Problem Set 4

Due: Thursday, March 5, 2009 at 12:30 in class (i.e., SC 1105)
Version: **1.1**

Submission instructions: Submit each problem on a **separate sheet of paper**, put your name on each sheet, and write your discussion section time and day (e.g. Tuesday 10am) in the upper righthand corner. These details may sound picky, but they make the huge pile of homeworks much easier to grade quickly and more importantly, since we return them in the discussion sections, easier for you to get them back.

Also, write on each exercise the name/netid of your group members.

(Q1) Irregular.
[**Category**: Proof., **Points**: 10]

Let $\Sigma = \{\mathtt{1}, \#\}$ and let

$$L = \left\{ w \;\middle|\; w = x_1\#x_2\#...\#x_k \text{ for } k \geq 0, \text{ where } x_i \in \mathtt{1}^* \text{ and } x_i \neq x_j \text{ for } i \neq j \right\}.$$

Provide a direct proof that L is not a regular language.

(Q2) Inverse homomorphism.
[**Category**: Proof., **Points**: 10]

Let $h : \Sigma^* \rightarrow \Gamma^*$ be a homomorphism, and L be a regular language over Γ^*. Show that the inverse homomorphism language $h^{-1}(L)$ is regular. Here

$$h^{-1}(L) = \left\{ w \;\middle|\; h(w) \in L \right\}.$$

For example, for the language $L = \left\{ (\mathtt{aaa})^i \;\middle|\; i \geq 0 \right\}$, and the homomorphism $h(\mathtt{0}) = \mathtt{aa}$ and $h(\mathtt{1}) = \mathtt{aa}$, the inverse homomorphism language is

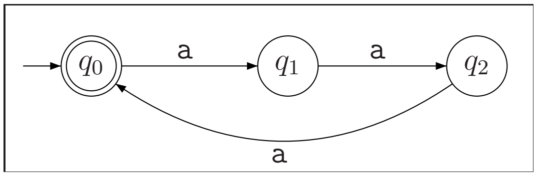
$$L' = \left\{ (\mathtt{0} + \mathtt{1})^i \;\middle|\; i \geq 0 \text{ is a multiply of } 3 \right\},$$

since for any $w \in L'$, we have that $h(w) \in L$. Note, that the inverse homomorphism is not a unique mapping. A word $w \in L$ might have several inverse words $w_1, \ldots, w_j \in L'$, such that $h(w_1) = h(w_2) = \cdots = h(w_j) = w$. For example, for the word $\mathtt{a}^6 \in L$, we have $h^{-1}(\mathtt{a}^6) = \{\mathtt{000}, \mathtt{001}, \mathtt{010}, \mathtt{011}, \mathtt{100}, \mathtt{101}, \mathtt{110}, \mathtt{111}\}$, since, for example, $h(\mathtt{000}) = h(\mathtt{101}) = \mathtt{a}^6$.

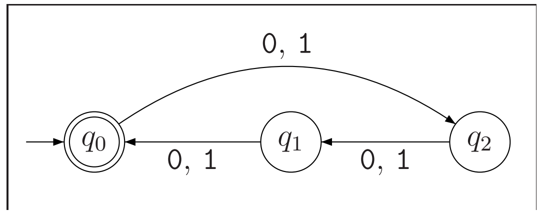
Similarly, it might be that a word $w \in L$ has no inverse in $h^{-1}(L)$. For example, $\mathtt{aaa} \in L$, but has no inverse in L', since $h(w)$ is always an even length word.

To prove that if L is regular then $h^{-1}(L)$ is regular, assume you are given a DFA $\mathtt{M}$ such that $L = L(\mathtt{M})$. Show how to modify this DFA $\mathtt{M} = (Q, \Gamma, \delta, q_0, F)$ into a DFA $\mathtt{D}$ for $h^{-1}(L)$. Describe formally the construction, and prove *formally* that $h^{-1}(L) = L(\mathtt{D})$. (Hint: If $w = w_1 w_2 \ldots w_k \in h^{-1}(L)$ then $h(w) = h(w_1)h(w_2)\ldots h(w_k) \in L$.)

For example, for the languages show above, we have the following two DFAs.



DFA for $L == \left\{ (\texttt{aaa})^i \;\middle|\; i \geq 0 \right\}$



DFA for $L' = h^{-1}(L)$

This implies that regular languages are closed under inverse homomorphism.

