Talen en Automaten

Retake Exam, Wed 1st Mar, 2017
18:00 – 21:00

This test consists of **5** problems over **6 pages**. Explain your approach, and **write your answers to the exercises on a separate folio (double pages) as indicated.** You can score a maximum of 100 points, and each question indicates how many points it is worth. The test is closed book. You are NOT allowed to use a calculator, a computer or a mobile phone. You may answer in Dutch or in English. Please write clearly, and do not forget to put on each page: your name and your student number.

Notation Throughout the test, we denote for any alphabet A, $w \in A^*$ and $a \in A$ by $|w|_a$ the number of a's in w, as it was introduced in the lecture.

Write your answers to Problems 1 and 2 on a separate folio (double page)

Problem 1.

Let $A = \{a, b, c\}$ be our finite alphabet and define $f : A^* \to A^*$ inductively as follows

$$f(\lambda) = \lambda \qquad f(a\,v) = a\,b\,f(v) \qquad f(b\,v) = b\,c\,f(v) \qquad f(c\,v) = c\,a\,f(v)$$

a) Give a word w for which $f(w) = abcabc$. **(3pt)**

Solution: ....

acb: $f(acb) = a\,b\,f(cb) = a\,b\,c\,a\,f(b) = a\,b\,c\,a\,b\,c$ □

b) Prove by induction the following property (for all $w \in A^*$) **(8pt)**

$$|f(w)|_a + |f(w)|_b = 2|w|_a + |w|_b + |w|_c.$$

Solution: ....

We show by induction on w that $|f(w)|_a + |f(w)|_b = 2|w|_a + |w|_b + |w|_c$.

- Induction Basis: $w = \lambda$. Here, we get $|f(\lambda)|_a + |f(\lambda)|_b = |\lambda|_a + |\lambda|_b = 0 + 0 = 2|\lambda|_a + |\lambda|_b + |\lambda|_c$.
- Induction Step: $w = xv$. For $v \in A^*$, assuming $|f(v)|_a + |f(v)|_b = 2|v|_a + |v|_b + |v|_c$ (the Induction Hypothesis IH), we need to prove $|f(xv)|_a + |f(xv)|_b = 2|xv|_a + |xv|_b + |xv|_c$ for all $x \in A$.

  Let $x \in A$. We need to distinguish three cases.

  1. If $x = a$, then we have
  $$|f(av)|_a + |f(av)|_b = |a\,b\,f(v)|_a + |a\,b\,f(v)|_b = 1 + |f(v)|_a + 1 + |f(v)|_b$$
  $$\stackrel{\text{IH}}{=} 2 + 2|v|_a + |v|_b + |v|_c = 2|a\,v|_a + |a\,v|_b + |a\,v|_c$$
  as required.
  2. If $x = b$, then
  $$|f(bv)|_a + |f(bv)|_b = |b\,c\,f(v)|_a + |b\,c\,f(v)|_b = |f(v)|_a + 1 + |f(v)|_b$$
  $$\stackrel{\text{IH}}{=} 2|v|_a + 1 + |v|_b + |v|_c = 2|b\,v|_a + |b\,v|_b + |b\,v|_c$$
  3. If $x = c$, then
  $$|f(cv)|_a + |f(cv)|_b = |c\,a\,f(v)|_a + |c\,a\,f(v)|_b = 1 + |f(v)|_a + |f(v)|_b$$
  $$\stackrel{\text{IH}}{=} 2|v|_a + |v|_b + 1 + |v|_c = 2|c\,v|_a + |c\,v|_b + |c\,v|_c$$

So in all cases $|f(xv)|_a + |f(xv)|_b = 2|xv|_a + |xv|_b + |xv|_c$.

This induction proves the desired equation. □

Problem 2.

Consider the following regular grammar for well-formed natural number expressions. The alphabet is $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and N is the start symbol. (Note that we avoid leading zeroes, so 34 and 304 are well-formed, but 034 is not.)

$$\begin{array}{lcl} N & \to & 0X \mid 1A \mid 2A \mid ... \mid 9A \\ A & \to & 0A \mid 1A \mid 2A \mid ... \mid 9A \mid \lambda \\ X & \to & \lambda \end{array}$$

a) Give a context free grammar for the language L of *well-formed arithmetic expressions*, that is expression involving natural numbers (see above), the binary operation $\oplus$ and brackets. **(10pt)**
So the alphabet is $\{0, 1, 2, 3, 4, 5, 6, 7, 8, \oplus, (,)\}$ and you need to describe the expressions where the "brackets match" and $\oplus$ is a binary operation. For example, $33 \oplus 33$, $((33) \oplus 34)$, $33 \oplus 34 \oplus 34$ and $((33))$ are well-formed, but $33 \oplus 4)$, $33) \oplus 4$, $33(\oplus 44)$ and $\oplus \oplus 33$ are not.

