Assignment #3

Gustavo Estrela de Matos
CSCE 433: Formal Languages and Automata

February 24, 2016

Question 1. Find the smallest string in $\{a, b\}^*$ not in the language corresponding to the given regular expressions.
(a) $b^*(ab)^*a^*$

(b) $(a^* + b^*)(a^* + b^*)(a^* + b^*)$

(c) $a^*(baa^*)^*b^*$

(a)

This regular expression accepts any string with lenght less or equal to 2.

To make a string that is not accepted we can put a b in the end of the string, considering the fact that this string can't be of format $b^*(ab)^*$ because it would be accepted. The string abb, for instance, is not accepted and it is minimal because any string with size less then abb is accepted.

(b)

We can see that this regular expression accepts any string with length less or equal to 3. Also, the look of a string accepted by this language is one that has 3 or less sequences of a's and b's; if we have more than 3 sequences the string won't be accepted. Therefore, the string $abab$ is not accepted and is minimal since any string with length less than 4 is accepted.

(c)

First let's observe that any string with size less or equal to 2 is accepted by this language.

To find a string not accepted by this regular expression we are going to exploit the fact that the term in the middle ($baa*$) repeats the string ba everytime that it is repeated.

Then, if we have a string that starts with bb the first (a^*) and second ($(baa*)^*$) sequences are not in this string, therefore this string should have just b's, then bba is not accepted by the language and it is minimal since every string with size less then 3 is accepted.

Question 2. Find a regular expression corresponding to each of the following subsets of $\{a, b\}^*$.
(a) The language of all strings that begin or end with aa or bb.

(b) The language of all strings with at most two occurrences of the substring ab.

(c) The language of all strings not containing the substring bba.

(a)

Let's split this subset in simpler subsets:

- the subset of strings that start with aa or bb: $(aa + bb)(a + b)^*$
- the subset of strings that end with aa or bb: $(a + b)^*(aa + bb)$

Therefore the subset we want is represented by the regular expression: $(aa + bb)(a + b)^* + (a + b)^*(aa + bb)$

(b)

We can think of this subset as "sequences of a's and b's (any string in $\{a, b\}^*$ is a sequence of a's and b's) that changes from a's to b's at most twice". Then the regular expression should look like $a^*b^*a^*b^*$ and since we could have strings starting with b and ending with a without creating any more ab we have that the regular expression is $b^*a^*b^*a^*b^*a^*$.

(c)

A string that does not contain bba is a string that has a sequence of a's and after a b appears we have three options following this b:

- a sequence of a's until the end of the string
- a sequence of b's until the end of the string
- another string that does not contain bba

Which allow us to see that the regular expression for this set is $(a^*b)^*(a * + b^*)$

Question 3. Use the pumping lemma to prove that the following languages are not regular.
(a) $L = \{a^n b^n c^{2n} \mid n \geq 0\}$

(b) $L = \{w \mid w \in \{0, 1\}^*$ **and** w **is a palindrome**$\}$

(c) $L = \{0^i 1^j \mid j = i \textbf{ or } j = 2i\}$

(a)

Assume that L is a regular language and let p be the pumping constant for L.

Take the string $s = a^p b^p c^{2p}$. Since $s \in L$ we know that, according to the pumping lemma, s can be written as $s = xyz$ where $y \neq \epsilon$, $|xy| \leq p$ and $xy^i z \in L$ for all integer $i \geq 0$.

Since $|xy| \leq p$ we have that $y = a^j$ for some integer $j > 0$ $(y \neq \epsilon)$ such that $s = a^i a^j a^k b^p c^{2p}$ with $i + j + k = p$. Then, using the pumping lemma we know that $s' = xy^0 z$ should be in L, but $xy^0 z = a^{i+k} b^p c^{2p}$ and since $j > 0$ then $i + k \neq p$ and $s' \notin L$ which is a contradiction.

Assuming that L is a regular language led us to a contradiction, therefore L is not regular.

(b)

Assume that L is a regular language and let p be the pumping constant for L.

Take the string $s = 0^p 1^p 1^p 0^p$. Since $s \in L$ we know that, according to the pumping lemma, s can be written as $s = xyz$ where $y \neq \epsilon$, $|xy| \leq p$ and $xy^i z \in L$ for all integer $i \geq 0$.

Since $|xy| \leq p$ we have that $y = 0^j$ for some integer $j > 0$ $(y \neq \epsilon)$ such that $s = 0^i 0^j 0^k 1^p 1^p 0^p$ with $i + j + k = p$. Then, using the pumping lemma we know that $s' = xy^0 z$ should be in L, but $xy^0 z = 0^{i+k} 1^p 1^p 0^p$ and since $j > 0$ then $i + k \neq p$, so s' is not palindrome and $s' \notin L$ which is a contradiction.

Assuming that L is a regular language led us to a contradiction, therefore L is not regular.

(c)

Assume that L is a regular language and let p be the pumping constant for L.

Take the string $s = 0^p 1^{2p}$. Since $s \in L$ we know that, according to the pumping lemma, s can be written as $s = xyz$ where $y \neq \epsilon$, $|xy| \leq p$ and $xy^i z \in L$ for all integer $i \geq 0$.

Since $|xy| \leq p$ we have that $y = 0^j$ for some integer $j > 0$ $(y \neq \epsilon)$ such that $s = 0^i 0^j 0^k 1^{2p}$ with $i + j + k = p$. Then, using the pumping lemma we know that $s' = xy^0 z$ should be in L, but $xy^0 z = 0^{i+k} 1^{2p}$ and since $j > 0$ we know that $i + k < p$ therefore $i + k \neq p$ and $i + k \neq 2p$, then $s' \notin L$ which is a contradiction.

Assuming that L is a regular language led us to a contradiction, therefore L is not regular.