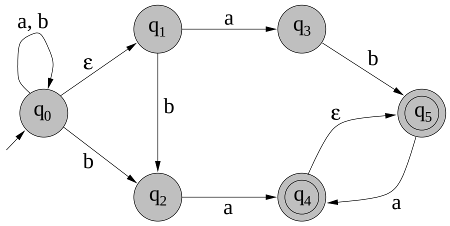
15-453 Formal Languages, Automata, and Computation

Assignment 3

Due: Friday, February 11

Problem 1 (15 points)

Using the method described in class, construct a regular expression for the language recognized by the finite automaton below. Present each step of the process. You may simplify the regular expressions using simple equivalences and omit empty expressions.



Problem 2 (20 points)

For each of the following languages, determine whether it is regular or not. Be sure to prove your answer.

1. $\{w \mid w \text{ is the unary notation for a number that is a multiple of } 7\}$
2. $\{w \mid w \text{ is the decimal notation for a number that is a multiple of } 7\}$
3. $\{xyz \mid x, y, z \in \{0,1\}^*, \text{ and } x + y = z\}$

Problem 3 (20 points)

Perl has the following constructs to denote what the Perl manual calls "regular expressions". For each of these, indicate how you would translate it to the primitives given in class for regular expressions, and if this is not possible, prove this fact. For the purpose of the exercise, assume that Perl's alphabet is $\{0, 1, 2, a, b, c, d, e\}$.

<table>
<tr><th>Expression</th><th>Stands for</th><th>Example match</th></tr>
<tr><td><code>.</code></td><td>any character</td><td><code>.</code> ⟶ <code>a</code></td></tr>
<tr><td><code>[a-c]</code></td><td>any character in this character range</td><td><code>[a-c]</code> ⟶ <code>b</code></td></tr>
<tr><td><code>[^a-c]</code></td><td>any character NOT in this range</td><td><code>[^a-c]</code> ⟶ <code>e</code></td></tr>
<tr><td><code>\d</code></td><td>a digit</td><td><code>\d</code> ⟶ <code>1</code></td></tr>
<tr><td><code>R+</code></td><td>R repeated at least 1 time</td><td><code>a+</code> ⟶ <code>aaaa</code></td></tr>
<tr><td><code>R{3}</code></td><td>expression R repeated 3 times</td><td><code>a{3}</code> ⟶ <code>aaa</code></td></tr>
<tr><td><code>R{3,5}</code></td><td>expression R repeated between 3 and 5 times</td><td><code>a{3,5}</code> ⟶ <code>aaaa</code></td></tr>
<tr><td><code>R{4,}</code></td><td>R repeated at least 4 times</td><td><code>a{3,}</code> ⟶ <code>aaaaaa</code></td></tr>
<tr><td><code>\2</code></td><td>the string that was matched by the second expression in parentheses</td><td><code>(\d+).\1</code> ⟶ <code>12c12</code></td></tr>
<tr><td><code>R1|R2</code></td><td>alternation (R1 or R2)</td><td><code>a|b</code> ⟶ <code>b</code></td></tr>
<tr><td><code>R*</code></td><td>repeat any number of times</td><td><code>a*</code> ⟶ empty string</td></tr>
</table>

Problem 4 (25 points)

Let $M = (Q, \Sigma, \delta, q_0, F)$ be a deterministic finite automaton, and let w be any string in $L(M)$ of length at least $|Q|$. Show that there are strings x, y, and z such that $w = xyz$, $|y| \geqslant (|w| - |Q| + 1)/|Q|$, and $xy^n z \in L(M)$ for each $n \geqslant 0$.

Problem 5 (20 points)

Construct a context-free grammar that generates the language of all strings in $\{a, b\}^*$ that have equal number of occurrences of a and b. Prove that your answer is correct.