Discussion 5: More on non-deterministic finite automatas

17 February 2009

Questions on homework 4?

Any questions? Complaints, etc?

1 Non-regular languages

1.1 $L(0^n1^n)$

Lemma 1.1 *The language $L_1 = \left\{ 0^n1^n \;\middle|\; n \geq 0 \right\}$ is not regular.*

Proof: We remind the reader that we already saw that the language $L(a^nb^n) = \left\{ a^nb^n \;\middle|\; n \geq 0 \right\}$ is not regular. As such, assume for the sake of contradiction that L_1 is regular, and consider the homomorphism $h(0) = a$ and $h(1) = b$. Since regular languages are closed under homomorphism, and L_1 is regular, it follows that $h(L_1)$ is regular. However, $h(L_1)$ is the language

$$h(L_1) = \left\{ h(0)^n h(1)^n \;\middle|\; n \geq 0 \right\} = \left\{ a^nb^n \;\middle|\; n \geq 0 \right\} = L(a^nb^n),$$

which is a contradiction, since $L(a^nb^n)$ is not regular. ∎

1.2 $L(\#a + \#b = \#c)$

Let $\Sigma = \{1, 2, 3\}$, consider the following language

$$L_2 = \left\{ x \;\middle|\; \#_a(x) + \#_b(x) = \#_c(x) \right\},$$

where $\#_h(x)$ is the number of times the character h appears in x, for $h = a, b, c$.

1.2.1 Direct proof

Lemma 1.2 *The language $L_2 = \left\{ x \;\middle|\; \#_a(x) + \#_b(x) = \#_c(x) \right\}$, is not regular.*

Proof: Assume for the sake of contradiction, that L_2 is regular, and let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that accepts it. For $i \geq 0$, let $q_i = \delta(q_0, a^i)$ be the state M is in, after reading the string a^i (i.e., a string of length i made out of as). We claim that if $i \neq j$, then $q_i \neq q_j$. Indeed, If there are $i \neq j$ such that $q_i = q_j$, then we have that M accepts the string c^j, if we start from q_j, since M accepts the string $\mathtt{a}^j\mathtt{c}^j$. But then, it must be that M accepts $\mathtt{a}^i\mathtt{c}^j$. Indeed, after M reads $\mathtt{a}^i$ it is in state $q_i = q_j$, and we know that it accepts $\mathtt{c}^j$, if we start from q_j. But this is a contradiction, since $\mathtt{a}^i\mathtt{c}^j \notin L_2$, for $i \neq j$.

This implies that M has an infinite number of states, which is of course impossible. ■

1.2.2 By closure properties

Here is another proof of Lemma 1.2.

Proof: Assume, for the sake of contradiction, that L_2 is regular. Then, since regular languages are closed under intersection, and the language $\mathtt{a}^*\mathtt{c}^*$ is regular, we have that $L_3 = L_2 \cap \mathtt{a}^*\mathtt{c}^*$ is regular. But L_3 is clearly the language

$$L_3 = \left\{ \mathtt{a}^n\mathtt{c}^n \;\middle|\; n \geq 0 \right\},$$

which is not regular. Indeed, if L_3 was regular then $f(L_3)$ would be regular (by closure under homomorphism), which is false by Lemma 1.1, where $f(\cdot)$ is the homomorphism mapping $f(\mathtt{a}) = \mathtt{0}$ and $f(\mathtt{b}) = \epsilon$ and $f(\mathtt{c}) = \mathtt{1}$. ■

1.3 Not too many as please

Lemma 1.3 *The language*

$$L_4 = \left\{ \mathtt{a}^n x \;\middle|\; n \geq 1, x \in \{\mathtt{a},\mathtt{b}\}^*, \text{ and } x \text{ contains at most } 2n \text{ } \mathtt{a}\text{'s} \right\}..$$

is not regular.

We first provide a direct proof.

Proof: Assume for the sake of contradiction that L_4 is regular, and let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting L_4. Let q_i be the state that M arrives to after reading the string $\mathtt{a}^i\mathtt{b}$. Now, the word $\mathtt{a}^i b \mathtt{a}^{2i} \in L_4$, and as such, $\delta(q_i, \mathtt{a}^{2i})$ is an accepting state. Similarly, $\mathtt{a}^i b \mathtt{a}^{2j} \notin L_4$ if $j > i$, since this string has too many as in its second part. As such $\delta(q_i, \mathtt{a}^{2j})$ is not accepting if $j > i$.

We claim that because of that $q_i \neq q_j$ for any $j > i$. Indeed, $\delta(q_i, \mathtt{a}^{2j})$ is not an accepting state, but $\delta(q_j, \mathtt{a}^{2j})$ is an accepting state. Thus $q_i \neq q_j$, but this implies that M has an infinite number of states, a contradiction. ■

1.3.1 Proof using the pumping lemma

Using the pumping lemma we can show that L_4 is not regular.

Proof: By contradiction, if L_4 is regular, then it must satisfy the pumping lemma. Let p be the pumping length guaranteed by pumping lemma. Looking at definition of L_4, the word $\mathtt{a}^p b \mathtt{a}^{2p} \in L_4$. Now we must have this decomposition: $\mathtt{a}^p \mathtt{b} \mathtt{a}^{2p} = xyz$, where $|y| \neq 0$ and $|xy| \leq p$, such that for all $k \geq 0$, $s_k = xy^kz \in L_4$. But it is easy to see $s_0 \notin L$. Indeed, s_0 has a run of $p - |y| < p$ of as in its prefix, but after the b, it has a string of length $2p$ made out of b, which is not in L_4. A contradiction. ■

1.4 A Trick Example (Optional)

Consider the following language.

$$L_7 = \left\{ wxw^R \;\middle|\; w, x \in \{0,1\}^+ \right\}.$$

Is it regular or not? It seems natural to think that it is not regular. However, it is in fact regular. Indeed, L_7 is the set of all strings where the first and last character are the same, which is definitely a regular language.