Solution: ....

$$\begin{array}{lcl} S & \to & (S) \mid S \oplus S \mid N \\ N & \to & 0 \mid 1A \mid 2A \mid ... \mid 9A \\ A & \to & 0A \mid 1A \mid 2A \mid ... \mid 9A \mid \lambda \end{array}$$

□

b) Prove that L is not regular. **(10pt)**

Solution: ....

Suppose L is regular. Then there is a pumping number, say k. Consider the word $w = (^k\, 0\,)^k \in L$ (k opening brackets, 0 and then k closing brackets). According to the Pumping Lemma, there are x, y, z such that $|xy| \le k$, $|y| \ge 1$ and $x\, y^i\, z \in L$ for each $i \ge 0$. Then y consists of a non-zero sequence of left-brackets: $y = (^p$ for some $p \ge 0$. Taking $i = 2$ we get that $(^{k+p}\, 0\,)^k \in L$, which is a contradiction. So L is not regular. □

c) Give a regular grammar for the language L' of well-formed arithmetic expressions over natural numbers (see above) without brackets. **(8pt)**
So the alphabet is $\{0, 1, 2, 3, 4, 5, 6, 7, 8, \oplus\}$. For example, $33 \oplus 33$ and $33 \oplus 34 \oplus 34$ are well-formed, but $33\oplus$ and $\oplus \oplus 33$ are not.

Solution: ....

$$\begin{array}{lcl} S & \to & 0P \mid 1P \mid \ldots \mid 9P \mid 1A \mid 2A \mid \ldots \mid 9A \\ A & \to & 0A \mid 1A \mid \ldots \mid 9A \mid 0P \mid 1P \mid \ldots \mid 9P \\ P & \to & \oplus S \mid \lambda \end{array}$$

Or

S	$\to$	$0X \mid 0P \mid 1P \mid \ldots \mid 9P \mid 1A \mid 2A \mid \ldots \mid 9A$
X	$\to$	λ
A	$\to$	$0A \mid 1A \mid \ldots \mid 9A \mid 0P \mid 1P \mid \ldots \mid 9P \mid \lambda$
P	$\to$	$\oplus S$

Or

S	$\to$	$0P \mid 1A \mid 2A \mid \ldots \mid 9A$
A	$\to$	$0A \mid 1A \mid \ldots \mid 9A \mid \lambda \mid \oplus S$
P	$\to$	$\oplus S \mid \lambda$

□

d) Give a regular expression for the language L' (of the previous item). **(8pt)**

Solution: ....

First give a regular expression for natural numbers: $e_N := 0 + (1+2+3+4+5+6+7+8+9)(0+1+2+3+4+5+6+7+8+9)^*$. Then the regular expression for L' is $e_N(\oplus e_N)^*$. □

Write your answers to Problems 3, 4, 5, 6 on a separate folio (double page)

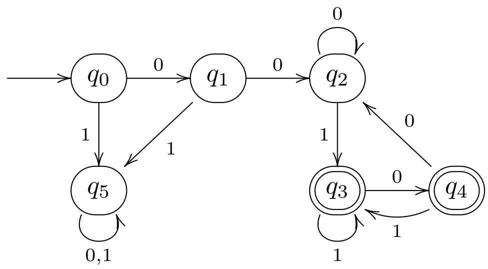
Problem 3.

Consider the language L over the alphabet $A = \{0, 1\}$, defined by **(8pt)**

$$L = \{w \mid w \text{ starts with } 00 \text{ and does } \textit{not} \text{ end with } 00\}$$

Give a DFA that accepts L. Explain your answer.

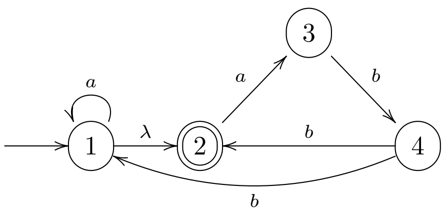
Solution: ....



Explanation: we use q_0, q_1, q_2 to read the first two 0's, and move to the trap state q_5 if we encounter a 1. If we arrive in q_2, q_3 or q_4 then the first two symbols are 00. In q_2 the last two symbols are 00, in q_3 the last symbol is a 1, and in q_4 the last two symbols are 10. □

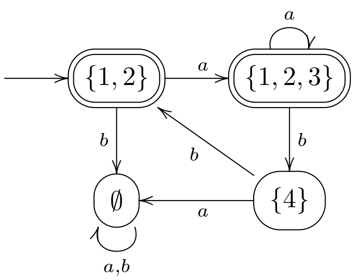
Problem 4.

