Homework 3: Non-regular languages

CSE 30151 Spring 2020

Due Friday, 2020/02/21 at 5:00pm

Instructions

- Create a PDF file (or files) containing your solutions. You can write your solutions by hand, but please scan them in the library or using a smartphone to get them into PDF.
- Please name your PDF file(s) as follows to ensure that the graders give you credit for all of your work:
  - If you're making a complete submission, name it `*netid*-hw3.pdf`, where *`netid`* is replaced with your NetID.
  - If you're submitting some problems now and want to submit other problems later, name it `*netid*-hw3-123.pdf`, where `123` is replaced with the problem numbers you are submitting at this time.
- Submit your PDF file(s) in Sakai. Don't forget to click the Submit button!

Problems (10 points each)

1. **Regular expressions vs. Unix regular expressions.** Regular expressions and Unix regular expressions have some superficial differences, but also some deeper ones that affect the class of languages recognized.

   (a) Unix regular expressions do not have $\emptyset$, because it's not really needed. Prove that if L is a nonempty regular language, it can be described by a regular expression without $\emptyset$. Hint: First use induction to prove the related statement: If α is a regular expression, there is regular expression α' such that $\mathcal{L}(\alpha) = \mathcal{L}(\alpha')$ and either $\alpha' = \emptyset$ or α' does not use $\emptyset$.

   (b) Unix regular expressions have *backreferences*.[1] Give an example of a Unix regular expression that uses backreferences to describe a nonregular language, and prove that this language is not regular. We want you to get

[1] `http://www.regular-expressions.info/backref.html`

practice writing a non-regularity proof, so although you may use Examples 1.73–77, do not simply cite one of them; please write out a full proof.

2. **Binary addition.** This problem is about two ways of representating addition of binary natural numbers. We consider 0 to be a natural number, we allow binary representations of natural numbers to have leading 0s, and we consider ε to be a binary representation of 0. When adding numbers, we do not allow overflow, so, for example, $1111 + 0001 = 0000$ is false.

   (a) [Problem 1.32] Let

   $$\Sigma_3 = \left\{ \begin{bmatrix} \mathtt{0} \\ \mathtt{0} \\ \mathtt{0} \end{bmatrix}, \begin{bmatrix} \mathtt{0} \\ \mathtt{0} \\ \mathtt{1} \end{bmatrix}, \begin{bmatrix} \mathtt{0} \\ \mathtt{1} \\ \mathtt{0} \end{bmatrix}, \begin{bmatrix} \mathtt{0} \\ \mathtt{1} \\ \mathtt{1} \end{bmatrix}, \begin{bmatrix} \mathtt{1} \\ \mathtt{0} \\ \mathtt{0} \end{bmatrix}, \begin{bmatrix} \mathtt{1} \\ \mathtt{0} \\ \mathtt{1} \end{bmatrix}, \begin{bmatrix} \mathtt{1} \\ \mathtt{1} \\ \mathtt{0} \end{bmatrix}, \begin{bmatrix} \mathtt{1} \\ \mathtt{1} \\ \mathtt{1} \end{bmatrix} \right\},$$

   that is, an alphabet of eight symbols, each of which is a 3-tuple of binary digits. Thus, a string over Σ_3 gives three rows of binary digits. Show that the following is regular:

   $$B = \{w \in \Sigma_3^* \mid \text{the bottom row of } w \text{ is the sum of the top two rows}\}.$$

   Hint: Since it's easier to think about addition from right to left, design an automaton for B^R first, then convert it into an automaton for B.

   (b) [Problem 1.53] Let $\Sigma = \{\mathtt{0}, \mathtt{1}, +, =\}$, and prove that the following is not regular:

   $$ADD = \{x \mathtt{=} y \mathtt{+} z \mid x, y, z \in \{\mathtt{0}, \mathtt{1}\}^* \text{ and } x = y + z \text{ is true}\}.$$

3. **Two similar but different languages** [Problem 1.49].

   (a) Let $B = \{1^k w \mid w \in \{0,1\}^* \text{ and } w \text{ contains at least } k \text{ } \mathtt{1}\text{s, for } k \geq 1\}$. Show that B is a regular language. Hint: Try out some strings to see what does and doesn't belong to B, in order to find another simpler way of thinking about B.

   (b) Let $C = \{1^k w \mid w \in \{0,1\}^* \text{ and } w \text{ contains at most } k \text{ } \mathtt{1}\text{s, for } k \geq 1\}$. Prove that C is not a regular language.