not be in Chomsky Normal Form.)

(3) (12 points) Consider the following languages. For each of the languages, say whether the language is (1) regular, (2) context-free but not regular, or (3) not context-free. Explain your answer (give a DFA or argue why one exists, give a CFG or PDA, apply appropriate pumping lemma or give pairwise distinguishable strings).

(a) $L_1 = \{x\#y\#x^R\#y^R \mid x, y \in \{\mathtt{a},\mathtt{b}\}^*\}$, $\Sigma = \{\mathtt{a},\mathtt{b},\#\}$

(b) $L_2 = \{w \in \Sigma^* \mid \text{every prefix of } w \text{ contains at most as many } \mathtt{a}\text{'s as } \mathtt{b}\text{'s}\}$, $\Sigma = \{\mathtt{a},\mathtt{b}\}$
Recall that s is a prefix of w if $w = st$ for some string t

(c) $L_3 = \{w \in \{\mathtt{0},\mathtt{1},\mathtt{2}\}^* \mid w \text{ contains the same number of } \mathtt{0}\text{'s}, \mathtt{1}\text{'s and } \mathtt{2}\text{'s}\}$

(d) $L_4 = \{x\#y \mid y \text{ is an anagram of } x \in \{\mathtt{a},\mathtt{b}\}^*\}$, $\Sigma = \{\mathtt{a},\mathtt{b},\#\}$
Recall that y is an anagram of x if x and y both have the same number of occurrences of every symbol, but these symbols may or may not appear in the same order. For example, `silent` is an anagram of `listen`

(4) (8 points) Context-free grammars are sometimes used to model natural languages. In this problem you will model a fragment of the English language using context-free grammars. Consider the following English sentences:

`The girl met the boy.`

`The girl that the teacher knows met the boy.`

`The girl that the teacher that the staff saw knows met the boy.`

`The girl that the teacher that the staff that the reporter interviews saw knows met the boy.`

This is special type of sentences built from a subject (`the girl`), a relative pronoun (`that`) followed by another sentence, a transitive verb (`met`) and an object (`the boy`).

(a) Give a context-free grammar G that models this special type of sentence. Your grammar should generate the sentence in lower case letters and without punctuation (`the girl met the boy`).

(b) Is the language of G regular? If so, write a regular expression for it. If not, prove using the pumping lemma for regular languages or pairwise distinguishable strings.

(c) Can you give an example of a sentence that is in G but does not make sense in common English?