Let $A = \{a, b\}$ and consider the following NFA-λ M over A. **(10pt)**



Use the powerset construction to give a DFA D with $\mathcal{L}(D) = \mathcal{L}(M)$. Indicate clearly from which states in M a state in D originates.

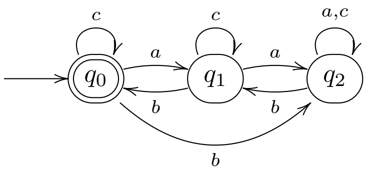
Solution: ....



□

Problem 5.

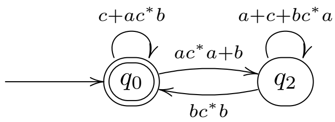
Let $A = \{a, b, c\}$ and the DFA M over A given by **(10pt)**



Use the procedure from the lecture to construct a regular expression e such that $\mathcal{L}(M) = \mathcal{L}(e)$. Show each intermediate step.

Solution: ....

Remove q_1 (one can also start by removing q_2):



and we obtain the expression $(c + ac^*b + (ac^*a + b)(a + c + bc^*a)^*bc^*b)^*$. □

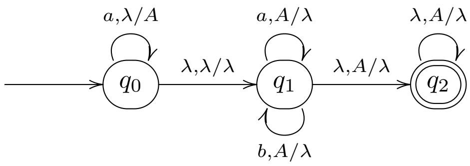
Problem 6.

Let M be the PDA with

$$\begin{array}{lll}
Q = \{q_0, q_1, q_2\} & \delta(q_0, a, \lambda) = \{\langle q_0, A\rangle\} & \delta(q_0, \lambda, \lambda) = \{\langle q_1, \lambda\rangle\} \\
\Sigma = \{a, b\} & \delta(q_1, a, A) = \{\langle q_1, \lambda\rangle\} & \delta(q_1, b, A) = \{\langle q_1, \lambda\rangle\} \\
\Gamma = \{A\} & \delta(q_1, \lambda, A) = \{\langle q_2, \lambda\rangle\} & \delta(q_2, \lambda, A) = \{\langle q_2, \lambda\rangle\} \\
F = \{q_2\} & &
\end{array}$$

a) Draw a state diagram for M. **(4pt)**

Solution: ....



□

b) Show that *aab* is accepted by M, but that *aabb* is not. **(5pt)**

Solution: ....

For *aab*, we have

$$\begin{aligned}(q_0, aab, \lambda) &\to (q_0, ab, A)\\ &\to (q_0, b, AA)\\ &\to (q_1, b, AA)\\ &\to (q_1, \lambda, A)\\ &\to (q_2, \lambda, \lambda)\end{aligned}$$

hence it is accepted.

For *aabb*, first observe that reading a b can only happen in q_1 by popping an A, and every A on the stack in q_1 has been pushed in q_0 by reading an a. Hence, the only possibility for reading *aabb* is by reading *aa* in q_0 (pushing two A's on the stack) and subsequently reading *bb* in q_1 (popping two A's). Thus any potential computation reading *aabb* starts by

$$\begin{aligned}(q_0, aabb, \lambda) &\to^* (q_0, bb, AA)\\ &\to (q_1, bb, AA)\\ &\to^* (q_1, \lambda, \lambda)\end{aligned}$$

But q_1 is not accepting, and there is no further transition possible. Hence *aabb* is not accepted.

□

c) Give a precise description of $\mathcal{L}(M)$ using set notation. **(8pt)**

Solution: ....

$$\mathcal{L}(M) = \{a^i w \mid w \in \{a, b\}^* \text{ and } |w| < i\}$$

where $|w|$ is the length of a word w.

□

d) Construct a CFG G such that $\mathcal{L}(M) = \mathcal{L}(G)$. Explain your answer. **(8pt)**

Solution: ....

For instance:

S	$\to$	aT
T	$\to$	$aTa \mid aTb \mid aT \mid \lambda$

T generates words of the form $a^n w$, with $|w| \leq n$. Hence S generates words of the form $a^{n+1} w$ with $|w| \leq n$. These words form the language $\mathcal{L}(M)$.

Or

$$\boxed{\begin{array}{lcl} S & \to & aSa \mid aSb \mid A \\ A & \to & aA \mid a \end{array}}$$

S generates words of the form $a^n Aw$, with $|w| = n$, and A generates a^m with $m \geq 1$. Hence S generates words of the form $a^p w$ with $|w| < p$. These words form the language $\mathcal{L}(M)$.

$\square$