Formal languages, grammars, and automata
Assignment 1, Wednesday, Nov. 12, 2014

Exercises with answers

1. **(5 points)** Consider the following languages over $\Sigma = \{a, b\}$: $L_1 = \mathcal{L}((a + b)^*)$, $L_2 = \mathcal{L}((a^*b^*)^*)$, and $L_3 = \mathcal{L}(((ab^*)^*))$. Prove that precisely two of these languages are equal.

   Answer Info: ....

   L_1 consists of the words $x_1 \dots x_n$ with each $x_i \in \{a, b\}$ and $n \geq 0$. (NB. $n = 0$ gives the empty word, λ.) This means that L_1 consists of all words over $\{a, b\}$.

   L_2 consists of the words $v_1 \dots v_n$ with $n \geq 0$ and each v_i being of the form $a \dots ab \dots b$: a (possibly empty) sequence of a's followed by a (possibly empty) sequence of b's.

   $L_1 \subseteq L_2$, because, given $x_1 \dots x_n \in L_1$, we can choose $v_1 := x_1, \dots, v_n := x_n$ in the description of L_2, so $x_1 \dots x_n \in L_2$.

   $L_2 \subseteq L_1$, because, L_1 consists of all *words over $\{a, b\}$.*

   Conclusion: $L_1 = L_2$.

   L_3 consists of the words $v_1 \dots v_n$ with $n \geq 0$ and each v_i being of the form $ab \dots b$: one a followed by a (possibly empty) sequence of b's.

   Now $L_1 \not\subseteq L_3$, because $b \in L_1$, but $b \notin L_3$. For $b \notin L_3$: note that every $w \in L_3$ that is not λ contains an a.

   End Answer Info ....

2. Consider the languages $L_1 = \mathcal{L}((abba)^*)$, $L_2 = \mathcal{L}(a(bba)^*)$, $L_3 = \mathcal{L}((a(bba)^*)^*)$.

   (a) Show that each of these languages is different.

   (b) For which pairs L_i, L_j (with $i \neq j$) doe we have $L_i \subseteq L_j$? Prove your answer.

   (c) For which of these languages do we have $L_k\, L_k \neq L_k$? Prove your answer.

   Answer Info: ....

   (a) *i. λ and abbaabba are in L_1 and L_3, but not in L_2. So $L_2 \neq L_1$, $L_2 \neq L_3$. (That they are in L_1 and L_3 is straightforward. That $\lambda \notin L_2$ as well. That abbaabba $\notin L_2$ follows from the fact that the words of L_2 in length order are a, abba, abbabba, abbabbabba, ..., so abbaabba $\notin L_2$.)*

   ii. a is in L_2 and L_3 but not in L_1. So $L_1 \neq L_3$.

   (b) $L_2 \subseteq L_3$, because L_3 is the Kleene-$*$ of L_2, and therefore contains it by definition. $L_1 \subseteq L_3$. Proof: The words in L_1 are exactly the words $(abba)^n$ for $n \geq 0$. The words in L_3 are exactly the words

   $$a(bba)^{m_1}a(bba)^{m_1} \dots a(bba)^{m_n}$$

   for $n \geq 0$ and $m_i \geq 0$ for all i $(1 \geq i \geq n)$. So, a word in L_1 is also in L_2, because we can take $m_1 = m_2 = \dots = m_n = 1$ in the pattern for the words of L_3.

   (c) $L_2\, L_2 \neq L_2$, because $aa \in L_2L_2$, but $aa \notin L_2$. For the others, $L_1L_1 = L_1$ and $L_3L_3 = L_3$, because their regular expressions are both of the shape $(-)^*$.

   End Answer Info ....

