11-711: Algorithms for NLP

Homework Assignment #2: Formal Language Theory

Out: September 18, 2007
Due: October 2, 2007

Problem 1

Use the Pumping Lemma for regular languages to show that the following language over $\Sigma = \{a\}$ is not regular:

$$L = \left\{ w \mid |w| = 2^i, i > 0 \right\}$$

Problem 2

Let L be a language, and let $E = \{uv \mid v \in L, u \in \Sigma^*\}$. In other words, E is the language containing strings that have a suffix in L. Prove that if L is regular, then so is E.

Problem 3

Suppose G is a CFG and we are given a word $w \in L(G)$ such that $|w| = n$. What is the *exact* derivation length of w in G (that is, the number of derivation steps in which w is derived) if:

1. G is in Chomsky Normal Form (CNF)
2. G is in Greibach Normal Form (GNF)

Explain your answers in both cases.

Problem 4

For each of the following statements, answer whether the claim is **true** or **false**, and give a *short* (two- to three-sentence) example and/or explanation.

1. $L = L_1 \cap L_2$. If L_1 and L are regular languages, then L_2 must also be a regular language.
2. $L = L_1 \cup L_2$. If L_1 and L are regular languages, then L_2 must also be a regular language.
3. Let $L_1 \subset L_2$. If L_1 is not regular, then L_2 must also be not regular.

4. $L = \bigcup_{i=1}^{\infty} L_i$. If all of the L_i are regular languages, then L is also a regular language.

5. Let L be the union of *all* the context-free languages over an alphabet Σ. Then L is a regular language.

Problem 5

Show that if L is a CFL that does not accept the empty word ϵ, then there exists a PDA $M = (Q, \Sigma, \Gamma, \delta, q, Z, \emptyset)$ that accepts L by empty stack and has the following two properties:

(1) M has a single state q (i.e. $|Q| = 1$).

(2) If $(q, \gamma) \in \delta(q, a, X)$ for some $a \in \Sigma \cup \{\epsilon\}$ and $X \in \Gamma$, then $|\gamma| \leq 2$. (In other words, at each single move, the machine can replace the top element in the stack by at most two elements.)

Hint: Look at a grammar in CNF that accepts L, and construct a PDA that will simulate a leftmost derivation of the input. You must prove the correctness of your construction!