Languages and Automata

Assignment 1, Tue 4th Feb, 2020

Handing in your answers: There are two options:

1. Brightspace. Before submitting, make sure:
   - the file is a PDF document
   - your name and student number are included in the document (they might be printed).
2. Post box, located in the Mercator building on the ground floor. There will be boxes labelled with *LnA* and the corresponding group teacher's name. Put your work in the post box corresponding to your group. Before putting your solutions in the post box make sure:
   - your name and student number are written clearly on the document.

   There will be 1 box, the *Uitleverbak*, for work that hasn't been picked up at the exercise hours.

Deadline: Fri 7th Feb, 2020, 17:00 (in Nijmegen!). This deadline is strict: submission in brightspace will close at that time.

Goals: After completing these exercises successfully you should be able to carry out definitions and proofs by induction on words, you should be able to write a regular expression for a simple regular language and be able to grasp what language a regular expression denotes.

There are 3 mandatory exercises, worth **10 points** in total. There is 1 more, extra hard, 'fun exercise'. Be aware that this exercise is just for fun, you cannot earn any points with it.

1 Removing Letters

Let $A = \{a, b, c\}$.

a) Define, by structural induction, a function **(2pt)**

$$f \colon A^* \to A^*$$

that removes all occurrences of the letter a. For instance, we should have $f(abcbab) = bcbb$ and $f(bc) = bc$.

b) Show by induction that for all $w \in A^*$, we have: $f(f(w)) = f(w)$. **(2pt)**

2 Regular Expressions (I)

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

a) Give a word that is in all three languages, and a word that is in none of them. **(1pt)**

b) Show that each of the languages L_1, L_2 and L_3 is different. **(1.5pt)**

c) Consider the language L, given by

$$L = \mathcal{L}((b^*a)^* + b^*)$$

Is the language L equal to $\{a, b\}^*$? Justify your answer. **(1pt)**

3 Regular Expressions (II)

a) Let L be the language given by

$$L = \{ w \in \{a, b\}^* \mid |w|_b \text{ is even and } w \text{ does not end with a } b \}$$

Give a regular expression for the language L and explain your answer. **(1pt)**

b) Here is a very practical application of what we've learned so far! Consider a building with three floors: a ground floor, a basement (below it) and a first floor (above it). We would like to install a lift in this building, but this, of course, requires a careful description of proper lift behaviour.

We use words to describe movements of the lift. If the lift goes up one floor, we denote this by U, and if it goes down one floor, we denote this by D. We can then describe a path of the lift through the building as a word in $\{U, D\}^*$. We assume that the lift starts on the ground floor. For instance, $UDDU$ means the lift first goes up to the first floor (the first U), then down entirely to the basement (DD, that's two floors) and then up again to the ground floor (the last U). The empty word λ simply means that the lift does nothing. A word such as UU is wrong: it would mean that the lift goes up two floors starting from the ground floor, meaning it would fly out of the building. Similarly, DD is wrong as we can not go below the basement.

Let $L \subseteq \{U, D\}^*$ be the language of all words that are 'correct', meaning that the corresponding path of the lift keeps it within the building. Give a regular expression for L and explain your answer. **(1.5pt)**

4 Fun Exercises – Regular Expressions

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!]

b) Show that the language

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

is regular.

c) We say, that two regular expressions are equal, if they generate the same language. Symbolically: $e_1 = e_2$ iff $\mathcal{L}(e_1) = \mathcal{L}(e_2)$. There is also an order on regular expressions, given by

$$e_1 \leq e_2 \Leftrightarrow e_1 + e_2 = e_2$$

using the equality of generated languages.

Show that $e_1 \leq e_2$ iff $\mathcal{L}(e_1) \subseteq \mathcal{L}(e_2)$.

d) Show that for any regular expression e the inequality $1 + ee^* \leq e^*$ holds.