(Q3) Irregularity via closure properties.
[**Category**: Proof., **Points**: 10]

Use (*only*) closure properties to show that the following languages are not regular, using a proof by contradiction. You can use languages that you saw in class that are not regular as a starting (or ending) point to your arguments.

(a) $L_1 = \left\{ \texttt{a}^n \texttt{b}^m \texttt{c}^n \;\middle|\; n, m \geq 0 \right\}$

(b) $L_2 = \left\{ \texttt{a}^n \texttt{b}^n \texttt{c}^n \;\middle|\; n \geq 0 \right\}$.

(c) $L_3 = \left\{ \texttt{a}^n \texttt{b}^{3n} \;\middle|\; n \geq 0 \right\}$.
(Hint: Use the result from question 2.)

(d) $L_4 = \left\{ \texttt{a}^n \texttt{b}^n \texttt{a}^n \;\middle|\; n \geq 0 \right\}$.
(Hint: Use the result from question 2 and (b).)

(e) $L_5 = \left\{ w \in (\texttt{a} + \texttt{b})^* \;\middle|\; \#_{\texttt{a}}(w) = \#_{\texttt{b}}(w) \right\}$, where $\#_c(w)$ is the number of times the character c appears in w.

(f) $L_6 = \left\{ w \in (\texttt{a} + \texttt{b})^* \;\middle|\; \#_{\texttt{a}}(w) \neq \#_{\texttt{b}}(w) \right\}$.

(Q4) Palindromes are irregular.
[**Category**: Proof., **Points**: 10]

A ***palindrome*** is a string that if you reverse it, remains the same. Thus `tattarrattat`[1] is a palindrome. Let us consider the empty string ϵ to be a palindrome.

(a) Give a direct proof (without using the pumping lemma) that L_{pal}, the language of all palindromes over the alphabet $\{\texttt{a}, \texttt{b}\}$ is not regular. Your proof should show that any DFA for this language has an infinite number of states.

(b) (Hard?) Prove using only closure properties that L_{pal} is not a regular languages.

(Q5) CFGs are as strong as regular languages.
[**Category**: Proof., **Points**: 10]

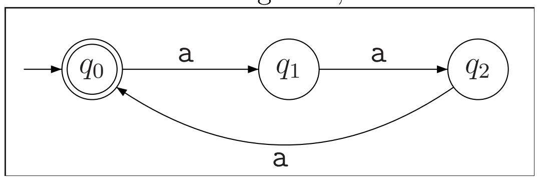
Let L be a regular language recognized by a DFA $\texttt{M} = (Q, \Sigma, \delta, q_0, F)$. We claim that we can construct a CFG for L as follows. We introduce a variable X_i for every state $q_i \in Q$. Given a transition $q_i \xrightarrow{\texttt{c}} q_j$ in $\texttt{M}$, we introduce the rule

$$\mathsf{X}_i \rightarrow c\mathsf{X}_j$$

[1] Which is the longest palindromic word in the Oxford English Dictionary is tattarrattat, coined by James Joyce in Ulysses for a knock on the door.

in the grammar. We also introduce the rule $\mathsf{X}_i \to \epsilon$, if $q_i \in F$. Finally, we set X_0 to be the starting variable for the resulting grammar. Let $G = G(\mathtt{M})$ denote this grammar.

For example, the grammar for the following DFA,



is the following grammar

(G)
$$\begin{array}{ll} \Rightarrow & \mathsf{X}_0 \to \mathsf{a}\mathsf{X}_1 \mid \epsilon \\ & \mathsf{X}_1 \to \mathsf{a}\mathsf{X}_2 \\ & \mathsf{X}_2 \to \mathsf{a}\mathsf{X}_0 \end{array}$$

Prove that in general this construction works. That is, prove formally that for any DFA $\mathtt{M}$, we have that the language of $L_G = L(G(\mathtt{M}))$ is equal to $L(\mathtt{M})$.

(Hint: Do not use induction in your proof. Instead, argue directly about accepting traces and derivations of words.)

(Q6) What * has to do with it, anyway?
[**Category**: Proof., **Points**: 10]

(Extra credit – really hard.)

Let $L \subseteq \left\{ \mathsf{0}^i \,\middle|\, i \geq 0 \right\}$ be an arbitrary language. Prove, that the language L^* is regular.

(Note, that this is quite surprising if L is not regular.)

(Hint: Consider first the case when L contains two words. Then prove it for any finite L. Finally, prove it for the general case. Credit would be given only for a solution for the general case, naturally.)