3. Give a regular expression for the following languages and explain your answer.

(a) **(5 points)**

$$\{w \in \{a,b,c\}^* \mid |w| \geq 3\}.$$

Answer Info: ....
$(a+b+c)(a+b+c)(a+b+c)(a+b+c)^*$. *Explanation: A word in the language is of the shape $x_1x_2x_3w$ with x_1, x_2 and x_3 in $\{a,b,c\}$ and w and arbitrary word over $\{a,b,c\}$. That is x_1, x_2 and x_3 are in $\mathcal{L}(a+b+c)$ and $w \in \mathcal{L}((a+b+c)^*)$.*
Obvious alternatives $(a+b+c)^*(a+b+c)(a+b+c)(a+b+c)$, $(a+b+c)(a+b+c)(a+b+c)(a^*b^*c^*)^*$
End Answer Info ....

(b) **(5 points)**

$$\{w \in \{a,b\}^* \mid w \text{ begins with } b \text{ and } |w|_b \text{ is even}\}.$$

Answer Info: ....
$ba^*ba^*(a^*ba^*ba^*)^*$. *Explanation: A word in the language should start with a b and contain at least 2 b's, so it is of the shape $ba^{n_1}ba^{n_2}w$ with $n_1, n_2 \geq 0$ and w having an even number of b's. A word with an even number of b's is of the shape $a^{m_1}ba^{m_2}ba^{m_3}b\ldots a^{m_{2p}}ba^{m_{2p+1}}$ for $p \geq 0$, $M_i \geq 0$. This is captured by the regular expression $(a^*ba^*ba^*)^*$.*
Obvious alternatives: $ba^*ba^*(ba^*ba^*)^*$, $(ba^*ba^*)^*ba^*ba^*$, $b(a^*ba^*b)^*a^*ba^*$.
End Answer Info ....

(c) **(5 points)**

$$\{w \in \{a,b\}^* \mid bb \text{ doesn't occur in } w\}.$$

Answer Info: ....
$a^*(baa^*)^*(b+\lambda)$. *Explanation: If w is a word in the language, then every b in w should be directly followed by an a* except for a possible last b!. *So w starts with an arbitrary number of a's, and then, possibly a sub-word ba, followed again by an arbitrary number of a's, and so forth, ending with a possible single b In conclusion, w is of the shape $a^pbaa^{m_1}\ldots baa^{m_n}x$ with $x \in \{b,\lambda\}$, $p \geq 0, n \geq 0$ and each $m_i \geq 0$.*
End Answer Info ....

4. [These exercise are hard now. They show that the subject is non-trivial. Later we will learn methods to solve this more easily]

   (a) Show that the language

$$\{w \in \{a,b\}^* \mid aa \text{ occurs exactly twice in } w\}.$$

is regular.
[Hint. Beware of the string *aaa*!]

Answer Info: ....
A possible answer is:

$$(b^*(abb^*)^*aaab^*(abb^*)^*(a+\lambda)) + (b^*(abb^*)^*aabb^*(abb^*)^*aab^*(abb^*)^*(a+\lambda)).$$

This solution borrows from 3(c) the regular expression that does not *contain aa, either at the end (expression $b^*(abb^*)^*(a+\lambda)$) or at the beginning (expression $b^*(abb^*)^*$). The expression $(b^*(abb^*)^*aaab^*(abb^*)^*(a+\lambda))$ covers the words that contain aaa, the expression $(b^*(abb^*)^*aabb^*(abb^*)^*aab^*(abb^*)^*(a+\lambda))$ covers the words that contain two separate sub-words aa.*
Another possible solution (from Twan): $b^*(abb^*)^*aa(a+bb^*(abb^*)aa)(bb^*a)^*b^*$
End Answer Info ....

(b) Show that the language

$$\{w \in \{a,b\}^* \mid |w|_a \text{ and } |w|_b \text{ are even}\}$$

is regular.

Answer Info: . . . .
A possible answer is:

$$(b(aa)^*b + (a + b(aa)^*ab)(bb + ba(aa)^*ab)^*(a + ba(aa)^*b))^*.$$

How to find this answer will be explained in Lecture 3.
End Answer Info . . . .